EXAMINATION OF THE CLOSED CONTROL LOOPS ON THE BASIS OF SOME CHARACTERISTICS IN THE STEP RESPONSE OF THE OPEN CONTROL LOOPS

(PRELIMINARY REPORT)

By

A. FRIGYES

Department of Process Control, Technical University, Budapest

(Received September 9, 1969)

The conventional method of designing closed control loops is based upon the knowledge of the frequency-transfer properties of the blocks. The designing procedures developed in the practice of engineering have their specifications for the Nyquist-plot of the open control loop. E.g. they require that the phase-margin be at least 30° to 45° or that the gain margin be at least 6 dB. These specifications are shown by the Bode-plot in Fig. 1. Whether the design is made for the phase-margin φ_t or the gain-margin a_t, that section of the Nyquist-plot is decisive in which the amplitude ratio is a value near the unit (middle frequency range). The low frequency range of the Bode-plot ($\omega \ll \omega_c$) has little influence upon the stability of the closed control loop, it only determines the steady-state error and — on the other

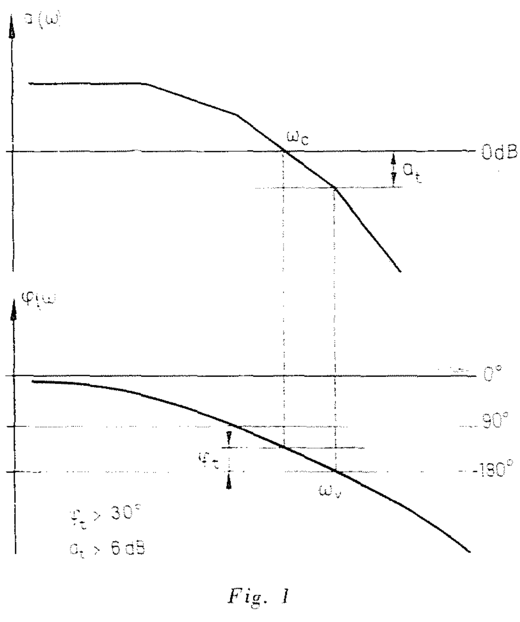


Fig. 1

hand — it is completely justified that the section of high-frequency range ($\omega \gg \omega_c$) is not taken into consideration at all.

To obtain the middle frequency range we draw the 0 dB axis on the log magnitude plot of the Bode-diagram and with it we determine the angular cut-off frequency ω_c. Its environment will be the "middle frequency range". The stability requirements, however, determine the middle frequency range, — whether the design is made under the consideration of the phase margin or of the gain-margin. In both cases, that part will be the middle frequency range (i.e. there will the amplitude ratio be about the unit) where the phase curve intersects the value —180°. The plot of the phase curve does not depend on the open loop gain. Thus it determines where the middle frequency range has to be placed, where the 0 dB axis of the amplitude plot has to be drawn, i.e. how great the open loop gain has to be.

Let us examine some properties of the phase plot, which will be important for the further considerations.

1. At $\omega \to 0$ the phase plot starts from the value k (—90°) where k means the multiplicity of the poles at $s = 0$ in the transfer function of the open loop $W(s)$ ($k \geq 0$, because $k < 0$ would mean that $W(s)$ has zeros in $s = 0$, which cannot practically occur). The amplitude curve of the Bode-plot starts from $\omega \to 0$ with a slope of $-k \cdot 20$ dB/decade.

2. At $\omega \to \infty$ the phase plot goes to the value l.(—90°) where l is the difference between the order of the polynomials in the denominator and numerator resp. of $W(s)$. The final slope (at $\omega \to \infty$) of the amplitude plot of the Bode-diagram is $-l \cdot 20$ dB/decade.

3. If the phase plot of a system lies above —180° in the whole frequency range, the system is structurally stable.

4. If the phase plot of a system lies below —180° in the whole frequency range, the system is structurally unstable.

5. The phase plot of a system at the limit of structural stability is —180° at any place (this is the only system which has an open loop with a transfer function $W(s) = K/s^2$).

6. Between the value $\varphi(\omega_x)$ belonging to ω_x of the phase plot and the slope of the amplitude plot there is a relation formulated by BODE, i.e.:

$$\varphi(\omega_x) = \frac{\pi}{2}\left|\frac{da}{du}\right|_{u=0} + \frac{1}{\pi}\int_{-\infty}^{\infty}\left\{\left|\frac{da}{du}\right| - \left|\frac{da}{du}\right|_{u=0}\right\}\ln\coth\left|\frac{u}{2}\right| du\,,$$

where a is the logarithm of the amplitude ratio, and $u = \ln \omega/\omega_x$.

