Graphical Analysis of Control Systems

WALTER R. EVANS
ASSOCIATE AIEE

Synopsis: The purpose of this paper is to demonstrate some graphical methods for finding the transient response of a control system. A simple position follow-up system is considered for convenience although the method is applicable in the same form for higher order systems or those in which only empirical frequency data is known. The basic procedure is to find the roots of the differential equation which correspond to the exponential transient terms which dominate the response. Doctor Profos[1] of Switzerland points out that the plot of the function which describes the system from error to output is a function of a complex variable of which frequency is the imaginary part and damping is the real part. The Nyquist plot is thus one line of a conformal map with the root of the equation being the value of the variable which makes the function equal to -1. Any line of plot can be calculated for systems with known functions with essentially the same ease as the Nyquist plot by use of some graphical tricks. The amplitude of any transient term is determined from the plot once the root is known by use of a theorem of operational calculus. The development possibilities of the subject seem to be very great as suggested by several topics not yet investigated.

Review of Fundamentals

A quadratic system will first be analyzed in order to emphasize the important concepts in finding any transient response. Consider the position follow-up system shown in Figure 1.

The differential equation relating the output to the error is

$$K\epsilon=\left(1+T_M\frac{d}{dt}\right)\frac{d}{dt}\theta_o(t) \tag{1}$$

K is the output speed corresponding to a unit error. T_M is the time constant of motor acceleration, other delays are neglected. But input is the output plus the error.

$$\theta_i(t)=\theta_o(t)+\epsilon(t)=\left[1+\frac{1}{K}\left(1+T_M\frac{d}{dt}\right)\frac{d}{dt}\right]\theta_o(t) \tag{2}$$

Consider the input to be a unit step and note that the steady state value of output will also be unity. Assume that the output transient can be represented by exponential terms, so that $\theta_o(t)=Ae^{st}$ is substituted into the differential equation, and the common factor Ae^{st} cancelled.

$$0=1+\frac{1}{K}(1+T_Ms)s \tag{3}$$

Note that s appears at each point where d/dt had occurred before. This equation in s is an algebraic one, and any value of s which satisfies it represents an exponential term which can exist in the transient.

Anticipating the fact that s will replace d/dt when an exponential solution is assumed, the system itself can be more conveniently represented by the block diagram of Figure 2. The function in a block represents the ratio of its output to its input. The relationship between θ_i and θ_o can now be set up directly.

$$\frac{\theta_i}{\theta_o}=\frac{\theta_o+\epsilon}{\theta_o}=1+\frac{\epsilon}{\theta_o}=1+\frac{1}{K}(1+T_Ms)s \tag{4}$$

For this case of a quadratic equation the roots can be found directly by completing the square and solving for s.

$$s=-\frac{1}{2T_M}\pm j\sqrt{\frac{K}{T_M}-\left(\frac{1}{2T_M}\right)^2}=-\sigma_n\pm j\omega_n \tag{5}$$

The oscillatory case is taken because it is typical of the response of a fast control system. The transient solution is the sum of two exponentials, one for each root.

$$A_1e^{(-\sigma_n+j\omega_n)t}+A_2e^{(-\sigma_n-j\omega_n)t} \tag{6}$$

But this can be converted to a cosine function using the relation

$$e^{j\omega t}=\cos\omega t+j\sin\omega t \tag{7}$$

Select as the new undetermined constants the amplitude A and the phase angle ϕ

$$Ae^{-\sigma_nt}\cos(\omega_nt-\phi) \tag{8}$$

The constants are determined from the initial conditions that the output is zero and its rate of change is zero at time zero. The complete solution for output then becomes:

$$\theta_o(t)=1-\left(\frac{1}{\cos\phi}\right)e^{-\sigma_nt}\cos(\omega_nt-\phi);\quad \text{in which } \tan\phi=\frac{\sigma_n}{\omega_n} \tag{9}$$

