IFAC

WARSZAWA 1969

INTERNATIONAL FEDERATION
OF AUTOMATIC CONTROL

Direct Control Problems

Dead Time and Distributed Systems, Realy Systems, etc.

Fourth Congress of the International Federation of Automatic Control
Warszawa 16–21 June 1969

TECHNICAL SESSION
4

NOT

Organized by
Naczelna Organizacja Techniczna w Polsce

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

Direct Control Problems

Dead Time and Distributed Systems, Realy Systems, etc.

TECHNICAL SESSION No 4

FOURTH CONGRESS OF THE INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL
WARSZAWA 16 – 21 JUNE 1969

Organized by
Naczelna Organizacja Techniczna w Polsce

Contents

Wydawnictwa Czasopism Technicznych NOT
Warszawa, ul. Czackiego 3/5 - Polska

Zakład Poligraficzny WCT - NOT

A PREDICTIVE CONTROL SCHEME FOR DEAD-TIME PROCESSES USING A LEARNING METHOD OF PROCESS IDENTIFICATION

M. S. Beck — University of Bradford, England.
P. R. Birch — University of Bradford, England.
N. E. Gough — University of Bradford, England.
A. Plaskowski — Research Centre for Automatic Control in the Chemical Industry, Warsaw.

1. Introduction

The theoretical advantages of predictive control loops for use in processes possessing large dead-times are well-known. Mathematical verifications of the effect of dead-time on stability have been given by Qin Yuan-Xun et al[1] and Choksy[2] . Oetker[3] has shown that for a prescribed state of stability to be obtained with three-term control, it is necessary to lower the gain to such a degree that control may become unsatisfactory. Buckley[4] lists several methods for overcoming these problems and Weiss[5] has given a comprehensive bibiliography. All theory seems to point to the use of an accurate model for predicting optimal control signals in the face of load disturbances. However, the way in which the model is to be determined and implemented has been a matter of conjecture. Few successful applications have been reported.

It is the opinion of the authors that, since accurate prediction requires an on-line computer model, the potential of the model should be used to the full by the introduction of adaption and optimisation. A time domain model is used rather than a frequency domain model for the following reasons -

a). There is no need to transform between the time domain and the frequency domain which is a complex procedure for a high-order model.

b). The time domain model can be completely specified by the impulse response as discussed in section 4.

This paper describes the work which is being carried out at Bradford on a pilot-scale process controlled by an Argus 400 digital computer. Predictive control is employed by operating the model as faster than real time and a feedback loop is used to remove the drift in the process output which is caused by differences between the model and process dynamic responses. The way in which the model accuracy affects the process response has been shown by Wheater[6] for four types of dead-time compensation.

Results showed that the sensitivity to model inaccuracy tends to increase with the adoption of more sophisticated methods of predictive control. Consequently, experiments are being carried out using a new learning method of process identification employing pseudo-random sequence perturbations, so that the model can be periodically updated. A hill-climbing procedure then optimises the model performance and gives an optimal control sequence which is blended with the feedback control signal. Also, optimal feedback control settings are obtained by the use of a procedure which involves a new method of computing controllability using the mean square error criterion.

2. The Experimental Plant

Figure 1 shows the pilot-scale water heating process which incorporates the facilities for demonstrating feedback/predictive computer control. This process is representative of a broad class of industrial processes involving load disturbances and dead-time, including for example, rotary solids driers[7].

Depending on the settings of solenoid valves SV1 and SV2, either hot or cold water is fed through a section of pipe which constitutes an input time delay, to a stirred vessel. Here the water is heated by a 3kw electrical heater and passes along the outlet pipe to drain. The feedback control loop involves the measurement of the tank temperature T2 using a thermocouple; the output signal from the computer controls the power to the heater using phase angle control of two thyristors in inverse parallel configuration. As mentioned above, time delay in the control loop is a serious problem on many processes. In order to study this situation thermocouple T3 is used instead of T2 for control purposes and the section of pipe after the vessel then constitutes a measurement time delay.

Poor control associated with the dead-time in the feedback control loop is ameliorated by measuring the load temperature disturbances at T1 and computing a predictive control signal. This signal is then blended with the feedback control signal. Load disturbances are produced by switching the solenoids from a random sequence generator which uses a novel arrangement of Geiger-Müller tubes fired by cosmic radiation (see appendix 1).

In order to measure the dynamic properties of the heater and stirred vessel without seriously affecting the normal process operation, provision is made for modulating the heater voltage with a $\pm$ 1% pseudo random binary sequence perturbation. A maximum length sequence of length $L = 2^7-1 = 1=7$ is generated by a seven stage shift register[8]. The sequence generator output is connected to relay RL1 which shortcircuits a resistor in the heater power supply.

