Fundamentals of automatic controls

The term automatic control is typically related not only with the verification of a particular system, but also on the possible actions such that the system can behave in a desired way.

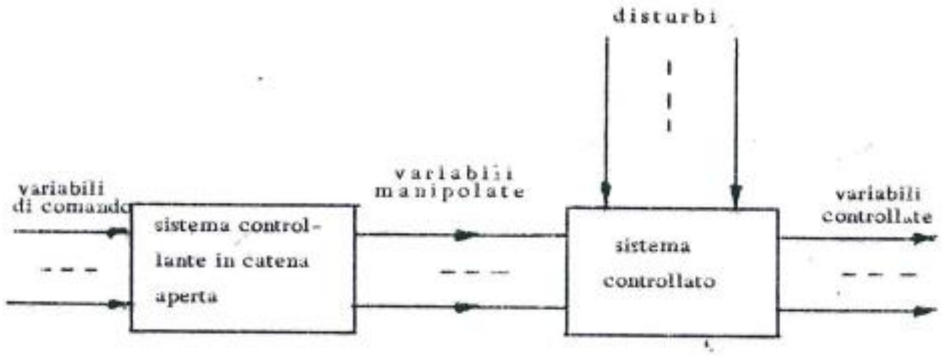
Generally speaking, one typically talks about a control system which consists of the physical system to be controlled (which is normally existent) and the controller which has to be designed.

Normally, the physical system has one or more than one inputs u and one or more outputs y, which we want to be as close as possible to their reference signals y_{des}.

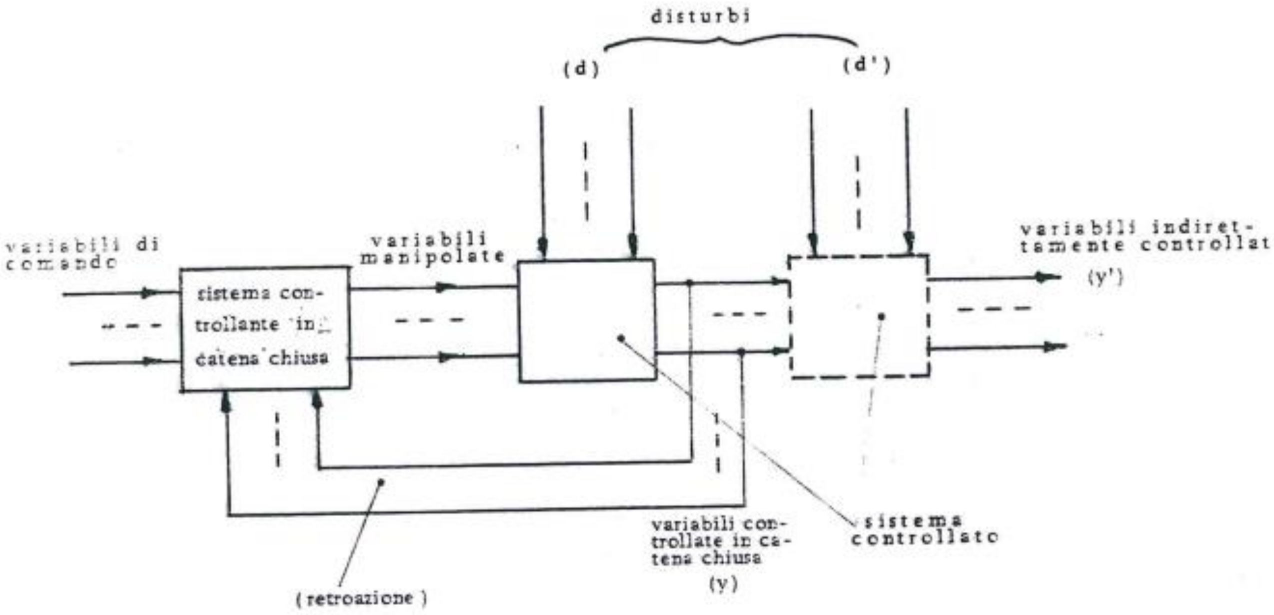
Typically we divide the inputs in two main categories: the ones which one can act on (which will be indicated as inputs u in a more strict sense) and the ones which cannot be controlled in any way, which will be called disturbances and will be indicated with the letter d.