From the above formula it is evident that the phase angle really depends on the slope da/du of the amplitude plot in the way that $\varphi(\omega_x)$ is decisively influenced by the da/du slope values in the environment of ω_x ($u = 0$). The influence of those values of da/du which are farther from ω_x is considerably reduced by the factor ln coth $|u/2|$ of the integrand.

7. A monotonously decreasing phase plot — unless the system is structurally stable or unstable — intersects the —180° line once, and thus the cut-off frequency ω_c and also the value K_{crit} of the open loop gain can be marked univocally. In this case the closed loop is stable if $K < K_{crit}$, and unstable if $K > K_{crit}$.

8. The intersection of a monotonously increasing phase plot with the —180° line — if there is such an intersection at all — determines the open loop gain (K_{crit}) as well in accordance with the previous paragraph. In this case, however, the closed loop will be stable if $K > K_{crit}$.

9. If the phase plot intersects the —180° line twice, there will be two critical open loop gains, $K_{crit\,1}$ and $K_{crit\,2}$. Depending on whether the phase plot between these two points of intersection is convex or concave, the closed system is stable in the range of

$$K_{crit\,1} < K < K_{crit\,2}$$

or

$$K < K_{crit\,1} \quad \text{and} \quad K > K_{crit\,2}$$

respectively, in any other case it will be unstable.

10. In such type 0 control systems where there are many equal or almost equal time constants, but none of them are much larger, the K_{crit} is relatively small. On the other hand, if the time constants considerably differ from each other, the K_{crit} will be large. In the first case the phase plot usually cuts the —180° line steeply, in the latter case less steeply. Generally the phase plot changes steeply where the second derivate of the log magnitude plot is large.

11. The cut-off frequency gives approximate information of the settling time t_s (i.e. the time in which, on the influence of the step function, the controlled system approaches the steady state value to the required extent). The settling time can be computed from the inequality

$$\frac{\pi}{\omega_c} < t_s < \frac{3\pi}{\omega_c} .$$

As ω_c is in the region of the intersection of the phase plot and the —180° line, the frequency ω_r belonging to this point of intersection determines approximate value of the settling time.

In many cases the step response of the controlled process is the starting point to the design of the control loop. In such cases the transfer function must be produced with some identification method so as to obtain the Bode-plot for the work of designing. The identification methods known nowadays require a considerable work of computation which, in addition, gives results of limited accuracy. At the registration of the transient function errors, measurement inaccuracies and noises — normally usual and tolerated in the practice of engineering — appear which, associated to the errors made at the numerical computation (mainly the repeated differentiations and integration), distort the transfer function resulting from identification, since false poles and zeros are obtained which can be screened out with great difficulties only. A further basic problem is given by the fact that the time range and the frequency range are connected by infinite limit integral formulas. The numerical computation, however, is unable by its character to reveal properly those relations between zero and infinite value for which the mathematical analysis gives exact formulas and which are used as important theorems at the work of designing on the basis of mathematical analysis and function transformation (Laplace—Fourier). Also the fact cannot be disregarded that the frequency method of designing control systems assumes such quality characteristics secondary from the viewpoint of the behavior in time (phase margin, gain margin, cut-off frequency, etc.) which are in close, but not univocal relation with the quality characteristics of the time range (overshoot, setting time, etc.).

Thus it is justified to ask whether it is worth carrying out the function transformation from the time domain to the frequency domain, which requires much of computation and includes considerable sources of error, and then making the design on the basis of the secondary quality characteristics of the frequency domain which is no fully univocal with the primary characteristics of the time domain.

Also such a designing procedure can be applied which uses quality characteristics given in the time domain and being in direct connection with the step response of the open control loop, even if they are secondary, supposed that they are not in less certain connection with the time and overshoot of the closed control loop than the characteristics of the frequency method are. The simplest tasks are to set the open loop gain, then to compensate by introducing the effect PI, PD.

Let us consider first the task of determining the open loop gain:

The step response $v(t)$ of the open control loop, except for a constant multiplying factor in it, is given. This means that neither the scale of $v(t)$ nor the position of the amplitude plot in the Bode diagram is given. Here the setting of the open loop gain is done by displacing the amplitude curve in the way that the cut-off frequency, i.e. the unit value of the amplitude ratio is at the frequency determined by the phase plot.

At designing directly in the time domain, the corresponding procedure would mark out such a time t_c to which some definitively given value of the step response belongs. In other words, the diagram is scaled e.g. in the way to get $v(t_c) = 1$. The problem is whether there is any function in the time domain which plays a role at marking out t_c, similar to the role of the phase curve of the Bode-diagram at choosing ω_c. There are such functions, e.g.

$$\varkappa(t) = \frac{t \cdot v'(t)}{v(t)}$$

$$\gamma(t) = \frac{t \cdot v(t)}{v_i(t)},$$

where

$$v_i(t) = \int_0^t v(t)\, dt .$$

Further examinations are needed to ascertain which of the above — and possibly other — functions is the most suitable to mark out the "middle-frequency" part of $v(t)$ in the time domain, i.e. that part which, similarly to the middle frequency range of the Bode-diagram, has a decisive influence upon the stability conditions and the choice of the open loop gain, and which determines the method of compensation and the parameters of the compensation block.

