Course Notes

Spacecraft Attitude Dynamics and Control

Politecnico di Milano
Prof. Franco Bernelli Zazzera

Part 3:
Introduction to state-space control and attitude control

Note for the reader

These short notes are in support of the course "Spacecraft attitude dynamics", they are not intended to replace any textbook. Interested readers are encouraged to consult also printed textbooks and archival papers.

Franco Bernelli Zazzera

Index

Introduction

The main tasks of a control system are:
- stabilise an unstable system;
- to increase the stability of a stable system but with a rather slow transient.

It is always very important, when studying any system, to be able to identify the quantities (states or measurements) that are of interest. Especially since such quantities are often not directly possible but must be deduced from a smaller number of parameters. We must therefore identify the system's inputs and outputs. It can be seen that in general we need to deal with Multi-Input-Multi-Output (MIMO) systems. The main difficulty encountered in moving from the analysis of Single-Input-Single-Output (SISO) systems to MIMO systems is not in the number of inputs and outputs but in moving from 1-input/output to multiple inputs/outputs. Extending the analysis from a 2-input/output system to an n-m-input/output system is then relatively simple.

We have therefore somewhat defined
- the variables to be controlled: i.e. what we have called the state of the system;
- the control variables: what we have called the input of the system.

It is now necessary to have a mathematical model of the physical system. In the most general case you will have non-linear differential equations. For a non-linear system, it is not possible to calculate the transfer function. As far as the analysis we will be dealing with is concerned, we will always try to go back to a linear type of analysis so, in the case of non-linearities, it will be necessary to perform two fundamental operations
1) analysis of the equilibrium states;
2) linearisation of the system around an equilibrium state in order to obtain a linear system that can be treated analytically. It should be noted that linearisation around one of the "few" equilibrium states that a system can assume is not a reduction of control capacity because in practice one will want to control a system to keep it in a very precise equilibrium state and not to make it wander between different equilibrium states.

After having, if necessary, linearised the equations that mathematically describe the behaviour of the physical system (and bear in mind that these equations, however detailed they may be, will always be a simplification of what happens in reality), it will be possible to design the controller that deals with stabilising (or increasing the stability of) the physical system. At this point it is necessary to carry out checks on the controller designed. The steps to be followed are roughly as follows:
1) verification of operation on the linear mathematical model of the physical system;
2) verification of operation on the non-linear mathematical model of the physical system (only if it has been necessary to linearise the equations of motion);
3) verification of operation on the real system, which will also involve differences with the non-linear model, if any, that is certainly more accurate than the linear model adopted for the project.

It should be noted that designing the controller after having carried out these operations implies limits on the initial conditions that the system may have when the control actively intervenes. If the initial conditions deviate much from the considered equilibrium, the linearised model of the equations of motion will no longer be valid (i.e. it will deviate from the non-linear model as it increases) and, consequently, the controller we have designed will probably be ineffective in its work. It will also be our task to understand up to what values of (for the example we are considering) the controller can still work. At this point we assume that we have designed and verified (at least analytically) the

desired controller. The problem of its practical implementation arises. There are basically two methods of introducing the controller into the physical system:
1) Analogue controller;
2) Digital controller.

In the first case, it will be necessary to construct an electric circuit with resistors, capacitors, op-amps, etc., which has, as a whole, the transfer function of the controller we have analytically designed. One problem with this type of controller is that in order to modify its characteristics (this must always be taken into account), there are two different ways of doing so
- detaching and reattaching certain elements from the circuit;
- introducing variable elements from the outset (such as variable resistors) in order to be able to intervene without damaging the circuit. It should be noted that in this case, the motion of the physical system itself could in some way influence the parameters of the controller. It is up to us to ensure that this phenomenon can be to the advantage or disadvantage of the control.

An alternative to analogue systems are digital control systems. These systems are much more adaptable to changing operating conditions for the simple reason that the control calculations are typically performed by software. The main problem with this type of system is that, due to their binary nature, they are inherently discrete and therefore require a suitable interface with the system to be controlled. In short, this means that the measurements read by the physical system and the actions exercised on the physical system by the digital controller are, of necessity, timed according to a precise schedule. In order for this to work correctly (the physical system is continuous), two converters will be needed (in reality, other elements will be needed, but for now, this is fine), one to convert the analogue quantities into digital form upstream of the controller (A/D) and one to do the reverse operation downstream of the controller (D/A), before intervening on the physical system.
Because of this different way of interacting with the system, it may happen that a controller designed according to "continuous logic" does not have the same effect when implemented in digital form (we will see later the cases in which this is possible or not). In these cases, therefore, it is more convenient to design the controller directly with methods that take into account the discretisation linked to digital control.

The important thing is to bear in mind when designing the control system that it must also be physically feasible. The controller we have designed tells us the extent of the actions that must take place according to the measurements of the outputs of the physical system. At this point we need to go into a little more detail. How are these forces applied? In general, we will have a device (actuator) which is distinct from the physical system to be controlled. This device will, in turn, be a system described analytically by a mathematical model of some kind. For now, it is sufficient to note that there may be situations in which the force required by the controller (i.e. the action which will be the input of the system to be controlled) is not available (saturation of the actuators) or arrives at the physical system with a certain delay. In order to be aware of the possibility of such phenomena, it is necessary to know in sufficient detail the dynamic behaviour of the device used, which we have so far assumed to be ideal (i.e. with a transient = 0 s and regime equal to that required). It should also be noted that the dynamic behaviour of an actuator, by partly influencing the dynamics of the system to be controlled (since it will necessarily have to come into contact with the system in order to intervene), could also render ineffective the controller we have designed based solely on the dynamics of the physical system. The same applies to the devices (transducers) used to measure the quantities required by the controller (i.e. the system outputs).

Finally, one thing we must not forget is disturbances, understood as noise in the measurements taken or disturbances acting directly on the physical system. This means that the controller we have designed, apart from being strictly valid only for the linearised mathematical model of the physical system, does not take into account the effects just listed. Typically, the influence of actuators and transducers and noise is introduced in the controller verification phase. As it happens, although they can all be modelled, introducing them at the controller design stage would be an unnecessary complication. Therefore, when it is not necessary for other reasons, these elements will only be taken into account after the controller has been designed.

Modeling physical systems

Linearization

All (or almost all) dynamical systems can be described, in the most general case, by non-linear, time-varying equations which, symbolically, look as follows.

$$\begin{cases} \dot{x} = f(x, u, t) \\ y = g(x, u, t) \end{cases}$$

If a system has one or more equilibrium configuration, then it is possible to linearize the dynamics in the vicinity of the equilibrium state and express the system dynamics in state-space form. First the equilibrium configuration (state and control input) needs to be evaluated by solving

$$0 = f(\bar{x}, \bar{u})$$

Then, the state and control matrices of the state-space model can be evaluated as

$$A = \left.\frac{\partial f}{\partial x}\right|_{\substack{x=\bar{x} \\ u=\bar{u}}}; \quad B = \left.\frac{\partial f}{\partial u}\right|_{\substack{x=\bar{x} \\ u=\bar{u}}}$$

Typically, we will act on dynamic systems described by equations of the following type.

$$\begin{cases} \dot{x} = A \cdot x + B \cdot u \\ y = C \cdot x + D \cdot u \end{cases}$$

These equations are a major simplification of the reality of things. The two most obvious simplifications are:
- linearity of the equations;
- constancy of the coefficients (i.e. time invariant).

However, these equations are used very often for the following reasons:
- they are easy to solve;
- they still give good results.

In practice, it is a matter of performing a linearisation of the equations around their equilibrium point. This operation allows in fact to:
- study the stability of the equilibrium points of the system;
- evaluate the response of the system (in equilibrium) to small perturbations.

One might think of using linear time-varying equations, but this type of system is not well suited to control synthesis (i.e. designing a system to control them) and, in general, their simulation is quite complex.
To get around this problem (if it is possible), it is assumed that the variations in time (of the mathematical model of the physical system) are sufficiently slow with respect to the phenomenon being studied.
Then, as already said, one studies linear time-invariant systems that are obtained from the linearisation of a system.

It is however necessary to remember that, behind the equations we are using, there are much more complex equations which, in turn, cannot describe the real behaviour of the system.

The importance of the effect of sensors

Let us consider again the system described in state-space. The quantities within the matrices are often measured by means of appropriate sensors. It is therefore important to remember that these sensors have their own dynamics which may or may not influence that of the physical system in which they are "introduced". Typically, they will introduce a time delay.
For example, a sensor can be described by an expression such as:

$$\begin{cases} \dot{x}_a = A_a \cdot x_a + B_a \cdot y_a \\ a = c_a \cdot x_a \end{cases}$$

There are some cases in which sensor dynamics can be neglected. It should also be borne in mind that a design that is too refined (and possibly expensive in computing resources) may not make much sense. For example, if our calculations foresee the use of an electrical resistor of 45.2279 , we will never be able to follow these results in the same way, for the simple reason that such a resistor cannot be found on the market and it is not at all convenient to have it custom-made (also because probably already in the order of 10^{-2} the precision in the production of a resistor can be very poor).

The time response of the system

Let us suppose for now that we have the linearised equations describing the physical system we are studying, in the well-known form

$$\begin{cases} \dot{x} = A \cdot x + B \cdot u \\ y = C \cdot x + D \cdot u \end{cases}$$

What should we do now? Generally, two steps are taken:
- stability study according to Liapunov;
- study of the response.

Stability is studied with respect to initial conditions of the system. That is, we are interested in seeing whether the perturbed motion is stable or not. Remember that there are both concepts of simple stability and asymptotic stability. The general integral of a differential system, is usually composed of two parts:
- general integral of the associated homogeneous system;
- particular integral.

It is interesting to note that stability is only associated with the first part of this solution since the particular integral describes the system's transient from the perturbed initial conditions. A system is stable when the real part of the eigenvalues is negative or null. In the case where it is null, it is possible for the system to be stable or unstable depending on the multiplicity of these null eigenvalues. The system can be stable only in the case in which the multiplicity of the eigenvectors is equal to the multiplicity of the eigenvalues. The solution of the state-space system is given, in general, by an expression of the type:

$$x = \Phi \cdot x_0 + \int_0^t \Phi(t-\tau) \cdot B \cdot u(\tau) d\tau$$

where Φ (known as the transition matrix) is the solution of:

$$\dot{\Phi} = A \cdot \Phi$$
$$\rightarrow \Phi = e^{A \cdot t}$$

Substituting this solution into the state-space model, we observe that we obtain an identity, that confirms this is a solution of the initial system.

The system output is then

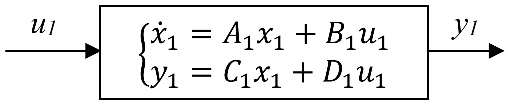
$$y = C \cdot \Phi(t) \cdot x_0 + C \int_0^t \Phi(t-\tau) \cdot B \cdot u(\tau) d\tau + D \cdot u(t)$$

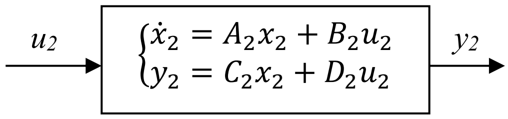
The integral present in the solution is called convolution integral. Its generic form is

$$\int f(t-\tau) g(\tau) d\tau$$

Systems connected in series, parallel, feedback

Let us now suppose that we have two systems described in the state space form:

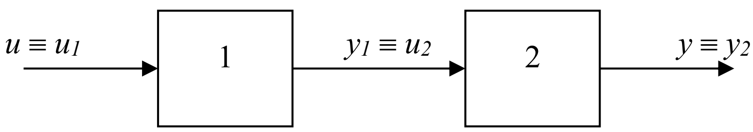




We want to see what the equations of a system made up of these two systems look like, connecting systems in three different ways:
- series;
- parallel;
- feedback.

Series

The series of the two systems is represented in the following block diagram.

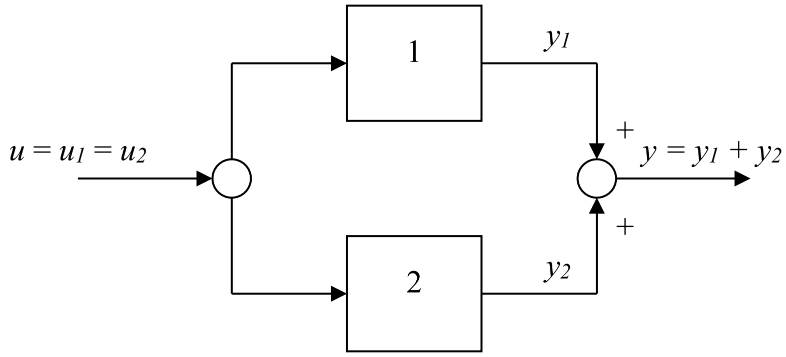


The global system obtained is represented by the following equations that describe the state-space model of two connected linear systems.

$$\begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{Bmatrix} = \begin{bmatrix} A_1 & 0 \\ B_2C_1 & A_2 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} + \begin{bmatrix} B_1 \\ B_2D_1 \end{bmatrix} \{u_1\}$$
$$y = [D_2C_1 \quad C_2] \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} + [D_2D_1]\{u_1\}$$

Parallel

Let us now see what is achieved by linking them in parallel. The block diagram shows that they have the input in common and the output of the overall system is the sum of the two outputs.

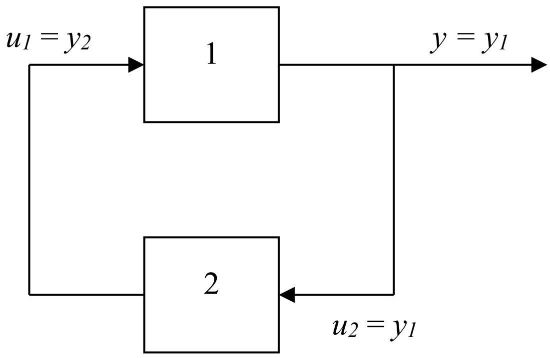


In this case the state-space model is as follows.

$$\begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{Bmatrix} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \{u\}$$
$$y = [C_1 \quad C_2] \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} + [D_1 + D_2]\{u\}$$

Feedback loop

This is the typical case of control in which the output of the system to be controlled is fed back in order to process it and eventually intervene on the system. The block diagram shows that in the analysis that will be carried out, the output that is of most interest is the one of the system that is controlled, while the system on the feedback loop will process the output of the other system.



The determination of the equations of the system is more complex because there is a "circular reference" which complicates the explication of the equations. The substitutions that must be made in order to continue with the calculation are:

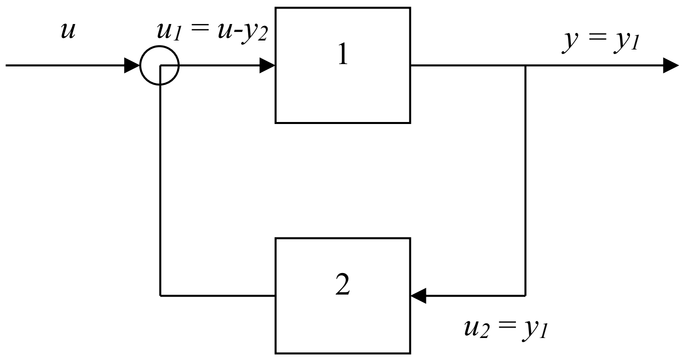
$$u_1 = y_2$$
$$u_2 = y_1$$
$$y = y_1$$
$$u = 0$$

The final state-space model is as follows.

$$\begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{Bmatrix} = \begin{bmatrix} A_1 + B_1(I - D_2D_1)^{-1}D_2C_1 & B_1(I - D_2D_1)^{-1}C_2 \\ B_2C_1 + B_2D_1(I - D_2D_1)^{-1}D_2C_1 & A_2 + B_2D_1(I - D_2D_1)^{-1}C_2 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix}$$

$$y = \begin{bmatrix} C_1 + D_1(I - D_2D_1)^{-1}D_2C_1 & D_1(I - D_2D_1)^{-1}C_2 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix}$$

It is observed that there is no input *u* because the output of one system becomes the input of the other. In a more general case, the system can be subject to a feedback control in addition to a reference input.



In such configuration, the state-space model becomes.

$$\begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{Bmatrix} = \begin{bmatrix} A_1 - B_1D_2(I + D_1D_2)^{-1}C_1 & -B_1C_2 - B_1D_2(I + D_1D_2)^{-1}D_1C_2 \\ B_2(I + D_1D_2)^{-1}C_1 & A_2 - B_2(I + D_1D_2)^{-1}D_1C_2 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} + \begin{bmatrix} B_1 - B_1D_2(I + D_1D_2)^{-1}D_1 \\ B_2(I + D_1D_2)^{-1}D_1 \end{bmatrix} \{u\}$$

$$y = \begin{bmatrix} (I + D_1D_2)^{-1}C_1 & -(I + D_1D_2)^{-1}D_1C_2 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} + \begin{bmatrix} (I + D_1D_2)^{-1}D_1 \end{bmatrix} \{u\}$$

Global vs. individual system stability

It is interesting to see how the individual stability of the two systems is "transported" once they are connected in the ways just seen. Looking at the matrix A of the different interconnected systems we can deduce that:

- *Systems in series*: we note that, although there is an extra-diagonal term, if we evaluate the eigenvalues of the matrix we obtain the set of eigenvalues of A_1 and A_2 since the matrix is triangular. Therefore, if two stable systems are connected in series, the stability of the overall system is guaranteed.
- *Systems in parallel*: in this case even the dynamics of the two systems are completely decoupled and the global eigenvalues will be nothing but the set of eigenvalues of system 1 and system 2. Even for systems connected in parallel, individual stability is maintained in the global system.
- *Feedback systems*: In this case, a more complex relationship between the two systems can be observed, the dynamics of which are strongly linked. It can therefore be deduced that the global eigenvalues of the loop system will generally not be equal to the eigenvalues of the two individual systems. This is a disadvantage (stability analysis is also required for the global system), but it also makes it possible to have a stable global system even when one of the two systems is unstable or even both unstable. This consequence is an advantage for those who are about to design a control system, especially if the system to be controlled is unstable.

Control proportional to the system state

We want to see control theory developed in the time domain using models in the state space. The system model we will refer to is the linear one

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx + Du \end{cases}$$

Let assume that the generic control *u* is formulated as

$$u = k \cdot x$$

such that the dynamics of the controlled system becomes, symbolically:

$$\dot{x} = (A + Bk)x$$

This is the state matrix of a feedback system on which a state-proportional control is applied.
The characteristic equation (the one that makes it possible to calculate the eigenvalues, which are essential for stability analysis) of this, in symbolic form, is

$$|sI - (A + Bk)| = 0$$

from which we obtain a polynomial of degree equal to the number of states of the system, which can be used to calculate the gains *k* to correct the stability of the system at will.

Pole-placement control

Controllability canonical form (first companion form, FCF)

Let us now suppose that we have the transfer function of a generic system, given by the expression:

$$\frac{y(s)}{u(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1} s + b_n}{s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n}$$

From this transfer function (which links an input to an output) we want to deduce the model of the system in the space of states in order to design its proportional control to the states. We will have to proceed by processing the transfer function in order to obtain the matrices *A*, *B*, *C* and *D* that describe the system in state space. Note that the denominator of the transfer function is none other than the characteristic polynomial of the matrix *A*.
The transfer function can also be written as follows:

$$\frac{z(s)}{u(s)}\frac{y(s)}{z(s)} = \frac{1}{s^n + a_1 s^{n-1} + \cdots + a_n}(b_0 s^n + b_1 s^{n-1} + \cdots + b_n)$$

where *z(s)* is a generic function of *s* that we do not define for now. From this expression, we can write

$$\frac{z(s)}{u(s)} = \frac{1}{s^n + a_1 s^{n-1} + \cdots + a_n}$$
$$u(s) = z(s)(s^n + a_1 s^{n-1} + \cdots + a_n)$$

Similarly

$$\frac{y(s)}{z(s)} = (b_0 s^n + b_1 s^{n-1} + \cdots + b_n)$$

from which, performing the anti-transform (L^{-1}), we obtain an expression of *u(t)* as a linear combination of the function *z* and all its derivatives up to the *n*-th order

$$u(t) = D^n z(t) + a_1 D^{n-1} z(t) + \cdots + a_n z(t)$$

If we now define, in a completely arbitrary way (but with a precise intention), the state of the system as the set of *z* and its time derivatives up to the order $n-1$

$$x = \begin{Bmatrix} D^{n-1} z(t) \\ D^{n-2} z(t) \\ \vdots \\ z(t) \end{Bmatrix}; \quad \dot{x} = \begin{Bmatrix} D^n z(t) \\ D^{n-1} z(t) \\ \vdots \\ D^1 z(t) \end{Bmatrix}$$

we can write the model of the system states in the usual form $x = Ax + Bu$. In fact, by expressing $D^n z(t)$ the following is obtained

$$D^n z(t) = u(t) - a_1 D^{n-1} z(t) - a_2 D^{n-1} z(t) - \cdots - a_n z(t)$$

From which the structure of matrices *A* and *B* is obtained

$$A = \begin{bmatrix} -a_1 & -a_2 & -a_3 & \cdots & -a_{n-1} & -a_n \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}; \quad B = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

We must now find a way to identify, in association with the matrices *A* and *B*, the matrices *C* and *D* which describe the output of the system. To this end, we extract the two remaining terms, thanks to which we can write:

$$y(s) = z(s)(b_0 s^n + b_1 s^{n-1} + \cdots + b_n)$$

In analogy to what was done to find *A* and *B*, by doing the anti-transform, we obtain

$$y(t) = b_0 D^n z(t) + b_1 D^{n-1} z(t) + \cdots + b_n z(t)$$

From which matrices *C* and *D* are otained

$$C = [b_1 - a_1 b_0 \quad b_2 - a_2 b_0 \quad \ldots \quad b_n - a_n b_0]; \quad D = [b_0]$$

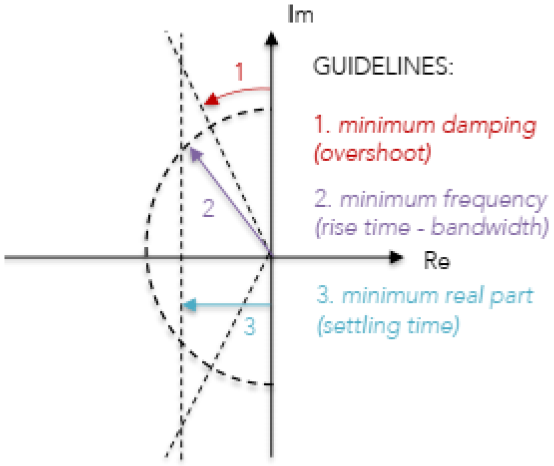
We have therefore managed to obtain a model of the system's states from its generic transfer function. The characteristic aspect of this model (in which the matrix *A* contains the coefficients of the characteristic polynomial of the system in the first row) makes it particularly interesting and we give it the name First Companion Form (FCF) or Controllability Canonical Form. The convenience of this form lies in the simplicity of its deduction from the transfer function. If we then introduce a control proportional to the state, we can note that the matrix *A+Bk* is particularly simple and becomes

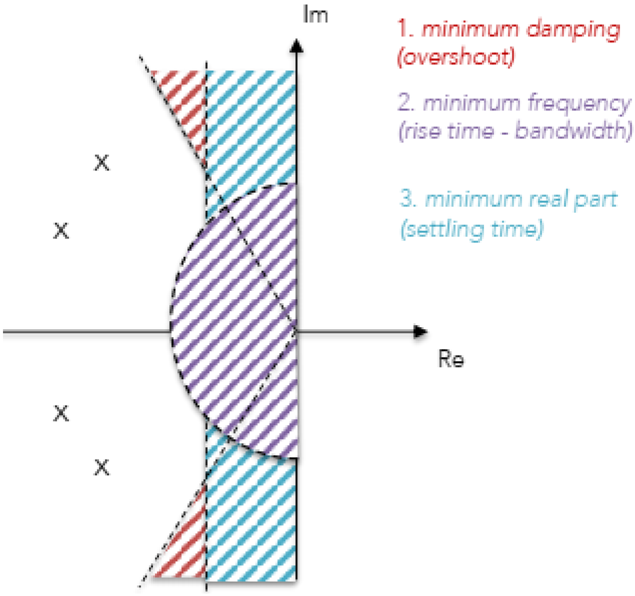
$$A + Bk = \begin{bmatrix} k_1 - a_1 & k_2 - a_2 & \cdots & k_{n-1} - a_{n-1} & k_n - a_n \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

It can be seen that, because of the aspect it assumes, each gain modifies only one of the coefficients of the characteristic polynomial, guaranteeing complete freedom in positioning the poles of the system. We have thus obtained a very interesting result of a rather general nature (at least for linear time-invariant systems): if it is possible to write the system in the FCF, through a feedback proportional to the state one can place the poles of the system (in a closed loop) where one wishes in order to have the best possible dynamic response. The problem with this form of the state model is that the state is unlikely to have any physical meaning and is therefore difficult to interpret in practice. The state will usually have to be transformed through a state transformation to obtain the system as a function of a more meaningful state. We shall see, however, that it is also possible to circumvent this problem.

Where to place the closed-loop poles?