Basically, there are two main control strategies:

- the open loop control, according to which the time profile of the inputs u will be chosen only on the basis of the desired behavior of y, no matter the action of the disturbances (or accounting for them with some prediction logics)



- The closed loop control in which the time profile of the inputs u will be chosen according to the real behavior of the outputs y, which depends both on u and on d. Such strategy is typically performed by means of the so-called feedback action.



The main difference between the two approaches is the following:

- The first relies on the model (as a matter of fact, this kind of strategy is always called model-based control strategy). The possible uncertainties are:
    - parametric uncertainties (i.e. the numbers which have to be inserted into the equations describing the process can never be known perfectly)
    - model uncertainties (the equations describing the physical process are always derived with some approximations)

So, in this first kind of control actions, we will assume that both parametric and model uncertainties will not affect the system behavior in a significant way. Moreover, we will suppose that the disturbance can be predicted in an effective way or that its action on the system will not affect it.

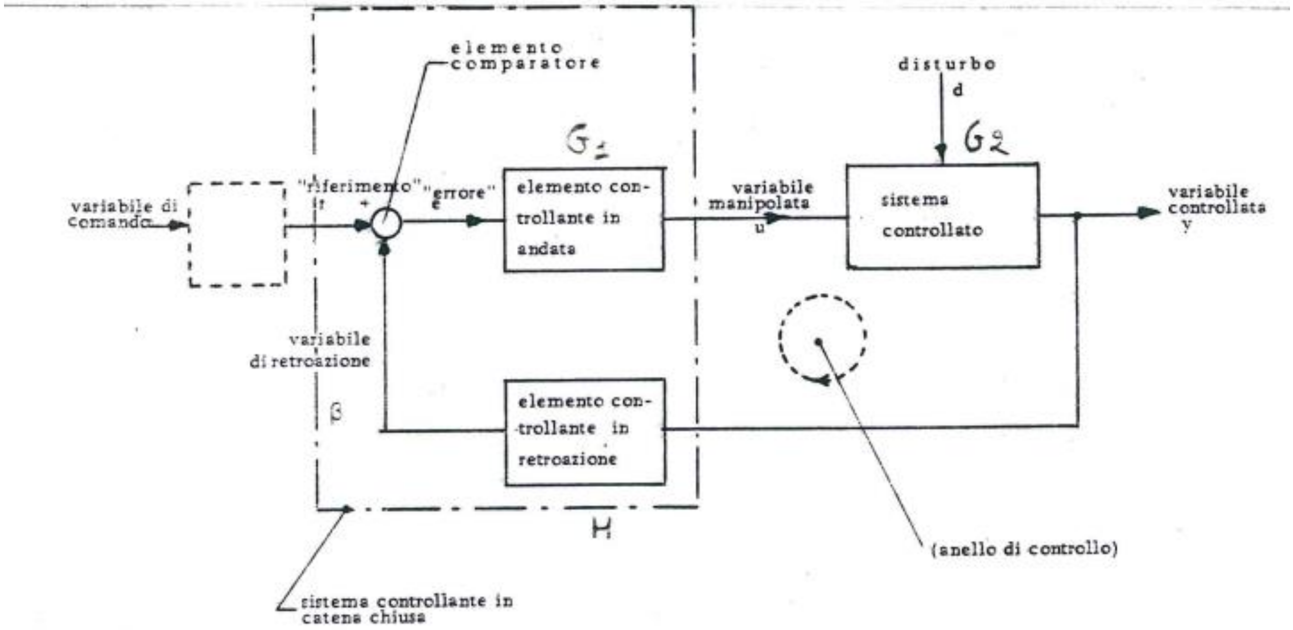
Let's do an example. Suppose we are driving our car with a speed equal to 70 km/h and we see that the speed limit is 50 km/h. So the problem is to define a suitable action in order to decrease the speed.

So, y is the actual speed, y_{des} is the value of the speed limit, u is the force we can act on the brake, d is the air resistance and the friction.

The open loop control strategy would consist of defining a suitable mathematical model of the problem (e.g. the equations of the constant acceleration motion) and calculate the force necessary to reduce the speed to 50 km/h within a distance lower than the one between our car and the speed limit.

