9 February 1962, Volume 135, Number 3502 SCIENCE

CURRENT PROBLEMS IN RESEARCH

Automatic Process Control

Coupling large-scale computers with process systems makes possible the fully automatic process plant.

Ernest F. Johnson

If we define *control* in the broadest sense as the organization of activity for a purpose, and if we set aside in this discussion any teleological consideration of the physical universe, we must conclude that the earliest controlled processes were living organisms. The purpose of these primal control systems was only survival, but with increasing diversification of species and ultimately with the emergence of the human species and subsequent social organization, varieties of control systems and subsystems with equally varied specific purposes have proliferated almost without limit. These systems include not only biological systems but mechanical systems; physical, chemical, and nuclear process systems; and sociological systems, including economic and political systems. Although we shall focus our attention on automatic process control systems, all control systems have the same basic characteristics, and knowledge about any particular class of control systems is applicable in principle to all other classes of control systems.

Process control systems are systems which involve, under controlled conditions, physical processes such as the flow of fluids and the transfer of thermal energy, and chemical and nuclear processes such as the manufacture of ammonia and the fission of uranium for power generation. If human operators manipulate the valves of other control elements, the control is called manual control, but if machines do the manipulating without direct human assistance, the control is called automatic process control (*1*).

The author is professor of chemical engineering at Princeton University, Princeton, N.J.

Characteristics of Control Systems

The common characteristics of all control systems can be identified by considering the simplest kind of automatic process control system. Figure 1 shows a well-stirred tank which is being used to heat a fluid by means of steam condensing in the jacket on the tank. The specific purpose of this system is to hold the temperature of the effluent at the value fixed by the set-point adjustment in the controller, regardless of fluctuations in feed temperature or flow rate.

It is impossible to hold the temperature of the effluent absolutely constant and equal to the set point, but, if the control system has been designed well, disturbances to the system will be ironed out quickly. Indeed, the best simple gauge of the control performance of the system is the manner in which the system variables recover from disturbances.

Thus, a fundamental characteristic of the system is the fact that it is a dynamic system—that is, its behavior varies in time, and this temporal behavior provides an index of the system's performance.

A second characteristic of control systems is that they are information processing systems. They obtain information, digest information, and generate information. Just as we use process-flow sheets to keep track of the flows of material and energy, as in Fig. 1, so may we use a signal-flow diagram to keep track of the flow of information in a control system. The signal-flow diagram for the heat-exchange process in Fig. 1 is shown in Fig. 2. There are a variety of types of signal-flow diagrams. The one shown here is a simple block-type diagram.

All signals on this diagram are designated by θ. The set point or command signal θ_r is fed to a summer, where it is compared with the measure θ_c, which in this case is the actual temperature of the effluent at the point of measurement. The difference between the set point and the signal θ_{c_1} is called the deviation, or error signal. It is the forcing signal to elements in the controller. These elements in turn generate controlling actions, which may depend on amplifying the error signal, or on integrating or differentiating the error. The signal from the controller θ_o forces the final control elements, which in our example consist of a valve motor or actuator and a valve which moderates the flow of steam to the jacket on the tank. A disturbance signal corresponding to a change in steam pressure in the line ahead of the control valve is shown entering the system as θ_{u_1} at a summing point ahead of the process unit. A change in feed flow rate or temperature is shown entering the control system as θ_{u_2} at a point between the process unit, which comprises the thermal transfer from jacket to stirred fluid, and the time-delay element, which is the contribution to the performance of the system made by the discharge line from the tank to the point of temperature measurement.

A significant feature of the signal-flow diagram is the fact that all the elements of the control system are