Knowledge of the input time delay is also necessary for accurate predictive control. This can be inferred from the water flowrate and turbine flowmeter F1 is used for this purpose.

3. The Control System

3.1. Predictive - Exploratory Control

It has been shown earlier by Beck and Gough[7,9] that effective disturbance reductions can be achieved using a search technique on a model. Referring to Figure 2, which shows a block diagram of the process described in section 2 together with the control system, the principle features are:

a). The load disturbance, n, is sampled with period T secs.

b). The response of the process is predicted using an impulse response model which excludes the process dead-time. By omitting the time delay, a sequence of optimal control signal estimates can be computed & and held until the time to actuate the control element occurs. In order to do this, the system response y, is regarded as a linear combination of the outputs of the process and load transfer functions, Gp and G_1 respectively. For a control signal u the model response is thus:

$$Y(s) = U(s)\ G_p(s) + N(s)\ G_1(s) \qquad 3.1$$

where upper case letters are used to denote Laplace transform variables. In fact the transfer functions are not required since the response is computed using impulse responses, h_p and h_1. Thus, the response due to the load disturbance is given by the scalar convolution:

$$y_1(kT) = T \sum_{m=o}^{k} n(kT-mT)\ h_1(mT) \qquad 3.2$$

where k is the sampling instant. Experiments have shown that if the sampling interval T satisfies the inequality $T_1 > 10T$ where T_1 is the dominant system time constant, then the summation gives an adequate representation of the system response.[9]

c). A cost function performance criterion, g, is computed based on the response of the model and m economic or safety requirements of the process, x_i, which may be time varying:

$$g = f\left[y(kT), x_i\right], i=1,\ldots\ldots,m \qquad 3.3$$

d). A fast search is carried on the model using an iterative decision algorithm in order to minimise the cost function.

A preliminary simulation of this scheme using a simple cost-function showed[9] that a suitable stepping sequence is:

$$u^j = u^{j-1} - \Delta^j\ \mathrm{sign}\ (g^j - g^{j-1})\ \mathrm{sign}\ \Delta^{j-1} \qquad 3.4$$

where u^j is the sub-optimal control signal and Δ^j is a variable scaling factor. As a result, an optimal control sequence $u^*(o)$, $u^*(1)$,.....,$u^*(k)$ was obtained which minimised the criterion g. Since the control power is limited in practice u may appear in the cost function and the signal is also constrained to lie between the high and low limits

$$u_c^*(k) \leq u^*(k) \leq u_H^*(k) \qquad 3.5$$

For a simulation of a first order system, it was found that optimal control sequence estimates could be computed in an average time of 11.4msecs and a maximum time of 24.9 m sec* using an Argus 400 process control computer[9].

e). The control signal estimates are then stored for the duration of the input dead-time and blended with the direct feedback control signal. This trimming action overcomes drift due to mismatch between process and model characteristics and reduces the effect of unmeasured process disturbances.

3.2. The Feedback Loop

In the work described in this paper, an incremental control signal is produced using a standard control algorithm, reproduced by permission of Ferranti Limited, of the form:

$$U_k = K_o \left[(y'_k - y'_{k-1}) + K_{-1} \cdot \frac{y'_k + y'_{k-1}}{2} + (1-\lambda') \left[K_1 (\bar{y}_k - \bar{y}_{k-1}) \quad U^d_{k-1} \right] \right] \qquad 3.6$$

where y'_k is the process output at sampling instant k

U_k is the feedback control signal

Ko = 100/P.B

P.B = Proportional band

$K_1 = T/T_I$

T_I = Integral action time constant

$K_1 = T_D/T$

T_D = Derivative action time

λ' = a smoothing constant

$$\bar{y}k = (1 - \lambda)\, y'_k + \lambda' \bar{y}_{k-1} \qquad 3.7$$

U^d_{k-1} is the last full value of the derivative term.

In this particular application the derivative term is not used for the feedforward predictive control scheme, although the 3 term standard routine would be used with other control loops. The reasons for not using the derivative term are -

i). Power limitations are included in the feedforward control model (section 3.1d). Derivative action tends to produce large short -

term power requirements, which cannot be allowed for in the predictive control algorithm, and would invalidate the optimising procedure.

ii). In any case there is no real advantage in including derivative action in a system where transient errors should be eliminated by feedforward control.