The closed loop control strategy consists of continuously monitoring both the actual speed measurement and the speed limit and making a comparison between them in order to decide how to act on the brake until our speed becomes equal or lower than the limit value.

The great majority of control systems is typically performed with the second strategy. The main advantage is that we do not need a detailed model of the system, as the continuous feedback action allows to correct the control input u in order to make the output y track its reference signal y_{des}. The principle scheme of the closed loop strategy is depicted in the figure below.



The elementary theory of automatic controls typically refers to linear and time invariant systems, which produce linear differential equations with constant coefficients of the kind:

$$\begin{cases} \sum_{i=0}^{n} a_i y^{(i)}(t) = \sum_{i=0}^{m} b_i u^{(i)}(t) \\ y^{(i)}(0) = y_{0i} \quad i = 0..n-1 \end{cases} \tag{1.1}$$

Such equations can be transformed using the Laplace transform theory to produce the following algebraic relationship between the output and the input Laplace transforms, obtained starting from null initial conditions (i.e. $y_{0i}=0$).

$$\sum_{i=0}^{n} a_i s^i Y(s) = \sum_{i=0}^{m} b_i s^i U(s) \tag{1.2}$$

that is to say:

$$Y(s) = G(s)U(s) \tag{1.3}$$

having indicated with G the following quantity:

$$G(s) = \frac{\sum_{i=0}^{m} b_i s^i}{\sum_{i=0}^{n} a_i s^i} \tag{1.4}$$

which is typically addressed as **transfer function**.

Referring to the above figure, in the following will we indicate with G_1 the transfer function of the controller, with G_2 the transfer function of the system to be controlled (between the control input u and the output y) and with G_d the transfer function accounting for the disturbance d.

As a consequence, one can write:

$$\begin{aligned} Y(s) &= G_2(s)U(s) + G_d(s)D(s) \\ U(s) &= G_1(s)E(s) \\ E(s) &= Y_{des}(s) - H(s)Y(s) \end{aligned} \tag{1.5}$$

After some algebraic manipulations, one can obtain the following closed loop relationships.

$$\begin{aligned} Y(s) &= F(s)R(s) + F_d(s)D(s) \\ F(s) &= \frac{G_1(s)G_2(s)}{1+G_1(s)G_2(s)H(s)} \\ F_d(s) &= \frac{G_d(s)}{1+G_1(s)G_2(s)H(s)} \end{aligned} \tag{1.6}$$

As can be noticed examining (1.6), the feedback action changes the transfer function between u and y and between d and y. The new closed loop transfer functions are equal to the open loop ones divided by the same denominator $1+G_1(s)G_2(s)H(s)$. From a physical point of view, this means that the feedback action is able to change the system dynamics. This opens two kind of problems:

1. the analysis of a closed loop control system
2. the design of suitable transfer functions G_2 and H such that the resulting system satisfies certain properties. This problem is typically called control system synthesis.

The three main issues for a control system design are:

1. the steady-state behavior
2. the stability
3. the transient behavior

Steady-state behavior

With this term one typically refers to the system regime behavior when forced by step inputs

Recalling that the final value theorem states that:

$$\lim_{t\to\infty} y(t) = \lim_{s\to 0} sY(s) \tag{1.7}$$

if the time domain limit exists and recalling that :

$$\ell\{1(t)\} = \tfrac{1}{s} \tag{1.8}$$

it results:

$$\lim_{t\to\infty} y(t) = F(0)r + F_d(0)d \tag{1.9}$$

having indicated with r and d the reference and disturbance step amplitudes . As a consequence, recalling (1.6), the controller should act as:

$$F(0)=\frac{G_1(0)G_2(0)}{1+G_1(0)G_2(0)H(0)}\rightarrow 1$$
$$F_d(0)=\frac{G_d(0)}{1+G_1(0)G_2(0)H(0)}\rightarrow 0 \qquad (1.10)$$

in order to make the output as close as possible to its reference signal and to nullify the disturbance action.