The following numerical values are selected for convenience

$$K=2/\text{seconds}\quad T_M=1\text{ second} \tag{10}$$

Note that if T_M were equal to one-tenth second, the value of θ_i/θ_o given in equation 4 would be the same if s and K were both made ten times larger. Thus, the results of these problems can be shifted into any range of values with which the reader may be normally accustomed. Substituting the foregoing values gives

$$\theta_o(t)=1-1.07e^{-0.5t}\cos(1.32t-21°) \tag{11}$$

Graphical Plot to Locate Real Roots

The consideration of an additional delay in the control system raises the degree of the equation from second to third. In this case, consider the delay to be the time constant of the inductive build-up of current in the field of the generator supplying the motor. Setting up the ratio θ_i/θ_o

$$\frac{\theta_i}{\theta_o}=1+\frac{\epsilon}{\theta_o}=1+\frac{1}{K}(1+T_Gs)(1+T_Ms)s \tag{12}$$

The previous values of $K=2, T_M=1$ will be kept and T_G taken as one-fourth second.

Paper 48-85, recommended by the AIEE basic sciences committee and the joint subcommittee on servomechanisms and approved by the AIEE technical program committee for presentation at the AIEE winter general meeting, Pittsburgh, Pa., January 26–30, 1948. Manuscript submitted August 11, 1947; made available for printing December 29, 1947.

WALTER R. EVANS, assistant professor, department of electrical engineering, Washington University, St. Louis, Mo.

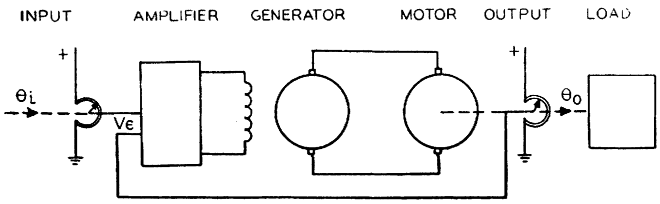


Figure 1. Position follow-up system

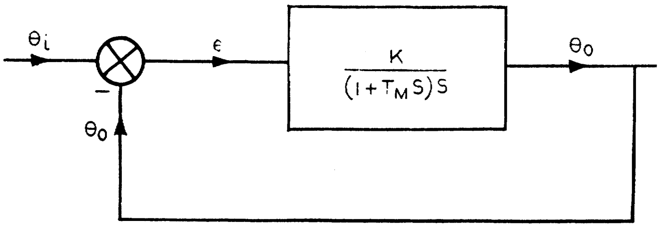


Figure 2. Simplified block diagram

Complex roots such as those which arose in the quadratic system always occur as a conjugate pair; therefore this third degree equation must have either one or three real roots. In finding the real value of s which makes the function of s of equation 12 equal to zero, one of the reasons for the simplicity of finding roots of a known function will become apparent. The factors of ϵ/θ_o are already known therefore, make $\theta_i/\theta_o = 0$ by making $\epsilon/\theta_o = -1$. In guessing s to be a negative real quantity $-\sigma$, each factor of ϵ/θ_o is a real quantity and can be plotted against $-\sigma$ as shown in Figure 3.

The range of $-\sigma$ from 0 to -1 gives a product of factors which is negative, but the magnitude is a fraction so that a root cannot exist in this region. The region beyond $-\sigma=-4$ certainly contains a root. A guess of $-\sigma=-5$ gives a result of $\epsilon/\theta_o = -2^1/_2$. A second guess of $-\sigma=-4^1/_2$ suggests itself to make $\epsilon/\theta_o=-1$ because the factor $1+T_G S$ is then cut in half whereas the other factors are nearly constant. The new result is $\epsilon/\theta_o=-63/64$ with further correction probably not justified by accuracy of the data. Note the calculation of ϵ/θ_o is very rapid by maintaining the identity of the factors, and that guesses of values of roots are immediately suggested.

