

STABILITY AND OPTIMAL LOOPGAIN OF SIMPLE CONTROL CIRCUITS

M. VAN TOL

Philips' Research Laboratories, Eindhoven, The Netherlands

The intention of this paper is to bring forward once again the relation between the loopgain, the various time-constants of a control circuit and its stability. For the case of an idealized feedback circuit, where the process merely consists of a series of uncoupled RC-filters with different time-constants and where the controller has only pure proportional action, this relation can be expressed in a very simple formula: $A = \tau_1/\tau_2$.

Unfortunately this formula cannot be found in most textbooks on automatic control. This is indeed a great pity, for with the help of it all normal means used in control practice to get better control quality can be derived in a rather simple and elegant way. It makes it possible to arrange all these means in a coherent and surveyable scheme. Therefore this way of dealing with the problem has, in my opinion, a great didactical value. This does not mean, however, that it cannot be used in practice. On the contrary, it has originated from practical experience and it has been used in our laboratory many times already, with remarkable success.

DERIVATION OF THE BASIC FORMULA

The process is assumed to consist of some elements with different time-constants, and unity gain (Dynamic characteristics: $y_i/x_i = 1/1 + j\omega\tau_i$). The controller is a linear amplifier with gain A (Fig. 1).

The frequency characteristics of each process element, when drawn on double logarithmic paper, can be approximated by two straight lines, with an intersection point at $\omega\tau_i = 1$. The phase difference between input and output signals varies from 0° to -90° , and is exactly -45° for the frequency $\omega = 1/\tau_i$ (Fig. 2).

The frequency response function of the completed open loop now can easily be constructed by adding the functions of the different components together. Again, this function can be approximated by straight lines (Fig. 3).

Now it is supposed that the time-constants of the elements are really different (at least a factor of 5 between the successive τ 's). Then the phase difference as a function of ω too can easily be plotted through the points:

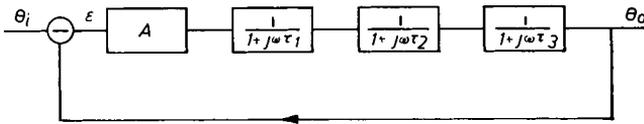


Fig. 1.—Block diagram of a simple closed loop.

-45° at $\omega\tau_1 = 1$, -135° at $\omega\tau_2 = 1$, -225° at $\omega\tau_3 = 1$, etc., with sufficiently good approximation.

The dynamic behaviour of the closed loop, expressed as a function of θ_i and of a disturbance D at the output of the process, can be written as:

$$\theta_o = \frac{AG}{1 + AG} \theta_i + \frac{1}{1 + AG} D,$$

G being $1: (1 + j\omega\tau_1) (1 + j\omega\tau_2) (1 + j\omega\tau_3)$, etc.

This means that, for minimum offset and minimum influence of disturbances, the loopgain A should be as high as possible. But of course this factor A cannot be chosen arbitrarily large for reasons of stability. It is quite clear that at a frequency where $\varphi = -180^\circ$, the gain should be diminished to less than unity (see Figs. 3 and 4). The higher the line of unity gain lies in Fig. 3, the better is the stability (less overshoot at step-function disturbances), but also, when this line comes higher and higher, A becomes smaller and smaller. So a compromise should be found here. Good results will be reached in choosing a so-called phase margin of 45° . This means that the loopgain equals 1 at a frequency where the phase difference is $180^\circ - 45^\circ = 135^\circ$. The overshoot at step disturbances then is only some 16 per cent.

From Fig. 5 it can be seen that, when choosing this phase margin of 45° , the unity gain line intersects the frequency response function exactly in the second breakpoint. So the loopgain A should be chosen just high enough, so that the gain has dropped to unity at $\omega_2 = 1/\tau_2$.

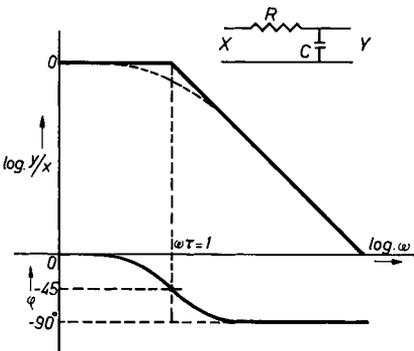


Fig. 2.—Logarithmic frequency and phase characteristics of a simple RC-filter.