This is typically achieved the imposing that:

$$\left|G_1(0)G_2(0)\right|\gg 1$$
$$H(0)\simeq 1 \qquad (1.11)$$

which can be obtained for example defining the controller transfer functions such as:

$$G_1(s)G_2(s)H(s)=\frac{K_h}{s^h}g(s)$$
$$g(0)=1, h\geq 0 \qquad (1.12)$$

Comments:

- the transfer function H typically describes the delay time due to the measurement system that allows to perform the comparison between the reference signal and the actual output. Such delay is often negligible and so we can assume that H=1.
- The transfer function G_2 describing the physical system is typically "proper", i.e. the denominator degree is greater or equal the numerator one. From now on, the denominator polynomial roots will be indicated as poles, while the numerator polynomial roots will be indicated as zeros.
- Examining (1.12), it readily follows that if the transfer function G_2 does not contain any poles in the origin, these will be inserted into G_1. But, as it is well-known that the division by s corresponds to and integration in the time domain, one typically calls a transfer function containing a pole in the origin "integral regulator".
- As will be clarified in the stability example, a pole in the origin could create problems of stability; such problems can be solved adding a zero in the function G_1. This corresponds to construct a controller transfer function of the kind $k_1\frac{s+a}{s}=k_1+\frac{k_1 a}{s}$, which basically contains an integral and a proportional term. For this reason, these kinds of regulators will be indicated as proportional/integral (PI).

Stability issues

Consider an autonomous nonlinear dynamical system

$$\dot{x}=f(x(t)),\quad x(0)=x_0,$$

where $x(t)\in\mathcal{D}\subseteq\mathbb{R}^n$ denotes the system state vector, $\mathcal{D}$ an open set containing the origin, and $f:\mathcal{D}\rightarrow\mathbb{R}^n$ continuous on $\mathcal{D}$. Suppose f has an equilibrium at x_e so that $f(x_e)=0$ then

- This equilibrium is said to be Lyapunov stable, if, for every $\epsilon > 0$, there exists a $\delta > 0$ such that, if $\|x(0) - x_e\| < \delta$, then for every $t \geq 0$ we have $\|x(t) - x_e\| < \epsilon$.
- The equilibrium of the above system is said to be asymptotically stable if it is Lyapunov stable and there exists $\delta > 0$ such that if $\|x(0) - x_e\| < \delta$, then $\lim_{t \to \infty} \|x(t) - x_e\| = 0$.
-
- The equilibrium of the above system is said to be exponentially stable if it is asymptotically stable and there exist $\alpha > 0, \beta, \delta > 0$ such that if $\|x(0) - x_e\| < \delta$, then $\|x(t) - x_e\| \leq \alpha \|x(0) - x_e\| e^{-\beta t}$, for all $t \geq 0$.

Conceptually, the meanings of the above terms are the following:

Lyapunov stability of an equilibrium means that solutions starting "close enough" to the equilibrium (within a distance δ from it) remain "close enough" forever (within a distance ϵ from it). Note that this must be true for any ϵ that one may want to choose.

Asymptotic stability means that solutions that start close enough not only remain close enough but also eventually converge to the equilibrium.

Exponential stability means that solutions not only converge, but in fact converge faster than or at least as fast as a particular known rate $\alpha \|x(0) - x_e\| e^{-\beta t}$.

An equilibrium point x_0 is unstable if there exists a neighborhood U such that, for any neighborhood V centered in x_0, one can always find an initial position $x \in V$ such that x becomes sufficiently far from the equilibrium point in order to go out from U

Stability for linear systems

In the linear time invariant systems, if the poles belong to the right half plane (of the complex plane), the system is unstable.
If such poles belong to the left half plane the system is said to be asymptotically stable.
Finally, if the poles belong to the imaginary axis of the complex plane with multiplicity equal to one, the system is said to be weakly stable which means that the system impulse response does not tend to zero but can be upper bounded by a real positive number.

Transient behavior:

For linear time invariant systems, the transient behavior can be easily analyzed simply by evaluating its close loop poles. In particular,

- the poles real part is an indication of the transient speed (i.e. how much time does the system require to approach its final value?)
- the poles imaginary part gives an indication of the kind of transient response (i.e. if the imaginary part is zero, then transient will be non-oscillating, while the amplitude of such imaginary part will give an indication of the rate of oscillation).

So consequently, the control system analysis consists of investigating the stability, transient and steady-state properties of the system.

The synthesis of the control system consists of designing the structure and parameters of the transfer functions G1 and H so that the final system will be stable and able to lead the output to its desired value with an "acceptable" transient behavior.