Recalling the general characteristics of the time response as a function of the poles of the system, preliminary relationship between characteristics of the step response (rise time, settling time, overshoot) and the location of closed-loop poles can be deduced. Note that this is rather straightforward for typical second-order response, while for higher order systems it is not so obvious. Also, it must be remarked that the response is also affected by the zeros.





The concept of controllability

We have seen that a proportional control of the states of a system allows us to position all the poles of the system through the relation:

$$\begin{aligned} &\dot{x} = Ax + Bu \equiv (A + Bk)x \\ &u = kx \end{aligned}$$

We have also seen that the positioning of the poles is particularly simple if the system is represented by the model expressed in the FCF, although we will often have to find a state transformation that allows us to transform the matrix A into the form of the FCF.
This operation is not always possible. For this reason we define the concept of controllability.

Controllability of a system: a system is controllable if it is possible to bring its state from an initial (finite) value to a final (finite) value in a limited time, using only the control variable.

Pole placement is possible if and only if the system is controllable. Tools are therefore needed to establish the controllability of a system in a rigorous manner. There is an algebraic condition that can also be applied to non-linear systems. Having defined the matrix

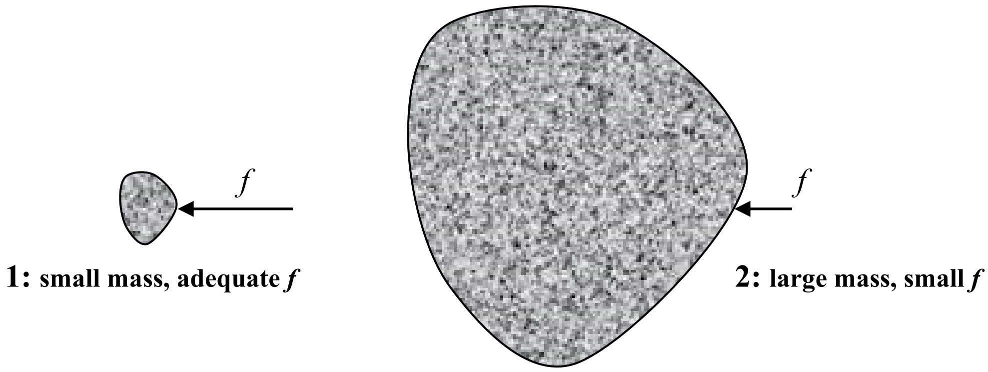
$$P = \int_0^t \Phi(\lambda)B(\lambda)B^T(\lambda)\Phi(\lambda)d\lambda$$

if P is full rank, the system is controllable. Considering linear systems, the analysis can be simplified by looking at the rank of the following matrix:

$$Q = [B \quad AB \quad A^2B \quad \cdots \quad A^{n-1}B]$$

If Q is full rank (equal to the number of states in the system), the linear system (characterised by the matrices A and B) is controllable. Note that the dimensions of Q are given by:
- rows: number of states of the system;
- columns: product of the number of states and inputs.

It is interesting to observe that as the number of inputs increases (i.e. the actions that can be exercised on the system in order to control it), the columns of B and, consequently, of Q increase, so that the latter is more likely to have maximum rank. In other words, it means that a system is easier to control if one acts on it with a greater number of actions. It is also interesting to see that Q depends not only on the matrix that determines the free dynamics of the system (A), but also on the matrix that introduces the effect of the control actions into the system (B). One thing missing from the analytical verification of controllability just introduced is the quantitative aspect. In fact, it is possible to verify whether a system is controllable or not, but it is not possible to assess how controllable a system can be. An example that can make the importance of the quantitative aspect clear is the following comparison.



System 1 (a small isolated mass) is certainly controllable. System 2 is also controllable (it is a system with the same dynamics as system 1 but with different physical characteristics), but it has a problem: the large mass will move slowly under the effect of the small force. This means that, although both are controllable, system 2 is "less" controllable than system 1. One of the first resources we call upon to distinguish these cases is common sense. There is also the possibility of identifying parameters (of little physical significance) that indicate in some way how controllable a system is. However, these cannot be deduced from a theorem, but are defined case by case by intuition. If, for example, the matrix Q is square and the system is controllable (i.e. det(Q) 0), one can try to see how much the determinant of Q is worth. An example may be clearer. Let us consider a clamped beam.

If a force in the vertical direction allows me to control, for example, the position of the tip of the beam (and if Q is square), it is to be expected that the determinant of Q will be closer to zero in the case of force 2 because this is the case in which more action is required (compared to the case with force 1) to obtain the same result.

The concept of observability

In addition to controllability characteristics, a system must also fulfil other requirements. We define the concept of observability.

Observability of a system: a system is observable if a record of the outputs y can be used to determine the value of the state x.

This characteristic is fundamental because in order to apply proportional control to the states, it is necessary to know the value of state x of the system, and this can only be known from the system output y.
Similarly to what has been done for controllability, it is possible to define a matrix

$$N = [C^T \quad A^T C^T \quad \cdots \quad A^{T^{n-1}} C^T]$$

If N is full rank (also equal to the number of states of the system), the system is observable. The dimensions of N are given by:
- - rows: number of states;
- - columns: product of the number of states and the number of outputs.

Here again, the quantitative aspect can be deduced more on the basis of intuition than mathematical rigour. One can certainly observe, in the case of a square N-matrix, the value of the determinant. It is clear that this method makes more sense if several measurement possibilities are considered for the same system in order to compare the different determinants of the different N obtained.
As can be seen, N depends not only on the matrix describing the dynamics of the free system (A), but also on the matrix establishing the state combination provided by the system's outputs (C).

Note that both the matrix Q and the matrix N are calculated in practice by a recursive method of the type B, $A \cdot B$, $A \cdot AB$, $A \cdot A^2 B$, … and C^T, $A^T \cdot C^T$, $A^T \cdot A^T C^T$, ecc.

Partially controllable systems

There are systems that are only partially controllable in the sense that only part of their state can be controlled by the inputs under consideration. Alternatively, some rather complex systems may require a subdivision of the control into subsystems. In this case the controllability analysis can be done on the subsystems and not on the complete system. But the controllability analysis can also be used to identify, once a control system is known, which states are not controllable. Suppose we know a state transformation T that allows us to write the system in the following form:

$$\begin{Bmatrix} \dot{\bar{x}}_C \\ \dot{\bar{x}}_N \end{Bmatrix} = \begin{bmatrix} A_{CC} & 0 \\ A_{NC} & A_{NN} \end{bmatrix} \begin{Bmatrix} \bar{x}_C \\ \bar{x}_N \end{Bmatrix} + \begin{bmatrix} B_C \\ 0 \end{bmatrix} \{u\}$$
$$x = T\bar{x}$$

where controllable and non-controllable states are separated. It has been said that to control a state means to bring it from one condition to another (both limited) in a finite time. Suppose then that we want to bring the state of the system up to its regime ($x = 0$). Then, by setting the derivative of the state equal to zero and solving for x_C and x_N, we obtain:

$$\bar{x}_C = -{A_{CC}}^{-1} B_C u$$
$$\bar{x}_N = -{A_{NN}}^{-1} A_{NC} \bar{x}_C$$

These show that, once the steady state behaviour of the state controlled by u (i.e. x_C) has been decided, the steady state behaviour of the non-controllable state (x_N) cannot be decided and will depend on the value assumed by x_C. It is clear that in order to obtain the system in the form just seen it is necessary to find the appropriate matrix T.

Pole placement without the FCF

In this paragraph we will consider again the case of a single input evaluated obviously through a linear combination of the state (proportional control to states). The problem with pole placement in the FCF is, as we have already seen, the little physical meaning that the state takes on in that particular form. Since the model of a physical system is usually written by choosing a state that has some meaning, it would be nice to be able to perform pole placement (i.e. determine the k matrix of gains that gives $u = kx$) without having to go through the representation of the system in the FCF. Suppose we have a generic system whose state has some comprehensible physical meaning:

$$\dot{x} = Ax + Bu$$

Consider the linear transformation:

$$x = T\bar{x}$$

That transforms the original system into its FCF:

$$\dot{\bar{x}} = \bar{A}\bar{x} + \bar{B}u$$
$$\bar{A} = T^{-1}AT$$
$$\bar{B} = T^{-1}B$$

Let us now say that we do the pole placement by introducing the matrix k that multiplies the transformed state $\bar{x}$. Then

$$u = \bar{k}\bar{x}$$
$$\dot{\bar{x}} = \left(\bar{A} + \bar{B}\bar{k}\right)\bar{x}$$

The fact that the system is transformed does not affect the correction of the stability of the system since we know that the poles of the system are always the same (the matrix A, even if transformed, always has the same eigenvalues). The problem arises due to the fact that it is the state of the FCF that is multiplied by $\bar{k}$. At this point we can take two paths

1. transform the $\bar{k}$ to adapt it to the state of the initial system (but this requires knowledge of the matrix T);
2. calculate the k suitable for the initial system directly, without going through the FCF.

Evaluation of the transfromation matrix T

If we write the controllability matrix Q and then substitute the expressions of A and B as a function of A, B and T, we can see that the:

$$\bar{Q} = T^{-1}Q$$

At this point it should be noted that it is possible to calculate both Q and $\bar{Q}$ because the starting system is certainly known (i.e. A and B) and, once the characteristic polynomial of A is known, A and B are also known from which we can deduce Q. Then it is possible to calculate the transformation matrix T and the gain matrix k

$$T^{-1} = \bar{Q}Q^{-1}$$
$$k = \bar{k}T^{-1}$$

The operations to be performed to find k by this method are:
- place the poles in the FCF (which involves evaluating the characteristic polynomial of A) and find k;
- calculate the matrix Q, invert it and multiply it with Q (also calculated) to obtain T^{-1};
- multiply k by T^{-1} to finally obtain the desired k.

This method is not very convenient because it involves the inversion of the matrix Q, which numerically always presents problems due to the possible ill-conditioning of Q.

Avoid using the FCF

Let us consider the coefficients of the characteristic polynomial of the open-loop system (a_i) and those desired for the closed-loop system ($\breve{a}_i$). The gain k for the FCF is computed as

$$\bar{k}_i = \breve{a}_i - a_i$$

Let us now consider the characteristic polynomials in an open loop ($\alpha(s)$) and closed loop ($\alpha_C(s)$). Since a property of the characteristic polynomials of square matrices is that "every matrix is a solution of its characteristic polynomial" it can be written that:

$$\alpha(\bar{A}) \equiv \bar{A}^n + a_1\bar{A}^{n-1} + \cdots + a_{n-1}\bar{A} + a_n I = 0$$

The same holds if we replace $\bar{A}$ with A (note that $A^0 = I$). If instead we substitute $\bar{A}$ in the polynomial α_C (that is not its characteristic polynomial) we have

$$\alpha_C(\bar{A}) = \bar{A}^n + \breve{a}_1\bar{A}^{n-1} + \cdots \breve{a}_{n-1}\bar{A} + \breve{a}_n I \neq 0$$

From $\alpha(\bar{A})$ we make explicit $\bar{A}^n$ and substitute in $\alpha_C(\bar{A})$, obtaining:

$$\alpha_C(\bar{A}) = (\breve{a}_1 - a_1)\bar{A}^{n-1} + (\breve{a}_2 - a_2)\bar{A}^{n-2} + \cdots + (\breve{a}_n - a_n)I \neq 0$$

in which the coefficients at the second member are the k_i. Unfortunately, we cannot extract the gain vector k since its elements each multiply a matrix.

Then, introduce a vector e_n^T defined as:
- it is a row vector of order n;
- it is composed by zeros with the exception of position n (indicated as subscript) where it values 1.

Then, e_n^T is

$$e_n{}^T = [0 \quad 0 \quad \cdots \quad 0 \quad 1]$$

Recalling the structure of $\bar{A}$, it is observed that its product with e_n^T is

$$e_n{}^T \cdot \bar{A} = [0 \quad 0 \quad \cdots \quad 1 \quad 0] \equiv e_{n-1}{}^T$$

Implementing this property recursively, it is then

$$e_n{}^T \cdot \bar{A}^m = e_n{}^T \overbrace{\bar{A} \cdot \bar{A} \cdot \ldots \cdot \bar{A}}^{m \text{ times}} = e_{n-m}{}^T$$

$$\underbrace{e_n{}^T \bar{A}}_{e_{n-1}{}^T}, \qquad \underbrace{e_n{}^T \bar{A} \cdot \bar{A}}_{e_{n-2}{}^T}, \qquad \underbrace{e_n{}^T \bar{A} \cdot \bar{A} \cdot \ldots \cdot \bar{A}}_{e_{n-m}{}^T}$$

Therefore, pre-multiplying the polynomial α_C by e_n^T the following is obtained:

$$e_n{}^T \cdot \alpha_C(\bar{A}) = \underbrace{e_1{}^T(\breve{a}_1 - a_1)}_{[\breve{a}_1 - a_1 \quad 0 \quad \cdots \quad 0 \quad 0]} + \underbrace{e_2{}^T(\breve{a}_2 - a_2)}_{[0 \quad \breve{a}_2 - a_2 \quad \cdots \quad 0 \quad 0]} + \cdots + \underbrace{e_n{}^T(\breve{a}_n - a_n)}_{[0 \quad 0 \quad \cdots \quad 0 \quad \breve{a}_n - a_n]}$$

In synthesis

$$e_n{}^T \cdot \alpha_C(\bar{A}) = \bar{k}$$

Now introduce the state transformation T, i.e. substitute $\bar{A}$ with its expression function of A ($T^{-1}AT$) and substitute $\bar{k}$ with its expression function of k (kT):

$$(e_n{}^T \alpha_C(T^{-1}AT) \equiv) \quad e_n{}^T T^{-1} \alpha_C(A) T = kT \quad (\equiv \bar{k})$$

From which the expression for k is obtained as

$$k = e_n{}^T T^{-1} \alpha_C(A)$$

Recalling now the expression of T^{-1} and considering that, for the particular structure of Q the relation $e_n^T \cdot Q = e_n^T$ holds, it is finally possible to write

$$k = e_n{}^T Q^{-1} \alpha_C(A)$$

which we note is independent of FCF. Since we do not want to invert the matrix Q, we make the following substitution, in order to use the well-known numerical methods for solving linear systems that allow us to avoid the inversion of Q, with all its advantages

$$e_n{}^T Q^{-1} = b^T$$

and b is computed by solving the linear system

$$Q \cdot b = e_n$$

Thus, the gain matrix k is obtained, after solving for b, as

$$k = b^T \cdot \alpha_C(A)$$

We were therefore able to see that the positioning of the poles of a system (as long as it is controllable!) can also be done without resorting to FCF, which requires rather complex and numerically critical calculations.

The advantage of using more than one actuator (input)

Until now, we have dealt with systems with only one input. If we move on to control with more than one actuator, we can immediately emphasise that the control possibilities increase (which is intuitive but is confirmed by the controllability analysis introduced previously). In fact, the matrix B has more than one column and there is a greater probability of obtaining a Q with maximum rank. In practice, the advantage, as long as the system is controllable, is certainly in the possibility of positioning the poles. Moreover, it may be possible to intervene on other aspects of the system (such as minimising the effort of controlling).

The state observer

In order to apply a control proportional to the state ($u = kx$) it is necessary to know the entire state, otherwise such control is impossible. We know that the only information that "comes out" of the physical system is that which is obtained from the y-outputs. It will therefore be necessary to be able to pass from the measurements of the outputs y (which may or may not be equal to the state or a part of it) to the value of the state x. We have already said that for this to be possible, the system must be observable. Once this property has been verified, it is necessary to find a way to deduce x once the outputs have been measured. A first trivial case that is impossible to encounter (but is best illustrated) is that in which the linearized system is proper and has non-singular square matrix C. In this case the state can be obtained directly from

$$x = C^{-1} y$$

The fact that this case never occurs is related to its physical meaning. A matrix C that is square and invertible but different from a diagonal matrix implies that the outputs y measure linear combinations of the state variables. If it were possible to measure a linear combination of states then it is presumably easier to measure the states directly (i.e. C is diagonal and y coincides, barring any 'amplifications' or

'reductions', with x). Therefore the case of non-diagonal and non-singular square C is, for these practical considerations, very rare to find. However, the case of a diagonal C-matrix is also very rare because the systems have, typically:
- many states;
- few possible measures.

This should not discourage because we have seen that, for the system to be observable, it is sufficient that the observability matrix N is of maximum rank. If the available y are few, we will have to find a way to deduce all the states from this few information. This is the purpose of the systems known as observers. An observer is a fictitious system which, by processing the outputs of the real system y, allows us to deduce its state x. The observer will be characterised by its own state ($\hat{x}$) which will follow a certain dynamics. To describe the evolution of this dynamics, we consider the following mathematical model:

$$\begin{cases} \dot{\hat{x}} = \hat{A}\hat{x} + \hat{B}u + Ly \\ \hat{y} = \hat{x} \end{cases}$$

The observer receives the measurements y and introduces them into his own dynamics through the matrix L. We want the observer's state ($\hat{x}$) to be representative of the state of the real system (i.e. as close to it as possible). Thus, the size of $\hat{x}$ and x will be the same. We will now see how the model of the observer is derived from the model of the true system. We recall that the equations of the true system are

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx + Du \end{cases}$$

The observation error is defined as

$$e = \hat{x} - x$$

Deriving the error with respect to time and substituting the dynamics of the real system and the observer, we obtain the dynamics of the observation error which, reworked properly, becomes

$$\dot{e} = \hat{A}e + \left(\hat{A} - A + LC\right)x + \left(\hat{B} - B + LD\right)u$$

The observation error must be (according to our requirements) zero once the (observation) transient has been exhausted ($e = 0$). From this we deduce two conditions that allow us to express the observer matrices as a function of the system matrices and L:

1. In order for the error to be null, it is necessary that the two terms in x and u in the system are null, so we deduce the following two necessary conditions:

$$\begin{cases} \hat{A} = A - LC \\ \hat{B} = B - LD \end{cases}$$

2. If the system is not asymptotically stable, the steady-state error can never be zero. Then a new condition is that the matrix $\hat{A}$ must be characteristic of an asymptotically stable system.

What needs to be done is then:
- - find an L that makes the matrix $\hat{A}$ asymptotically stable;
- also calculate, after finding the appropriate L, the matrix$\hat{B}$.

We must therefore see how the matrix L can be calculated so as to guarantee the asymptotic stability of $\hat{A}$. Recall that from $A+Bk$ (i.e. by introducing a control proportional to the system state) it is possible to place the poles of the system as desired. In other words, this means that we can make the matrix $A+Bk$ asymptotically stable. Note that $\hat{A}$ is in a similar form:

$$\hat{A} = A - LC$$

The difference is that the matrix L premultiplies C instead of multiplying it to the right as in $A+Bk$. However, the problem is only apparent because we can transpose $\hat{A}$ and obtain:

$$\hat{A}^T = A^T - C^T L^T$$

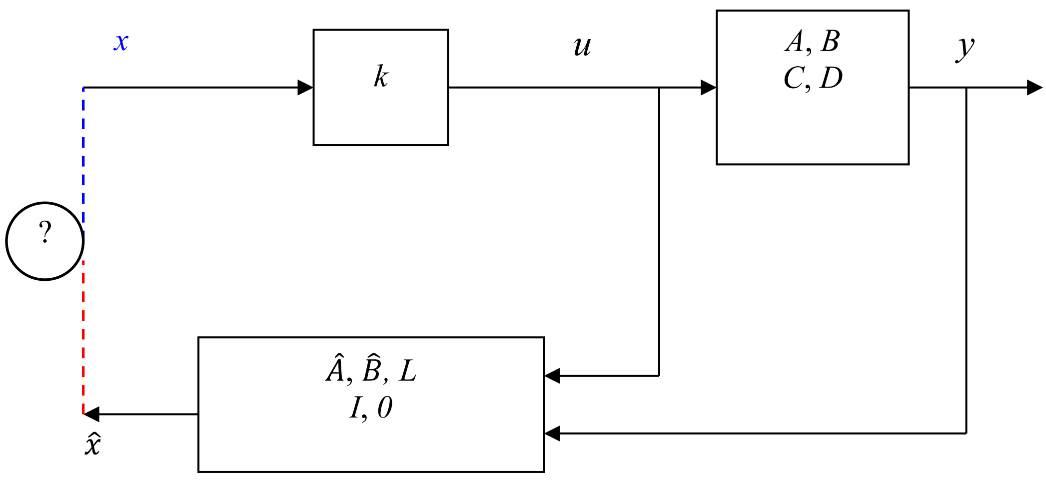
which is therefore correct and allows us to follow the methods seen above for positioning the A^T poles. Note that transposing a matrix does not change its eigenvalues. We can therefore deduce the L^T through the positioning of the poles of A^T. The operations to be done in order to identify the mathematical model of the observer are therefore

1. the positioning of the poles of the matrix A^T is done and the matrix L^T is computed;
2. L^T is transposed in order to obtain the matrix L;
3. Calculate (obviously, the characteristics of the true system A, B, C and D are known) the matrices $\hat{A}$ e $\hat{B}$.

We have then found the mathematical model of a fictitious system (the observer) that allows us to deduce the state of the real system from its outputs. Note that if the real system is a proper system, matrix $\hat{B}$ coincides with matrix B due to the absence of matrix D.

Controlled system and observer

Observe the following block diagram.



We have already shown (by analysing the dynamics of the observation error) that, after a certain observation transient, the observer's state must coincide with the true state of the system. However, it is inevitable that the input to the control block (k) is the observed state and not the true state. How can we be sure that the observer has completed his transient (or, in other words, that it can accurately track the variations of the true state)? If in fact this were not true, the pole placement that is being performed through block k, is not the one that was previously designed because it processes a state that is not that of the true system. Consequently, it is not certain that the system behaves in a stable way (and therefore it is likely that the observer will never "see" the true state).

Let us consider the two systems (true and observer):

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx + Du \end{cases}$$
$$\begin{cases} \dot{\hat{x}} = (A - LC)\hat{x} + (B - LD)u + Ly \\ \hat{y} = \hat{x} \end{cases}$$

In which the control *u* is:

$$u = k\hat{x}$$

Rewrite in matrix form:

$$\begin{Bmatrix} \dot{x} \\ \dot{\hat{x}} \end{Bmatrix} = \begin{bmatrix} A & Bk \\ LC & A - LC + Bk \end{bmatrix} \begin{Bmatrix} x \\ \hat{x} \end{Bmatrix}$$

Having already substituted Ly with its equivalent function of the real system
The system describes the dynamics of both systems, true and observer. If we assume that we have computed the matrices *k* and *L* by the methods described in the previous paragraphs, it should be noted that:
- we have guaranteed the stability of the true system through the control of the states, neglecting the possible influence of the observer;
- we have designed the observer of the true state without worrying about the possible influence of the control system.

Recalling the properties of loop systems, we want to see if the eigenvalues of the overall system are the union of the eigenvalues of the two systems (true system in a closed loop with the controller and observer). This would in fact guarantee the validity of the design method adopted (designs of *k* and *L* "decoupled").

In order to see that in reality the dynamics of the two systems do not influence each other (i.e. the eigenvalues of the are just the union of the eigenvalues of the two separate systems!) we make a state transformation on the

$$\hat{x} = x + e$$

we have substituted for the observed state the observation error whose dynamics is known. The system thus obtained is

$$\begin{Bmatrix} \dot{x} \\ \dot{e} \end{Bmatrix} = \begin{bmatrix} A + Bk & Bk \\ 0 & A - LC \end{bmatrix} \begin{Bmatrix} x \\ e \end{Bmatrix}$$

It can be seen that the state matrix
- is a triangular matrix for which the contribution to the eigenvalues comes only from the main diagonal;
- the two sub-matrices on the main diagonal are precisely those that we have made asymptotically stable, the first through the positioning of the poles of *A* with the proportional control *k* and the second through the positioning of the poles of A^T by means of L^T.

It therefore means that the eigenvalues of the global system are precisely the union of
- eigenvalues of the system subjected to proportional control (i.e. the eigenvalues of $A+Bk$);
- eigenvalues of the observer ($\hat{A} = A - LC$).

The above shows that it is possible to position the poles of the real system (i.e. design the gain matrix *k*) without worrying about the number of known states. If states are missing from the measurements on the real system, it is possible to design (independently of the controller) an observer that provides the whole system state. When the two systems are coupled (as shown in the previous diagram) we are sure that the stability of the global system is maintained since its eigenvalues are the union of the eigenvalues of the two designed systems.

Some considerations on the observer

Let us consider a real system and the design of a proportional control system. We have already shown that it is very important to bear in mind that the availability of control action is not infinite (i.e. the actuators saturate). Therefore, when designing the k, care should be taken because the larger it is, the sooner actuator saturation can occur. This is why it is often important to keep the gain low in order to avoid such saturation.

If we now take into account that k is the means by which we reposition the poles of the open-loop system (because the new poles are those of the $A+Bk$ matrix), this limitation translates into a limited freedom of repositioning. This in turn limits the speed of the system's response. This is a physical limitation that shows that it is impossible to design control systems that are too fast (of course, it depends on the actuators, but they are always characteristic of mechanical systems).