Acceptable controller settings are determined using a procedure which computes the normalised integral - square - error criterion of process controllability. By this means the best trimming action of the feedback loop on the predictive signal is obtained without fear of instability. This program, Epton and Gough[10,11], is described in Appendix 2. The I.S.E. criterion is set up as a function of process parameters and includes time delays written as Padé polynomial approximations. Avoidance of the complex standard integral tables is achieved by solving the integral in the general case using a matrix formulation. The criterion is plotted as a function of proportional, derivative or integral constants and clearly shows the values of minimum controllability and the onset of instability. Figure 3 shows a typical set of controllability curves plotted as a function of controller gain for a second order process with varying time delays [12].

4. Process Identification

4.1. Direct Correlation

Process identification has been obtained by using a pseudo-random binary sequence perturbation and correlating the sequence x with the tank temperature, (T2),[8,13]. The correlation equation is

$$h(m\tau) = C \sum_{k=k_p-2L}^{k_p} y(k)x(k-m) \qquad 4.1$$

where $h(m\tau)$ is the impulse response.

C is a constant.

$\tau = \lambda/2$ = sampling interval, where λ is the basic sequence time interval.

k_p is the present sampling instant.

The sampling period τ is chosen as half the basic sequence time interval in order to satisfy the Nyquist sampling theorem[14].

Figure 4 shows an impulse response relating changes in heater power to tank temperature, measured using a $\pm$ 2% voltage perturbation and an integration period of approximately 7 times the major process time constant.

On-line correlation does present some difficulties since in order to directly measure the open-loop impulse response the correlation must be completed when the feedback control signal is inoperative. In many processes there are periods of time when the process state is sufficiently close to the optimal state for the feedback control loop to be disconnected. This condition can be detected by the computer and the identification then carried out. Since the predictive control is still in operation during the identification period, the process state should not deviate far from the optimal state. If unmeasured process disturbances occur, the computer can detect the resultant deviation in the process state, automatically terminate the identification and close the feedback loop. Experiments are being carried out to implement such a scheme which enables the computer to decide the most convenient times for system identification.

4.2. The Learning Method of Process Identification

It is thought that more accurate and rapid identification can be achieved using a method devised by Beck[15]. The essential feature of this learning method is to take an initial process model, prepared from an estimate of the dynamic properties of the process, and to update this model using experimental data. The updating procedure corrects for any inadequacies in the initial model and allows periodic updating of the model for time-varying processes. A particular advantage is that the experimental procedure has only to determine part of the process model. Thus, the signal-to-noise ratio is better than that obtained using a direct experimental correlation where the whole model must be determined. Alternatively, a more rapid identification can be achieved for the same signal-to-noise ratio.

Refer to figure 5. The process has an impulse response $h(\gamma)$ and an input $x(t)$ comprising the normal process signal $n(t)$ and P.R.B.S. test input $x^{i}(t)$. $x(t)$ is monitored and the initial process model $\bar{h}(\gamma)$ is used to compute the response $q(t)$ to $x(t)$ using the convolution

$$q(t) = \int_{\gamma=o}^{\xi} x(t-\gamma)\, \bar{h}(\gamma)\, d\gamma \qquad 4.2$$

This is expressed as a scalar convolution which involves matrix multiplication in order to obtain the q vector

$$[Q] = [X][\bar{H}] \qquad 4.3$$

The period of integration is taken to be

$$\xi = 5\, T_1 + \tau_p \qquad 4.4$$

where τ_p is the process time delay

T_1 = dominant time constant of process

Next, the measured process output y(t) is correlated with the model output q(t) using

$$\phi_{qy}(\beta) = \frac{1}{\zeta_1} \int_{t=o}^{\zeta_1} y(t)\, q\,(t-\beta)dt \qquad 4.5$$

which again is achieved digitally by matrix multiplication

$$\Phi_{QY} = \left[Y \right] \left[Q \right]_\beta \qquad 4.6$$

Also, the autocorrelation of the model output is given by

$$\phi_{qq}(\beta) = \frac{1}{\zeta_1} \int_{t=o}^{\zeta_1} q(t)\, q(t-\beta)dt$$

or $$\left[\Phi_{QQ} \right] = \left[Q \right] \left[Q \right]_\beta \qquad 4.7$$

These correlation functions enable the updating model $\hat{h}(\gamma)$ to be compensated

$$\phi_{qy}(\beta) = \int_{\gamma}^{\zeta} \hat{h}(\gamma)\, \phi_{qq}\, (\beta - \gamma)\, d\gamma$$

or $$\left[\Phi_{QY} \right] = \left[\hat{H} \right] \left[\Phi_{QQ} \right] \qquad 4.8$$

and $\hat{H}$ is found by matrix inversion using Gaussian elimination.