From a mathematical point of view, in a linear time invariant system, this correspond to find out the zeros of a n-degree polynomials for many different values of certain parameters.

Example

Let us consider the case of a synchronous generator. As well known, the no-load stator voltages form a three phase symmetrical system, as in the following equation:

$$\begin{aligned} e_{AA'}(t) &= \omega\phi_{max}(I_e)\sin(\omega t + \delta) \\ e_{BB'}(t) &= \omega\phi_{max}(I_e)\sin\left(\omega t + \delta - \frac{2}{3}\pi\right) \\ e_{CC'}(t) &= \omega\phi_{max}(I_e)\sin\left(\omega t + \delta - \frac{4}{3}\pi\right) \end{aligned} \quad (1.13)$$

being ω the electric speed of the rotor, $\phi_{max}(I_e)$ the maximum value of the magnetic flux function of the current I_e flowing in the rotor winding and δ a suitable phase angle with respect to a given reference.

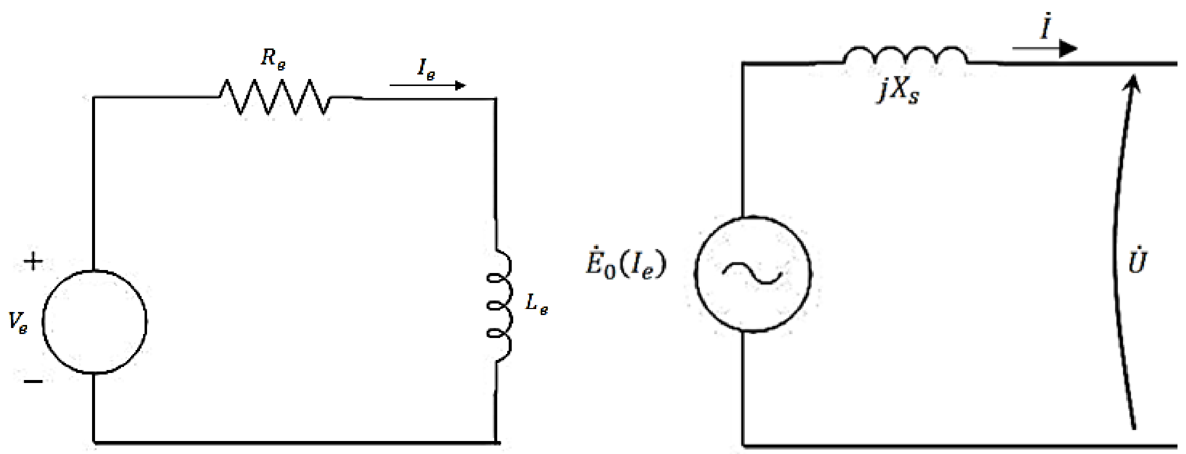
As a consequence, the corresponding phasor is given by:

$$\dot{E}_0 = \frac{\omega\phi_{max}(I_e)}{\sqrt{2}} e^{j\delta} \quad (1.14)$$

so, as one can notice

1. the amplitude of the generated voltages is mainly dependent on the rotor current (and so on the voltage Ve which feeds the rotor DC circuit)
2. the frequency of the generated voltages is function of the rotor speed.

The resulting equivalent one phase circuit is plotted in the following figure, in which Xs is the synchronous reactance and U is the voltage at the machine stator terminals.

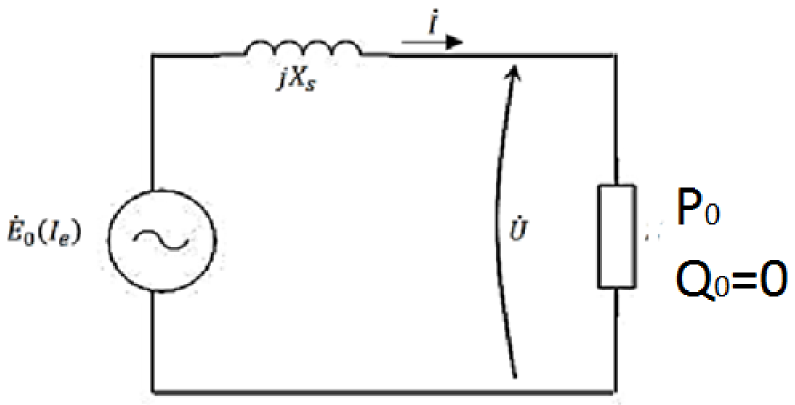


The above considerations imply that if one wants to change the voltage generated by the machine, it is necessity to act on the voltage V_e, while, in order to adjust the voltage frequency, it is necessary to change the machine speed. This can be achieved by acting on the prime mover power P_m, as the motion equation states that:

$$P_m - P = M\frac{d\omega}{dt} \tag{1.15}$$

being P the electromagnetic power, that is to say the power which is injected into the electric network connected to the machine itself.

Now, let us consider the following very simple situation in which a synchronous generator feeds a load by means of a connection cable (with negligible resistance and reactance).



Suppose that the load requires an active power P with power factor equal to one.

So, knowing the value of the current I_e, it is possible to evaluate the load voltage U and the powers generated by the machine, as follows:

$$\dot{I} = \frac{\dot{E}_0 - \dot{U}}{jX_s} \tag{1.16}$$

$$P_0 + jQ_0 = \dot{A} = 3\dot{U}\dot{I}^* \tag{1.17}$$

choosing the load voltage as phase reference and indicated with δ the machine voltage phase, one has:

$$\begin{aligned}\dot{A} &= 3\dot{U}\dot{I}^* = 3\dot{U}\left(\frac{\dot{E}_0 - \dot{U}}{jX_s}\right)^* = 3\frac{\dot{U}\dot{E}_0^* - \dot{U}\dot{U}^*}{-jX_s} = \\ &3\frac{UE_0e^{-j\delta} - U^2}{-jX_s} = 3\frac{UE_0\left(\cos(\delta) - jsin(\delta)\right) - U^2}{-jX_s} \\ &= 3\frac{UE_0}{X_s}sin(\delta) + j3\frac{UE_0\cos(\delta) - U^2}{X_s}.\end{aligned} \tag{1.18}$$

and so

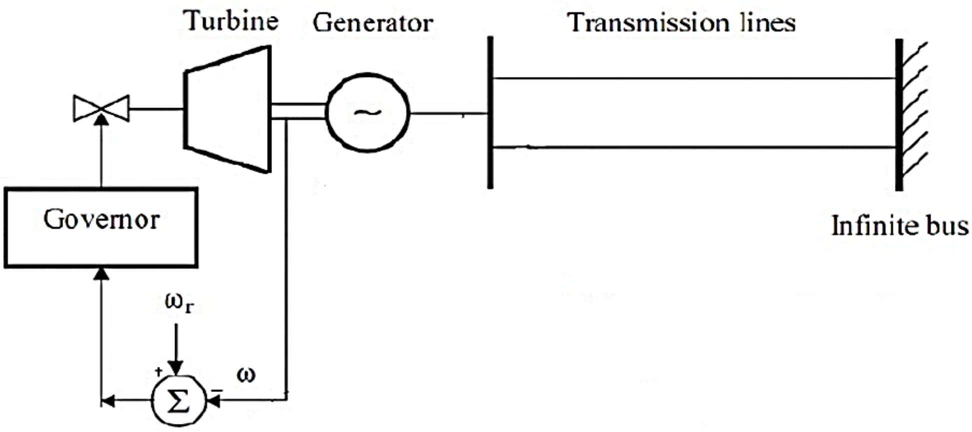
$$\begin{cases} P_0 = 3\dfrac{UE_0}{X_s}sin(\delta) \\ Q_0 = 3\dfrac{UE_0\cos(\delta) - U^2}{X_s} \end{cases} \tag{1.19}$$