Let us now turn to the observer. Its design (i.e. the determination of the L-matrix) also involves the positioning of the poles (of the A^T-matrix), but there is a notable difference with respect to control systems: the observer is a fictitious system calculated by a computer, so the displacement of the poles can be as large as one likes (as far as possible!). This means that the speed of the observer can be "chosen" in such a way that the observation transients are negligible with respect to the variations of the state of the system, i.e. in such a way that the observer follows the state of the system without problems and almost as if it had no dynamics of its own (i.e. as if it were always at steady state).

Let us now consider the moment when the system is started up. The conditions prior to start-up will be:

- the physical system will have its state x;
- the observer will have a state equal to zero.

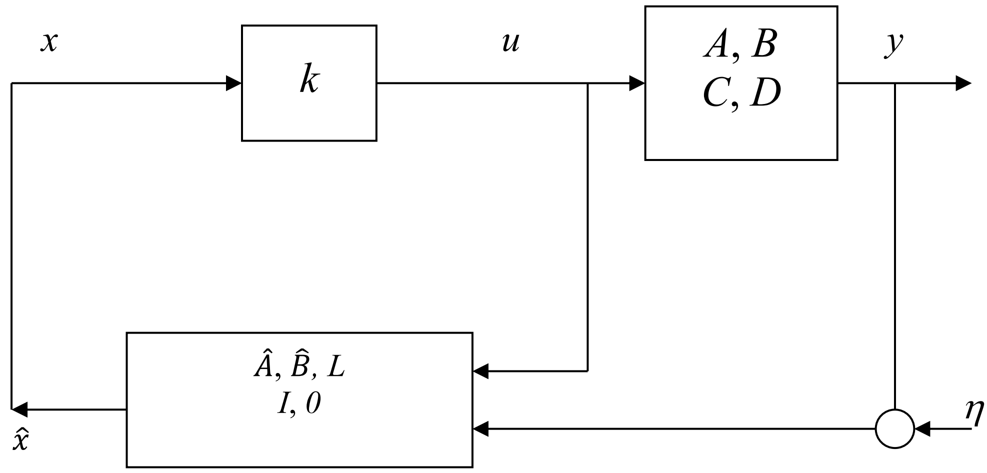
It therefore means that when the control/observation system is switched on, the observation error will be $e = -x$. If the observer is too fast (which usually implies low "damping" values), overshoots of the observed state may occur before the observed state reaches steady state (and properly "locks on" to the true state).

The problem is that these overshoots also enter the control system (which processes the observed state and not the true state) and could damage the true system.

The trick used to get around this problem is to first switch on the observer (to make it "lock" the true state) and then, once the observation transient is over, switch on the control system as well. Unfortunately, this trick can only be used if the true system is already stable.

Note that the speed of the observer is not established by a precise rule. It will depend on the physical system and the actuators considered. A somewhat empirical rule is that the response speed of the observer should be about 10 times that of the system "under observation".

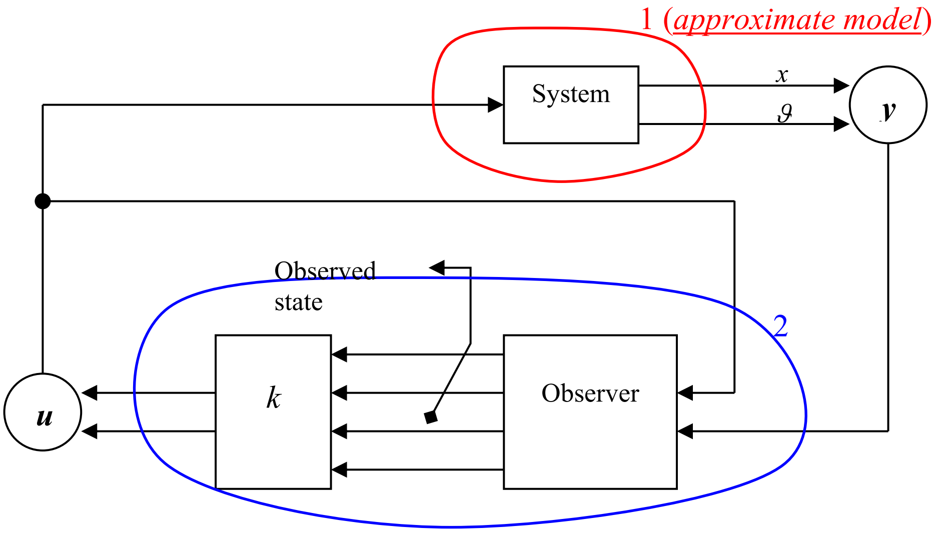
Another aspect to be considered on the speed of the observer is the presence of disturbances (which are usually high frequency, i.e. they have very fast variations) in the acquisition branch of the outputs y :



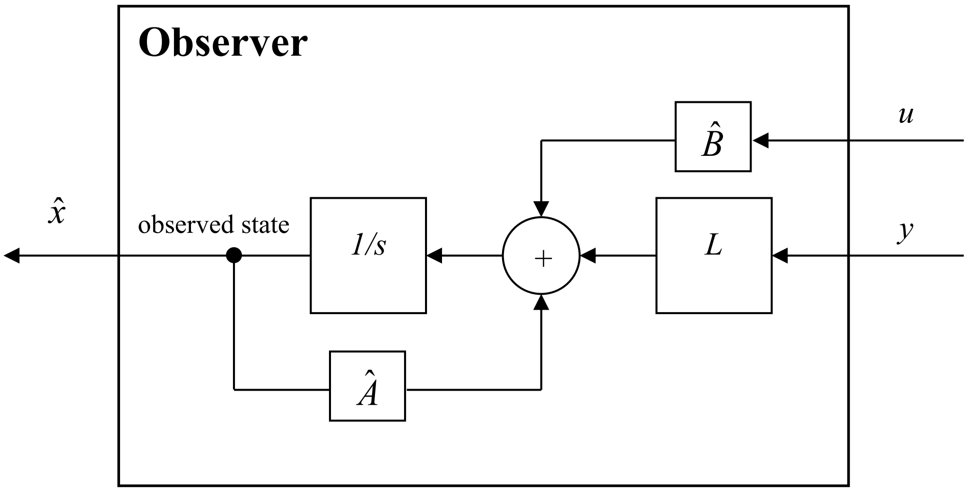
Let us assume that the system's output y coincides with its state x (as far as we are concerned, this is fine). If the observer is very fast, the observed state will be equal to ($x + \eta$) because it will also be able to follow disturbance oscillations. Since this status then goes into controller k, you may find yourself having a control action even when you don't need it. ($x = 0$ but $\hat{x} = \eta$ and thus $u = k \cdot \eta$!).
One might think of introducing a low-pass filter so that disturbances are eliminated before they enter the observer. In this case, however, it may be convenient to make the observer slower so that the disturbances are naturally filtered out by his "slowness". Furthermore, the speed of the observer is determined not only by the disturbances, but also by the physical characteristics of the real system, which leads us to think that filtering by means of an observer is more efficient than any other filter, whose cut-off frequency is certainly less related to the characteristics of the system. Obviously, in the absence of noise, it is preferable to have a fast observer, but this is a difficult case to find in practice because noise is everywhere.

Details for the observer design

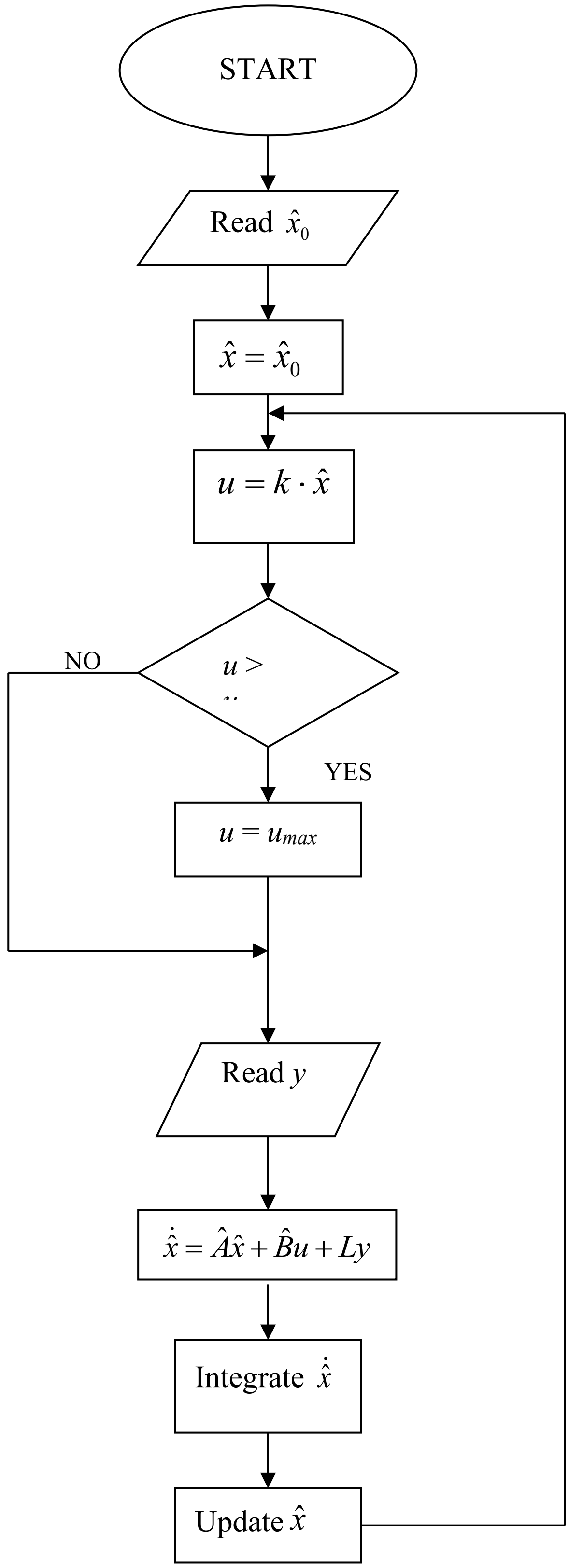
It should be noted that the mathematical model of the real system, which is used to evaluate the system's dynamics, differs from the real behaviour of the system the more simplifying the hypotheses considered are. Its "uncertain" outputs will enter the observer which, on the contrary, is described exactly by the model because it is a fictitious system whose dynamics were chosen by us in the design phase. In turn, the outputs of the observer will enter the gain block k (control proportional to the observed state) which will allow the stabilising intervention on the real system (assuming that the observed state deviates little from the real one). The whole can be schematised in the following block diagram.



Block 1 is the actual physical system while block 2 (the compensating block) is often implemented on the computer. So let's suppose we do the observer on a computer with appropriate software. The block diagram of the observer would become as follows.



The flow chart describing the operations (cyclical and performed at each output measurement, at the appropriately chosen sampling rate) carried out from the moment the compensator block is switched on is shown here below. Note that it also presents a logical control on the intensity of the output in case the limits of the actuators are already known.



As far as actuators and sensors are concerned, we have assumed that they are contained in the representative block of the physical system. If these are not ideal, it will be necessary to introduce their dynamics in order to be able to understand their influence on the dynamics of the system "under control". However, we have already underlined the fact that if the bandwidth of a sensor is considerably wider than that of the real system, it is possible to consider it as an ideal sensor (i.e. without dynamics or, in other words, with zero-order dynamics).

The observer and the real system

So far we have assumed that we have a perfect knowledge of the model of the system to be compensated. There are a number of experimental methods that allow us to refine our knowledge of poorly known systems.

1. Transfer function. It is possible to deduce the transfer function experimentally by relating the system's outputs to suitable known inputs. However, this method is more related to the inputs and outputs and it is difficult to infer the physical characteristics of the system (the system looks like a black box of which we only know the ratio between the input and the outputs; the state of the system is totally unknown).
2. State space model. The state space model contains (if the states are appropriately chosen) the physical parameters of the system (e.g. masses, elastic or viscous friction constants, etc.). If through a (well-designed) observer we "look" at the state of the real system, we can see how close the analytical model of the system (obviously approximated) we used for the design is to that of the real system. In this way it is possible to "calibrate" the analytical model in order to have a better performance of the compensator.

This means that, if the model of a physical system is known (at least approximately), it is possible to do more than just predict its output as a function of a certain input.

Reduced state observer

Let us now suppose that we have a system of which only a portion of the state can be observed. We then partition the system into two blocks

$$\begin{cases} \dot{x}_1 = A_{11}x_1 + A_{12}x_2 + B_1 u \\ \dot{x}_2 = A_{21}x_1 + A_{22}x_2 + B_2 u \\ y = C_1 x_1 \end{cases}$$

Let us assume that x_1 is completely observable while the x_2 portion of the state is not. We have seen that the observer requires a certain amount of computing power. In order to reduce the computational effort of the observer (and thus increase its speed), one could think (for such a system) of designing a "partial" observer that observes only the "invisible" state (x_2). Let's try to see if, proceeding as for the global observer, we reach a good solution. Considering the second block of the model, we write the observer's dynamics for the part x_2 of the system

$$\begin{aligned} &\dot{\hat{x}}_2 = (A_{21} - L_2 C_1)\hat{x}_1 + A_{21}\hat{x}_2 + B_2 u + L_2 y \\ &\rightarrow \hat{x}_1 = {C_1}^{-1} y \\ &\dot{\hat{x}}_2 = A_{21}\hat{x}_2 + B_2 u + A_{21}{C_1}^{-1} y \end{aligned}$$

This is achieved by:
- designing a full-state observer;
- writing the reduced-state observer as a portion of the full-state observer.

It has already been shown that the stability of the observer is guaranteed by making the matrix A-LC (representative of the dynamics of the observer) asymptotically stable through an appropriate choice of the coefficients of the matrix L. Unfortunately, in this case we cannot decide the stability of the reduced order observer because the matrix describing its dynamics is A_{22} and there is no way to modify its poles. It follows that the dynamics of the reduced order observer will inevitably be comparable to that of the true system (A_{22} being a portion of A of the true system). We do not want

the observer's velocity to be fixed. It therefore means that constructing the reduced order observer as a sub-block of the global observer is not the best approach for what we expect to achieve.
In order to shape the dynamics of the reduced order observer we introduce an auxiliary dynamics (dynamics added to the dynamics of the global observer) described by the system

$$\dot{z} = Fz + Gu + Hy$$

Also, assume the following dynamics of the reduced-order observer

$$\dot{\hat{x}}_2 = Ly + z$$

At this point we identify (as done for the full state observer) the dynamics of the observation error

$$\begin{aligned}\dot{e} &= \dot{\hat{x}}_2 - \dot{x}_2 \\ &= L\dot{y} + \dot{z} - \dot{x}_2\end{aligned}$$

we are interested in writing the error in terms of e, x_1 and u. We then substitute

$$\begin{aligned}x_2 &= \hat{x}_2 - e \\ z &= \hat{x}_2 - Ly\end{aligned}$$

After collecting the common coefficients for the different variables present, we obtain

$$\begin{aligned}\dot{e} = {} & (A_{22} - LC_1A_{12})e + (LC_1A_{11} - FLC_1 + HC_1 - A_{21})x_1 + \\ & +(LC_1A_{12} - A_{22} + F)\hat{x}_2 + (LC_1B_1 + G - B_2)u\end{aligned}$$

Now, in analogy to what has been done for the full state observer, we impose the coefficients of x_1, $\hat{x}_2$ and u equal to zero in order to make the error independent from them. This allows us to obtain the three matrices *F, G* and *H* of the auxiliary dynamics and, imposing the zero error at steady state (i.e. the asymptotic stability on the matrix of the coefficients of *e*), the matrix *L* of the observer.

The expressions sought are:

$$\begin{aligned}F &= A_{22} - LC_1A_{12} \\ G &= B_2 - LC_1B_1 \\ H &= FL + (A_{21} - LC_1A_{11}){C_1}^{-1}\end{aligned}$$

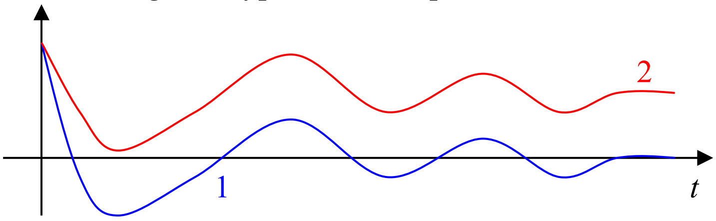
As a result of what has been said, we note that the asymptotic stability is imposed on the matrix *F* (equivalent to the matrix of the coefficients of *e*) or on the dynamics of the auxiliary system (just as before it was imposed on the dynamics of the observer). A summary of the operations to do to deduce the reduced order observer of a system is the following:
1. place the poles of the matrix A_{22} through the matrix *L* (obviously one must always resort to the trick of transposition to adjust things). In this way the matrix *F* characteristic of the dynamics of the introduced auxiliary system is obtained (it must be asymptotically stable).
2. determine the matrices *G* and *H*.

In this way the following can be decided
- the best possible dynamics (by means of *F*) of the system being integrated for observation (*z*);
- the best combination of *z* and *y* (through *G* and *H*) that allows me to obtain the portion of the state that cannot be observed directly (x_2).

The block scheme of the operations to be carried out is analogous to the one of the full state observer with the difference that the dynamics to be integrated is the one of the state z from which then the evolution of $\hat{x}_2$ is deduced.

Optimal control

Let us now look at a different philosophy of control system design. The optimal control, instead of aiming at obtaining only an asymptotically stable system (in closed loop), tries to guarantee stability together with the optimisation of some quantity linked to the system to be controlled. It is therefore clear that "optimal" is not to be understood as "best" but as the possibility of optimising something. The goodness of the control will depend on what you decide to optimise. In classical control theory, the problem is studied through the system's transfer functions (defined in the frequency domain) which allow only one output to be fed back to one input. In the state-space domain, where we were able to bind multiple outputs to multiple inputs by means of state-proportional control. Again, however, when it came to establishing the stability of the system (positioning of the poles), we returned to the frequency domain (where the poles of the system are defined).
Let us now consider the following two hypothetical responses.



From the point of view of oscillations and damping they are indistinguishable. What changes is the value they take on at steady state. Assuming that the correct value of the response is zero, we can also say that response 1 is better than response 2. In the frequency domain, one method of noticing this difference is the final value theorem. However, this only helps us if we are designing a controller in the frequency domain, whereas it cannot help us in the case of state space control. The step forward we now propose is to solve the control problem by reasoning exclusively in the time domain. In this case we could, for example, "adjust" the problems of response 2 by evaluating the integral in a certain time interval and minimising it, i.e. evaluating the gain matrix k of the proportional state control in such a way that this integral is minimised.

$$\int_0^{t_f} x \cdot dt$$

Since we have always considered the state x to be a vector, it is more correct if written as

$$\int_0^{t_f} x^T x \cdot dt$$

where in practice the norm of x is optimised. Minimising this without taking into account the amount of control (u) doesn't make much sense because it would result in an exaggerated control request (through k) which would certainly guarantee an excellent response behaviour but would result in a control expense/performance ratio which would be too high. So let's rewrite it introducing also the inputs (indices of the control expenditure)

$$\int_0^{t_f} (x^T x + u^T u) \cdot dt$$

This is more appropriate. However, it would be nicer to be able to decide which quantity to give more importance to (either x or u). We then arrive at the classical form found for the optimal control:

$$\int_0^{t_f} (x^T Q x + u^T R u) \cdot dt$$

in which Q and R allow us to shift the focus of the minimisation between x and u. If Q is larger than R we favour the response to control cost (high expenditure/performance ratio) and vice versa.

The cost function to minimise

Finally, we need to take into account the end of the control transient (which we assume occurs at time t_f) with a term outside the integral:

$$J_0 = \frac{1}{2} x_f{}^T S x_f + \frac{1}{2} \int_0^{t_f} (x^T Q x + u^T R u) \cdot dt$$

This is the cost function we want to minimise. The matrix S has roughly the same meaning as Q and R and, together with these two, it guides us in the process of minimisation. Let us try to deduce some properties of the above matrices:
- they are certainly square;
- they have no reason not to be symmetric.

There are two reasons for them to be symmetrical: one mathematical and one, more important, of a physical nature. We write $u^T R u$ by splitting u down into two elements:

$$[u_1 \quad u_2] \begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = R_1 u_1{}^2 + (R_2 + R_3) u_1 u_2 + R_4 u_2{}^2$$

As can be seen, what ultimately matters in the mixed product $u_1 u_2$ is the sum of the two extra-diagonal terms. Therefore, from a mathematical point of view, there is no reason for R_2 and R_3 to be different. From a physical point of view, optimising by considering R_2 different from R_3 would mean imposing, on the same contribution ($u_1 u_2$), two different weights. We must therefore minimise J_0. Note that the minimisation cannot be done freely because there is a link between x and u (the equations of dynamics). We take this link into account through appropriate multipliers. Then the complete expression to be minimised is

$$\min_u J = \frac{1}{2} x_f{}^T S x_f + \frac{1}{2} \int_0^{t_f} [x^T Q x + u^T R u + 2\lambda^T (Ax + Bu - \dot{x})] \cdot dt$$

Note that we minimise in terms of u since it is the quantity we can use to intervene in the system. We must then impose grad(J) = 0. In other words, we will have to derive J with respect to the different variables (x, u and λ) and set all these derivatives equal to zero. The different terms of grad(J) are as follows:

$$\frac{\partial J}{\partial \lambda} = \frac{1}{2} \int_0^{t_f} 2 \frac{\partial}{\partial \lambda} [\lambda^T (Ax + Bu - \dot{x})] dt = 0$$

$$\rightarrow \frac{\partial}{\partial \lambda} [\lambda^T (Ax + Bu - \dot{x})] = 0$$

we derive for λ^T. We then obtain the constraint imposed by the equations of dynamics

$$\dot{x} = Ax + Bu$$

So the first condition we have found is that the minimum of J can exist as long as the constraint given by the equations of dynamics is respected.

$$\frac{\partial J}{\partial u} = \frac{1}{2}\int_0^{t_f} \frac{\partial}{\partial u}[u^T R u + 2\lambda^T B u]dt = 0$$
$$\rightarrow \frac{\partial}{\partial u}[u^T R u + 2\lambda^T B u] = 0$$

Now, derive with respect to u^T, to obtain

$$2B^T\lambda + 2Ru = 0$$
$$\rightarrow u = -R^{-1}B^T\lambda$$

A clarification is now necessary with regard to the R-matrix. For it to be invertible, it must be positive definite (we have already seen that it will also be symmetric). To say that R is positive definite means to say that in the cost functional J the contribution of u (i.e. of the control cost) is always present and is only null when u is null (i.e. when there is no control). This mathematical requirement has a very precise physical meaning: if one wants to impose a control on the system (through a certain u) it is indispensable to put a limit to this control (which is done for sure because, once u is there, having to be R positive definite, the control action is surely minimised). We note that anyway u is not known because it depends on λ (which we do not know). Note also that since R and B are constant (linear time-invariant systems), it will inevitably be variable (otherwise the control u would be constant and would not make much sense).
The last term of the gradient of J is given by:

$$\frac{\partial J}{\partial x} = \int_0^{t_f} \frac{\partial}{\partial x}\left[\frac{1}{2}x^T Q x + \lambda^T A x + \dot{\lambda}^T x\right]dt = 0$$
$$\rightarrow \frac{\partial}{\partial x}\left[\frac{1}{2}x^T Q x + \lambda^T A x + \dot{\lambda}^T x\right] = 0$$

where we have made an integration by parts to eliminate the term in x

$$\int_0^{t_f} \lambda^T(-\dot{x})dt = [-\lambda x]_0^{t_f} + \int_0^{t_f} \dot{\lambda}^T x dt$$

And the term outside the integral sign ($-\lambda x_f + \lambda x_0$) is constant and is not derivated (as done for the term x_f in J).
To obtain a meaningful condition, we transpose the term in brackets and derive for x^T instead of x. We obtain the following condition:

$$\dot{\lambda} + A^T\lambda + Qx = 0$$
$$\rightarrow \dot{\lambda} = -A^T\lambda - Qx$$

In summary:

$$\dot{x} = Ax + Bu$$
$$u = -R^{-1}B^T\lambda$$
$$\dot{\lambda} = -A^T\lambda - Qx$$

It is remarked that the third equation describes the dynamics of the multipliers which are directly linked to the state. The second one instead allows to obtain, indirectly through λ, the control action u

to be exercised on the system. If we manage to integrate the third equation we have solved the control problem.
We have seen that R must be positive definite. As far as Q is concerned, however, it can also be the case that it is positive semi-definite. The meaning is that one can accept large values of the state (because if Q is positive semi-definite, it will have null terms that will prevent the whole state from being minimised). The important point is to have a limited control action (and this is guaranteed by R).