Finally, the initial model and the updating model are combined to give the updated model

$$\bar{h}_u(\gamma) = \int_{\tau}^{\gamma} \bar{h}(\tau)\, \hat{h}\,(\gamma-\tau)d\tau$$

or $$\left[\bar{H}_u \right] = \left[\bar{H} \right] \left[\hat{H} \right] \qquad 4.9$$

The above equations (4.3, 4.6, 4.7, 4.8 and 4.9) have been programmed using a conversational language to a remote computer connected by telephone lines to the user terminal. A flow-diagram is shown as Appendix 3. (This method of programming has been particularly useful for the rapid development of the program).

Some results of using the learning method for a typical process are shown in figures 6 and 7. A P.R.B.S. test signal of length L = 15 (fig.6a) was applied to a noise-free process having the impulse response shown in fig. 6b. Fig. 6c shows the resultant process output. A rough representation of this process was used as an initial model (fig. 6d). The program output included the autocorrelation, cross-correlation, and updating model as well as the final updated model, all shown in fig. 6.

Fig.7 shows corresponding results for the same process with spurious noise (fig. 7e) added to the process output.

The diagrams cleary show that the leaning method has been successful in forming a reasonably accurate updated model in both the noise free and noisy cases. Values of a dispersion cofficient (D.C.) for the appropriate data are shown in table I.

Table I. Performance of Learning Method

	Noise Free	With Spurious Noise
D.C. of process output from noise free value	0	$I.4 \times IO^{-2}$
D.C. of updated model from true process impulse response	$4.9 \times IO^{-5}$	$2.I \times IO^{-2}$

Where

$$D.C. = \sqrt{\frac{\sum_{i=o}^{P}(x(i) - x_t(i))^2}{(P+I)\sum_{i=o}^{P}[x_t(i)]^2}}$$

x(i) =Observed values of variable

$x_t(i)$ =True values of variable

P = Number of observations

5. Discussion

This paper indicates how some of the new and powerful control techniques may be applied to a linear process with dead time where accurate control is essential. The computer storage requirements for the implementation of the basic scheme using combined feedback/predictive control and simple correlation identification are approximately as follows:

Predictive control with simple cost function	2000 words
Feedback control	200
Correlation identification (127 length sequence)	1500
TOTAL	3700 words

Calculation times for an Argus 400 computer are:

Predictive Control	(average	11.4 m.sec.
	(maximum	24.9 m. ec.
Feedback control		0.7 m.sec.
Correlation Identification		2 seconds whenever an identification is required

Since many of the routines could be held in a backing store so that only a small immediate access store is required, the cost of storage will be quite modest. In addition these routines can be used with a large number of control loops since the calculation time per loop is quite small.

The above basic scheme should be satisfactory for many purposes. However, it is anticipated that a fuller scheme would use a multidimensional cost function requiring a general minimisation procedure, the learning method of identification requiring matrix inversion and the integral-square-error program for determining optimal feedback control settings. The computer storage and calculation time requirements would then be several times larger than those of the basic scheme. This type of work is best suited to a hierarchical computer system and the Bradford University Argus 400 computer is being linked to a large ICT 1909 scientific computing installation to implement such a scheme. This high-speed data link offers the facilities of conversational mode and fast data transference between the computers, using direct store access. Hence the Argus 400 will carry out the basic operations of closed loop control and alarm monitoring and will interrupt the ICT computer, which will then carry out the higher level operations of identification, prediction and optimisation.

Acknowledgements

The authors would like to thank G. Dean of Ferranti Limited for preparing the three-term-control routine and F. Dadachanji of De.La Rue Ltd., for permission to publish the I.S.E.curves.

Appendix 1

The Random Binary Sequence Generator

Figure A1 shows a schematic diagram of the random sequence generator. Two Geiger-Müller tubes are used in a coincidence circuit in such a way that the instrument only gives an output pulse if cosmic particles strike the tubes G1 and G2 successively within a certain time, this time being equal to the time constant of a monostable multivibrator. The output of the monostable is gated (using a logical NAND) with the signal from the second tube, G2. By varying the time constant of the monostable, the average switching rate of the sequence can be varied from 3 - 1500 secs. The gated signal operates a bistable unit and the output finally switches a relay.

Tests have been made on the instrument in which the observed frequency distribution of the sequence was compared with a Poisson distribution. The goodness of fit was tested using a chi-squared distribution and the results gave us reason to believe that the distribution follow a Poisson distribution.

Appendix 2

General Procedure for Evaluating the Controllability of Time Delay Feedback Control Systems[10,11].