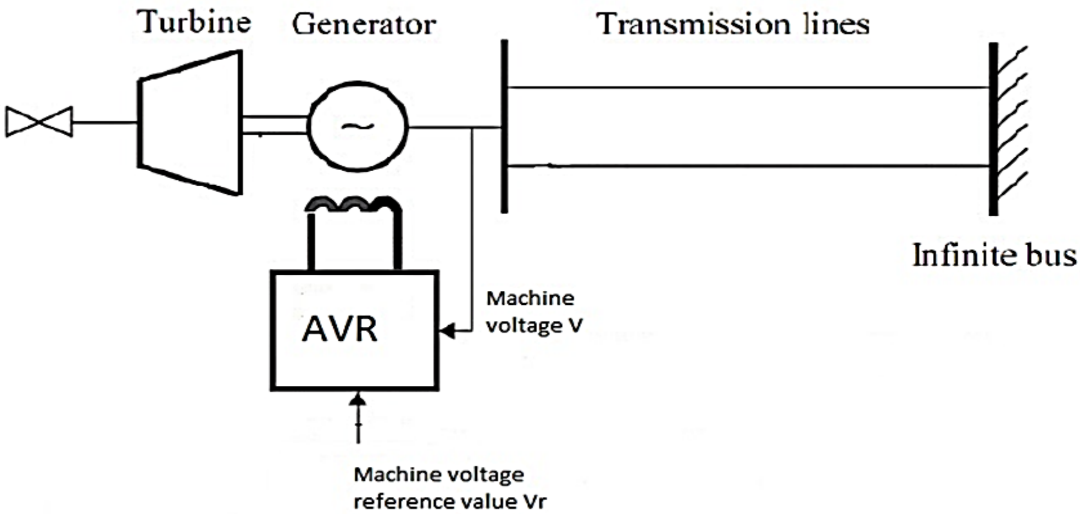
Taking the square of both members and summing up, one can obtain a 4° equation that can be solved in the unknowns U and δ.

Now, let us suppose that the load active power request decreases from P_0 to P. What happens to the system?

- The first immediate consequence is that the speed of the machine increases, and so the frequency does (as it is apparent examining (1.15) in the hypothesis of no changes in the prime mover power)
- the second consequence is that also the voltage at the load bus changes, and the current flowing in the sequence changes.

So, if we do not do anything, both the frequency and the voltage provided to the load change, which could be unacceptable. This claims for the definition of a suitable automatic control system in order to adjust both voltage and frequency. So, with reference to the previous notations, the output y consists of U and ω, the disturbance is the load power P, and the two system inputs are P_m and V_e. The two control systems that perform respectively the frequency and voltage regulations are called governor and automatic voltage regulator (AVR) (see figures below).





In order to apply the theory previously presented, we'll focus of our attention on the speed regulation.

If the pre-disturbance condition is a balance one, then the initial value of the prime mover power P_{m0} must be equal to P_0. In this case, the equations above do not allow to find out the initial value of the machine speed, which will be supposed to be equal to the desired angular frequency ω_0.

So, introducing the variations of all the quantities with respect to the pre-disturbance conditions, one can write:

$$\begin{cases} \Delta P_m - \Delta P = M \dfrac{d\Delta\omega}{dt} \\ \Delta\omega(0) = 0 \end{cases} \tag{1.20}$$

being

$$\begin{aligned} \Delta P_m &= P_m - P_{m0} \\ \Delta P &= P - P_0 \\ \Delta\omega &= \omega - \omega_0 \end{aligned} \tag{1.21}$$

with positions (1.21), the differential equation describing the system dynamics has null initial conditions and so can be treated as explained before.

The system transfer functions are given by:

$$G_2(s) = -G_d(s) = \frac{1}{Ms} \tag{1.22}$$

If H=1, one has:

$$\begin{aligned} Y(s) &= F(s)\Delta\omega_{des}(s) + F_d(s)\Delta P(s) \\ F(s) &= \frac{G_1(s)\dfrac{1}{Ms}}{1+G_1(s)\dfrac{1}{Ms}} \\ F_d(s) &= -\frac{\dfrac{1}{Ms}}{1+G_1(s)\dfrac{1}{Ms}} \end{aligned} \tag{1.23}$$

Being $\Delta\omega_{des}(s) = 0$, as we do not want the frequency to change.

Setting

$$G_1(s) = k \tag{1.24}$$

one has:

$$
\begin{aligned}
F(s) &= \frac{k\frac{1}{Ms}}{1+k\frac{1}{Ms}} = \frac{k}{Ms+k} \\
F_d(s) &= -\frac{\frac{1}{Ms}}{1+k\frac{1}{Ms}} = -\frac{1}{Ms+k}
\end{aligned}
\tag{1.25}
$$

Examining (1.25), it is apparent that the system results to be asymptotically stable for any k>0.