The Riccati equation

Continuing with the optimal control, the system we have to solve is

$$\begin{Bmatrix} \dot{x} \\ \dot{\lambda} \end{Bmatrix} = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix} \begin{Bmatrix} x \\ \lambda \end{Bmatrix}$$

The problem with what is obtained is that λ is not defined before completing the optimisation. To circumvent the problem we can impose a constraint on the multipliers λ

$$\lambda = Px$$

In this way, an approximation is made by saying that the multipliers λ are directly proportional to the state of the system. It is also possible to write a direct link between the state x and the control action u

$$u = -R^{-1}B^T Px$$

Rewrite J_0 substituting the expression for u.

$$J_0 = \frac{1}{2} x_f{}^T S x_f + \frac{1}{2} \int_0^{t_f} x^T (Q + P^T B R^{-1} B^T P) x dt$$

We must now determine P. Substituting then $\lambda=Px$ in the equation describing the dynamics of λ, we obtain an equation describing the dynamics of P which contains inside it also the constraint given by the dynamics of x

$$\begin{gathered} \dot{x} = Ax + Bu \\ \dot{P}x + P\dot{x} = -APx - Qx \\ u = -R^{-1}B^T Px \\ \dot{P} = -A^T P - Q - PA + PBR^{-1}B^T P \end{gathered}$$

The differential equation for matrix P is known as Riccati equation. It is the solution of this equation that allows us to solve the control problem. The solution P is a symmetric and positive definite matrix. Note that if P is a solution of the Riccati equation, it is guaranteed that the controlled system with the resulting k is stable.

<u>Boundary conditions for the integration of the Riccati equation</u>
The Riccati equation, being differential, requires boundary conditions. It also suggests to us that P may vary in time and so we may have, at least in principle, a variable gain. We are interested in

relating the initial and final values of P to the other parameters involved in our equations. We start from the following identity:

$$x_f{}^T P_f x_f - x_0{}^T P_0 x_0 = \int_0^{t_f} \frac{d}{dt}(x^T P x) dt$$

Let's take the time derivative of the term in brackets

$$\frac{d}{dt}(x^T P x) = \left(\dot{x}^T P x + x^T \dot{P} x + x^T P \dot{x}\right)$$

The three terms can be expressed in terms of A, B, P, Q and R leading to

$$\frac{d}{dt}(x^T P x) = -x^T (Q + PBR^{-1}B^T P) x$$

Recalling the expression of J_0, the following holds:

$$J_0 = \frac{1}{2} x_f{}^T S x_f - \frac{1}{2} x_f{}^T P_f x_f + \frac{1}{2} x_0{}^T P_0 x_0 = \frac{1}{2} x_f{}^T \left(S - P_f\right) x_f + \frac{1}{2} x_0{}^T P_0 x_0$$

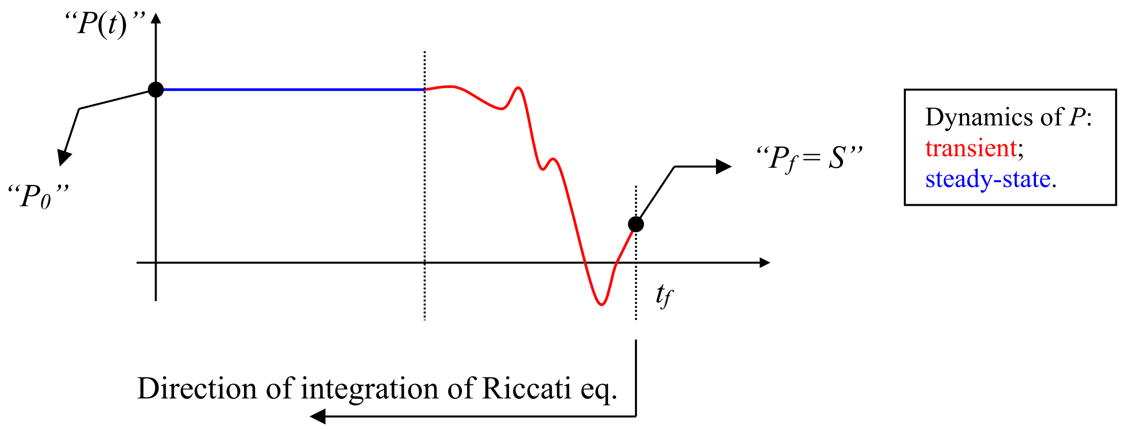
Since our aim is to minimise J_0 and since we know that P is symmetric and positive definite (so the term P_0 can never be negative) we deduce that the most favourable condition is that:

$$P_f = S$$

So these are the boundary conditions we were looking for the solution of the Riccati equation.

Integration of the Riccati equation

In the decision of the different weights of the optimisation, we can also decide the duration of the transient by imposing the time t_f (indeed, it is necessary that we assume for it a value). To obtain the $P(t)$ we integrate backwards in time the Riccati equation from the time t_f to the time t. This means that P, evolving backwards in time, will have a transient between t_f and a slightly earlier time (as in the following figure) and then it will be constant at its value P_0.



In practice it is often assumed that t_f is larger than the characteristic time of the transient of the system under control. The reason is that in this way the transient of P can be neglected (because by now the system to be controlled will already be at steady state) and it can be assumed that P is constant. This hypothesis allows to simplify considerably the calculations since it's no more necessary to integrate the Riccati equation but it's enough to solve it algebraically (in this case it's called Algebraic Riccati Equation). If we assume a sufficiently large t_f, the cost function can be also written:

$$J = \int_0^\infty (x^T Q x + u^T R u) dt$$

where steady state values of x and u are zero, so the term x_f vanishes.

Alternatives to optimising the system state (standard problem)

For some systems it may not be the best thing to optimise the behaviour of state x. For example in those cases where the physical meaning of the state is not very clear and therefore it is not possible to identify well what to optimise. An alternative may be the following:

$$J = \int_0^\infty (y^T Q y + u^T R u) dt$$

where the decision was made to optimise the output instead of the system state (obviously the limit on the control action is always there). This is a transformation of the standard problem that we have already seen and solved. The cases we can run into in solving the new problem depend on the system which may or may not be proper (i.e. D may or may not be null). If $D = 0$ the problem is easily solved because:

$$\begin{aligned} &y = Cx \\ &\rightarrow y^T Q y = x^T (C^T Q C) x \end{aligned}$$

so that matrix Q of the standard problem is substituted by C^TQC, then the optimal control problem is solved to obtain k. If $D \neq 0$ the problem is slightly more complex. In fact the cost function becomes

$$J = \int_0^\infty [x^T (C^T Q C) x + u^T (D^T Q D + R) u + x^T C^T Q D u + u^T D^T Q C x] dt$$

Note that the mixed terms are each the transpose of the other and, since they give a scalar (doing all the products), they will be the same number. Therefore we also collect the mixed contribution and rewrite the cost function as follows:

$$\begin{aligned} &J = \int_0^\infty [x^T \bar{Q} x + u^T \bar{R} u + 2 x^T \bar{W} u] dt \\ &\bar{Q} = C^T Q C \\ &\bar{R} = D^T Q D + R \\ &\bar{W} = C^T Q D \end{aligned}$$

We observe that the aspect of this functional is quite general (dealing with linear systems) because it contains two "pure quadratic" terms and one "mixed quadratic" term. Through this functional (and through the weight matrices) we can decide to optimise not only the state and the control action, but also the interaction between the two (which could also have, for some systems, an interesting physical meaning). We can anyway go back to an expression analogous to the standard one even if the mixed term is present. Let's suppose that the control u is composed of two terms (i.e. we make a change of variables):

$$u = \bar{u} - \bar{R}^{-1}\bar{W}^T x$$

If we replace this expression in the cost function, we obtain

$$J = \int_0^\infty \left(x^T \bar{\bar{Q}} x + \bar{u}^T \bar{\bar{R}} \bar{u}\right) dt$$
$$\bar{\bar{Q}} = \bar{Q} - \bar{W}\bar{R}^{-1}\bar{W}^T$$
$$\bar{\bar{R}} = \bar{R}$$

So we managed to show that, in the seen cases, the problem of the optimal control can be always recast as the standard problem. The solution of the problem however is not yet finished because the matrix of the gains that we obtain is now relative to a system different from the initial one (because of the change of variables). In fact the solution reads

$$\bar{u} = -\bar{\bar{R}}^{-1} B^T \bar{\bar{P}} x$$

which provides the control $\bar{u}$ and not the one of the initial system. In other words, we obtain the following system

$$\dot{x} = Ax + B\bar{u} - B\bar{R}^{-1}\bar{W}^T x = \bar{\bar{A}} x + \bar{\bar{B}} \bar{u}$$
$$\bar{\bar{A}} = A - B\bar{R}^{-1}\bar{W}^T$$
$$\bar{\bar{B}} = B$$

The gain matrix needs to be transformed to fit the original system. For this it's sufficient to do the inverse change of variables, leading to

$$u = \cdots = \left(-\bar{\bar{R}}^{-1} \bar{\bar{B}}^T \bar{\bar{P}} - \bar{R}^{-1}\bar{W}^T\right) x$$

This shows the value of k to be applied to the state in order to obtain the required optimum control of the initial system.

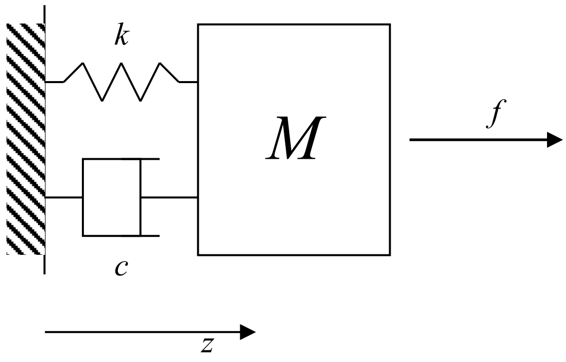
The important thing is to have seen that it is also possible to introduce in the functional J a mixed term (allowing to optimise in some way the contribution of x and u together) because it is always possible to go back to the standard form.

Criteria for the definition of weights Q and R

Having seen how to solve the optimal control problem from an analytical point of view (by solving the Riccati equation), we need to see how to define the optimisation weight matrices.

System energy

Let us look at an example. Consider the spring-mass-damper system in the figure.



Its dynamics is described by the equations

$$\begin{Bmatrix} \ddot{z} \\ \dot{z} \end{Bmatrix} = \begin{bmatrix} -\frac{c}{M} & -\frac{k}{M} \\ 1 & 0 \end{bmatrix} \begin{Bmatrix} \dot{z} \\ z \end{Bmatrix} + \begin{bmatrix} \frac{1}{M} \\ 0 \end{bmatrix} \{f\}$$

The total energy of the system is composed of the kinetic energy of the mass plus the elastic energy of the spring

$$E = \frac{1}{2} M \dot{z}^2 + \frac{1}{2} k z^2$$

This expression can also be written in matrix form

$$E = [\dot{z} \quad z] \begin{bmatrix} \frac{M}{2} & 0 \\ 0 & \frac{k}{2} \end{bmatrix} \begin{bmatrix} \dot{z} \\ z \end{bmatrix}$$

which looks just like the term in x of the cost function J ($x^T Q x$). Selecting then the matrix Q equal to the matrix multiplying the state vector in the expression of the energy, the optimal control guarantees the minimization of the total energy of the system. Setting then R = 1 it's possible to optimise (minimise) the energy spent for the control of the system. Wanting to be more general, one can consider a $R = \rho$ in order to move more or less weight on the minimization of the control expenditure.

Unfortunately, it is not always so easy to find a possible expression of Q and R. Remaining in fact in the case of the total energy of the system, it could be possible to have a non-linear expression of E so that, in order to write it in matrix form, it will be necessary to linearise it and then write a Q that will be close to the matrix of E only if we are close to the position of linearisation.

Limiting the maximum values of states and control

We may wish to optimise control according to a different criterion. For example, we may wish to establish the "freedom of movement of states" and/or controls by imposing the maximum value they can assume. This can be achieved by defining the matrices Q and R as follows.

$$Q = diag\left(\frac{1}{{x_{iMax}}^2}\right); \quad R = diag\left(\frac{1}{{u_{iMax}}^2}\right)$$

In this way, for the states that have more freedom of movement (and for the controls that can intervene "more") the weight will be lower than that of the states that are more "constrained" and for this reason the optimal control will guarantee a more energetic action on those more limited states. Weights of this kind can be used also try to compare quantities with different dimensions (or when different units of measure are used for quantities of the same nature). With this method it's then possible to control a system without knowing anything about its energy but giving exclusively more or less importance to its states. It should be noted that a disadvantage of this type of weight is the impossibility of knowing the behaviour of the system during the transient. For example, there may be too many oscillations. To solve this problem, the system's response must be analysed once a first solution has been computed (and therefore in closed loop) and, in the event of a poor response, Q and R must be adjusted and the response checked again.

Emulating the behavior of a target system

Another interesting method of determining Q and R depends on a good knowledge of the mathematical model of the system to be controlled. Assuming that the open loop system is described by the generic model:

$$\dot{x} = Ax + Bu$$

It could be interesting to make this system behave like another similar system but described by equations with different coefficients (i.e. behaving better)

$$\dot{x}_m = A_m x + B_m u$$

where obviously matrices A_m and B_m will be different from A and B. What you want to optimise (minimise) then is the difference between these two systems, i.e.

$$e = \dot{x} - \dot{x}_m = (A - A_m)x + (B - B_m)u$$

To obtain the minimisation of e, it is necessary to define the cost function as follows

$$J = \int_0^\infty (e^T Q e + u^T R u) dt$$

The new cost function (replacing the expression of e as a function of x and u) is given by

$$\begin{aligned} J = \int_0^\infty (x^T[(A - A_m)^T Q (A - A_m)]x + \\ +u^T[(B - B_m)^T Q (B - B_m) + R]u + \\ +2x^T[(A - A_m)^T Q (B - B_m)]u) dt \end{aligned}$$

With this we can minimize the "distance" between the behaviour of the real system and the one of the system we chose as model. Note that in this case, we could also consider the initial R as null because the term in $(B - B_m)$ appears, which guarantees anyway a constraint on the control action. The meaning of $R = 0$ in this case is that, since there is a constraint on the action u anyway, one wants to bring the two systems closer "whatever it takes". In this type of optimal control it is noted that the weight matrices Q and R play a secondary role (because the main purpose is to "bring" the two systems as close as possible) and one could take both equal to the identity matrix (i.e. giving the same weight to all errors on the state). In this way the weight matrices will depend exclusively on the difference between the model of the true system and the model that we would like the system to follow.

Minimum guaranteed damping

To guarantee to the closed loop system a certain degree of damping it is possible to modify the cost function *J*. The modification to introduce is:

$$J = \int_0^\infty e^{2\alpha t}(x^T Q x + u^T R u)dt$$

Minimizing this cost function means being able to subtract from the bracketed term a contribution opposite to $e^{2\alpha t}$ which results in a damping of the state (and of the control). We notice however that this cost function has a different form from the one of the standard problem and, as far as we have seen up to now, it is not even equal to the form containing the mixed term. To bring it back to the standard problem, we make the following substitution:

$$x = e^{-\alpha t}\bar{x}$$
$$u = e^{-\alpha t}\bar{u}$$

It is clear that this transformation must also be taken into account in the equations of the system's dynamics. The equations that describe the dynamics of the new state *x* are

$$\dot{\bar{x}} = \alpha e^{\alpha t} x + e^{\alpha t}\dot{x}$$

That can be also written as

$$\dot{\bar{x}} = (\alpha I + A)\bar{x} + B\bar{u}$$

With the appropriate change of state and control variable, the cost function assumes the well known standard form. The solution (matrix *k*) is then obtained considering the new state

$$\bar{u} = k\bar{x}$$

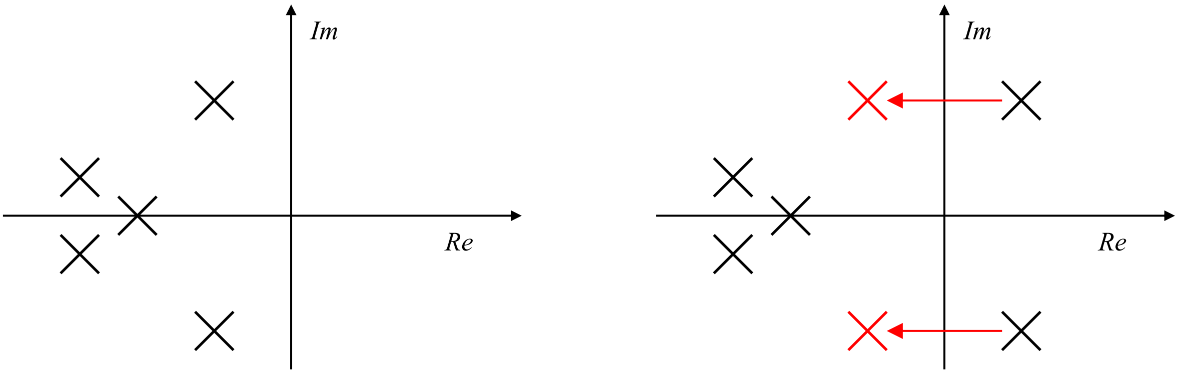
In this case, however, returning to the original state is very simple because both *x* and *u* undergo the same transformation and, in applying the inverse transformation, the contribution of the exponentials cancels and we obtain therefore the same matrix *k* obtained for the "auxiliary" state

$$u = kx$$

The gain matrix *k* obtained applying the optimal control to the modified system is then the same that we have to consider for the true system. The solution guarantees the stability of the state $\bar{x}$ because it is obtained by solving a Riccati equation. It follows that, if this state is stable, so will be the state *x* (given by $e^{-\alpha t}\bar{x}$) which will have, in addition, an exponential decay of order α (in the worst case in which $\bar{x}$ is constant). This method therefore allows us to impose a minimum damping on the state of the system by simply modifying (in practice) the matrix *A* by adding to it the contribution αI where α is the desired degree of exponential damping.

Closed-loop stability

We have already said that, through optimal control, it is not possible to know where the poles of the closed-loop system will be placed. Let's see what happens (qualitatively speaking) to the poles of a system in the case of $Q = 0$ and $R > 0$. In particular, we want to compare a system with stable poles with one with unstable poles. The following figure summarises the consequences in the two cases.

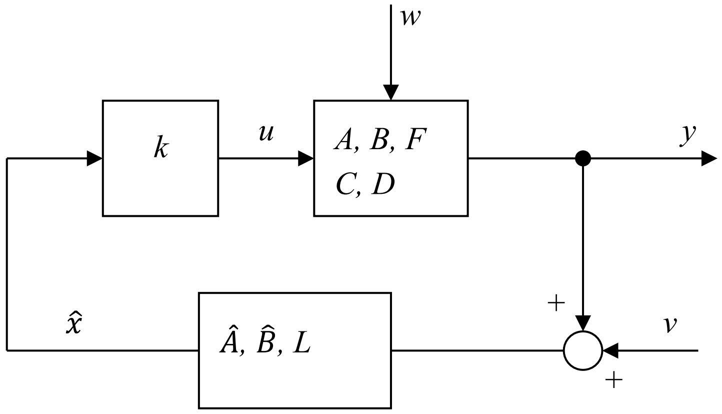


Stable open loop system: since there are no constraints on the movement of the state, free movement is already fine and therefore the smallest control that can be applied is $u = 0$!

Unstable open loop system: since the solution of the control problem guarantees a stable system, the fact of accepting any movement of the state as long as it is stable, implies a control u that places the closed-loop poles in symmetrical position of the unstable poles with respect to the *Im* axis.

Optimal observer

We have seen how to design (by placing the poles) any linear observer that allows us to see the whole state of a physical system. If we are in the presence of disturbances, the concept of an optimal observer may arise. The optimisation of the observer can be done instead in the time domain, following a method similar to the one seen for optimal control. Let's consider the following scheme, representative of the situation we want to study:



Assume the following model for the system, where both states and measurements are corrupted by noise

$$\begin{cases} \dot{x} = Ax + Bu + Fw \\ y = Cx + v \end{cases}$$

which is particularly simple when w and v are white noise. The matrix F is an analogue of B for the disturbances w. The observer, assuming it is represented by a linear system, will be described by

$$\begin{cases} \dot{\hat{x}} = \hat{A}\hat{x} + \hat{B}u + Ly \\ \hat{y} = \hat{x} \end{cases}$$

where $\hat{A}$ and $\hat{B}$ are of appropriate dimensions. Write the dynamics of the error between the true state and the observed state

$$\dot{x} - \dot{\hat{x}} \equiv \dot{e} = Ax + Bu + Fw - \hat{A}\hat{x} - \hat{B}u - Ly$$

and, noticing that

$$e = x - \hat{x} \quad \rightarrow \quad \hat{x} = x - e$$
$$y = Cx + v$$

reformulate the error dynamics as

$$\dot{e} = \left(A - LC - \hat{A}\right)x + \left(B - \hat{B}\right)u + \hat{A}e + Fw - Lv$$

By analogy with the development of the linear observer, in order to make the dynamics of e independent of x and u we will impose

$$\hat{A} = A - LC;$$
$$\hat{B} = B$$

Then

$$\dot{e} = (A - LC)e + Fw - Lv$$

Assuming now that w and v are white noises, we can say that their algebraic sum will be another white noise and so we'll pose:

$$\eta = Fw - Lv$$

such that

$$\dot{e} = (A - LC)e + \eta$$

Note that noise η is continuously forcing the observation error. One can then think of applying an optimisation (i.e. a minimisation) on the variance of e. In this way, the error, although forced by the noise w and v (and thus η) will be distributed around its mean value (hopefully equal to zero) with a minimum variance (i.e. all the values it takes on can be considered very close to the mean value = 0). For this, it is necessary to calculate the statistical indicators of e. The optimisation can then be applied. We will then look for the matrix L that allows us to minimise the variance of e, which, in general, will also be a matrix. Let us call the variance

$${\sigma_{ee}}^2 = P$$

The dynamics of the variance is represented by the Lyapunov equation

$$\dot{P} = (A - LC)P + P(A^T - C^T L^T) + Q_\eta$$

where Q_η is the variance of η (${\sigma_{\eta\eta}}^2$) that is given by

$$Q_\eta = \int \big(Fw(t) - Lv(t)\big)\big(Fw(t) - Lv(t)\big)^T dt$$

Knowing the variance matrices W, V associated to the noise signals w and v , and the cross-variance X, Q_η is computed as:

$$Q_\eta = FWF^T + LVL^T - FXL^T - LX^T F^T$$

In most cases v and w are considered uncorrelated, that means that the cross-variance matrix X is zero. The Lyapunov equation then becomes

$$\dot{P} = (A - LC)P + P(A^T - C^T L^T) + FWF^T + LVL^T - FXL^T - LX^T F^T$$

This equation has the same structure as the Riccati equation used to solve the optimal control problem. We then look for a solution by carrying out an analogy with the results obtained for optimal control. We observe in particular that the solution of the optimal problem in the case of the control was provided by:

$$k = -R^{-1} B^T P$$

P is the solution of the Riccati equation. Let us then try to relate this to the solution we are looking for (i.e. the matrix L that optimises the control by minimising the variance of the error). Let us suppose that the solution for L is given by an expression analogous to the one of the optimal control

$$L^T = V^{-1} CP$$

in which the following analogy is implemented

$$-R \leftrightarrow V$$
$$B^T \leftrightarrow C$$

Note that *L* is transposed because, while *k* multiplies the state of the system (and so has a number of columns equal to the number of states), *L* will have the number of rows equal to the number of states of the system (because $y = Cx + v$). If we also assume that X is null, we obtain, after a simplification:

$$\dot{P} = AP + PA^T + FWF^T - PC^TV^{-1}CP$$

which can is perfectly equivalent to Riccati's equation if we consider the following associations

$$-A^T \leftrightarrow A$$
$$-Q \leftrightarrow FWF^T$$

To find the *L* that optimises the observation, I can then (assuming $X = 0$) solve the Riccati equation and then substitute *P* in the $L^T = V^{-1}CP$.

The term *X* can be seen as the cross term in the optimal control problem. Another way to take it into account is to see it as a perturbation of the optimal condition. Suppose, for example, that we decompose *P* and *L* into the optimal solution and perturbation:

$$P = \hat{P} + U$$
$$L = \hat{L} + Z$$

where *U* and *Z* are the perturbations of the solutions, respectively, $\hat{P}$ ed $\hat{L}$. The equation obtained by substituting these into the Lyapunov equation is made up of two groups of separable terms and, putting these two groups in order, looks as follows:

$$\begin{aligned}\dot{\hat{P}} + \dot{U} = {} & \left(A - \hat{L}C\right)\hat{P} + \hat{P}\left(A^T - C^T\hat{L}^T\right) + FWF^T + \hat{L}V\hat{L}^T - FX\hat{L}^T - \hat{L}X^TF^T + \\ & +\left(A - \hat{L}C - ZC\right)U + U\left(A^T - C^T\hat{L}^T - C^TZ^T\right) + ZVZ^T + \\ & -Z\left(C\hat{P} - V\hat{L}^T + X^TF^T\right) - \left(\hat{P}C^T - \hat{L}V + FX\right)Z^T\end{aligned}$$

Since $\hat{P}$ is, by definition, solution of the equation

$$\dot{\hat{P}} = \left(A - \hat{L}C\right)\hat{P} + \hat{P}\left(A^T - C^T\hat{L}^T\right) + FWF^T + \hat{L}V\hat{L}^T - FX\hat{L}^T - \hat{L}X^TF^T$$

they cancel out from the above equation. Furthermore, the terms $Z\left(C\hat{P} - V\hat{L}^T + X^TF^T\right)$ and $\left(\hat{P}C^T - \hat{L}V + FX\right)Z^T$ need to be zero since they are not symmetric and would not be compatible with the definition of Riccati equation. Therefore, the following must hold

$$\hat{P}C^T - \hat{L}V + FX = 0$$
$$\rightarrow \hat{L} = \left(\hat{P}C^T + FX\right) \cdot V^{-1}$$

This provides the solution for $\hat{L}$ as function of $\hat{P}$. Note that this solution reverts to the previous one in case *FX* is missing. This guarantees that *U* is a solution of a Riccati equation (represented by the remaining terms of the equation). Consequently, *U* will surely be positive definite and this implies that $\hat{P}$ is really a minimum (because $\hat{P} + U$ will surely be greater) and therefore an optimal solution (as we had originally assumed). What we have done so far shows that if we are able to solve the optimal control problem, we can, by applying the appropriate substitutions, solve in the same way the optimal observer problem.