This procedure considers a feedback control system which includes process and load transfer functions and time delay of any order, together with a three-term control equation. All the process parameters are fed to the computer as data and the following operations are carried out in order to compute the I.S.E. criterions of controllability:

a). The error-to-load transfer function H(s) is expressed as the ratio of two polynomials in complex pulsatance, S. To do this, the process and load factors are multiplied together. Polynomial multiplication and addition routines are required.

b). Time delays are written as Padé polynomial approximations of any order and these are multiplied into the polynomials, giving the I.S.E. in the standard form:

$$\text{I.S.E.} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \left| H(s) \right|^2 ds \qquad \text{A1}$$

where

$$H(s) = c(s)\,/d(s)$$

for $c(s) = \sum_{k=o}^{n-1} C_k s^k$ and $d(s) = \sum_{k=o}^{n} d_k s^k$

c). Standard tables for high order integrals are not available and hence a matrix formulation is used to solve the above integral. To do this the c(s) and d(s) coefficients are built up into two matries, [C] and

$[D]$ respectively. Theory shows[10,11] that these matrices are related by an equation of the form

$$[C] = [D][A] \qquad \text{A2}$$

and the value of the nth order integral is given by

$$I_n = \alpha_{n-1} / d_n \qquad \text{A3}$$

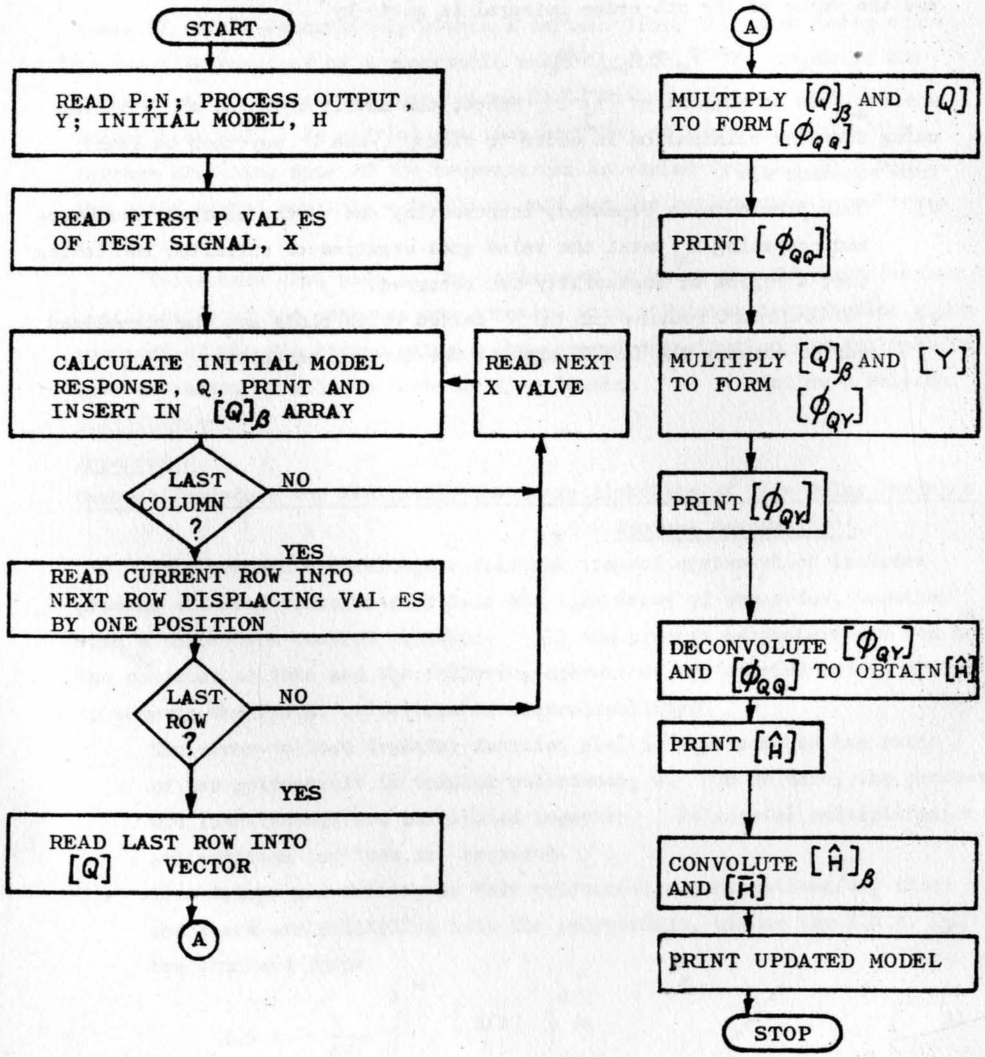
where α_{n-1} is an element of $[A]$. Hence, the matrix equation A2 is solved using Gaussian elimination in order to find α_{n-1} and I_n may then be found from equation A3.

d). This procedure is repeated, incrementing one of the control parameters and computing I_n until the value goes negative or infinite, indicating that a region of instability has occurred.

e). A Graphplot routine can be called which plots out the normalised curves showing the I.S.E. against the control parameter.