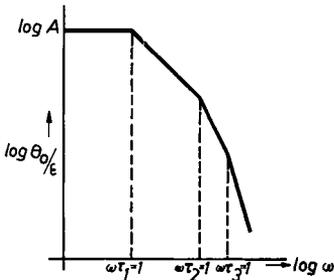


Fig. 3.

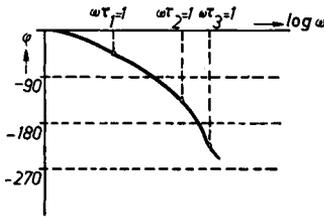


Fig. 4.

Figs. 3 and 4.—Logarithmic frequency and phase characteristics of an open control loop.

Now the frequency characteristic has, between ω_1 and ω_2 , a slope of minus 1, so the triangle PQR is isosceles. But this means, that $PR = QR$, so $\log A - \log 1 = \log \omega_2 - \log \omega_1$,

$$A = \frac{\omega_2}{\omega_1} = \frac{\tau_1}{\tau_2}$$

This is a very simple rule, stating that for excellent stability the loopgain A should be equal to, or smaller than, τ_1/τ_2 , τ_1 being the largest time-constant, τ_2 the largest but one. Of course this simple rule is only true for a simple idealized type of process and under the restriction that the third time-constant is sufficiently smaller than the second one. But in practice it gives a very good stronghold for a first approximation. On the other hand, this relation between the two main time-constants and the permissible loop-gain gives a powerful means for a good understanding of the measures that can be taken to improve the quality of an existing control system.

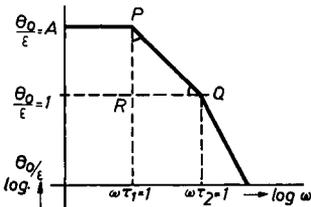


Fig. 5.—Logarithmic frequency response function of an open loop.

APPLICATION OF THE BASIC FORMULA

When the stability of a feedback loop is not sufficient, or when the system oscillates, it is always possible to lower the loopgain to a value where the stability is better. But in most cases the loopgain has then become so small that the main goal of the automatic controller is not reached: minimizing the offset. So a larger loopgain really is necessary. This means, however, according to the rule: $A = \tau_1/\tau_2$, that the “distance” between τ_1 and τ_2 should be made larger than it was originally.

Now it turns out that all normal means used in control practice to improve the stability and quality of control loops can be derived from this philosophy: all these means result in an enlargement of τ_1 or a reduction of τ_2 .

A. The simplest method to enlarge the distance between τ_1 and τ_2 of course is to enlarge τ_1 , the largest time-constant in the process. Sometimes this is possible and it is by far the easiest way to allow for more loopgain, and in this way improve the control behaviour. The speed of response of the system will not be lowered by this measure.

B. In some cases it is possible to make τ_2 smaller (the largest time-constant but one). This of course can give good results only when τ_3 is sufficiently smaller than τ_2 . In general this measure will make the system faster.

When it is not possible to alter τ_1 or τ_2 substantially, it will be necessary to use superficial means to obtain the same result:

C. One of the means of realizing an apparently larger τ_1 is the use of a so-called “phase-lag network”. This is an electrical (or pneumatical) network with a frequency response function:

$$\frac{y}{x} = \frac{1 + j\omega\tau_B}{1 + Bj\omega\tau_B}.$$

When choosing $\tau_B = \tau_1$, the factor

$$\frac{1}{1 + j\omega\tau_1},$$

representing the influence of the largest time-constant in the open-loop characteristic, is replaced by a factor

$$\frac{1}{1 + Bj\omega\tau_1}.$$

So the system will behave as if the largest time-constant were $B\tau_1$. This means that the loopgain may be raised by a factor B as well (Fig. 6).

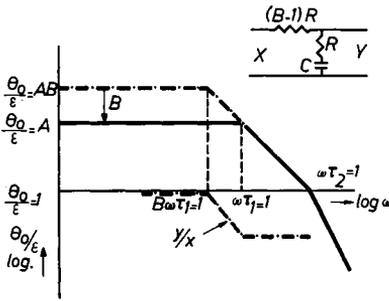


Fig. 6.