Divide by the factor $s+4^1/_2$ to reduce the third degree polynomial to a quadratic. Solve for the complex roots of the quadratic as before to find $s=-0.25 \pm j1.30$. Apply initial conditions to determine the constants to find that the total response is given by

$$\theta_o(t)=1-0.08e^{-4.5t}-1.036\epsilon^{-0.25t}\times \cos(1.30t-27°) \qquad (13)$$

The main effect of considering the second delay is to cut down the damping rate of the oscillation.

Vector Plot of Error–Output Ratio For Sinusoidal Signals

A standard method of counteracting the effect of a time delay is to insert in the amplifier a circuit whose output includes the derivative of error. The form of its output/input function is shown in the block diagram of Figure 4.

The determination of the roots of such a system is generally agreed to be a tedious job by present methods and is infrequently done. The frequency response method instead has been developed to a fine point as indicated by the many recent articles on the subject.[2] In this method the feed-back loop is broken and a sinusoidal signal is impressed. The output is determined in magnitude and phase angle as a function of frequency. This information can be obtained by any one of several calculating methods or by direct laboratory tests. A convenient way to show the results is to plot the vector ratio of error/output. The plot for the foregoing system is shown in Figure 5 with the frequency identified by numbers in parenthesis.

For a stable system, the locus of ϵ/θ_o must swing outside the -1 point, physically meaning that the output/error ratio is a fraction at the frequency such that the output is in phase with the error. The vector ratio of θ_i/θ_o is greater than the ϵ/θ_o ratio by 1 and is therefore a vector to the curve with tail at the -1 point. Increasing the gain to $K=6$ contracts the curve as shown bringing it inside the -1 point indicating an unstable system. An intermediate value of gain $K=3$ contracts the curve in toward -1 point so that one would expect that the system was closer to becoming unstable, or having less damping. A quantitative value of damping, however, is now primarily a matter of experience based on systems with known vector plots and known transient performance. It is precisely at this point that the key idea of Profos becomes effective.

Determination of Principle Roots From Vector Plot

The key idea of P. Profos[1] is to consider the vector plot to be the base line from which the complex roots of the main damped sinusoidal term can be determined. Recall from the simple quadratic system that these roots make the ϵ/θ_o ratio equal -1. These roots could be determined if a plot of the function of s could be made as a function of both $-\sigma$ and ω.

One recognizes a somewhat similar situation in plotting an electrostatic field in which the complex variable involves flux and voltage. The patterns for lines of constant voltage and lines of