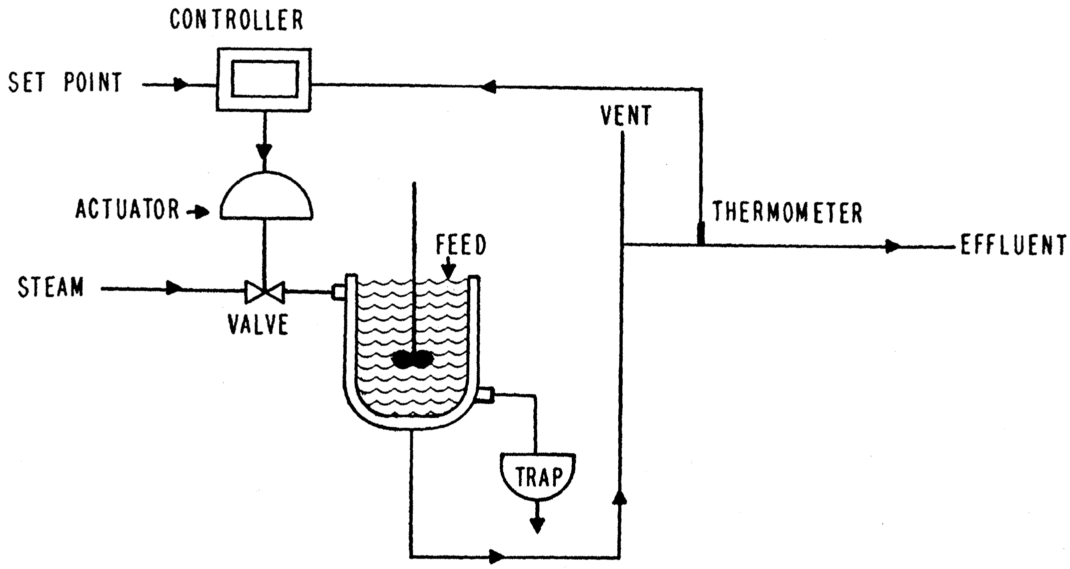
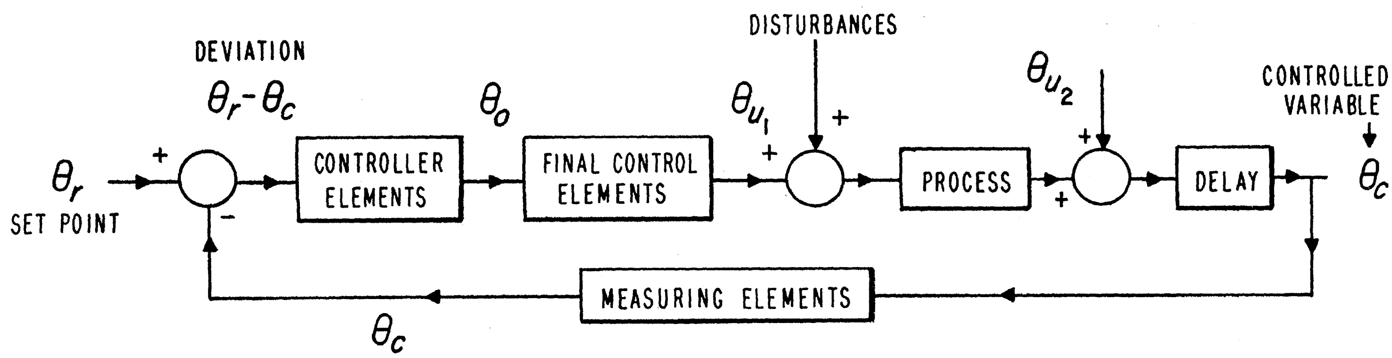


Fig. 1 (left). A simple automatic process control system.

Fig. 2 (below). Signal-flow diagram for simple control system.



shown as blocks of roughly the same size despite the fact that one block might correspond to a process unit larger by many magnitudes than some other system element. For example, the measuring element might be a high-speed thermocouple 2 millimeters in diameter and the process vessel might be as large as a house. Yet, so far as the control system is concerned, each of these disparate elements contributes as much as the others to the over-all dynamic behavior of the system. Any appraisal of control performance requires a scrutiny of the whole system and of the manner in which all elements bear on the system's behavior. A systems approach is required, and this requirement is the third basic characteristic of all control systems.

The fourth characteristic of automatic control systems is that they are inevitably feedback systems. Figure 2 shows this characteristic in that the controlled variable is fed back to the controller to ensure the generation of appropriate control action. Economics usually dictates that process control systems involve continuous feedback, since a system lacking feedback would require precise—hence costly—calibration of control actions so that all foreseeable influences on the system could be accommodated. In the long view all control systems must be feedback systems. Even clocks, which often are cited as examples of systems lacking feedback, require intermittent feedback in the form of an occasional resetting of the hands.