Appendix 3

Single-Stage Learning: Program Flowchart



References

1. Qin Yuan-Yun et al. Effect of Time Lags on Stability of Dynamic Systems. IFAC Mosow, 1960.
2. Choksy, N.H. Time Lag Systems. Progress in Control Engineering, Vol. 1, Heywood and Co., 1962.
3. Oetker, R. On the Control of Sectors with Dead time, IFAX Moscow 1960.
4. Buckley, P.S. Automatic Control of Processes with Dead time. IFAC Moscow 1960.
5. Weiss, R. Transportation Lag - Annotated Bibliography, IRE Trans. Auto. Control. AC - 4 (1) 1959.
6. Wheater, W.M. How Modelling Accuracy Affects Control Response. Control Eng. Oct, 1966.
7. Beck, M.S. and Gough, N.E. A Model Reference Adaptive Control System for Driers and Other Dead-Time Processes. Inst.J.Control Vol. 5, 55, 1967.
8. Briggs, P.A.N., Hammond, R.H., Hughes, M.T.G., Plumb, G.O Correlation Analysis of Process Dynamics using Pseudo-Random Binary Test Perturbations, Proc.Instn.Mech.Engrs. 1965, 179,3H p.37.
9. Beck, M.S., Gough, N.E. Some Further Contributions to the Above (Ref. 7) The Use of a Dynamic Model. Int.J.Control Vol.6, 5, 1967.
10. Epton, J.B.A., Gough, N.E., Evaluation of the Integral-Square-Error Definite Integral - Program JE20. Computer Program Interchange Scheme, The Chemical Engineer, No. 211, Sept. 1967.
11. Epton,J.B.A., Gough, N.E., Program JE21 - A Fortran Program for Evaluating the Mean-Square-Error Definite Integral from Process Parameters and for Plotting Normalised Controllability Curves. Report JENG/B School of Control Engineering, University of Bradford,1967.
12. Dadachanji, F.M. The Relative Performance of Electronic and Pneumatic Controllers. M.Sc. Thesis, Bradford University, School of Control Engineering, 1967.
13. Harvey, P.M. Cross-Correlation Program A1. Report. School of Control Engineering, University of Bradford, 1968.
14. Hutchinson, A.W., Shelton,R.J., Measurement of Dynamic characteristics of Full-Scale Plant using Random Perturbing Signals: An Application to a Refinery Distillation Column. Trans.Instn.Chem.Engrs.Vol.45, No. 8, 1967, pp T334-T342.
15. Beck, M.S., A Learning Method for Identification of Process Dynamics.