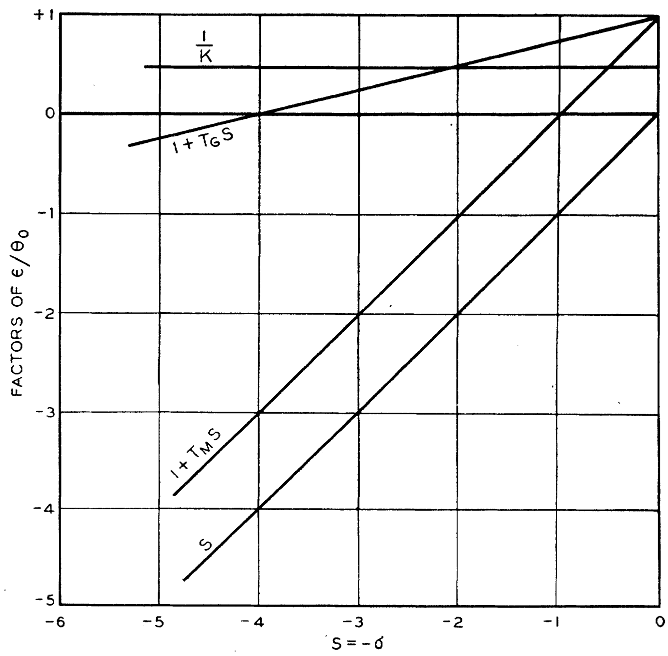


Figure 3. Plot to locate real roots

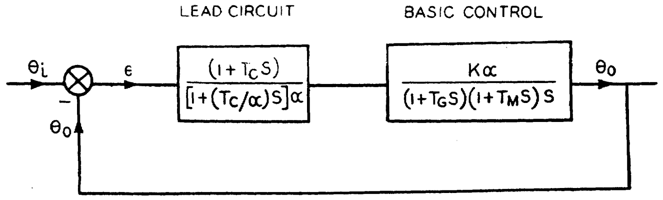


Figure 4. Block diagram of system. Lead circuit added to amplifier. Generator and motor delays considered

flux is known to be a grid of curvilinear squares. If this flux plotting property can be justified for ϵ/θ_o, the root values of $-\sigma_n$ and ω_n can be determined by sketching. The justification is that ϵ/θ_o is a function of s which has a particular derivative with respect to s for any value of s in the region of interest. Thus in Figure 6 the change of ϵ/θ_o along the curve for an increment $\Delta j\omega$ means that the derivative must be located 90 degrees clockwise from this increase. If the change in variable at this point were $-\Delta\sigma$ instead, the change in the ϵ/θ_o function would be opposite in direction to the derivative. The magnitude of the change of the function would be the same for equal small changes $-\Delta\sigma$ or $\Delta j\omega$. Thus for fairly large, but equal changes in $-\sigma$ and ω, a set of curvilinear squares will be formed as indicated by the dotted lines.

Note that this complex plot can be based on an original vector plot obtained from laboratory data as well as from one which is calculated. This is valid however, only for systems which are linear in the test range and the results are applicable only to this linear range. This restriction is necessary because the justification of the curvilinear square pattern is that the derivative of ϵ/θ_o is dependent only on the nature of the signal, not the amplitude.

One can sketch such a grid with more confidence after having calculated a few systems with known functional form. This can be done conveniently by some graphical tricks, as shown in the next section.

Calculation of Complex Plot

First consider the facts about present calculation methods as a function of frequency alone which simplify that task.

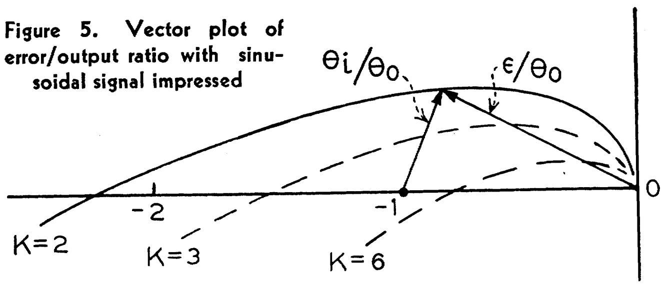


Figure 5. Vector plot of error/output ratio with sinusoidal signal impressed

The ratio of output to input for any one time delay is a vector quantity specified by amplitude and phase angle. The output to input ratio for several time delays in series is then the product of their amplitudes and a sum of their phase angles.

$$(1+j\omega T_1)(1+j\omega T_2)=A_1e^{j\phi_1}A_2e^{j\phi_2}=A_1A_2e^{j(\phi_1+\phi_2)} \quad (14)$$

If one keeps track of amplitudes logarithmecally, the product can be taken by adding logarithms.

$$\log A_1A_2=\log A_1+\log A_2 \quad (15)$$

The logarithmic scale commonly used is the decibel scale defined by the equation below.

$$\text{decibel}=20\log_{10}A \quad (16)$$

The new problem introduced is to express any one term $1+T_S$ as an amplitude and phase angle. Several schemes are of course possible, but the following is believed to be the most convenient. Consider the term to be factored as shown below with the T factors saved for later consideration.

$$1+Ts=T\left(\frac{1}{T}+s\right) \quad (17)$$

The term s can be located as a vector from the origin. The complete vector is one with tail at the $-1/T$ point and head at the s point. This vector can be measured by a protractor pivoted at the $-1/T$ point with scale in line with the s point. The amplitude is desired in decibels so the scale is so marked. This scale can be checked by noting the magnitude of 1 is marked as 0 decibel and the magnitude of 2 as 6 decibels in Figure 7.