Feedback is economically advantageous, but it exacts a penalty by introducing the possibility of instability in the control system even though none of the elements in the system is inherently unstable. This tendency toward instability in feedback loops is the fifth characteristic of all control systems.

Dynamic Response

Typical dynamic responses of a feedback control system are shown in Fig. 3 for the case where the system is forced, by a sudden change in the set point θ_r, from some initial steady condition to a new steady condition *A* units greater than the initial condition. This kind of forcing is called step forcing or constant forcing, and the transient responses to it indicate the dynamic performance of the system. Response curve No. 1 is overdamped; curve No. 2 is critically damped in that it is the fastest response without overshoot and without oscillation; and curve No. 3 is underdamped in that there is overshoot and some cycling but the cycling is damped and ultimately vanishes.

These three response curves could be obtained on a single control system by increasing the amplification of the error signal entering the controller elements, curve No. 1 resulting from the least amplification. If the amplification in the controller is increased beyond that corresponding to curve No. 3, a level of sensitivity will be reached where the response cycles continuously at constant amplitude and frequency. This response condition is zero damping, the limiting condition of stability, and any further increase in sensitivity will result either in system runaway or in saturation or breakdown of some element in the system.

Frequency Response

Since the condition of zero damping involves the propagation of steady sine waves around the loop, we can determine the system parameters which fix this condition by observing the flow of sinusoidal signals through all the elements in the loop. In general, if a steady sine wave of modest amplitude is imposed on the forcing variable, the response variable will oscillate at the same frequency as the forcing variable but at smaller amplitude, and the response wave will lag somewhat behind the forcing wave. If we break the loop between the controller elements and the final control elements and cause the input to the latter to oscillate steadily, sinusoidal signals will flow from element to element, becoming progressively attenuated and shifted in phase. At the set point summer the phase angle will be shifted an additional 180 degrees because the sign of the signal is changed.

With increasing frequency the attenuation and phase shift will increase until, at the critical frequency, the total phase shift around the loop will be 360 degrees. If there is sufficient signal amplification in the controller elements at this frequency to offset the attenuation of signals in the rest of the loop, closing the loop will result in sustained oscillation of the system variables. This condition of the system is the limiting condition for stability, and the characteristics of the controller elements which produce this condition provide a simple basis for specifying practical controller characteristics for good, stable control.

The technique of analysis which makes use of the response to steady-state sinusoidal forcing is called frequency response analysis. Although the technique was used successfully as early as the middle of the 19th century by Angstrom in measuring the properties of thermally conductive systems, the application to control systems was not made until World War II. In the early days of the war the techniques of frequency response analysis and their applications to the synthesis of aircraft guidance systems, naval fire control systems, antiaircraft systems, and the like were developed to a high degree of practicability.

These guidance systems comprise a broad class of control systems called servomechanisms, which are essentially position-controlling systems wherein the principal forcing variable is the set point or command signal. The principal function of these control systems is to hold the controlled variable in close consonance with the set point, which ordinarily fluctuates widely with time.

Process control systems, on the other hand, comprise another class of control systems called regulators, for which the set point is usually steady, and the principal function of the system is to hold the controlled variable close to the set point despite disturbances imposed on the system. The techniques developed for analyzing and designing servomechanisms have been demonstrated to be quite applicable to regulators (*1*). For servomechanisms the forcing variable is θ_r (the set point), and for regulators the forcing variables are the disturbance inputs θ_{u_1} and θ_{u_2}.

Prior to World War II there were no simple, general procedures for designing control systems. A few British workers in the 1930's attempted some direct theoretical analyses of simple, idealized feedback systems, but they were unable to deal with any kind of realistic complexity because their analytical mathematics was unsuited to the task. They wrote differential equations for the systems under study and attempted to manipulate these equations to give the desired system response. Unfortunately, there are no direct analytical means for relating the characteristics of particular elements in the loop to the over-all temporal behavior of the loop, hence any modification of elements to improve over-all performance requires trial-and-error calculation. Furthermore, each trial requires that the differential equation for the closed-loop system be solved. Since the simplest real process control system can only be described by a differential equation of order 5 or greater, the formidability of this direct approach to the design of control systems is apparent.