Fig. 6.—Correction with a phase-lag network.

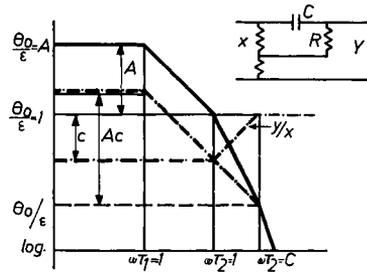


Fig. 7.

Fig. 7.—Correction with a phase-advance network.

D. In a similar way τ_2 can be reduced by the use of a phase advance network, with a frequency response function:

$$\frac{y}{x} = \frac{1}{C} \cdot \frac{1 + j\omega \tau_c}{1 + (j\omega \tau_c / C)}$$

This type of network sometimes is referred to as a damping network, because of its damping effect in the closed loop (when properly used). Here τ_c must be chosen equal to τ_2 , with the result, that in the open-loop characteristic τ_2 is replaced by τ_2/C , and the loopgain may be enlarged with a factor C (Fig. 7).

The results are essentially the same as when τ_2 really is made smaller, but there is one difficulty more: the filter gives a D.C. attenuation of a factor C , so that an additional amplification factor C is necessary. Sometimes this can give rise to noise problems.

In practice B and C have an order of magnitude of 10. Essentially the same effects as described under paragraphs C and D can be reached when integrating and differentiating action are incorporated in the controller:

E. In a PI-controller the deviation ϵ is not only amplified but in the same time it is integrated with respect to time. The response function of a PI-controller therefore is:

$$y = A \left[\epsilon + \frac{1}{\tau_i} \int \epsilon dt \right]$$

or, for sinusoidal signals:

$$\frac{y}{\epsilon} = A \left(1 + \frac{1}{j\omega \tau_i} \right) = A \frac{1 + j\omega \tau_i}{j\omega \tau_i}$$

Now integrators in practice are seldom real integrators (the amplification-

factor for D.C. signals is not infinite but $b \times$, $b \gg 1$). Therefore the frequency response function is in fact:

$$\frac{y}{\varepsilon} = A \frac{1 + j\omega \tau_i}{1/b + j\omega \tau_i} = Ab \frac{1 + bj\omega \tau_i}{1 + bj\omega \tau_i}.$$

This is exactly the same formula as can be derived for a phase-lag network. When choosing $\tau_i = \tau_1$, τ_1 is replaced virtually by $b\tau_1$. The difference with a phase-lag network is that b in most cases can be chosen considerably higher than B in phase-lag networks.

F. PD-controller. In principle here is $y = A (\varepsilon + \tau_d [d\varepsilon/dt])$ or, for sinusoidal signals: $y/\varepsilon = A(1 + j\omega \tau_d)$. Here, too, there is a practical limit, real differentiators stop differentiating at a certain high frequency, and therefore a more realistic expression is:

$$\frac{y}{\varepsilon} = A \frac{1 + j\omega \tau_d}{1 + j\omega \tau_d/c} \quad (c > 1).$$

This is the same function as was derived for a phase-advance network, again τ_d should be made equal to τ_2 . The only difference with a phase-advance network is that c in most cases is considerably higher than C . Both c and b are in the order of magnitude of $100 \times$.

In each individual case it will be necessary to decide which of the above-mentioned methods will be most suitable. Of course a combination of two or more is always possible for further improvement of the control quality. The best known combination is the PID-controller. Here τ_1 is made larger by the integrating action and τ_2 is made smaller by the differentiating effect. From the preceding discussion it will be clear that the right setting of a PID-controller should be $\tau_i = \tau_1$, $\tau_d = \tau_2$. This in itself is already a rule that can easily be observed and proves to give good results in many cases.

In this paper we have tried to make clear that all methods, well known in normal control practice, to improve the quality of the feedback loop, become natural and trivial means, when the formula $A = \tau_1/\tau_2$ (derived for an idealized case) is used as a starting point. For a proper appreciation of the value of the different methods possible and their mutual connection, this can be of great help to students in control theory.