Notes on the optimal observer

Let us see the differences in the behaviour of this observer with respect to the classical designed observer (i.e. L evaluated by positioning the poles of A). Let us summarise the expressions obtained that allow us to design the optimal observer (in the case where $X = 0$).

$$\begin{cases} \dot{\hat{x}} = \hat{A}x + \hat{B}u + Ly \\ \hat{y} = \hat{x} \end{cases}$$

$$\hat{A} = A - LC$$

$$\hat{B} = B$$

$$L = V^{-1}CP \quad \leftarrow \quad \dot{P} = AP + PA^T + FWF^T - PC^TV^{-1}CP$$

The good performances of the optimal observer in the presence of disturbances is due to the dependence of L on V^{-1}. This is seen from the following considerations.

Very noisy measurements (V with large elements)	Low noise measurements (V with small elements)
In this case, L will be small because it derives from the product of CP with V^{-1} which has small values. It follows that the observer will have a dynamics similar to that of the model considered for the observed system ($\hat{A} \to A$): "…the observer "trusts poorly" the measurements (because the contribution of y to his dynamics is damped by the small L) and relies more on the model we have assumed for the real system …"	In this case, L will be quite large. Thus the dynamics of the observer will be quite different from that of the model of the observed system. ($\hat{A} \neq A$): "…the observer "trusts greatly" the y-measurements and because of this it can be faster than the physical system providing a better observation …"

If we look at what has been said now, it seems that the contribution of noise w is not taken into account. This is actually not true because the matrix W (of the variances of the noises on the physical system) is contained within the Lypunov equation (which we then solve in P as if it were a Riccati equation). Therefore we can say that their contribution enters, indirectly, through the matrix P of the variances of the observation error. Bearing in mind that it is necessary to invert the matrix V, it will be appropriate to have sensors with similar "noise". In this way V will be well conditioned and will not present problems in the inversion.

Example of control system based on inertia and reaction wheels

We can now consider one example of control system with one IW along the pitch axis (z) and two RW along the roll and yaw axes (x and y). The Euler equations are in this case:

$$\begin{cases} I\underline{\dot{\omega}} + \underline{\omega} \wedge I\underline{\omega} + \underline{\dot{h}}_r + \underline{\omega} \wedge \underline{h}_r = \underline{M}_d \\ \underline{\dot{h}}_r = \underline{M}_c \end{cases}$$

We have 6 equations and each rotor is along one of the principal axes, so A is an identity matrix. Assuming we want to control the attitude and not the angular velocity, we must add equations for the attitude kinematics:

$$f(\underline{\alpha}, \underline{\dot{\alpha}}, \underline{\omega}) = 0$$

We can consider the case for which we refer the attitude to the LVLH frame. The nominal condition is a constant angular velocity around the pitch axis, one rotation per orbit, and a nominal angular velocity for the IW. The remaining nominal parameters are zero. The kinematic equations are then:

$$\begin{Bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{Bmatrix} = \begin{bmatrix} 1 & \alpha_z & -\alpha_y \\ -\alpha_z & 1 & \alpha_x \\ \alpha_y & -\alpha_x & 1 \end{bmatrix} \begin{Bmatrix} \dot{\alpha}_x \\ \dot{\alpha}_y \\ \dot{\alpha}_z + n \end{Bmatrix}$$

Since along x and y we have RW, $h_{rx}=h_{ry}=0$ in nominal conditions. The linearized equations are:

$$\begin{cases} I_x(\ddot{\alpha}_x - \dot{\alpha}_y n) + (I_z - I_y)(\mathrm{n}\dot{\alpha}_y + n^2\alpha_x) + \dot{h}_{\mathrm{rx}} + h_{\mathrm{rz}}(\dot{\alpha}_y + \mathrm{n}\alpha_x) = M_{\mathrm{dx}} \\ I_y(\ddot{\alpha}_y + \dot{\alpha}_x n) + (I_x - I_z)(\mathrm{n}\dot{\alpha}_x - n^2\alpha_y) + \dot{h}_{\mathrm{ry}} - h_{\mathrm{rz}}(\dot{\alpha}_x - \mathrm{n}\alpha_y) = M_{\mathrm{dy}} \\ I_z\ddot{\alpha}_z + \dot{h}_{\mathrm{rz}} = M_{\mathrm{dz}} \end{cases}$$

$$\begin{cases} \dot{h}_{rx} = -M_{cx} \\ \dot{h}_{ry} = -M_{cy} \\ \dot{h}_{rz} = -M_{cz} \end{cases}$$

Now we reformulate the problem inserting the control, getting to the equations:

$$\begin{cases} I_x(\ddot{\alpha}_x - \dot{\alpha}_y n) + (I_z - I_y)(\mathrm{n}\dot{\alpha}_y + n^2\alpha_x) + h_{\mathrm{rz}}(\dot{\alpha}_y + \mathrm{n}\alpha_x) = M_{\mathrm{cx}} \\ I_y(\ddot{\alpha}_y + \dot{\alpha}_x n) + (I_x - I_z)(\mathrm{n}\dot{\alpha}_x - n^2\alpha_y) - h_{\mathrm{rz}}(\dot{\alpha}_x - \mathrm{n}\alpha_y) = M_{\mathrm{cy}} \\ I_z\ddot{\alpha}_z = M_{\mathrm{cz}} \end{cases}$$

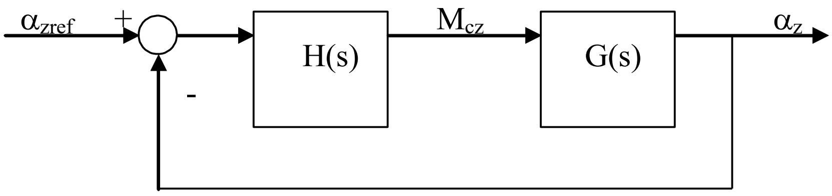
$$\underline{\dot{h}}_r = -\underline{M}_c$$

The design of the controller can be based on classical transfer function techniques. For the pitch axis (z) the transfer function is

$$s^2\alpha_z(s) = \frac{M_{cz}}{I_z}$$

$$G_z(s) = \frac{\alpha_z(s)}{M_{cz}} = \frac{1}{I_z s^2}$$

The block diagram of the controller is:



The closed loop transfer function becomes:

$$F(s) = \frac{H(s)G(s)}{1 + H(s)G(s)} = \frac{\alpha_z(s)}{\alpha_{\text{zref}}}$$

In order to define H(s), we should first define the performances required by F(s). In addition to the above scheme, we introduce the transfer function between a disturbance torque and the attitude angle, that is:

$$F_d(s) = \frac{G(s)}{1 + H(s)G(s)} = \frac{\alpha_z(s)}{T}$$

To evaluate H(s), we can assume a PD-like control, since the system dynamics is of second order:

$$M_{cz} = K_d s\alpha + K_p\alpha$$
$$H(s) = \frac{M_{cz}}{\alpha_z(s)} = \left(K_d s + K_p\right)$$

We then have:

$$F(s) = \frac{\left(K_d s + K_p\right)\frac{1}{I_z s^2}}{1 + \frac{\left(K_d s + K_p\right)}{I_z s^2}} = \frac{\frac{\left(K_d s + K_p\right)}{I_z s^2}}{1 + \frac{\left(K_d s + K_p\right)}{I_z s^2}} = \frac{K_d s + K_p}{I_z s^2 + K_d s + K_p}$$

This is actually a second order system characterized by:

$$\omega = \sqrt{\frac{K_p}{I_z}}$$
$$2\xi\omega = \frac{K_d}{I_z}$$

ω is the closed loop natural frequency of the pitch oscillations and ξ is the damping. The ideal damping would be around 0.7, but in order to reduce the control action a damping in the order of 0.3 can be accepted. The natural frequency can be related to the orbital frequency n, imposing that steady-state is reached within a specific fraction of orbit period. A choice of $\omega = 20n$ is commonly accepted as a good solution. With these assumptions, the steady state error angle can be evaluated by using the final value theorem:

$$\alpha_{z\infty} = \lim_{s\to 0} s\mathrm{F}_d\,(s)d(s)$$

where d(s) is the Laplace transform of the disturbing torque. Maximizing the disturbance, considering it as constant, we have:

$$d(s) = \frac{1}{s}$$

$$\alpha_{z\infty} = \lim_{s\to 0} F_d(s) = \frac{\frac{1}{I_z s^2}}{1 + \frac{\left(K_d s + K_p\right)}{I_z s^2}} = \frac{1}{I_z s^2 + K_d s + K_p} = \frac{1}{K_p}$$

This relation can also be used to tune the proportional gain in order to reach a specified steady state error. The performance of the IW can then be analyzed:

$$\underline{\dot{h}}_r = \underline{M}_c$$

Taking the Laplace transform of the IW with the control:

$$I_r s\omega_r = \left(K_d s + K_p\right)\alpha(s)$$

The above equation can be used to check if the IW reaches saturation.

To study the control for x and y axes we can make some simplifying assumptions in order to design the controller with Laplace transform techniques. Consider M_{cx}=0, and assume only α_y is measured, as could be done with a fixed head Earth sensor that does not provide α_x:

$$\begin{cases} I_x\left(s^2\alpha_x - s\alpha_y n\right) + \left(I_z - I_y\right)\left(ns\alpha_y + n^2\alpha_x\right) + h_{\mathrm{rz}}\left(s\alpha_y + n\alpha_x\right) = 0 \\ I_y\left(s^2\alpha_y + s\alpha_x n\right) + (I_x - I_z)\left(ns\alpha_x - n^2\alpha_y\right) - h_{\mathrm{rz}}\left(s\alpha_x - n\alpha_y\right) = M_{\mathrm{cy}} \end{cases}$$

The transfer function is then:

$$\begin{aligned}
&G_y(s) = \frac{\mathrm{A}s^2 + B}{\mathrm{C}s^4 + \mathrm{D}s^2 + E} \\
&A = I_x \\
&B = \left(I_z - I_y\right)n^2 + h_{\mathrm{rz}}n \\
&C = I_x I_y \\
&D = \left\{n^2 I_y\left(I_z - I_y\right) + I_y h_{\mathrm{rz}}n - I_x(I_x - I_z)n^2 - I_x h_{\mathrm{rz}}n + \left[\left(I_x + I_y - I_z\right)^2 - h_{\mathrm{rz}}^2\right]\right\} \\
&E = \left[(I_z - I_x)\left(I_z - I_y\right)n^4 - h_{\mathrm{rz}}(I_x - I_z)n^3 + h_{\mathrm{rz}}\left(I_z - I_y\right)n^3 + h_{\mathrm{rz}}^2 n^2\right]
\end{aligned}$$

The transfer function can be manipulated so that the terms depending on h_{rz} are explicitly taken into account. In this context, these should be considered constant, in the nominal condition.

$$\begin{aligned}
&\alpha_y\left[as^4 + bs^3 + cs^2 + ds + e + h_{rz}f(s)\right] = sh_{ry}g(s) \\
&\alpha_y H(s) = M_{cy}(s)
\end{aligned}$$

For this system a PD control might not be appropriate to impose specified performances, but a fourth order compensator can be designed and target performances reached.

Active Control of Spacecraft in the LVLH frame

$$\begin{cases} I_x\ddot{\alpha}_x + (I_z - I_y - I_x)n\dot{\alpha}_y + (I_z - I_y)n^2\alpha_x = 0 \\ I_y\ddot{\alpha}_y + (I_x + I_y - I_z)n\dot{\alpha}_x + (I_z - I_x)n^2\alpha_y = 3n^2(I_x - I_z)\alpha_y \\ I_z\ddot{\alpha}_z = -3n^2(I_y - I_x)\alpha_z \end{cases}$$

Simple proportional controllers can be used to control the pitch axis in the LVLH frame

$$I_z\ddot{\alpha}_z = T_3$$

The torque T_3can then be chosen to provide asymptotic stability for example:

$$T_3 = -k_1\alpha_z - k_2\dot{\alpha}_z$$

The closed-loop system can then be expressed as:

$$I_z\ddot{\alpha}_z + k_2\dot{\alpha}_z + k_3\alpha_z = 0$$

where $k_3 = k_2 + 3n^2(I_y - I_x)$

The eigenvalues related to this closed-loop system are:

$$\lambda = \frac{-k_2 \pm \sqrt{k_2^2 - 4k_3}}{2}$$

So we can see that (linear) asymptotic stability is achieved. It is also possible to use this information to reduce the tuning requirement since setting $k_2^2 = 4k_3$ provides the fastest system response (analogous to critical damping).

In the other two axis (yaw and roll) the equations are coupled. To design and tune a control law in this case it is useful to recast the equations into first order

$$\begin{cases} \dot{\alpha}_x = v_x \\ \dot{\alpha}_y = v_y \\ \dot{v}_x = -n(K_x - 1)v_y - n^2K_x\alpha_x + u_1 \\ \dot{v}_y = n(1 - K_y)v_x - n^2K_y\alpha_y + u_2 \end{cases}$$

With $x = [\alpha_x \quad \alpha_y \quad v_x \quad v_y]^T$ this can be expressed as $\dot{x} = Ax + Bu$ with $u = [u_1 \quad u_2]^T$and

$$B = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T$$

Firstly, controllability of the system muct be checked since the first order differnetial equation is 4 dimensional and we only have to control inputs. This is achieved through the computation:

$$C = [B \quad AB \quad A^2B \quad \dots \quad A^{n-1}B]$$

If the matrix C is full rank then the system is controllable.

If the system is controllable a simple continuous control law of the form:

$$\underline{u} = -K\underline{x}$$

Yields the following closed-loop system

$$\dot{\underline{x}} = (A - BK)\underline{x}$$

Then K is chosen such that the eigenvalues of the matrix $(A - BK)$ lie in the left hand side of the complex plane.

The gain matrix K can be determined by minimizing thr Linear quadratic performance index:

$$J = \frac{1}{2}\int_0^{\infty} (\underline{x}^T Q\underline{x} + \underline{u}^T R\underline{u})\, dt$$

Where Q is a symmetric and positive semidefinite weighting matrix and R a symmetric positive definite are weighting matrices then the optimal gain matrix is:

$$K = R^{-1}B^T P$$

With

$$0 = A^T P + PA - PBR^{-1}B^T P + Q$$

General control problem

We have seen that when the actuators are IW or RW we can write:

$$M = A\dot{h}_r$$

The question now is how to evaluate M in a general case. Assuming the satellite is inertial pointing, Euler equations assume a particularly simple form, since no nominal value of angular velocity is present. Equations are decoupled:

$$M_i = I\dot{\omega}_i i = 1,2,3$$

Linearizing the system, we have three second order decoupled equations:

$$M_i = I\ddot{\alpha}_i$$

Lack of a nominal angular momentum forces to have three independent actuators for the three axes; otherwise the system would not be controllable. Each equation represents a linear second order system, so that we can assume a simple PID control would allow obtaining the desired system performances:

$$M = f(\alpha) = PID(\alpha)$$

In some cases the integral term might not be required, in case a steady state error is allowed provided it is small or in case no constant component of disturbance torque is present.
When the satellite is far from equilibrium, its dynamics should include also the coupling terms due to angular velocities, so that we must write:

$$M = I\underline{\dot{\omega}} + \underline{\omega} \wedge I\underline{\omega}$$

To consider the coupling terms $\underline{\omega} \wedge I\underline{\omega}$ the control torque should be evaluated as:

$$M_c = I\underline{\dot{\omega}}$$

and then the nonlinear terms can be considered as a correction to the control torque as:

$$M = M_c + \underline{\omega} \wedge I\underline{\omega}$$

In case of a dual spin satellite the problem is formulated as:

$$\begin{aligned}
M &= I\underline{\dot{\omega}} + \underline{\omega} \wedge I\underline{\omega} + A\underline{\dot{h}_r} + \underline{\omega} \wedge \mathrm{Ah}_r && \rightarrow \quad \text{nonlinear dynamics} \\
0 &= I\underline{\dot{\omega}} + \underline{\omega} \wedge I\underline{\omega} + A\underline{\dot{h}_r} + \underline{\omega} \wedge \mathrm{Ah}_r && \rightarrow \quad \text{control equation} \\
M_c &= I\underline{\dot{\omega}} = \mathrm{PID}(\omega, \alpha) && \rightarrow \quad \text{pseudocontrol function} \\
M_c &= -\underline{\omega} \wedge I\underline{\omega} - A\underline{\dot{h}_r} - \underline{\omega} \wedge \mathrm{Ah}_r && \rightarrow \quad \text{actuator equation} \\
\underline{\dot{h}_r} &= A^*\left[-M_c - \underline{\omega} \wedge I\underline{\omega} - \underline{\omega} \wedge \mathrm{Ah}_r\right] && \rightarrow \quad \text{actuator command}
\end{aligned}$$

If, besides nonzero angular velocities, we consider also large angular rotations, the error angles definition depends on the sequence of rotations considered. A general solution of the control problem must then be sought for, starting from the definition of direction cosine matrix.

Call A_S the satellite attitude matrix in an inertial frame, and assume the target attitude is given by matrix A_T:

$$A_S = \begin{bmatrix} a_{11S} & a_{12S} & a_{13S} \\ a_{21S} & a_{22S} & a_{23S} \\ a_{31S} & a_{32S} & a_{33S} \end{bmatrix} = \begin{bmatrix} X_S \\ Y_S \\ Z_S \end{bmatrix}$$

$$A_T = \begin{bmatrix} a_{11T} & a_{12T} & a_{13T} \\ a_{21T} & a_{22T} & a_{23T} \\ a_{31T} & a_{32T} & a_{33T} \end{bmatrix} = \begin{bmatrix} X_T \\ Y_T \\ Z_T \end{bmatrix}$$

Each row of A_S and A_T represents one reference axis, either satellite or target. Our goal is to have:

$$A_S A_T^T = I$$

In actual conditions, the error in the attitude is expressed as:

$$A_S A_T^T = A_e$$

where A_e is the attitude error. We want now to relate the attitude error A_e with the control torque, in order to make A_e=I.

$$A_S A_T^T = \begin{bmatrix} X_S \\ Y_S \\ Z_S \end{bmatrix} \begin{bmatrix} X_T^T & Y_T^T & Z_T^T \end{bmatrix} = \begin{bmatrix} X_S X_T^T & X_S Y_T^T & X_S Z_T^T \\ Y_S X_T^T & Y_S Y_T^T & Y_S Z_T^T \\ Z_S X_T^T & Z_S Y_T^T & Z_S Z_T^T \end{bmatrix}$$

To reach the zero error, the extra diagonal terms must vanish to zero. $X_S Y_T^T = 0$ means that X_S and Y_T must become orthogonal, since this is the representation of the dot vector product. This condition can be obtained if the satellite rotates around its z body axis; so that we can say that the torque acting around the satellite body axis Z_S must take the form:

$$M_{zS} = f_z(X_S Y_T^T)$$

The actual control law is still undefined. The same must apply to the other extra diagonal terms of the attitude error matrix. If we want $X_S Z_T^T = 0$ the torque must be around the satellite body axis Y_S, in the general form:

$$M_{yS} = f_y(X_S Z_T^T)$$

Finally, to have $Y_S X_T^T = 0$ we need a torque around the satellite body axis X_S, so that:

$$M_{xS} = f_x(Y_S Z_T^T)$$

It is remarked that a similar result would be obtained by considering the terms below the diagonal, so that $M_{zS} = f_z(Y_S X_T^T)$ and the same for M_{yS} and M_{xS}.
To understand how the control functions can be designed, consider the case of small errors. In this case the error can be expressed in linearized form as:

$$A_S A_T^T = \begin{bmatrix} 1 & \alpha_Z & -\alpha_y \\ -\alpha_z & 1 & \alpha_x \\ \alpha_y & -\alpha_x & 1 \end{bmatrix}$$

In the case of decoupled motion or small rotations, we can write:

$$M_z = PD(\alpha) = K_{pz}\alpha_z + K_{dz}\dot{\alpha}_z$$

It is noticed that α_z corresponds to $X_S Y_T^T$ in the case of large errors, or to a_{12e}. We can then generalize the control function as:

$$M_{zs} = K_{pz}a_{12e} + K_{dz}\omega_z$$

The derivative should rigorously depend on the derivative of a_{12e}, or$X_S Y_T^T$, but since it is mostly used to introduce viscous damping any term with zero final value and sign opposite to the actual velocity can be used, so that ω_z can be a further candidate, in some cases easier to use for control. We can then evaluate K_p and K_d for the linear approximation of the dynamics and simply extend the validity of the control law as follows.

$$\begin{cases} M_x = K_{\text{px}}\alpha_x + K_{\text{dx}}\dot{\alpha}_x \\ M_y = K_{\text{py}}\alpha_y + K_{\text{dy}}\dot{\alpha}_y \\ M_z = K_{\text{pz}}\alpha_z + K_{\text{dz}}\dot{\alpha}_z \end{cases} \rightarrow \begin{cases} M_{\text{xs}} = K_{\text{px}}a_{23\text{e}} + K_{\text{dx}}\omega_x \\ M_{\text{ys}} = K_{\text{py}}a_{31\text{e}} + K_{\text{dy}}\omega_y \\ M_{\text{zs}} = -K_{\text{pz}}a_{21\text{e}} + K_{\text{dz}}\omega_z \end{cases}$$

The same can be designed adopting the terms below the diagonal, obtaining for each different case a different transient response in case of large initial errors:

$$M_{zs} = -K_{pz}a_{21e} + K_{dz}\omega_z$$

We can finally try to have an intermediate situation, for which:

$$\begin{cases} M_{xS} = f_x(Y_S Z_T^T - Z_S Y_T^T) \\ M_{yS} = f_y(X_S Z_T^T - Z_S X_T^T) \\ M_{zS} = f_z(X_S Y_T^T - Y_S X_T^T) \end{cases}$$

In this case, since

$$a_{12e} - a_{21e} = 2\alpha_z$$

we must divide by 2 the proportional gain if it was evaluated on the basis of a control proportional to α:

$$\begin{cases} M_{\text{xs}} = \dfrac{K_{\text{px}}}{2}(a_{23\text{e}} - a_{32\text{e}}) + K_{\text{dx}}\omega_x \\ M_{\text{ys}} = \dfrac{K_{\text{py}}}{2}(a_{31\text{e}} - a_{13\text{e}}) + K_{\text{dy}}\omega_y \\ M_{\text{zs}} = \dfrac{K_{\text{pz}}}{2}(a_{12\text{e}} - a_{21\text{e}}) + K_{\text{dz}}\omega_z \end{cases}$$

It is also remarked that a_{12e}-a_{21e} is proportional to the z component of the Euler axis of the attitude error matrix. It is likely that this final form of control will try to get the satellite rotating around the Euler axis of the attitude error, so that the transient response might be faster that in the previous formulations. One further option requires writing the terms a_{12} and a_{21} as a function of the quaternions, respectively $2(q_1q_2+q_3q_4)$ and $2(q_1q_2-q_3q_4)$, so that the control functions become:

$$\begin{cases} M_{xs} = \dfrac{K_{px}}{2}(4q_{1e}q_{4e}) + K_{dx}\omega_x = 2K_{px}q_{1e}q_{4e} + K_{dx}\omega_x \\ M_{ys} = \dfrac{K_{py}}{2}(4q_{2e}q_{4e}) + K_{dy}\omega_y = 2K_{py}q_{2e}q_{4e} + K_{dy}\omega_y \\ M_{zs} = \dfrac{K_{pz}}{2}(4q_{3e}q_{4e}) + K_{dz}\omega_z = 2K_{pz}q_{3e}q_{4e} + K_{dz}\omega_z \end{cases}$$

Notice that the quaternions are relative to the attitude error.
As the satellite approaches the target condition, the dynamics becomes automatically linear. If at steady state we want a nonzero value for one angular velocity component, around that specific axis the control function must be modified as:

$$M_i = K_{\mathrm{pi}}\alpha_i + K_{\mathrm{di}}(\dot{\alpha}_i - \bar{\dot{\alpha}}_i)$$
$$\Rightarrow M_{\mathrm{is}} = 2\mathrm{K}_{\mathrm{pi}}q_{\mathrm{ie}}q_{\mathrm{4e}} + K_{\mathrm{di}}(\omega_i - \bar{\omega}_i)$$

Angular velocity ω can be expressed as a function of quaternions. From the quaternion kinematics we know that:

$$\dot{q} = \frac{1}{2}\Omega q$$

This can be reformulated as:

$$\Omega q = \begin{bmatrix} q_4 & -q_3 & q_2 \\ q_3 & q_4 & -q_1 \\ -q_2 & q_1 & q_4 \\ -q_1 & -q_2 & -q_3 \end{bmatrix} \begin{Bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{Bmatrix} = Q\omega$$

and the angular velocity is:

$$\omega = 2Q^*\dot{q}$$

where Q^* is the pseudo inverse of Q, that is also equal to Q^T due to its structure. It is easy to verify that:

$$Q^T Q = I$$

Nonlinear Control Theory

In this section we describe the basics of nonlinear control theory. The theory is based on Lyapunov's second stability theorem and can be used to prove global stability of a system, as opposed to linear stability which can only prove global stability for linear systems. Lyapunov's second theorem can be applied to prove stability without the need for linearization of the nonlinear dynamics about an equilibrium point.