The case of a quadratic term in the ϵ/θ_o ratio frequently arises. Such a term can be broken into two factors having conjugate complex roots as shown below.

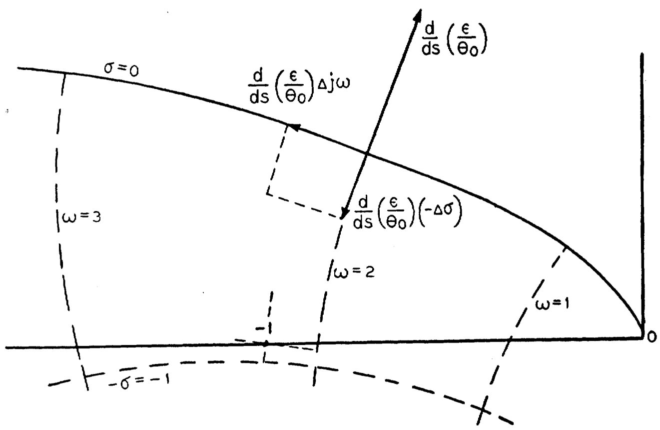
$$S^2+as+b=[S-(-\alpha+j\beta)][S-(-\alpha-j\beta)] \quad (18)$$

These conjugate roots are thus the pivot points for the protractor in making measurements to the s point.

The procedure is now simply one of tabulating decibels and angles for each term, adding decibels for the total decibels, and adding angles for the total angle of the ϵ/θ_o ratio. The over-all amplitude factor for the product can now be established by checking the special case of s is equal to 0. The value of ϵ/θ_o for $s=o$, eliminating K and s factors is 1 as shown in the block diagram. The sum of the decibels reading with protractor swung to the origin actually obtained is due to all of the factors accumulated in setting up the $1/T+s$ vectors. This decibel value should be subtracted from all sums obtained for other values of s. Division by the gain K is achieved by subtracting the decibels for that K from the net decibels previously obtained. The ϵ/θ_o vector is then known as an angle and decibels of magnitude and so can be plotted using the same protractor.

It is convenient to make the first calculation for constant values of $-\sigma$ and the uniform changes in $j\omega$. Thus a curvilinear square pattern should be formed with any mistake in calculating one point showing up by its failure to be in line with the grid set up by the other points. The plot for the system of Figure 4 is shown in Figure 6 for $T_C=1/2$ and $K=4$. The

Figure 6. Complex plot for system. Conformal map properties are indicated. Root at $s=-\sigma_n\pm j\omega_n=-0.95+j2.25$



value of s which makes $\epsilon/\theta_o=-1$ is $s=-\sigma_n\pm j\omega_n=-0.95\pm j2.25$. The amplitude of the transient having this root as well as its initial phase angle, needs to be determined before the transient can be plotted. Fortunately this is possible without need for finding the rest of the roots and substituting initial conditions by means of a theorem of operational calculus.

Amplitudes From Operational Calculus

The concept of s thus far presented has been simply that of being a complex number used in the exponential solution of the differential equation. This concept is sufficient for explaining the process of finding the roots to the equation. The symbol s, however, has a more potent significance as it is used in operational methods such as the Laplace Transform.[3] Those familiar with these methods of course realize that this is a long study in itself, but essentially only one fact need be used. The amplitude of any transient term is given in terms of its root by

$$A_1=\left.\frac{1}{s\dfrac{d}{ds}\left(\dfrac{\theta_i}{\theta_o}\right)}\right]_{s=s_1} \tag{19}$$

But the derivative of the function can be determined directly from the complex plot. It is vector with an angle of 84 degrees as shown in Figure 6. The magnitude is the change in ϵ/θ_o divided by the change in s. An average of several measurements in the region of the -1 point gives the value of 0.60.