Among the above functions e.g. it is $\varkappa(t)$ that has properties more or less similar to those mentioned in connection with the phase curve of the Bode-diagram. Such properties of $\varkappa(t)$ are as follows (to facilitate comparison, these properties of $\varkappa(t)$ are marked with the same figures as the corresponding properties of the phase curve):

1. If $t \to \infty$ the value of $\varkappa(t)$ goes to k, where k is that highest exponent of the polynom $c_k t^k + c_{k-1} t^{k-1} + \ldots + c_0$ to which the step response goes if $t \to \infty$. (Thus k is the multiplicity of poles at $s = 0$ in the corresponding transfer function.)

2. At $t = 0$, the value $\varkappa(t)$ starts from l, which is the lowest exponent of the Taylor series of the step response at $t = 0$, i.e. the lowest derivative of the step response at $t = 0$ differing from 0. (In other words l is the difference between the order of polynominals in the denominator and numerator of the transfer function).

3. If in the whole time domain $(0 \leq t \leq \infty)$ the value of $\varkappa(t)$ is smaller than 2, the closed control loop is structurally stable (Fig. 2).

4. If $\varkappa(t)$ is larger than 2, in the whole time domain the closed system is structurally unstable (Fig. 3).

5. In a system that is structurally at the limit of stability (the step response of the open control loop is $v(t) = ct^2$. $\varkappa(t) = 2$ in the whole time domain (Fig. 3).

6. The value of $\varkappa(t)$ depends on the slope of $v(t)$, and is proportional to it. In addition it is proportional to $t/v(t) = t/\int_0^t v'(t)dt$, i.e. it depends also