Lyapunov's Second Stability Theorem

Consider an autonomous nonlinear dynamic system described by:

$$\dot{\underset{\sim}{x}} = f(\underset{\sim}{x}), f(\underset{\sim}{x}^*) = 0$$

Where $\underset{\sim}{x}^*$ is an isolated equilibrium point. If there exists in some finite neighbourhood *D* , of the equilibrium point $\underset{\sim}{x}^*$a scalar function $V(\underset{\sim}{x})$ with continuous first partial derivative with respect to $\underset{\sim}{x}$ such that the following conditions hold

(i) $V(\underset{\sim}{x}) > 0$for all $\underset{\sim}{x} \neq \underset{\sim}{x}^*$in *D and* $V(\underset{\sim}{x}^*) = 0$
(ii) $\dot{V}(\underset{\sim}{x}) < 0$for all $\underset{\sim}{x} \neq \underset{\sim}{x}^*$in *D* except for $\dot{V}(\underset{\sim}{x}^*) = 0$

Then the system is said to be **asymptotically stable.** If *D* includes all possible states then the system is said to be **globally asymptotically stable.**

If $\dot{V}(\underset{\sim}{x}) = 0$for all $\underset{\sim}{x} \neq \underset{\sim}{x}^*$in *D*

then the system is **Lyapunov stable**.

The above theorem can be used to deduce if a feedback-control law is asymptotically stable or not. For example, lets take the case of designing a de-tumbling controller for a spacecraft. We have the equations of motion (ignoring environmental disturbances for now):

$$\dot{\omega}_x = \frac{I_y - I_z}{I_x}\omega_y\omega_z + \frac{u_x}{I_x}$$
$$\dot{\omega}_y = \frac{I_z - I_x}{I_y}\omega_x\omega_z + \frac{u_y}{I_y}$$
$$\dot{\omega}_z = \frac{I_x - I_y}{I_z}\omega_y\omega_x + \frac{u_z}{I_z}$$

Where $\underset{\sim}{u} = [u_x \quad u_y \quad u_z]^T$is the control torque input assumed to be given perfectly by an actuator. We can use the kinetic energy as a candidate Lyapunov function to determine suitable controls:

$$V(\underset{\sim}{\omega}) = \frac{1}{2}(I_x\omega_x^2 + I_y\omega_y^2 + I_z\omega_z^2)$$

then

$$\dot{V}(\underset{\sim}{\omega}) = I_x\omega_x\dot{\omega}_x + I_y\dot{\omega}_y + I_z\dot{\omega}_z$$

Then substituting in the dynamic equation (Euler equations plus the control) we get

$$\dot{V}(\underset{\sim}{\omega}) = \omega_x u_x + \omega_y u_y + \omega_z u_z$$

Then we need to define a control law such that $\dot{V}(\underset{\sim}{\omega})$ <0 to guarentee aymptotic stability i.e $\omega_x \to 0, \omega_y \to 0, \omega_z \to 0$. We can set $\underset{\sim}{u} = -k\underset{\sim}{\omega}$to yield global asymptotic stability where k is a tuning parameter and $\underset{\sim}{\omega} = [\omega_x \quad \omega_y \quad \omega_z]^T$. However,this vhoice is not unique and other choices exist such as a bang-bang type control (instantaneous switching)

$$\underset{\sim}{u} = -T\, \underset{\sim}{sgn}\, \omega_{max}$$

Where $\underset{\sim}{sgn}\, \omega = [sgn\, \omega_x \quad sgn\, \omega_y \quad sgn\, \omega_z]^T$where the sgn function returns 1 if the component is positive and -1 if it is negative.

For example, the kinetic energy of a rigid-body gives

$$E(\underset{\sim}{\omega}) = \frac{1}{2}(I_x\omega_x + I_y\omega_y + I_z\omega_z)$$

and with $\underset{\sim}{M} = -k\underset{\sim}{\omega}$ it can be easily shown that

$$\dot{E}(\underset{\sim}{\omega}) = -k(\omega_x^2 + \omega_y^2 + \omega_z^2)$$

Therefore, the kinetic energy in this case is a Lyapunov function and the equilibrium point (0,0,0) of the closed-loop system is globally asymptotically stable.

Lyapunov Control functions

In classical mechanics Lyapunov's second stability theorem is used to prove nonlinear stability of an autonomous nonlinear system. However, it can also be used to design control algorithms for systems of the form:

$$\dot{\underset{\sim}{x}} = f(\underset{\sim}{x}, \underset{\sim}{u}), f(\underset{\sim}{x}^*, \underset{\sim}{u}) = 0$$

In this case we choose a suitable Lyapunov candidate function that satisfies the condition:

(i) $V(\underset{\sim}{x}) > 0$for all $\underset{\sim}{x} \neq \underset{\sim}{x}^*$in D *and* $V(\underset{\sim}{x}^*) = 0$

Then differentiate this function with respect to time which will yield:

$$\frac{dV(\underset{\sim}{x})}{dt} = s(\underset{\sim}{x}, \underset{\sim}{u})$$

Then the control $\underset{\sim}{u}$is chosen such that

$$s(\underset{\sim}{x}, \underset{\sim}{u}) < 0$$

For all $\underset{\sim}{x} \neq \underset{\sim}{x}^*$. This provides a constructive approach for nonlinear control design. On differentiation of this Lyapunov function and for the nonlinear control system

$$J\frac{d\underset{\sim}{\omega}}{dt} = J\underset{\sim}{\omega} \times \underset{\sim}{\omega} + \underset{\sim}{u}$$

Yields

$$\dot{V}(\underset{\sim}{\omega}) = \underset{\sim}{\omega}^T \underset{\sim}{u}$$

Then we must choose a$\underset{\sim}{u}$such that $\underset{\sim}{\omega}^T \underset{\sim}{u} < 0$. Obviously, we can set $\underset{\sim}{u} = -k\underset{\sim}{\omega}$to yield global asymptotic stability. However, other choices exist such as

$$\underset{\sim}{u} = -T\, sg\underset{-}{n}\, \omega_{max}$$

Which also guarantees global asymptotic stability of the closed-loop system. This type of control is an on-off type controller that would be more suited to a thruster.

Slew maneuvers

$$\dot{\omega}_1 = \frac{I_2 - I_3}{I_1}\omega_2\omega_3 + \frac{u_1}{I_1}$$
$$\dot{\omega}_2 = \frac{I_3 - I_1}{I_2}\omega_1\omega_3 + \frac{u_2}{I_2}$$
$$\dot{\omega}_3 = \frac{I_1 - I_2}{I_3}\omega_2\omega_1 + \frac{u_3}{I_3}$$

$$\begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \\ \dot{\theta}_3 \end{bmatrix} = \frac{1}{\cos\theta_2}\begin{pmatrix} \cos\theta_2 & \sin\theta_1\sin\theta_2 & \cos\theta_1\sin\theta_2 \\ 0 & \cos\theta_1\cos\theta_2 & -\sin\theta_1\cos\theta_2 \\ 0 & \sin\theta_1 & \cos\theta_1 \end{pmatrix}\begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix}$$

Use a P(I)D controller $\underline{u} = -k_1\underline{e} - k_2\underline{\dot{e}} - k_3\int \underline{e}dt$

Closed-loop system

$$\ddot{\theta}_i + \frac{k_{1i}}{I_i}\dot{\theta}_i + \frac{c_{1i}}{I_i}\theta_i = 0$$

A slew motion is defined here as a controlled motion between two attitudes. The motion is also defined as a rest-to-rest motion. In this case the desired equilibrium point is $\underset{\sim}{\omega}_d = \underset{\sim}{0}$ and $I = A_d$. We can try the following Lyapnuov control function which clearly satisfirs $V(\underset{\sim}{\omega}_d = \underset{\sim}{0}, I = A_d) = 0$ and $V > 0$ for all other values

$$V = \frac{1}{2}\underset{\sim}{\omega}^T J\underset{\sim}{\omega} + k_2 tr(I - A)$$

then

$$\dot{V} = \underset{\sim}{\omega}^T \underset{\sim}{u} + k_2 tr(-\dot{A}) = \underset{\sim}{\omega}^T \underset{\sim}{u} + k_2 tr([\omega]^{\wedge} A)$$
$$= \underset{\sim}{\omega}^T(\underset{\sim}{u} + k_2(A^T - A))$$

It follows that the control

$$\underset{\sim}{u} = -k_1\underset{\sim}{\omega} - k_2(A^T - A)^V$$

asymptotically stabilizes the equilibrium point $\underset{\sim}{\omega}_d = \underset{\sim}{0}$ and $I = A_d$. If using quaternions to represent the attitude kinematics we have

$$\frac{d}{dt}\begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 0 & \omega_3 & -\omega_2 & \omega_1 \\ -\omega_3 & 0 & \omega_1 & \omega_2 \\ \omega_2 & -\omega_1 & 0 & \omega_3 \\ -\omega_1 & -\omega_2 & -\omega_3 & 0 \end{bmatrix}\begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix}$$

Then the control $\underset{\sim}{u} = -k_1\underset{\sim}{\omega} - k_2\underset{\sim}{q}_e$ with

$$\begin{bmatrix} q_{1e} \\ q_{2e} \\ q_{3e} \\ q_{4e} \end{bmatrix} = \begin{bmatrix} q_{4c} & q_{3c} & -q_{2c} & -q_{1c} \\ -q_{3c} & q_{4c} & q_{1c} & -q_{2c} \\ q_{2c} & -q_{1c} & q_{4c} & -q_{3c} \\ q_{1c} & q_{2c} & q_{3c} & q_{4c} \end{bmatrix}\begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix}$$

Asymptotically stabilizes the desired equilibrium $\hat{q}_d = [q_{1c} \quad q_{2c} \quad q_{3c} \quad q_{4c}]^T$. Proof of asymptotic stability can be shown by using the Lyapunov function

$$V = \frac{1}{2}\underset{\sim}{\omega}^T J\underset{\sim}{\omega} + 2k_2(1 - q_{4e})$$

Then on differentiating with respect to time we have

$$\dot{V} = \underset{\sim}{\omega}^T J\dot{\underset{\sim}{\omega}} + 2k_2(-\dot{q}_{4e})$$

Then substituting in the kinematics and dynamics we get

$$\dot{V} = \underset{\sim}{\omega}^T\underset{\sim}{u} + k_2(\underset{\sim}{\omega}_e{}^T\underset{\sim}{q}_e)$$

Since in this case $\underset{\sim}{\omega}_e = \underset{\sim}{\omega}$then we can write

$$\dot{V} = \underset{\sim}{\omega}^T(\underset{\sim}{u} + k_2\underset{\sim}{q}_e)$$

Then the control $\underset{\sim}{u} = -k_1\underset{\sim}{\omega} - k_2\underset{\sim}{q}_e$gives:

$$\dot{V} = -k_1\underset{\sim}{\omega}^T J\underset{\sim}{\omega}$$

Nonlinear control methods and tracking control

A general Lyapunov function for designing slew motions can be given as:

$$V = \frac{1}{2}\underset{\sim}{\omega} \cdot J\underset{\sim}{\omega} + 2k_2H(q_{4e})$$

Where $H(q_{4e})$ is a function of the scalar part of the quaternion error q_{4e}. Moreover, to satisfy the conditions of a Lyapunov function $H(\pm 1) = 0$and $H(q_{4e}) > 0$ for $q_{4e} \neq \pm 1$. This function is used to design a class of controls that depend on the control engineer's selection of $H(q_{4e})$

$$\begin{bmatrix} q_{1e} \\ q_{2e} \\ q_{3e} \\ q_{4e} \end{bmatrix} = \begin{bmatrix} q_{4d} & q_{3d} & -q_{2d} & -q_{1d} \\ -q_{3d} & q_{4d} & q_{1d} & -q_{2d} \\ q_{2d} & -q_{1d} & q_{4d} & -q_{3d} \\ q_{1d} & q_{2d} & q_{3d} & q_{4d} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix}$$

It follows on differentiation that

$$\dot{V} = \underset{\sim}{\omega} \cdot J\dot{\underset{\sim}{\omega}} + 2k_2 \frac{\partial H(q_{4e})}{\partial q_{4e}} \dot{q}_{4e}$$

And recalling the dynamic equations (assuming negligible disturbance torques for now) and the quaternion kinematic equations we have:

$$J\dot{\underset{\sim}{\omega}} = J\underset{\sim}{\omega} \times \underset{\sim}{\omega} + \underset{\sim}{u}$$
$$\frac{d\underset{\sim}{q}_e}{dt} = -\frac{1}{2}[\omega_e{}^{\wedge}]\underset{\sim}{q}_e + q_{4e}\underset{\sim}{\omega}_e$$
$$\frac{dq_{4e}}{dt} = -\frac{1}{2}\underset{\sim}{\omega}_e{}^T \underset{\sim}{q}_e$$

Then substituting these equations into $\dot{V}$yields:

$$\dot{V} = \underset{\sim}{\omega} \cdot (\underset{\sim}{u} - k_2 \frac{\partial H(q_{4e})}{\partial q_{4e}} \underset{\sim}{q}_e)$$

Then a simple slew controller can be selected

$$\underset{\sim}{u} = -k_1\underset{\sim}{\omega} + k_2 \frac{\partial H(q_{4e})}{\partial q_{4e}} \underset{\sim}{q}_e$$

Such that

$$\dot{V} = -k_1 \underset{\sim}{\omega} \cdot \underset{\sim}{\omega}$$

Now it is a matter of selecting the best performing potential $H(q_{4e})$. One example is to select

$$H(q_{4e}) = 1 - q_{4e}$$

Which gives

$$\underset{\sim}{u} = -k_1\underset{\sim}{\omega} - k_2\underset{\sim}{q}_e$$

This ensures that the zero error state $\underset{\sim}{\omega} = \underset{\sim}{0}, q_{4e} = 1$ is a stable equilibrium point according to our proof. However, $\underset{\sim}{\omega} = \underset{\sim}{0}, q_{4e} = -1$is unstable using this potential function and it can clearly be seen not to satisfy the condition of the Lyapunov function since $H(-1) > 0$. However, $\underset{\sim}{\omega} = \underset{\sim}{0}, q_{4e} = -1$also corresponds to the same zero error state. In other words,

$$q_{4e} = [0 \quad 0 \quad 0 \quad \pm 1]^T \rightarrow A_e = I_{3\times 3}$$

Thus, there is an ambiguity due to the non-uniqueness, whereby, the zero error represented by quaternions can be stable or unstable. This can lead to the problem of unwinding. Moreover, if our initial condition is $\underset{\sim}{\omega} = \underset{\sim}{0}, q_{4e} = -0.9$and we use the control corresponding to $H(q_{4e}) = 1 - q_{4e}$the the control will drive the system to the stable state $\underset{\sim}{\omega} = \underset{\sim}{0}, q_{4e} = 1$ and the system will perform a large rotation. It would be far more efficient to perform a small rotation to the error state $\underset{\sim}{\omega} = \underset{\sim}{0}, q_{4e} = -1$, but for this control it is unstable. Another choice is $H(q_{4e}) = 1 + q_{4e}$, but in this case $\underset{\sim}{\omega} = \underset{\sim}{0}, q_{4e} = -1$will be a stable equilibrium of the closed loop system and $\underset{\sim}{\omega} = \underset{\sim}{0}, q_{4e} = 1$will be unstable, so potentially problems of unwinding can occur. If we choose any of the following functions:

$$H(q_{4e}) = 1 - q_{4e}^2$$
$$H(q_{4e}) = 1 - sgn(\, q_{4e}(0))q_{4e}$$
$$H(q_{4e}) = arccos^2\, q_{4e}$$

Then $H(\pm 1) = 0$. In this case both $\underset{\sim}{\omega} = \underset{\sim}{0}, q_{4e} = 1$and $\underset{\sim}{\omega} = \underset{\sim}{0}, q_{4e} = -1$ which are physically the same error are both stable.

Eigenaxis Rotation slew maneuver

A final ideal control that we will use will be a slew that corresponds to an eigenaxis rotation. This control is particularly useful since it can be optimally tuned. We begin with a similar Lyaunov function to one that has been used before, but without the inclusion of the inertia matrix:

$$V = \frac{1}{2}\underset{\sim}{\omega} \cdot \underset{\sim}{\omega} + 2k_2 H(q_{4e})$$

Then on differentiation we have

$$\dot{V} = \underset{\sim}{\omega} \cdot \dot{\underset{\sim}{\omega}} + 2k_2 \frac{\partial H(q_{4e})}{\partial q_{4e}} \dot{q}_{4e}$$

And it follows from

$$J\dot{\underset{\sim}{\omega}} = J\underset{\sim}{\omega} \times \underset{\sim}{\omega} + \underset{\sim}{u}$$

And

$$\frac{dq_{4e}}{dt} = -\frac{1}{2}\underset{\sim}{\omega}_e{}^T \underset{\sim}{q}_e$$

That

$$\dot{V} = \underset{\sim}{\omega} \cdot (J^{-1}(J\underset{\sim}{\omega} \times \underset{\sim}{\omega} + \underset{\sim}{u}) - k_2 \frac{\partial H(q_{4e})}{\partial q_{4e}} \underset{\sim}{q}_e)$$

Then to guaranteed that $\dot{V} < 0$ we choose

$$\underset{\sim}{u} = \underset{\sim}{\omega} \times J\underset{\sim}{\omega} - k_1 J\underset{\sim}{\omega} + k_2 J \frac{\partial H(q_{4e})}{\partial q_{4e}} \underset{\sim}{q}_e$$

If we compare this to our previously derived slew control:

$$\underset{\sim}{u} = -k_1\underset{\sim}{\omega} + k_2 \frac{\partial H(q_{4e})}{\partial q_{4e}} \underset{\sim}{q}_e$$

The two differences are (i) the term $\underset{\sim}{\omega} \times J\underset{\sim}{\omega}$, which essentially cancels the dynamics and makes the system more responsive to the control. (ii) the matrix *J* now appears in each component of the control. In this case the closed loop system is:

$$\dot{\underset{\sim}{\omega}} = -k_1\underset{\sim}{\omega} + k_2\frac{\partial H(q_{4e})}{\partial q_{4e}}\underset{\sim}{q}_e$$

Assume the form of an eigenaxis rotation which is a rotation around a fixed axis $\underset{\sim}{e}$ of magnitude θ. In this case we can write

$$\underset{\sim}{\omega} = \underset{\sim}{\omega}_e = \dot{\theta}_e\underset{\sim}{x}, \underset{\sim}{q}_e = sin\frac{\theta_e}{2}\underset{\sim}{x}$$

Substituting into the closed form expression we have:

$$\ddot{\theta} + k_1\dot{\theta} - k_2\frac{\partial H(q_{4e})}{\partial q_{4e}}\theta = 0$$

So for example if we choose $H(q_{4e}) = 1 - q_{4e}$ then

$$\frac{\partial H(q_{4e})}{\partial q_{4e}} = -1$$

So

$$\ddot{\theta} + k_1\dot{\theta} + k_2\,sin\frac{\theta}{2} = 0$$

This reduction of the closed-loop dynamics can be used to tune the slew motion i.e. for small angles we have

$$\ddot{\theta} + k_1\dot{\theta} + k_2\frac{\theta}{2} = 0$$

And so the corresponding characteristic equation can be expressed as:

$$\lambda = \frac{-k_1 \pm \sqrt{k_1^2 - 2k_2}}{2}$$

So for optimal tuning we choose

$$k_1^2 = 2k_2$$

And only one parameter is required to be tuned.

Trajectory tracking

Slew motions are rest to rest motions, so when we design a slew control $\underset{\sim}{\omega}_d = \underset{\sim}{0}$and therefore in the Lyapunov function we set $\underset{\sim}{\omega} = \underset{\sim}{\omega}_e$. In the general case that we want to track a time dependent attitude trajectory, that is, a moving frame such as the LVLH frame, we can use the most general form of the Lyapunov function:

$$V = \frac{1}{2}\underset{\sim}{\omega}_e{}^T J\underset{\sim}{\omega}_e + 2k_2H(q_{4e})$$

Note that, as proved by differentiating $A_e = AA_d^T$, $\underset{\sim}{\omega}_e = \underset{\sim}{\omega} - A_e(q)\underset{\sim}{\omega}_d$

Note that

$$\dot{\underset{\sim}{\omega}}_e = \dot{\underset{\sim}{\omega}} - \frac{d}{dt}(A_e(q)\underset{\sim}{\omega}_d) = J^{-1}(J\underset{\sim}{\omega} \times \underset{\sim}{\omega} + \underset{\sim}{u}) - \frac{d}{dt}(A_e(q)\underset{\sim}{\omega}_d)$$

The term $\frac{d}{dt}(A_e(q)\underset{\sim}{\omega}_d)$ appears since the reference frame that we want to track is time dependent:

$$\dot{V} = \underset{\sim}{\omega}_e{}^T(J\underset{\sim}{\omega} \times \underset{\sim}{\omega} + \underset{\sim}{u} - J\frac{d}{dt}(A_e(q)\underset{\sim}{\omega}_d) - k_2\frac{\partial H(q_{4e})}{\partial q_{4e}}\underset{\sim}{q}_e)$$

Then since $\underset{\sim}{\omega}_e{}^T(J\underset{\sim}{\omega} \times \underset{\sim}{\omega}) \neq 0)$ the control

$$\underset{\sim}{u} = -k_1\underset{\sim}{\omega}_e + k_2\frac{\partial H(q_{4e})}{\partial q_{4e}}\underset{\sim}{q}_e + \underset{\sim}{\omega} \times J\underset{\sim}{\omega} + J\frac{d}{dt}(A_e(q)\underset{\sim}{\omega}_d)$$

Leads to

$$\dot{V} = -k_1\underset{\sim}{\omega}_e{}^T\underset{\sim}{\omega}_e$$

And so this control asymptotically tracks the moving frame (time-dependent attitude reference trajectory). Note that

$$\frac{d}{dt}(A_e(q)\underset{\sim}{\omega}_d) = \dot{A}_e(q)\underset{\sim}{\omega}_d + A_e(q)\dot{\underset{\sim}{\omega}}_d$$
$$= -[\omega_e]^{\wedge}A_e(q)\underset{\sim}{\omega}_d + A_e(q)\dot{\underset{\sim}{\omega}}_d$$

So we can write the control as:

$$\underset{\sim}{u} = -k_1\underset{\sim}{\omega}_e + k_2\frac{\partial H(q_{4e})}{\partial q_{4e}}\underset{\sim}{q}_e + \underset{\sim}{\omega} \times J\underset{\sim}{\omega} + J(A_e(q)\dot{\underset{\sim}{\omega}}_d - [\omega_e]^{\wedge}A_e(q)\underset{\sim}{\omega}_d)$$

The tracking control is far more complex than the slew motion control. It requires precise knowledge of the inertia matrix, and the angular acceleration $\dot{\underset{\sim}{\omega}}_d$of the moving frame. Note that for slowly evolving moving frames the term $A_e(q)\dot{\underset{\sim}{\omega}}_d - [\omega_e]^{\wedge}A_e(q)\underset{\sim}{\omega}_d$ will be very small so the control can be reduced to

$$\underset{\sim}{u} = -k_1\underset{\sim}{\omega}_e + k_2\frac{\partial H(q_{4e})}{\partial q_{4e}}\underset{\sim}{q}_e + \underset{\sim}{\omega} \times J\underset{\sim}{\omega}$$

Note that if you are using DCM the control can be derived in an analogous way with $H(q_{4e})$ replaced with $tr(I - A_e)$ in the Lyapunov function

$$\underset{\sim}{u} = -k_1\underset{\sim}{\omega}_e - k_2(A_e^T - A_e)^V + \underset{\sim}{\omega} \times J\underset{\sim}{\omega} + J(A_e\dot{\underset{\sim}{\omega}}_d - [\omega_e]^{\wedge}A_e\underset{\sim}{\omega}_d)$$

Previously a reference attitude has been designed based on a circular orbit around the Earth. However, a reference attitude can be designed in an arbitrary way. For example, we may be required to point at the Sun or Moon while in an elliptic orbit or even a Libration point orbit. Assuming that the pointing vector is aligned with the unit vector $\vec{r}_{sc}$with respect to the spacecraft and $\vec{v}_{sc}$its relative velocity then we can construct an orthonormal moving frame:

$$A_d = \begin{bmatrix} \vec{x}_1 = \vec{r}_{sc} \\ \vec{x}_2 = \vec{x}_3 \times \vec{x}_1 \\ \vec{x}_3 = \vec{r}_{sc} \times \vec{v}_{sc} \end{bmatrix}$$

$\vec{v}_{sc}$could be replaced with a Sun pointing vector, which may be more optimal for power generation.