The two transient terms involving the pair of complex roots can be converted into a single term as in the simple quadratic system. The conversion is made very rapid by noting that the amplitude terms are conjugates of each other. In taking the sum therefore their real parts add and their imaginary parts cancel. The result can thus be written as twice the real part of one of them.

$$2R\left[\frac{e^{-\sigma_n t}e^{j\omega_n t}}{ae^{j\alpha}be^{j\beta}}\right]=\frac{2e^{-\sigma_n t}\cos\ (\omega_n t-\alpha-\beta)}{ab} \tag{20}$$

in which

$$ae^{j\alpha}=s=-\sigma_n+j\omega_n$$

and

$$be^{j\beta}=\left.\frac{d}{ds}\left(\frac{\theta_i}{\theta_o}\right)\right]_{s=-\sigma_n+j\omega_n}$$

The other two roots to this fourth degree system are found to be real roots of -2.58 and -16.6. The derivatives can be found by any one of several methods but from the slope of the function they are -1.45 and $+1.66$, respectively. The complete solution including the steady state value of 1 becomes:

$$\theta_o(t)=1-1.36e^{-0.95t}\cos\ (2.25t-17°)+0.267e^{-2.58t}-0.036e^{-16.6t} \tag{21}$$

Procedure For Complicated Systems

A solution of a fourth degree equation is nothing new, but this method can readily be applied to higher order equations. Thus additional time delays or stabilization circuits will distort the frequency locus but the conformal map can still be sketched.

The quadratic roots found from the original plot represent the transient term which will frequently dominate the complete transient response. If a complete solution is desired, however, these known factors can be divided out of the θ_i/θ_o ratio. Thus, for any value of frequency, the angle and decibels readings for the protractors pivoted at the conjugate root points, with the scale at the frequency point, should be subtracted from those of original vector to obtain those of the new vector. Note that in general the portion of the original vector plot for higher frequency values will now be contracted into the -1 region. Thus, one would have to have frequency data in the region of the higher order roots in order to find them. Usually the corresponding transient term will be found to damp out rapidly so that they would be needed only to study the initial break away of the output. It is probable that other methods would be more applicable to studying this region.

The construction of a vector plot for a multiple loop system can be readily carried out on a completely vector basis. The multiplication process has already been shown, and the addition process is simply the familiar completion of a parallelogram. Many time saving tricks are possible however by shifting or rotating an entire plot with respect to its previous position. The predominance of feed-back signals over feed-forward signals makes the inverse plot more convenient, for starting at the output one can build up a diagram back to the input step by step.

Occasionally systems will have two pairs of quadratic roots in about the same frequency range. The Nyquist plot then circles the -1 point with the result that a plot built up from one side overlaps a plot built up from the other side. This situation can be handled by a process of successive approximations. Start with the pair of complex roots suggested by the plot built from one side and divide the function by these factors. The resultant plot will usually give good indication of the other pair of roots since overlapping is eliminated. The first pair of roots may now be determined more accurately by dividing the original plot by the factors corresponding to the second pair of roots.

For systems in which the damping is near critical, the behavior of the plot near the origin must be understood. A good example to work out is for a simple quadratic with roots of $-1\pm j^1$. The first "square" is shaped like a triangle with base from 0 to -1, and curved sides crossing at $-1/2+j^1$. The missing corner will be found to be at the midpoint of the 0 to -1 line.

Protractor measurements could be eliminated by substituting potentiometers at the pivot point which turn in resistance proportional to angle. The decibel readings could similarly be replaced with potentiometers which are turned as a logarithmetic function of the distance to the s point. Connecting each set of potentiometers in series will then give total angle and total decibels as resistances. These resistances could then be used as inputs to instrument servos which could actuate meters or locate the position of a pen.