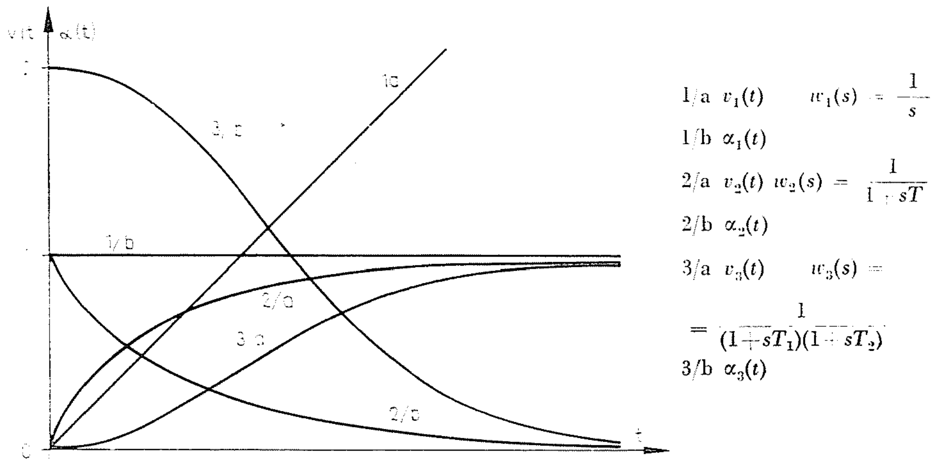


Fig. 2

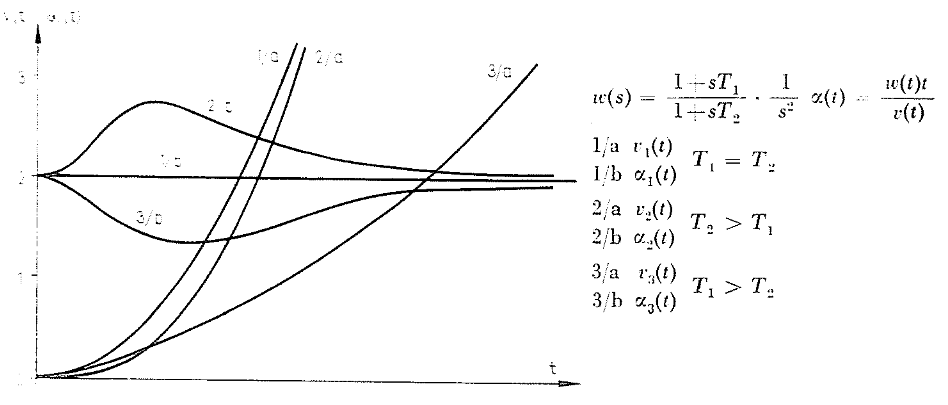


Fig. 3

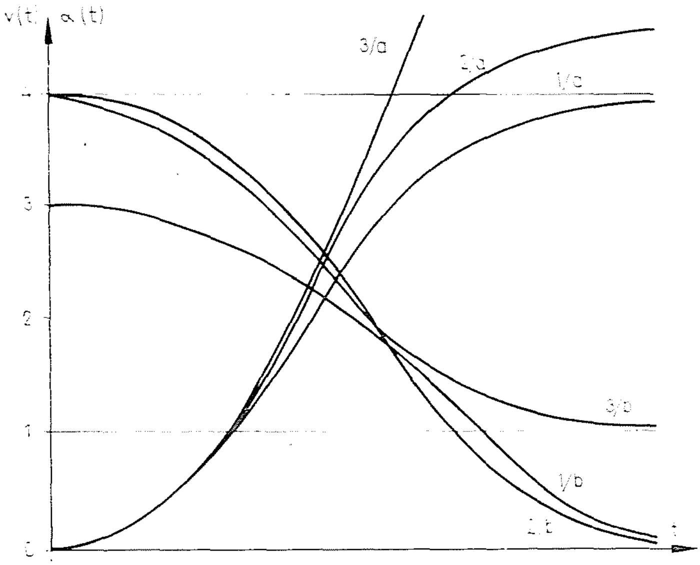
on the previous values of the derivate of the step response function (weight functions).

7. The monotonously decreasing $\alpha(t)$ — unless the system is structurally stable or unstable — intersects the line $\alpha = 2$ once.

If the step response of the open loop is such, that $\alpha(t)$ is decreasing in the whole time domain and the open loop gain (the scale of $v(t)$) is set in the way that the system is at the limit of stability ($K = K_{\text{crit}}$), the inequality

$$2.1 < \alpha(t_c) < 2.3$$

hold for that t_c with which $v(t_c) = 1$. In such a case the system is stable if $K < K_{\text{crit}}$ and unstable if $K > K_{\text{crit}}$ (Fig. 4).



1/a $v_1(t)$ $w_1(s) = \dfrac{K_{kr}}{\prod\limits_{i=1}^{3} (1+sT_i)}$
1/b $\alpha_1(t)$

2/a $v_2(t)$ $w_2(s) = \dfrac{K_{kr}}{\prod\limits_{i=1}^{4} (1+sT_i)}$
2/b $\alpha_2(t)$

3/a $v_3(t)$ $w_3(s) = \dfrac{K_{kr}}{S \prod\limits_{i=1}^{3} (1+sT_i)}$
3/b $\alpha_3(t)$

Fig. 4

8. The monotonously increasing $\alpha(t)$ — unless the system is structurally stable or unstable — intersects the line $\alpha = 2$. In this case there is such a K_{crit}, with which the system is stable if $K > K_{crit}$ and unstable if $K < K_{crit}$. The value of $\alpha(t_c)$ belonging to the t_c determined according to paragraph 7 is smaller than 2 and changes within broader limits.

9. If $\alpha(t)$ intersects the line $\alpha = 2$ twice, the closed system is stable in the ranges of the open loop gains of

$$K_{crit\,1} < K < K_{crit\,2}$$

or

$$K < K_{crit\,1} \text{ and } K > K_{crit\,2}$$

and it is unstable in any other case, depending on whether $\alpha(t)$ is convex or concave between the intersection points.

10. The value of $\alpha(t)$ changes more steeply at the time points t_c determined according to paragraph 7 in the case, when the time constants in the system are near to each other, than in the case if they differ considerably. Generally, at fixed values of $v(t)$ and $v'(t)$, the value of $\alpha(t)$ changes more rapidly, if the second derivate of $v(t)$ is larger.

11. The value of t_c determined according to paragraph 8 gives approximate information on the settling time t_s.

Summary

The conventional methods of closed control loop design are based upon the knowledge of the frequency-transfer characteristics of the systems.

The main disadvantage of these methods is the necessity to perform the required transformation from the time domain into the frequency domain which includes a number of error sources.

The direct time domain synthesis methods would be valuable tools in control system design for the engineers.

The first problem is to find some function in the time domain which points out the critical open loop gain like the phase curve of the Bode plot.

The paper presents some functions which show characteristics mentioned above.

The author acknowledges the assistance of Dr. P. Bakonyi.

References

BODE, H. W.: Network analysis and feedback amplifier design. Van Nostrand, New York—Toronto—London, 1955.

Prof. Dr. Andor FRIGYES, Budapest XI., Műegyetem rkp. 9. Hungary