The frequency response approach to control-system analysis and design became popular very soon after it was publicized because it avoids the two principal difficulties in the direct approach to design. With frequency response analysis the contribution of each element to the over-all behavior of the loop is immediately identifiable, and it is never necessary to solve the system equations.

Briefly, the use of frequency response analysis in the design of a control system involves determining the over-all open-loop frequency response characteristics of the system from the characteristics of all the individual elements and then adjusting the characteristics of appropriate elements so that a practical over-all characteristic is obtained.

The frequency response characteristics of an element consist of the magnitude ratio—the ratio of the amplitude of the response sine wave to the amplitude of the forcing sine wave—and the phase angle between the two waves, for the range of frequencies that is important to the system. For elements in series the magnitude ratios are multiplicative and the phase angles are additive, hence it is a simple matter to determine over-all open-loop frequency response characteristics from the individual element characteristics.

Frequency response characteristics may be determined experimentally or computed directly from the transfer functions for the individual elements or for the over-all system.

The transfer function of a process element is defined precisely as the ratio

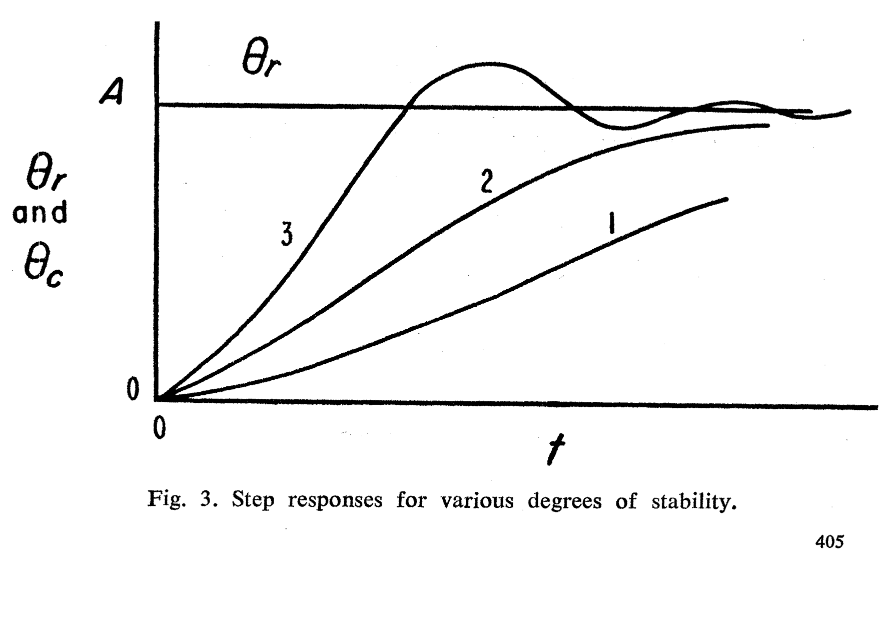


Fig. 3. Step responses for various degrees of stability.

of the Laplace transform of the response function to the transform of the forcing function. It is an algebraic expression in s, the Laplacian complex variable, which identifies conveniently all the dynamic characteristics of the element. Both the transfer function and the frequency response have meaning only for linear systems—that is, systems for which the dynamics are describable by linear differential equations. This restriction to linearity ordinarily is not a critical one because process control systems usually are subjected to small disturbances about fixed operating points. Under these conditions even inherently nonlinear systems may be approximated reliably by effecting a Taylor expansion of the differential equation about the normal operating points and dropping all nonlinear terms. For nonlinear systems subjected to large excursions in process variables the simple linearization is inadequate.

The desirable frequency response characteristics for good, stable control are expressed in terms of adequate margins of safety as compared to the limiting condition of stability. For example, the over-all magnitude ratio at the critical frequency, where the total phase angle around the loop is 360 degrees, should be set at less than 0.4. There are similar rules for phase angle and closed-loop magnitude ratio. The purpose of all these rules is to ensure a stable system response with a small amount of oscillation so that the system is not excessively sluggish. In addition to stability, however, there are two other factors to consider in designing control systems. One factor is the speed of response, and the other is the steady-state error.

The speed of response is related to the critical frequency of the system, and any alteration in the characteristics of process elements or of the controller elements which will increase the critical frequency will increase the speed of response. One simple procedure is to add a controlling action which is generated by the derivative of the error.