Disturbances attenuation

Let us now look at our previous control laws and see if the closed-loop systems are still guarenteed to be asymptotically stable. Using the simple example of a de-tumbing control $\underset{\sim}{u} = -k\underset{\sim}{\omega}$

The Euler equations now contain the control torque and the disturbance torques $J\frac{d\omega}{dt} = J\underset{\sim}{\omega} \times \underset{\sim}{\omega} + \underset{\sim}{u} + \underset{\sim}{d}$. Using the Lyapunov function:

$$V = \frac{1}{2}\underset{\sim}{\omega}^T J\underset{\sim}{\omega}$$

We have

$$\dot{V}(\underset{\sim}{\omega}) = \underset{\sim}{\omega}^T(\underset{\sim}{u} + \underset{\sim}{d})$$

It follows that

$$\dot{V}(\underset{\sim}{\omega}) = -\underset{\sim}{\omega}^T\underset{\sim}{\omega} - \underset{\sim}{\omega}^T\underset{\sim}{d}$$

Therefore, there is no guarentee that the spacecraft will de-tumble to exactly zero. However, lets augment our control to include a saturated term such that

$$\underset{\sim}{u} = -k_1\underset{\sim}{\omega} - k_2\, sgn(\,\underset{\sim}{\omega})$$

In this case we would have

$$\dot{V}(\underset{\sim}{\omega}) = -k_1\underset{\sim}{\omega}^T\underset{\sim}{\omega} - k_2\underset{\sim}{\omega}^T\, sgn(\,\underset{\sim}{\omega}) + \underset{\sim}{\omega}^T\underset{\sim}{d}$$

Defining $|\underset{\sim}{x}| = \underset{\sim}{x}^T\, sgn(\,\underset{\sim}{x})$ then we can say that $|\underset{\sim}{\omega}||\underset{\sim}{d}| \geq \underset{\sim}{\omega}^T\underset{\sim}{d}$then

$$\dot{V}(\underset{\sim}{\omega}) \leq -k_1\underset{\sim}{\omega}^T\underset{\sim}{\omega} - k_2|\underset{\sim}{\omega}| + |\underset{\sim}{\omega}||\underset{\sim}{d}|$$

Then the closed-loop system is asymptotically stable and the modified control is robust to disturbances if $k_2 > |\underset{\sim}{d}|$ i.e $-k_1\underset{\sim}{\omega}^T\underset{\sim}{\omega} - k_2|\underset{\sim}{\omega}| + |\underset{\sim}{\omega}||\underset{\sim}{d}| \leq 0$. So if we know the upperbound on the disturbances then we can design a robust control.

The reduced-attitude control problem

In some applications it may only be required to point in a prescribed direction. We define a desired pointing vector by $\Gamma_d$which is expressedIn such cases full 3-axis stabilization is not required and it may not be necessary to define the entire reference attitude on the pointing direction. Furthermore, in such cases it may not be necessary to determine the entire attitude but only a pointing vector. Define an orthonormal frame **N** such that the desired $\underset{\sim}{\Gamma}_d = [1 \quad 0 \quad 0]_N^T$. The current pointing vector can be expressed in the body frame as $\underset{\sim}{\Gamma}$then $\underset{\sim}{\Gamma} = A_{B/N}\underset{\sim}{\Gamma}_d$ then:

$$\dot{\underset{\sim}{\Gamma}} = \dot{A}_{B/N}\underset{\sim}{\Gamma}_d = -[\omega]^{\wedge} A_{B/N}\underset{\sim}{\Gamma}_d = -[\omega]^{\wedge}\underset{\sim}{\Gamma} = \underset{\sim}{\Gamma} \times \underset{\sim}{\omega}$$

Then define the error angle between the two vectors by:

$$\Gamma_d^T \Gamma = \cos\theta_e$$

We define the Lyapunov function

$$\boldsymbol{V} = \frac{\boldsymbol{1}}{\boldsymbol{2}}\langle \underset{\sim}{\boldsymbol{\omega}}, \boldsymbol{J}\underset{\sim}{\boldsymbol{\omega}}\rangle + \boldsymbol{k}_2(\boldsymbol{1} - \boldsymbol{\Gamma}_d^T\boldsymbol{\Gamma})$$

Which satisfies the first two requirements for a Lyapunov function. Then differentiating

$$\dot{\boldsymbol{V}} = \langle \underset{\sim}{\boldsymbol{\omega}}, \underset{\sim}{\boldsymbol{u}}\rangle - \boldsymbol{k}_2(\underset{\sim}{\boldsymbol{\Gamma}}_d^T \dot{\underset{\sim}{\boldsymbol{\Gamma}}}) = \langle \underset{\sim}{\boldsymbol{\omega}}, \underset{\sim}{\boldsymbol{u}}\rangle - \boldsymbol{k}_2 \underset{\sim}{\boldsymbol{\Gamma}}_d^T(\underset{\sim}{\boldsymbol{\Gamma}} \times \underset{\sim}{\boldsymbol{\omega}})$$
$$\dot{\boldsymbol{V}} = \langle \underset{\sim}{\boldsymbol{\omega}}, \underset{\sim}{\boldsymbol{u}}\rangle - \boldsymbol{k}_2 \underset{\sim}{\boldsymbol{\omega}}^T(\underset{\sim}{\boldsymbol{\Gamma}}_d \times \underset{\sim}{\boldsymbol{\Gamma}}) = \langle \underset{\sim}{\boldsymbol{\omega}}, \underset{\sim}{\boldsymbol{u}} + \boldsymbol{k}_2(\underset{\sim}{\boldsymbol{\Gamma}} \times \underset{\sim}{\boldsymbol{\Gamma}}_d)\rangle$$

Then the control

$$\underset{\sim}{u} = -k_1 \underset{\sim}{\omega} - k_2(\Gamma \times \Gamma_d)$$

Asymptotically converges to the desired poiting direction Γ_d.

Minimum time maneuvers

The adoption of the classical control techniques is feasible if the system performances are to be evaluated in terms of bandwidth, transient response or steady state error. In some cases, the optimal performances are provided in the form of minimum time or minimum fuel maneuvers. In these cases the controller has to be designed on different basis. To minimize maneuver time, the optimal performance requires minimizing the cost function that includes explicitly time and system dynamics (as a constraint):

$$min J = \int_{t_0}^{t_f} dt + \lambda \int_{t_0}^{t_f} (\dot{x} - Ax - Bu) dt$$

This problem has an exact, analytical solution only in some very special case. The rotation around one principal axis is one of these cases. For small rotations, the dynamics and the cost function become:

$$\begin{aligned}
&M = I\ddot{\alpha} \\
&\dot{\alpha} = \frac{d\alpha}{dt} \Rightarrow dt = \frac{d\alpha}{\dot{\alpha}} \\
&\Rightarrow J = \int_{\alpha_0}^{\alpha_f} \frac{1}{\dot{\alpha}} d\alpha
\end{aligned}$$

To minimize maneuver time, we must maximize $\dot{\alpha}$. We assume the available torque is limited to a maximum value:

$$u = \ddot{\alpha} = \frac{M}{I}$$

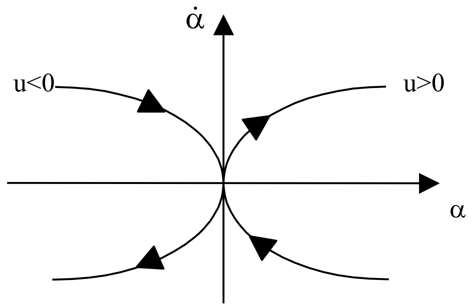
If the torque is sufficiently high the maneuver will be fast, but in cases in which the torque is limited the solution of the minimum time problem becomes of interest.
Integrate the dynamics to get:

$$\begin{aligned}
&\dot{\alpha} = \dot{\alpha}_0 + ut \\
&\alpha = \alpha_0 + \dot{\alpha}_0 t + \frac{1}{2} u t^2
\end{aligned}$$

The first equation is used to calculate time:

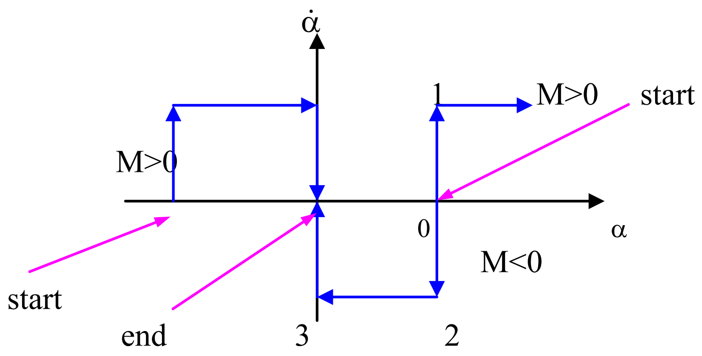
$$\begin{aligned}
&t = \frac{\dot{\alpha} - \dot{\alpha}_0}{u} \\
&\alpha = \alpha_0 + \frac{\dot{\alpha}_0(\dot{\alpha} - \dot{\alpha}_0)}{u} + \frac{1}{2}\frac{(\dot{\alpha} - \dot{\alpha}_0)^2}{u} \\
&\alpha - \alpha_0 = \frac{1}{u}\dot{\alpha}_0(\dot{\alpha} - \dot{\alpha}_0) + \frac{1}{2u}(\dot{\alpha} - \dot{\alpha}_0)^2
\end{aligned}$$

The phase plane plot $(\dot{\alpha}, \alpha)$ of the constant torque maneuver is then a family of parabolas, whose concavity depends on the torque limit:



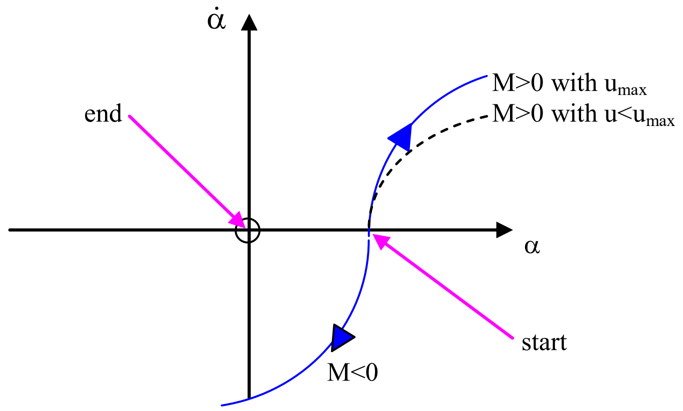
The arrows indicate that each parabola is followed with a specified direction. If for instance motion starts from the origin with M>0, then $\ddot{\alpha} > 0$ and $\dot{\alpha}$ will increase. By increasing the input u the parabolas are more open. Taking the extreme approximation of infinite torque, parabolas would be replaced by vertical lines, so that the maneuver would be impulsive, with a change in $\dot{\alpha}$ associated to a constant α. On the contrary, if u=0 we would have a phase portrait given by horizontal lines, with no change in $\dot{\alpha}$. In general, phase plane maneuvers are designed in order to have both $\dot{\alpha}_0$ and $\dot{\alpha}_f$ equal to zero, that is to say rest-to-rest maneuvers. In addition, since the origin of the phase plane is arbitrary, either α_0 or α_f is set to zero.

We can now design a maneuver in the phase plane assuming impulsive torques. This is the case of maneuvers performed by using high thrust propulsive systems.



Starting from point 0 to end in the origin, if we apply a positive torque we would reach point 1, but here $\dot{\alpha}$ is positive and we would de part from the desired attitude. We must then start with a negative torque, to reach point 2, switch off the controller to keep $\dot{\alpha}$ constant until point 3 is reached and then provide a positive torque to reach the target final attitude. Of course, should the initial attitude be negative all the maneuver has to be performed in the opposite way. Notice also that the vertical arcs of the phase plane are traced in almost zero time, since the torque is assumed infinite, and are equivalent to impulsive maneuvers. The total maneuver time depends then only on the horizontal arcs of the phase plane trace. In theory, we would like to have $\dot{\alpha}$ as high as possible to minimize maneuver time, so that the horizontal arc would be drawn in a short time. The major issue in this case is thruster synchronization, since with high $\dot{\alpha}$ even a small time error would mean to reverse the control (point 3 in the example) in a different point on the phase plane, so the target attitude would not be reached.

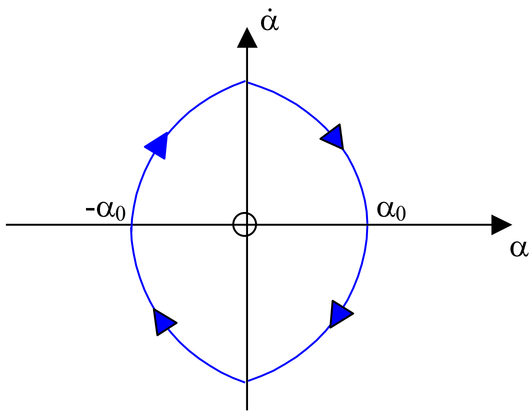
We can now consider a more realistic case, with bounded maximum torque.



We will consider only the parabolas corresponding to u_{max} so that $\dot{\alpha}$ is the maximum possible and time is minimum. The problem is to find the position in which the torque has to be switched in sign. If the maneuver is completed according to the control logic:

$$u = -u(\alpha)_{max}$$

once on the axis $\dot{\alpha}$ the sign of α changes so that the phase plane portrait would look like in the following figure:

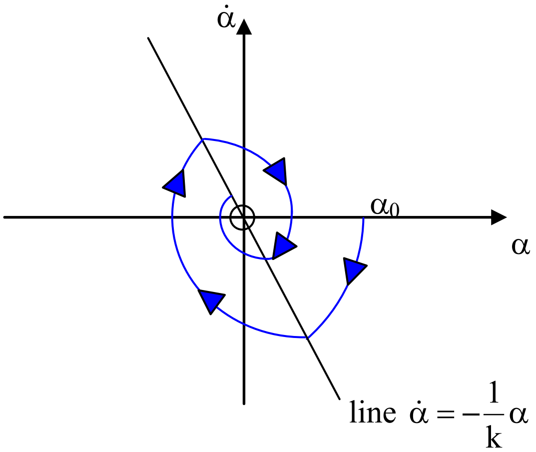


There is evidently a limit cycle, the system would behave like an undamped second order oscillator.

Change the control logic to:

$$u = -u(\alpha + k\dot{\alpha})_{max}$$

so that the switch in the sign of the control torque is along an inclined straight line. We would like k to be positive in order to have a negative inclination of the switching line:



Reaching the switching line, u changes its sign, so that the phase plane trace is switched to a parabola with reversed axis; as the number of torque switches increases, the trace gets closer and closer to the origin. k is then an index of the damping in the oscillations. However, rigorously, an infinite number of switchings are needed to reach exactly the origin in a general case.

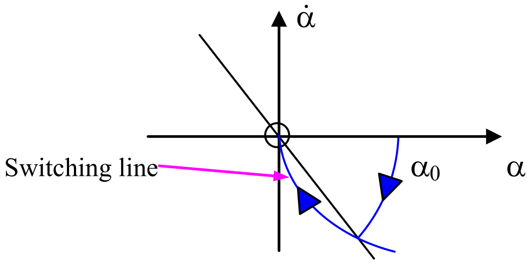
If we draw the two parabolas that pass through the origin, corresponding to positive and negative torque, since for $\dot{\alpha}_0$ equal to zero we have:

$$\alpha = \frac{1}{2u}\dot{\alpha}^2$$

we can consider the following switching curve:

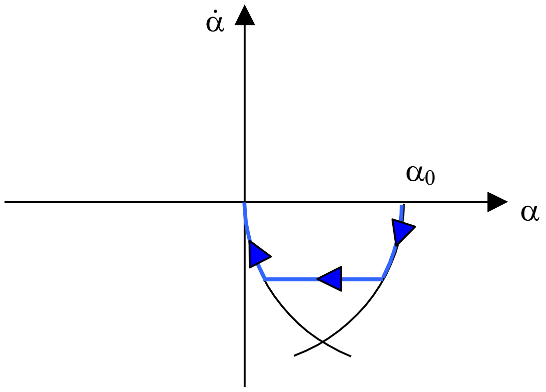
$$u = -u\left(\alpha - \frac{1}{2u}\dot{\alpha}|\dot{\alpha}|\right)_{max}$$

The phase plane portrait of the maneuver will then be:



It can be shown that this is the minimum time maneuver. If the initial and final velocities are zero, the satellite accelerates at the maximum level for half the rotation, then decelerates at same level for the second half of the rotation. The sign of the control torque becomes a function of α and $\dot{\alpha}$.

Finally, if we want to consider a minimum fuel maneuver we should fix a maximum maneuver time. This can be seen as a minimum time maneuver with one intermediate coast arc (at constant $\dot{\alpha}$) if the allowed maneuver time is greater than the minimum maneuver time for the same rotation.

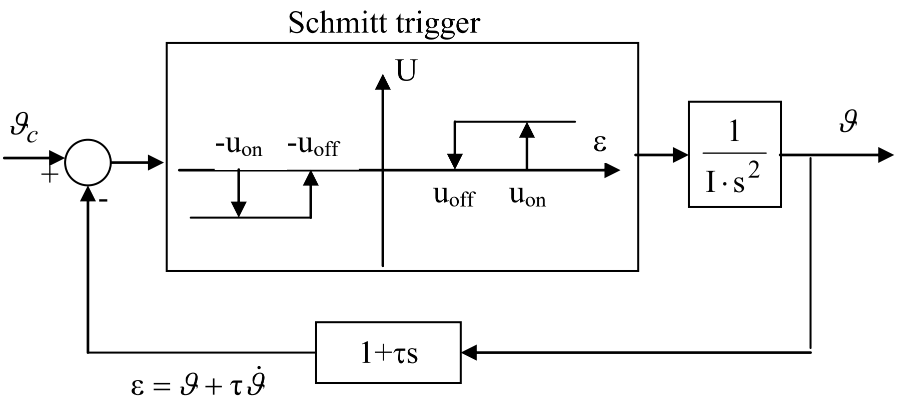


In this case, the switching from one parabola to the other occurs in a finite time. Notice that if t_{max} is equal to the minimum time we would reduce the coast arc to zero and find again the minimum time solution. Fixing t_{max} becomes equivalent to fixing $\dot{\alpha}_{max}$.

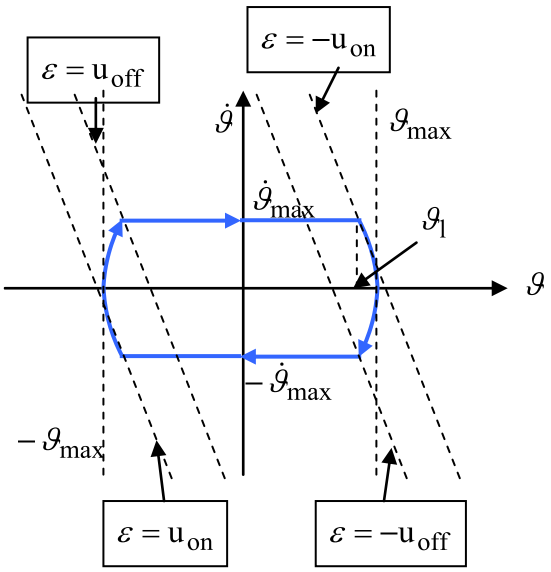
The process just show is valid only if the rotation is around one principal axis; in other cases, the complete set of Euler equations should be used as system dynamics (and optimization dynamic constraint) and a closed form solution can no longer be found.

Nonlinear control with constant thrust actuators

Considering the dynamic behavior of a satellite projected in the phase plane, it is possible to set up a nonlinear controller decoupled for each axis. The control is based on a combination of the angle ϑ and its derivative $d\vartheta$. In particular, a nonlinear switch called "Schmitt trigger" activates the controller on the basis of the value of a variable $\vartheta+\tau d\vartheta$. Assume, for example, that the value of $\vartheta+\tau d\vartheta$ is greater than a given limit u_{on}. In this case the actuators would be switched on until the same variable $\vartheta+\tau d\vartheta$ falls below a second limit u_{off}.



The values of u_{on}, u_{off} can be determined considering the maximum allowable angular error ϑ_{max}, the maximum admissible angular rate $d\vartheta_{max}$ and the time constant τ. With reference to the following figure, we must first of all consider the two parabolas passing from the points $(\pm\vartheta_{max},0)$, corresponding to the controlled dynamics with the maximum torque.



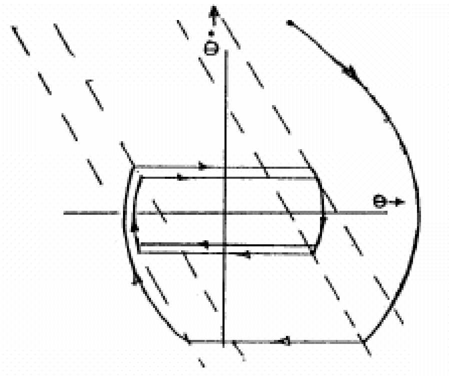
Intersect the two parabolas with the horizontal lines at $d\vartheta_{max}$, to identify ϑ_1, the angle error at which the controller must be switched on in order to prevent the error from getting larger than ϑ_{max}

$$\vartheta_1 = \vartheta_{\max} - \frac{d\vartheta_{max}^2}{2u_c}$$

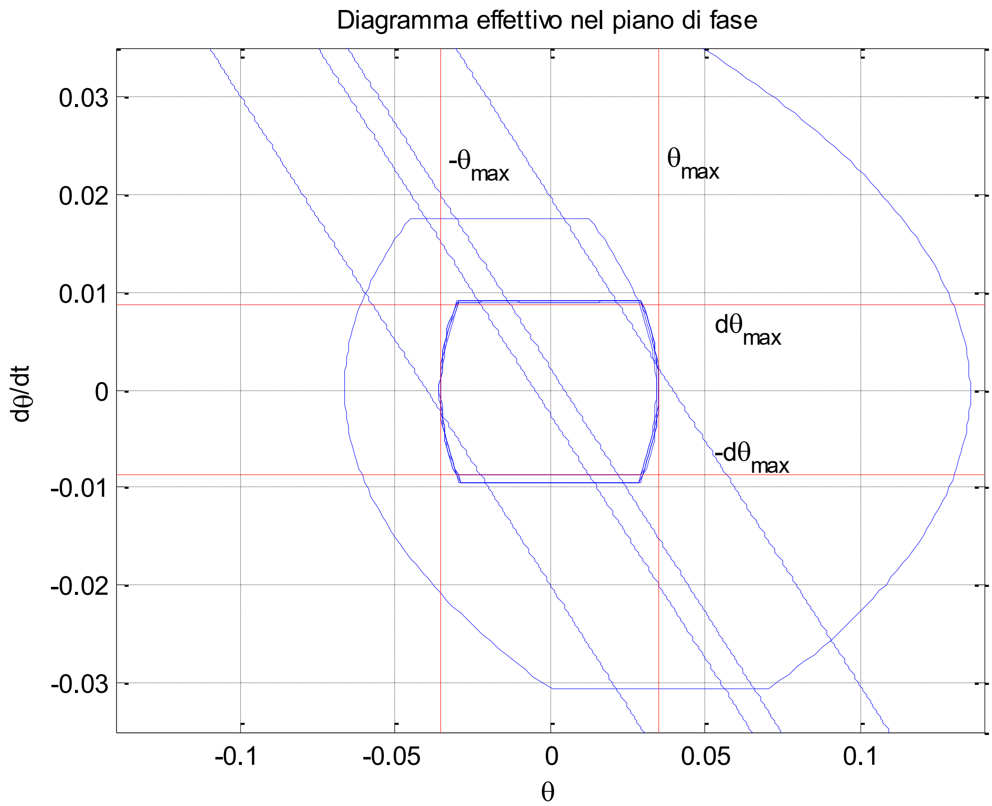
u_c = M/I is still the control command. The switching curve to activate the controller must intercept the point (ϑ_1, $d\vartheta_{max}$) and have a slope equal to $-1/\tau$. The value of τ can be tuned according to some performance requirements. The values of u_{on} and u_{off} are then evaluated as:

$$u_{on} = \tau d\vartheta + \vartheta_1$$
$$u_{off} = -\tau d\vartheta + \vartheta_1$$

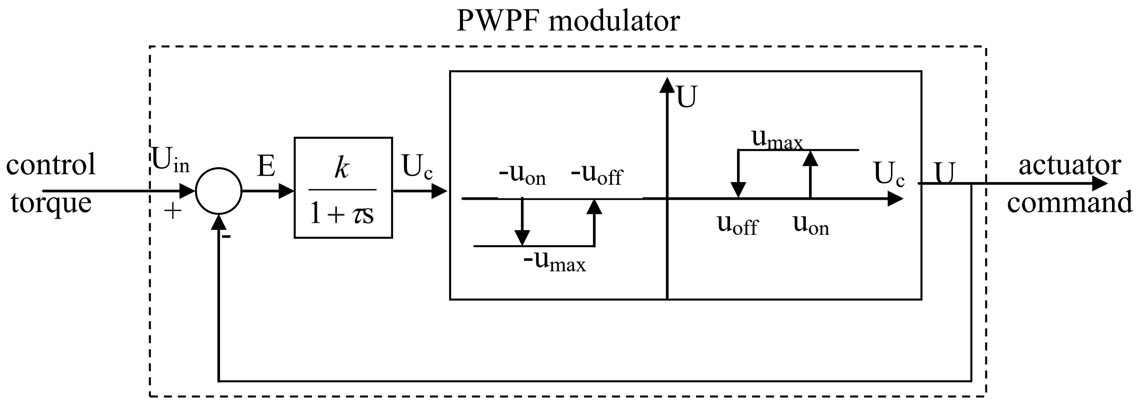
With symmetry considerations, the switching values for negative errors are determined. On the phase plane, in ideal conditions with no disturbance torque and sensor error, the satellite phase portrait must converge to a limit cycle bounded by the values $-\vartheta_{max}/+\vartheta_{max}$ and $-d\vartheta_{max}/+d\vartheta_{max}$.
The transient response for large initial errors will still converge to the same final limit cycle, provided the time constant τ is selected with the correct sign. The parameter τ has an influence on the way the phase portrait converges to the limit cycle.