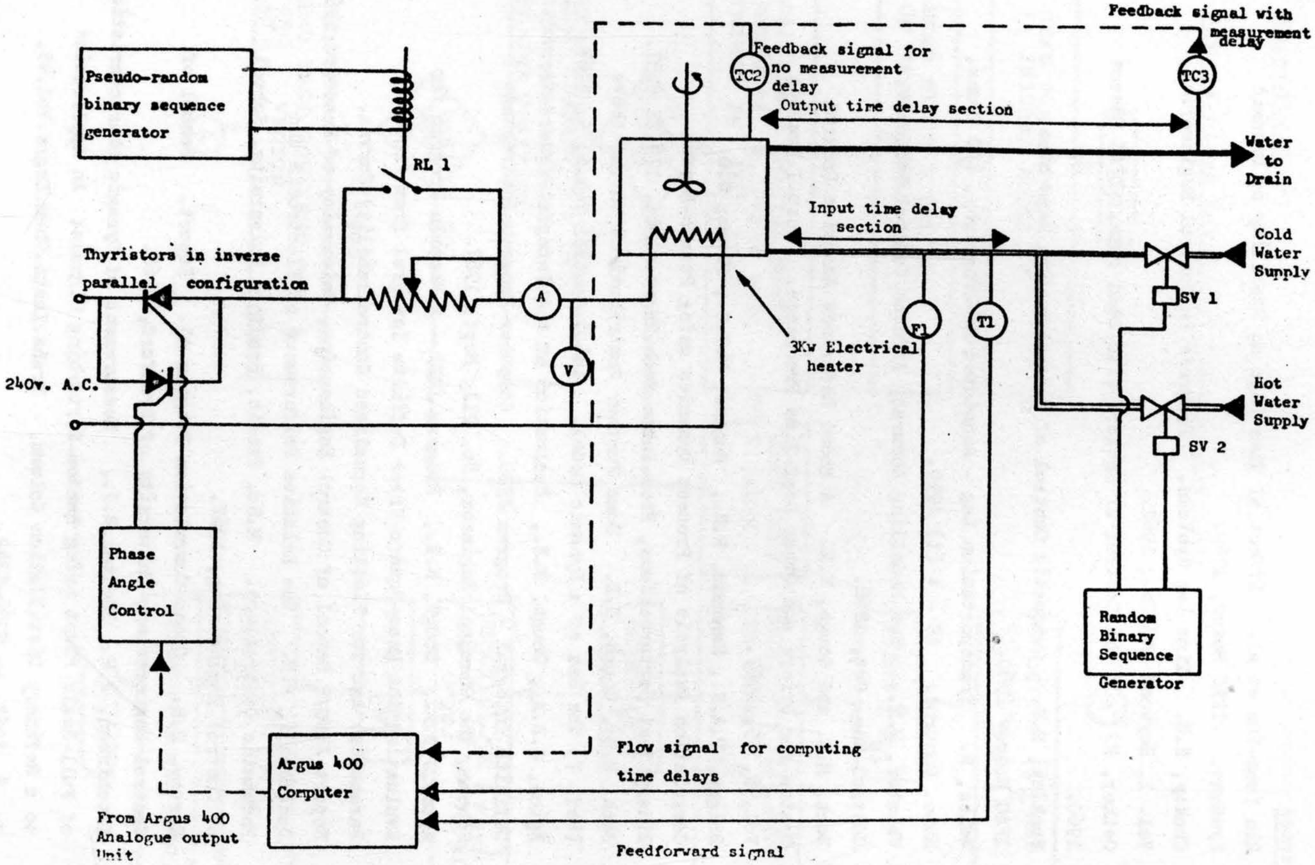


FIGURE 1. THE EXPERIMENTAL PLANT

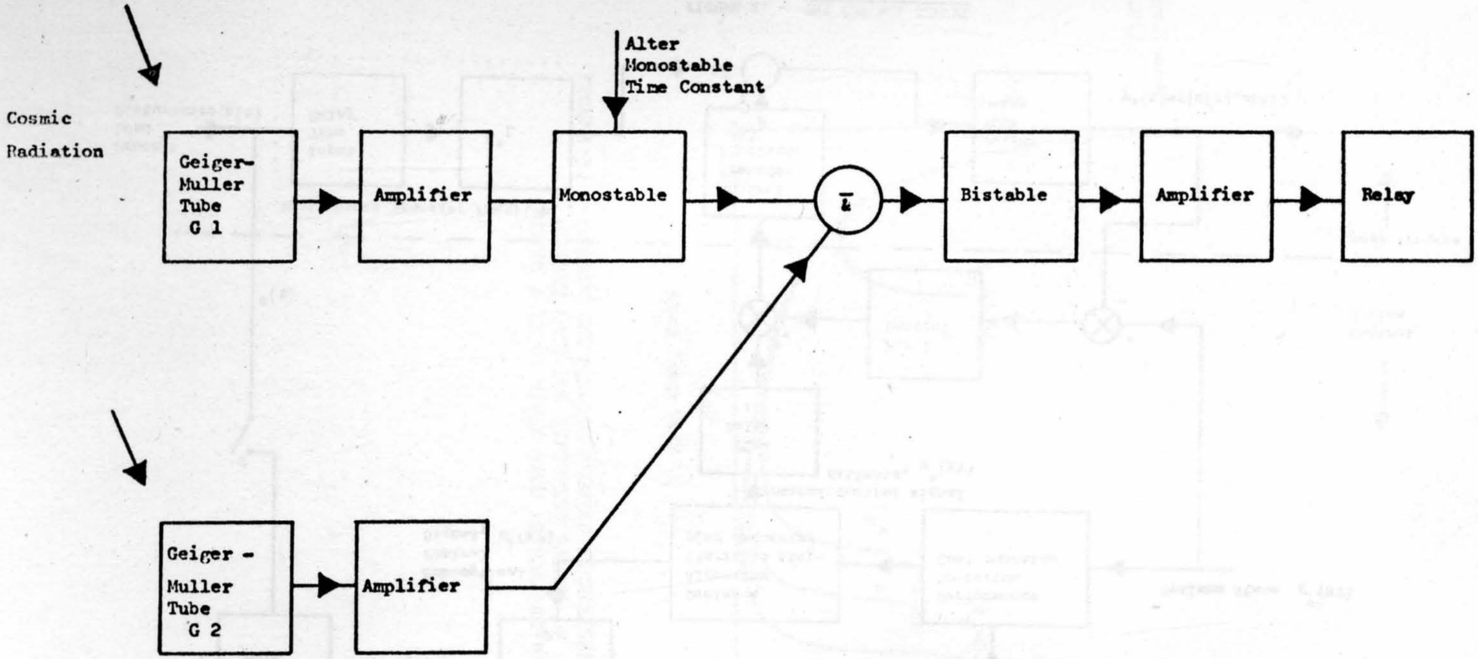