As far as the steady-state response is concerned, observing that:

$$
\begin{aligned}
F(0) &= 1 \\
F_d(0) &= -\frac{1}{k}
\end{aligned}
\tag{1.26}
$$

then

$$
\lim_{t\to\infty} \Delta\omega(t) = \Delta\omega_{des} - \frac{1}{k}\Delta P \tag{1.27}
$$

which means that the controller is not able to restore the speed (and so the frequency) to the initial value.

In order to verify the result, it is possible to solve the differential equation directly in the time domain:

$$
\begin{cases}
k\left(-\Delta\omega + \Delta\omega_{des}\right) - \Delta P = M\dfrac{d\Delta\omega}{dt} \\
\Delta\omega(0) = 0
\end{cases}
\tag{1.28}
$$

which provides:

$$
\Delta\omega(t) = \left(\Delta\omega_{des} - \frac{\Delta P}{k}\right)\left(1 - e^{-\frac{k}{M}t}\right) \tag{1.29}
$$

that confirms the above conclusions.

Suppose now to consider an integral controller, that is to say:

$$
G_1(s) = \frac{k}{s} \tag{1.30}
$$

in this case one has:

$$
\begin{aligned}
F(s) &= \frac{\frac{k}{s}\frac{1}{Ms}}{1+\frac{k}{s}\frac{1}{Ms}} = \frac{k}{Ms^2+k} \\
F_d(s) &= -\frac{\frac{1}{Ms}}{1+\frac{k}{s}\frac{1}{Ms}} = -\frac{s}{Ms^2+k}
\end{aligned}
\tag{1.31}
$$

whose poles are imaginary ($s = \pm i\sqrt{\frac{k}{M}}$) thus preventing the system from being asymptotically stable. As the time domain limit that not exist, is not possible to apply the final value theorem. The time domain differential equation is:

$$\begin{cases} k\int_0^t \left(\Delta\omega_{des}(\tau) - \Delta\omega(\tau)\right) d\tau - \Delta P = M\dfrac{d\Delta\omega}{dt} \\ \Delta\omega(0) = 0 \end{cases} \tag{1.32}$$

if the load power request change is a step, taking the time derivative of both members of (1.32), one has:

$$\begin{cases} k\left(\Delta\omega_{des} - \Delta\omega\right) = M\dfrac{d^2\Delta\omega}{dt^2} \\ \Delta\omega(0) = 0 \\ \dfrac{d\Delta\omega}{dt}(0) = \dfrac{-\Delta P}{M} \end{cases} \tag{1.33}$$

whose solution is:

$$\omega(t) = A\cos\left(\sqrt{\frac{k}{M}}t\right) + B\sin\left(\sqrt{\frac{k}{M}}t\right) + \omega_{des} \tag{1.34}$$

which permanently oscillates around the desired value (which is the physical meaning of the weak stability).

Suppose finally to consider the proportional integral controller, that is to say:

$$G_1(s) = k_0 + \frac{k}{s} \tag{1.35}$$

one obtains:

$$\begin{aligned} F(s) &= \frac{\left(k_0 + \frac{k}{s}\right)\frac{1}{Ms}}{1 + \left(k_0 + \frac{k}{s}\right)\frac{1}{Ms}} = \frac{k_0 s + k}{Ms^2 + k_0 s + k} \\ F_d(s) &= -\frac{\frac{1}{Ms}}{1 + \left(k_0 + \frac{k}{s}\right)\frac{1}{Ms}} = -\frac{s}{Ms^2 + k_0 s + k} \end{aligned} \tag{1.36}$$

The poles are:

$$s_{1,2} = \frac{-k_0 \pm \sqrt{k_0^2 - 4Mk}}{2M} \tag{1.37}$$

resulting negative for any k_0>0. As a consequence, now it is possible to make the system being asymptotically stable. The steady state behavior can be addressed by observing that:

$$\begin{aligned} F(0) &= 1 \\ F_d(0) &= 0 \end{aligned} \tag{1.38}$$

As a consequence:

$$\lim_{t\to\infty} \Delta\omega(t) = \Delta\omega_{des} \tag{1.39}$$

which shows that the PI regulator is able to restore the frequency.