Control systems with control actions generated only in proportion to the deviation of the controlled variable from the set point will tolerate a steady-state deviation. This steady-state error, or offset, is roughly inversely proportional to the amplification of signal in the controller. Since the stability criteria fix the maximum allowable amplification in the controller, it is necessary in many systems to add a control action produced by the time integral of the deviation. Empirical rules usually based on the critical frequency guide the addition of the integrating action.

Root Locus

The use of frequency response for designing control systems has the advantages of simplicity and convenience, but it does not permit designing to a specific quality of performance. For systems describable by linear ordinary differential equations it is possible to design for specific conditions of damping by scrutinizing the roots of the characteristic equation. This equation, which is the auxiliary algebraic equation for the differential equation, is obtained readily from the open-loop transfer function. Rapid graphical techniques devised by Evans (*2*) make it possible to sketch the locus of the dominant roots of the system for various controller amplifications. The dominant pair of conjugate roots determines the damping in the system, hence the controller gain necessary for a particular condition of damping can be found from the root locus.

The root-locus method is somewhat awkward to use for systems containing time delays, such as the system of Fig. 1, and also for distributed parameter systems. These latter systems are characterized by parameters which are spatially distributed rather than lumped at particular points in the system. For example, the system of Fig. 1 may be regarded as a lumped parameter system, but a tubular heat exchanger with temperatures varying along the length of the tubes would have to be treated as a distributed parameter system (*3*). The awkwardness in treating these systems arises from the fact that they have infinite numbers of roots, and only if a pair of roots clearly predominates can a controller selection be made confidently.

Damped Frequency Response

A technique for designing for specific conditions of damping which offers considerable promise but which has not been widely used is one based on the damped frequency response. By means of conformal mapping it is possible to sketch quickly the damped frequency response grid, provided the zero damped frequency response characteristics are known (*4*). The grid identifies lines of constant damping and constant damped frequency, and if they are plotted for the reciprocal of the transfer function for the entire system, exclusive of controller, it is possible to select controller characteristics directly from the plot.

Nonlinear Systems

Now, all of the foregoing methods of analyzing and designing control systems are strictly applicable only to linear systems. For nonlinear systems no general techniques of analysis are available. Some limited procedures have been studied in detail. One of these makes use of the describing function (*5*) in frequency response analysis. This function is merely the ratio of the fundamental component of the output of the nonlinear element to the amplitude of the sinusoidal input. If there is but one nonlinear element in the loop and if there are many linear elements, then all higher harmonics produced in the nonlinear element will be attenuated by the linear elements. Thus, the describing function may be used as the frequency response characteristics of the nonlinear element. It has significance, however, only for the particular signal levels at which it is obtained.

Another procedure for dealing with nonlinear systems or with systems containing nonlinear elements is the method of phase-plane analysis suggested by MacColl (*6*). This method is useful for systems which may be described by a particular form of second-order nonlinear differential equation, since the method depends on the properties of this second-order equation.

A third procedure makes use of the classical work of Lyapunov on the stability of nonlinear control systems (*7*). The fundamental theorems of Lyapunov state that a system will be stable if there can be found for that system a function having certain mathematical properties related to the equations of the system. Such a function is called a Lyapunov function. Another pair of theorems attributable to Lyapunov shows that the stability of a nonlinear system can be inferred from the first approximation obtained by expanding the system equations and retaining only the first-order terms. For linear systems the stability of the system depends on the roots of the characteristic equation obtained from the differential equation which describes the system. Conjugate complex roots contribute an oscillatory mode to the system's response, but if the real parts of these roots are negative, and if all real roots are negative, the system is stable. Conversely, if there are one or more real positive roots, or if the real part of one of the pairs of conjugate roots is positive, the system is unstable. According to Lyapunov's theorems these algebraic rules, when applied to the first approximation, can predict the stability of the nonlinear system regardless of the form of the higher-order terms in the total approximation to the nonlinear system. Thus, if the characteristic equation for the first approximation indicates instability, the nonlinear system will be unstable. This method of gauging nonlinear system stability is indeterminate, however, if there are one or more pairs of pure imaginary roots—that is, conjugate roots with zero real parts.