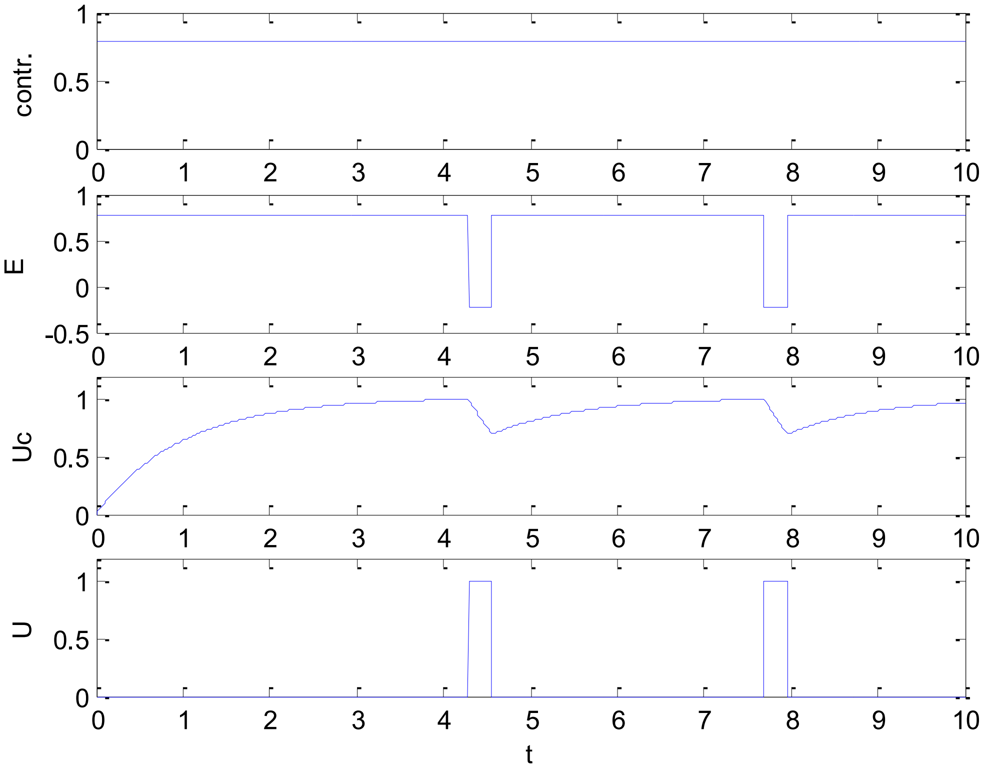
Considering the inevitable presence of sensor errors and delays in the activation of the actuators, the switching of the control will not be exactly on the desired switching lines. This means that the real limit cycle in the phase plane will be slightly different from the ideal one, as shown in the following example.



One alternative to the previous technique is based on the so called "PWPF modulator" (Pulse-Width-Pulse-Frequency Modulator). This varies both the width of the control pulse (the duration of the control action) and the frequency of the switchings. This is obtained as shown in the scheme below, with the integration and filtering of the desired control action until the switching threshold u_{on} is reached.



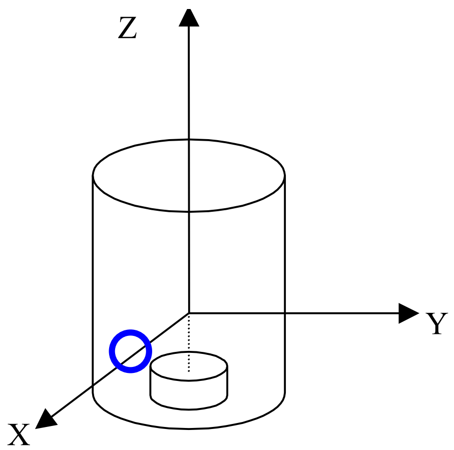
When the PWPF modulator provides a real control torque to the satellite, the torque is subtracted from the request of the controller before being passed to the integrator. Since in principle the supplied torque U is greater than the desired control U_{in}, the integrated signal will reverse its sign so that after a short period of time the PWPF modulator will switch the controller off. In this condition the initial behavior of the PWPF modulator is restored, until the following switch on of the controller. The time separation of two consecutive control pulses depends on the control law designed, while the duration of the control pulses depends on the maximum torque provided by the actuators U_c. To understand the principle of operation, assuming a constant control request U_{in}, the behavior of the PWPF modulator is the following.


contr.
E
Uc
U
t
0.5
-0.5
0 1 2 3 4 5 6 7 8 9 10

Passive damping systems

It is possible to make a distinction between attitude maneuver control and stability augmentation control. Maneuver control systems in general allow also controlling the stability of the system, while stability augmentation control systems do not allow controlling maneuvers, since they are passive systems. There are in fact some passive systems that can improve the stability of the satellite.
We can see two examples of passive or semi-active systems for the stability augmentation of nutation (conical motion of the spin axis) and libration (pendulum oscillations about the pitch axis) motion. These systems do not allow the execution of rotational maneuvers, in some cases they can even represent obstacles to maneuvers.
In general, nutation damping mechanisms base their effect on viscous damping created by the relative rotation of a viscous fluid within the spacecraft.

To model the damping mechanism, we can consider a dual spin satellite with nominal spin velocity around the z axis. Nutation appears as coupled oscillations around x and y axes, that can be damped with the aid of a fluid ring around x axis. Due to coupling of x and y equations, damping oscillations around x axis also damps oscillations around y axis.



Considering the fluid ring at the same rate of a reaction wheel, we can write the following equations:

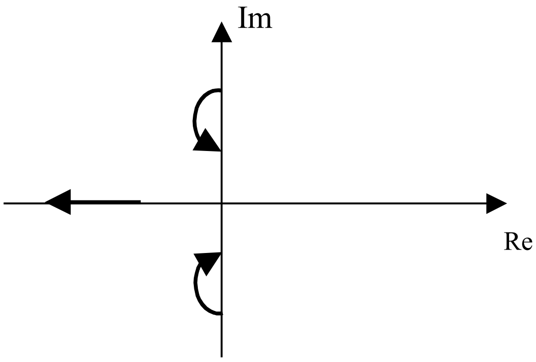
$$\underline{h} = \left(I_x\omega_x + I_f\omega_f\right)\underline{i} + I_y\omega_y\underline{j} + (I_z\omega_z + I_r\omega_r)\underline{k}$$

$$\begin{cases} I_x\dot{\omega}_x + \left(I_z - I_y\right)\omega_z\omega_y + I_r\omega_r\omega_y + I_f\dot{\omega}_f = 0 \\ I_y\dot{\omega}_y + (I_x - I_z)\omega_x\omega_z - I_r\omega_r\omega_x + I_f\omega_f\omega_z = 0 \\ I_z\dot{\omega}_z + I_r\dot{\omega}_r + \left(I_y - I_x\right)\omega_x\omega_y - I_f\omega_f\omega_y = 0 \\ I_r\dot{\omega}_r = 0 \\ I_f\dot{\omega}_f + c\left(\omega_x + \omega_f\right) = 0 \end{cases}$$

where c is the viscous damping of the fluid, which creates a relative torque between fluid ring and satellite.
Linearizing the equations we obtain:

$$\begin{cases} I_x\dot{\omega}_x + (I_z - I_y)\bar{\omega}_z\omega_y + I_r\bar{\omega}_r\omega_y + I_f\dot{\omega}_f = 0 \\ I_y\dot{\omega}_y + (I_x - I_z)\omega_x\bar{\omega}_z - I_r\bar{\omega}_r\omega_x + I_f\omega_f\bar{\omega}_z = 0 \\ I_f\dot{\omega}_f + c(\omega_f + \omega_x) = 0 \\ I_z\dot{\omega}_z + I_r\dot{\omega}_r = 0 \\ I_r\dot{\omega}_r = 0 \end{cases}$$

The last two equations represent the motion around axis z, while the first three represent a damped motion in which the damping depends on c.
The root locus, as a function of c, has the following trace on the complex plane:

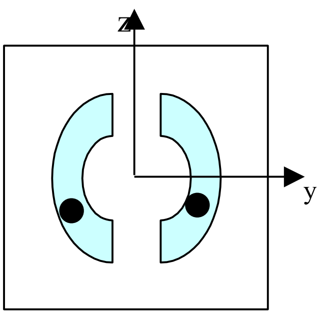


It can be clearly seen that an increase in the coefficient c does not mean a continuous increase in the system damping, there is in any case an optimal value of c that maximizes damping. If $\bar{\omega}_z$ is zero and $I_f \ll I_x, I_y, I_z$ then the optimal value of c is:

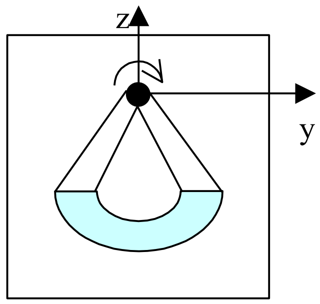
$$c = \frac{I_f I_r \omega_r}{\sqrt{I_x I_y}}$$

It can be seen that each satellite has its optimal value of c, depending on the inertia moments of the spacecraft. This means also that the optimal fluid is spacecraft-dependent. Since it is not always possible to adapt the fluid properties to the satellite, then optimality can be obtained by proper selection of the fluid quantity, through parameter I_f.

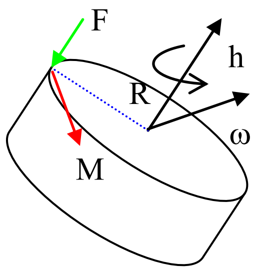
Popular damping configurations are realized through two partial fluid tubes containing a ball free to move inside the tube, increasing the damping:



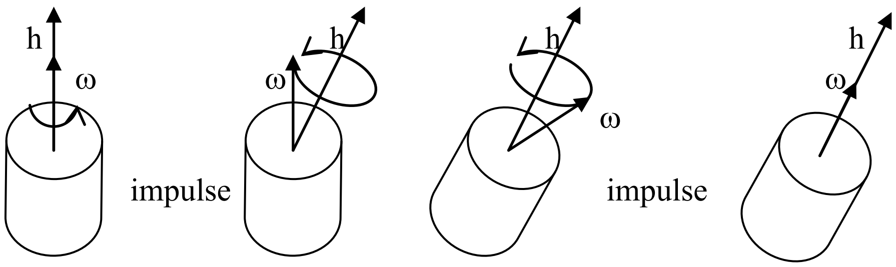
Alternatively, a pendulum configuration can be adopted:



A semi-active system to damp nutation, generated as a consequence of a spin axis reorientation maneuver, can act on $\underline{\omega}$ or on $\underline{h}$. To act on $\underline{h}$ a system based on thrusters can be set up, in order to align $\underline{\omega}$ with $\underline{h}$ and cancel nutation. This is realized by generating an impulse (activating the thruster) when the phase is correct, so that the torque M aligns $\underline{\omega}$ and $\underline{h}$:

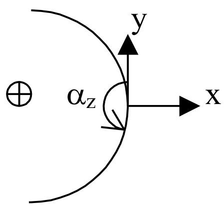


It is also possible to design a two-step maneuver, designed in such a way that at the end of the second step $\underline{\omega}$ and $\underline{h}$ are aligned:



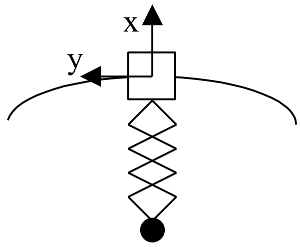
To damp libration oscillations, passive systems must damp α_z:

$$\ddot{\alpha}_z + 3K_{\ p}n^2\alpha_z = 0$$



Angular velocity is $n + \dot{\alpha}_z$. In this case a system based on absolute angular velocity is not useful, since this will never change sign if $n > \dot{\alpha}_z$ for any value of the angular rate $\dot{\alpha}_z$. So, a system based on $\dot{\alpha}_z$ only must be designed. For this, a permanent magnet immersed in a viscous fluid is suitable, if the permanent magnet tends to remain aligned with the radial direction.

Also for the libration damping a semi-active system can be designed:



Moving the tip mass along the radial direction we control the gravity gradient torque and so the libration oscillations.
One interesting semi-active damping mechanism is based on the conservation of angular momentum, in a configuration where the tip mass can be moved closer to the satellite body, starting from a given position. In the initial configuration, angular momentum is:

$$h_0 = I_z(n + \dot{\alpha}_{z0})$$

With the mass closer to the satellite angular momentum is instead:

$$h_1 = I_r(n + \dot{\alpha}_{z1})$$

Since h is constant:

$$I_z(n + \dot{\alpha}_{z0}) = I_r(n + \dot{\alpha}_{z1})$$

Furthermore, since $I_r<I_z$:

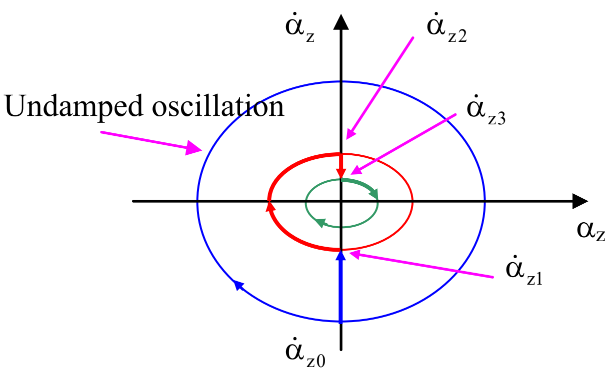
$$n + \dot{\alpha}_{z0} < n + \dot{\alpha}_{z1}$$
$$\dot{\alpha}_{z0} < \dot{\alpha}_{z1}$$

Assuming $\dot{\alpha}_{z0} < 0$ we have:

$$|\dot{\alpha}_{z1}| < |\dot{\alpha}_{z0}|$$

and the amplitude of oscillations, in the rotating orbit frame, is reduced.

Looking at the motion in the phase plane we have:



If the tip mass is repositioned at its original distance when $\dot{\alpha}_z$is positive, we have $\dot{\alpha}_{z3} < \dot{\alpha}_{z2}$. Notice that $\dot{\alpha}_{z2}$ is simply the opposite of $\dot{\alpha}_{z1}$, with same amplitude but changed in sign. So, considering negative $\dot{\alpha}_{z0}$ we have:

$$I_z(n - |\dot{\alpha}_{z0}|) = I_r(n - |\dot{\alpha}_{z1}|)$$

$$I_r(n + \dot{\alpha}_{z2}) = I_z(n + \dot{\alpha}_{z3}) \Rightarrow I_r(n + |\dot{\alpha}_{z1}|) = I_z(n + \dot{\alpha}_{z3})$$

If we impose $\dot{\alpha}_{z2}$ to be zero, then:

$$|\dot{\alpha}_{z1}| = \frac{(I_z - I_r)n}{I_r}$$

from which:

$$I_z(n - |\dot{\alpha}_{z0}|) = I_r\left(n - \frac{(I_z - I_r)n}{I_r}\right) = I_r n\left(\frac{2\mathrm{I}_r - I_z}{I_r}\right)$$

$$I_r = \frac{I_z(2\mathrm{n} - |\dot{\alpha}_{z0}|)}{2\mathrm{n}}$$

This provides the value of I_r, i.e. the closest position of the tip mass, required to damp the libration oscillations in only 2 maneuvers.

This maneuver, in the most favorable case, is completed in half libration oscillation, and in the worst case in 1.5 libration periods, that has a frequency in the order of $\sqrt{3}n(K_p \cong 1)$. Since there are 1.7 oscillations in one orbit, the overall maneuver is completed in more or less one orbit.

Spin rate damping

Active control with magnetic actuators

Active spin damping can be achieved with a feedback of the rate of change of the external magnetic field B_m on a set of magnetic torquers.

$$\underline{m} = -k_b \underline{\dot{B}_m},$$

This control is effective provided the spin rate of the satellite is sufficiently high, so that a change in the magnetic field is really caused by the spin and not by the change in satellite position along the orbit.

$$\underline{M_c} = \underline{m} \Lambda \underline{B} = -k_b \underline{\dot{B}_m} \Lambda \underline{B}.$$

The magnetic field can be represented in the principal axes of the satellite:

$$\underline{B} = A_{\text{rot}} \underline{B_{\text{orb}}}$$

and taking the derivative of the magnetic field in principal axes we have:

$$\underline{\dot{B}} = \dot{A}_{\text{rot}} \underline{B_{\text{orb}}} + A_{\text{rot}} \underline{\dot{B}_{\text{orb}}} \approx \dot{A}_{\text{rot}} \underline{B_{\text{orb}}}$$

The rate of change of the magnetic field is the sum of two terms, one due to the rotation of the satellite and one due to its change in position. We can assume that for high spin rates the term due to the change in position is negligible. Recalling that

$$\dot{A} = -[\omega\Lambda]A$$

we also have:

$$\underline{\dot{B}} = -[\omega\Lambda]\underline{B}$$

Evaluating the rate of change of the kinetic energy of the satellite $\dot{E}_k = \omega \underline{I} \dot{\omega}$, considering the satellite dynamics, the control law and the relation between magnetic field and its derivative we get to:

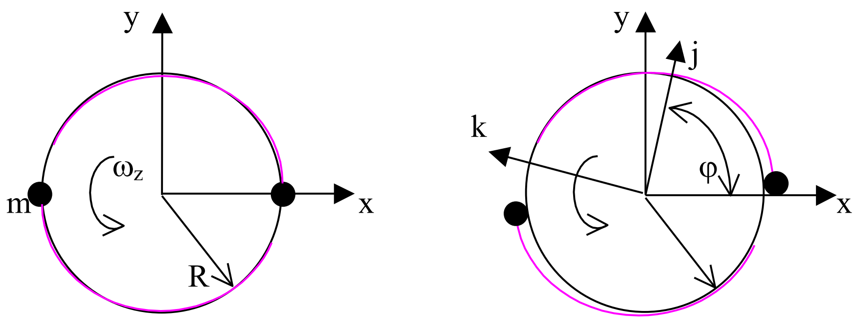
$$\dot{E}_k \approx \underline{\omega}^T \underline{M_c} = \underline{\omega}^T \left(\underline{m} \Lambda \underline{B} \right) = k_b \underline{\dot{B}}^T \left(\underline{\omega}^T \Lambda \underline{B} \right) = -k_b \underline{\dot{B}}^T \underline{\dot{B}},$$

which is negative, showing the spin rate of the satellite is reduced.

Passive control with yo-yo device

Passive control based on the so called yo-yo device is useful whenever, right after launch, the spacecraft must reset its spin speed due to launch. The mechanism is able to cancel the residual spin velocity, after which the spacecraft can start normal operation and precise control. The mechanism is based on the principle of conservation of angular momentum and kinetic energy, through a symmetric deployment of two small masses.
Assume as example a cylindrical shape and a spin axis aligned with the cylinder axis, labeled with z. If we have two tethers winded around the satellite, with a tip mass, the masses are fixed to the satellite

during launch and released after launch. As the masses are released, the tethers are deployed remaining tangent to the satellite lateral surface.



Consider the rotating frame (j,k) in which j defines the point of contact of the tether with the satellite; call φ the angle between the x axis and the j axis, identifying the deployment angle, and $\dot{\phi}$ the angular velocity of the rotating frame with respect to the body frame. The initial condition is:

$$h_0 = (I_z + 2mR^2)\omega_0$$
$$T_0 = (I_z + 2mR^2)\frac{\omega_0^2}{2}$$

where I_z is the inertia moment of the satellite only and mR^2 is the contribution to inertia due to one tip mass. In a generic time instant, the mass m has a velocity:

$$\underline{r} = R\underline{j} - R\phi\underline{k}$$
$$\underline{v} = \underline{\dot{r}} + \underline{\omega} \wedge \underline{r}$$
$$\omega = \omega_z + \dot{\phi}$$
$$\underline{v} = \begin{cases} R\phi\left(\omega_z + \dot{\phi}\right) \\ -R\dot{\phi} + R\left(\omega_z + \dot{\phi}\right) = R\omega_z \end{cases}$$

The angular momentum of one mass, aligned with z, will be:

$$h_m = \left(\underline{r} \wedge \underline{v}\right)m = \left(R^2\omega_z + R^2\left(\omega_z + \dot{\phi}\right)\phi^2\right)m$$
$$T_m = \frac{1}{2}m\underline{v} \cdot \underline{v} = \frac{m}{2}\left(R^2\omega_z^2 + R^2\left(\omega_z + \dot{\phi}\right)^2\phi^2\right)$$

Due to conservation of h and T with time, we can equate the initial condition to the generic condition to have:

$$h_0 = (I_z + 2mR^2)\omega_0 = I_z\omega_z + 2mR^2\left(\omega_z + \phi^2\left(\omega_z + \dot{\phi}\right)\right)$$
$$T_0 = (I_z + 2mR^2)\frac{\omega_0^2}{2} = \frac{I_z\omega_z^2}{2} + \frac{2mR^2}{2}\left(\omega_z^2 + \phi^2\left(\omega_z + \dot{\phi}\right)^2\right)$$

The two equations can be solved for the variables ω_z and $\dot{\phi}$, so that the deployment speed and the satellite angular velocity can be determined:

$$(I_z + 2mR^2)\omega_0 = (I_z + 2mR^2)\omega_z + 2mR^2\phi^2\left(\omega_z + \dot{\phi}\right)$$
$$(I_z + 2mR^2)\omega_0^2 = (I_z\omega_z^2 + 2mR^2)\omega_z^2 + 2mR^2\phi^2\left(\omega_z + \dot{\phi}\right)^2$$

To simplify the analysis, we can introduce a nondimensional parameter that indicates the ratio of the inertia of the masses to the overall system inertia:

$$a = \frac{(I_z + 2mR^2)}{2mR^2}$$

The equations become:

$$\begin{cases} a\omega_0 = a\omega_z + \phi^2(\omega_z + \dot{\phi}) \\ a\omega_0^2 = a\omega_z^2 + \phi^2(\omega_z + \dot{\phi})^2 \end{cases}$$
$$\Rightarrow \begin{cases} a(\omega_0 - \omega_z) = \phi^2(\omega_z + \dot{\phi}) \\ a(\omega_0^2 - \omega_z^2) = \phi^2(\omega_z + \dot{\phi})^2 \end{cases}$$
$$\Rightarrow \begin{cases} a(\omega_0 - \omega_z) = \phi^2(\omega_z + \dot{\phi}) \\ a(\omega_0 - \omega_z)(\omega_0 + \omega_z) = \phi^2(\omega_z + \dot{\phi})^2 \end{cases}$$
$$\Rightarrow \phi^2(\omega_z + \dot{\phi})(\omega_0 + \omega_z) = \phi^2(\omega_z + \dot{\phi})^2$$
$$\Rightarrow (\omega_0 + \omega_z) = (\omega_z + \dot{\phi})$$
$$\Rightarrow \dot{\phi} = \omega_0$$

We therefore see that the tether is deployed with constant angular velocity, equal to the initial satellite spin velocity. At this point we can substitute:

$$\phi = \omega_0 t (\phi_0 = 0)$$

so that the following is obtained:

$$a(\omega_0 - \omega_z) = \omega_0^2 t^2 (\omega_z + \omega_0)$$
$$\omega_z(\omega_0^2 t^2 + a) = a\omega_0 - \omega_0^3 t^2$$
$$\omega_z = \frac{a - \omega_0^2 t^2}{a + \omega_0^2 t^2}\omega_0$$

Now, since:

$$\phi = \omega_0 t$$
$$L = R\phi$$

with L indicating the length of the deployed tether, we can calculate:

$$\omega_z = \frac{a - R^2/L^2}{a + R^2/L^2}\omega_0$$

We can calculate either the time or the length of tether deployed that lead to ω_z=0:

$$t = \frac{\sqrt{a}}{\omega_0}$$
$$L = R\sqrt{a}$$

It is interesting to notice that the length of the tether does not depend on the initial velocity, so the system works correctly even in presence of large uncertainties in the spin rate. Notice also that, in order to have a non-zero final velocity there is a strong dependence on the initial velocity, so in this case errors are not corrected. This is the main reason why this system has been adopted only to cancel completely the spin.