Development Possibilities

What are the possibilities for new types of laboratory tests to give empirical data other than just frequency response? The problem would seem to be primarily one of achieving a steady state condition long enough to get a reading and not have the natural transient of the system present. An exponential build-up might work in that starting from zero the transient should not appear and the only reading necessary is the ratio of output to input, which should be constant during the build-up. An exponentially increasing sine wave would correspond to a complex value of s, though the reading here would have to include phase angle as well as magnitude.

The similarities of the conformal mapping properties of static fields and these functions of s seem to offer the most interesting possibilities. Conceivably the known system information could be set up as boundary conditions of an electrostatic field so that the pattern of equipotential lines and lines of flux would correspond to the system plot. The system plots of the type described in this paper

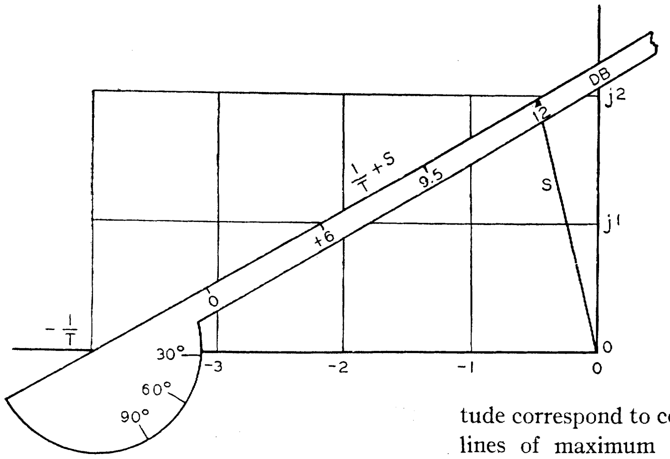


Figure 7. Vector determination of 1/T+s

however, have several values of s which makes the ϵ/θ_o equal to -1 where as only one voltage could exist at the -1 point for the field. Another system plot is possible however in which the roles of ϵ/θ_o and s are reversed. Any point on the plane would correspond to a value of s and ϵ/θ_o is plotted as loci of its constant real part and constant imaginary part.

A physical picture of the plot can be gained by considering a surveying contour map in which lines of constant altitude correspond to constant real part and lines of maximum slope correspond to lines of constant imaginary part. On this basis a quadratic function in the region of the roots has the shape of a horse's saddle. Imagine a view of the saddle from above with the origin at the horn of the saddle and the $-\sigma$ axis running straight back to the rear. One trace of zero imaginary part is along this axis and the other intersects it at right angles at the center of the saddle. Picture now the saddle to be immersed in water and gradually lifted out. The water level lines on the saddle give the locus for a real part of $\epsilon/\theta_o=-1$ as the gain of the servo is increased. In this situation the roots of the transient equation lie at the intersection of the locus for a real part of $\epsilon/\theta_o=-1$ and the locus for zero imaginary part. These roots therefore move down the saddle from the horn and the rear until they meet in the center for the critical damping case. Further increase in gain, or raising of the saddle, causes the roots to appear as a conjugate pair moving down each side toward the stirrups.

Will a static field be capable of mapping such a pattern? If so, what boundary conditions will be necessary? This seemingly never-ending chain of questions serves to keep the subject interesting, since one does not know what useful method might be encountered in the process of finding the answers.

References

1. A New Method for the Treatment of Regulation Problems, **P. Profos.** *Sulzer Technical Review* (New York, N. Y.), number 2, 1945.

2. Automatic Control Engineering (book), **E. S. Smith,** McGraw-Hill Company, New York, N. Y., 1944.

3. Transients in Linear Systems (book), **M. F. Gardner, J. L. Barnes.** John Wiley and Sons New York, N. Y., 1942.

No Discussion