Clearly, a major difficulty in applying Lyapunov's fundamental theorems is in finding Lyapunov functions for control systems under consideration. No straightforward procedures for seeking these functions have been worked out as yet, although there is a considerable body of practical experience among Russian control engineers.

The principal applications of Lyapunov's theorems to theoretical and practical control problems have been made in the Soviet Union. Letov (*7*) has described in some detail these applications and also the extension of the methods to the problem of designing systems to a specified quality of performance. This latter problem is called by the Soviet school of control engineers the second problem in control. The first problem is the problem of identifying the conditions of stability. Actually, the extension of Lyapunov's method to the second problem involves effecting a transformation so that the design to a particular specification of performance quality becomes in fact a stability problem.

The fourth procedure for treating nonlinear control problems is the use of machine computation. Both digital and analog computing equipment can solve nonlinear equations, but for many systems, particularly those having lumped parameters, analog computers are a convenient means of seeking favorable control system properties. Analog computers are not so convenient for distributed parameter systems because these systems effectively involve two independent variables and it is necessary to resort to approximation schemes. Whether analog or digital computation is used, a large and correspondingly costly machine is required for any real problem.

Computer Control Systems

In all control systems the controller elements exercise computing functions of one sort or other. For example, the summer takes a difference and the controller amplifier multiplies that difference (deviation) by an adjustable constant. Also, in some controllers the deviation may be integrated over time, or differentiated. Thus, all control systems may be regarded as computer control systems since they all contain computing elements.

Normally, the computing functions in controllers are generated by analogy. In systems involving multiple inputs and multiple outputs, however, it may prove advantageous to use a single digital controller to digest all feedback information and generate all control actions. This kind of installation has been made on a synthetic-ammonia plant and on a large-scale crude-petroleum fractionator.

For an elaboration of this installation, a relatively large-scale digital computer would be used to devise and refine continuously a mathematical description or model of the controlled process and, on the basis of the model, to generate control actions which would optimize simultaneously both the dynamic behavior and the steady-state operating levels in the plant. In a further elaboration, the computer would be put in control of an entire factory and ultimately in control of an entire corporate enterprise.

It is possible to have the computer conduct experiments on the process in order to construct a reliable mathematical model of the process. One method would be to insert low-level forcing signals at known frequencies and filter out the corresponding responses. Another method, less likely to be reliable but sound in principle, would be to monitor the random input and output signals and, by statistical correlations and cross correlations to derive the effective transfer functions for the system.

With the up-to-date model of the process and information on all current constraints on the system and on its operation stored in the computer memory, the computer for any disturbance or combination of disturbances will determine quickly the optimum strategy for holding the system on control and then execute the necessary action. The resulting control would be not only an optimizing control but also an adaptive control in the sense that the control system would adapt itself to changes in the process. Adaptation would result automatically from changes in the model.

Aris (*8*) and Kalman, Lapidus, and Shapiro (*9*) treat some aspects of the problem of fitting digital computation to the optimal control of process systems. Both Aris and Kalman *et al.* make use of dynamic programming to effect the optimization.

One problem that arises in using digital computers in process control loops is the wedding of elements producing continuous signals with a computer which can deal only with discrete signals (numbers). Although the effect of having to deal with sampled-data elements is a destabilizing one, the mathematics poses no special problems.

Current Problems

There are three broad current problem areas in the field of automatic process control.

As has been pointed out, one of the characteristic features of control systems is the fact that the individual elements in the system all contribute significantly to the over-all behavior. Hence, one critical problem is the description of process elements and other loop components in language that is pertinent to the design of the control system. In the final analysis such description requires the elucidation of the mechanisms of all the basic rate processes, including momentum transport, thermal energy transfer, mass transport, and chemical and nuclear reaction. Much of the current scientific research of the chemical engineer and the mechanical engineer, as well as of the chemist and physicist, is in some measure pertinent here. For example, an improved and quantitative description of flow turbulence in liquids would afford a sounder basis for defining the dynamics of mixing in the stirred reactor of Fig. 1.

Another critical current problem in automatic process control is the precise treatment of integrated control systems. Frequency response techniques have been used widely and quite successfully, but not for design to specific performance quality. The damped frequency response offers enough promise to merit further study. For nonlinear systems the approaches via the theorems of Lyapunov, with the extensions described by Letov, appear to be practical first steps.

The third current problem area in the field of automatic process control is concerned broadly with optimization at all levels of effort. What are practical criteria of optimality? What search procedures for finding optima are themselves optimum? The answers to these and similar questions must be given in greater refinement. Some aspects of these questions are of concern to economists, and some are of concern to mathematicians and computer programmers. But all of the questions concern the control engineer.

There is considerable activity in the field of automatic process control in this country and in Japan, France, England, and Germany. It is somewhat disquieting to know that in matters of automatic control theory the Soviet Union leads us and appears to be increasing the lead at the present time (*10*).

Future Trends

At the present time the fully automatic process plant where all operations, including direction and execution, are handled by machines does not exist. There are no major theoretical or technical barriers, but the details of application must be worked out in each case. With the continuing acceleration of activity in research, development, and application in the field of automatic process control, the automatic plant will soon become a reality. The long-range sociological effects will be enormous, but they can be good.

References

1. E. F. Johnson, *Advances in Chem. Eng.* **2**, 33 (1958).
2. W. R. Evans, *Control System Dynamics* (McGraw-Hill, New York, 1954).
3. W. C. Cohen and E. F. Johnson, *Ind. Eng. Chem.* **48**, 1031 (1956).
4. G. H. Farrington, *Fundamentals of Automatic Control* (Wiley, New York, 1951).
5. J. G. Truxal, *Automatic Feedback Control System Synthesis* (McGraw-Hill, New York, 1955).
6. L. A. MacColl, *Fundamental Theory of Servomechanisms* (Van Nostrand, Princeton, N.J., 1945).
7. A. M. Letov, *Stability in Nonlinear Control Systems*, J. G. Adashko, trans. (Princeton Univ. Press, Princeton, N.J., 1961).
8. R. Aris, *The Optimal Design of Chemical Reactors* (Academic Press, New York, 1961).
9. R. E. Kalman, L. Lapidus, E. Shapiro, *Chem. Eng. Progr.* **56**, No. 2, 55 (1960).
10. J. P. LaSalle and S. Lefschetz, *J. Math. Anal. and Appl.* **2**, 467 (1961).

An Econometric Model of Economic Development

The model used by Indian planners poses a paradox concerning the best allocation of scarce capital.

Martin Bronfenbrenner

Mathematical economics is nearly as old as economics itself—substantially older than, say, Adam Smith's *Wealth of Nations* (*1*). Yet it did not become intellectually respectable, at least in the Anglo-Saxon world, until the Great Depression of the 1930's. In retrospect, the founding of the Econometric Society and the appearance of its journal, *Econometrica,* in 1932 seem a kind of watershed within the economics profession. Mathematical economics still appears to most outsiders an esoteric and new-fangled sort of science or pseudoscience.

The earliest mathematical-economic studies were largely devoid of empirical content, just as the earliest statistical-economic studies were largely devoid of economic analysis. It was the econometricians (and their ancestors) who brought these two strands together. An early example, not otherwise particularly fortunate, was W. Stanley Jevons's sunspot theory of business fluctuations (*2*).

Most of the earlier econometric studies, Jevons's being by no means the sole exception, dealt mainly with the supply of or the demand for individual commodities one at a time. Later they developed into studies of a few related markets, and later still they reverted to Jevons-type grand schemes of "ambitious equation systems attempting to represent the dynamic properties of an entire economy" (*3*).

When an econometric study involves more than one equation (usually fitted statistically) at a time, the set of equations is called an economic model (*4*). To cite one simple case, a demand equation and a supply equation form a two-equation model of market price determination. When all the variables of an equation system or model refer to the same point in time, there is no need to "date" them and the model is called *static*. When different variables relate to different points in time, the model must be *dated* and is called *dynamic* (*5*). Suppose, for example, that the amount of wheat q_t^d demanded in period t depends on its price p_t during that period, but that, since wheat production requires time, the amount q_t^s supplied in period t depends on its price in the preceding period $(t-1)$. Then the three equations:

$$q_t^d = f(p_t)$$
$$q_t^s = g(p_{t-1})$$
$$q_t^d = q_t^s$$

The author is professor of economics in the School of Business Administration, University of Minnesota, Minneapolis.