FIGURE A 1. **SCHEMATIC DIAGRAM OF RANDOM SEQUENCE GENERATOR**

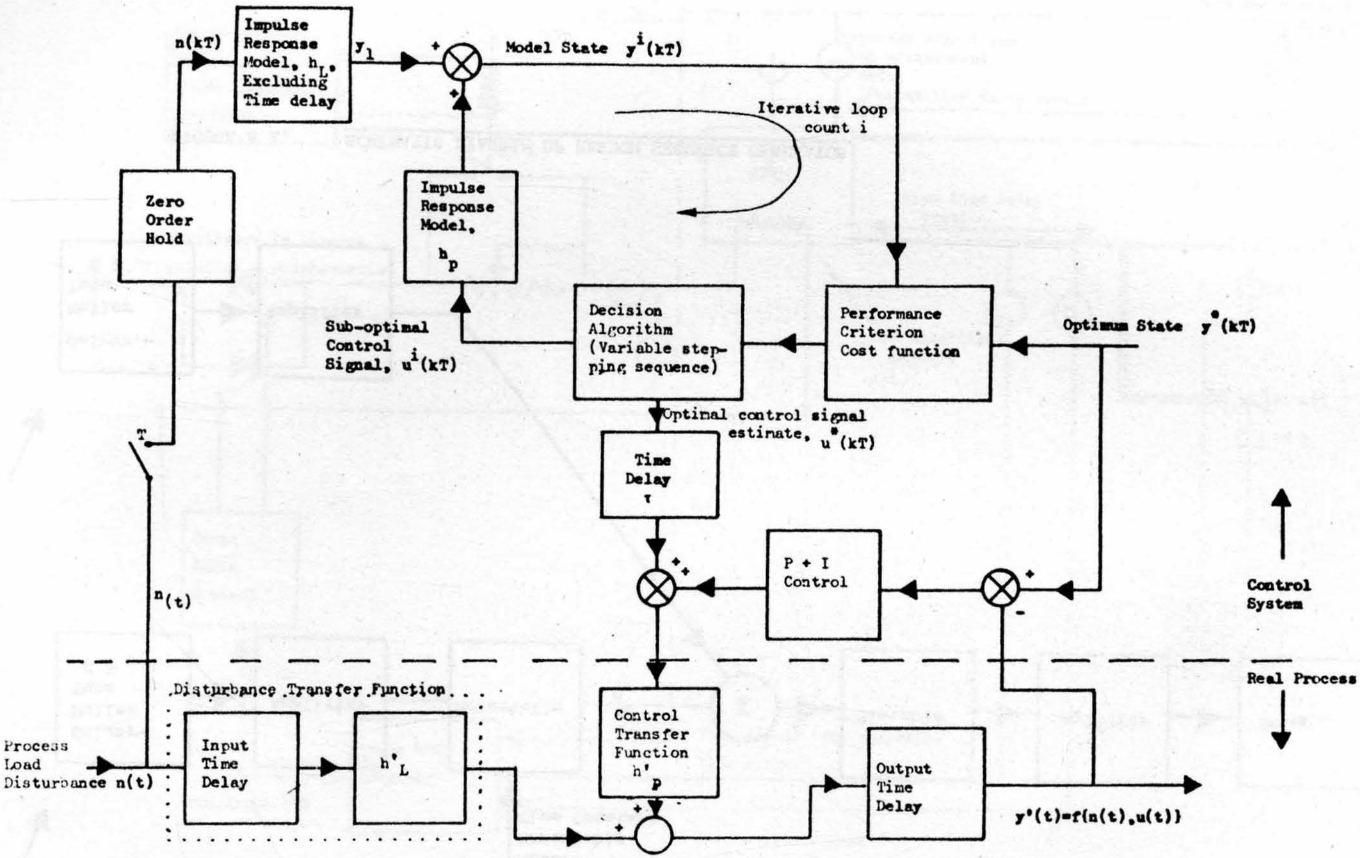


FIGURE 2. THE CONTROL SCHEME

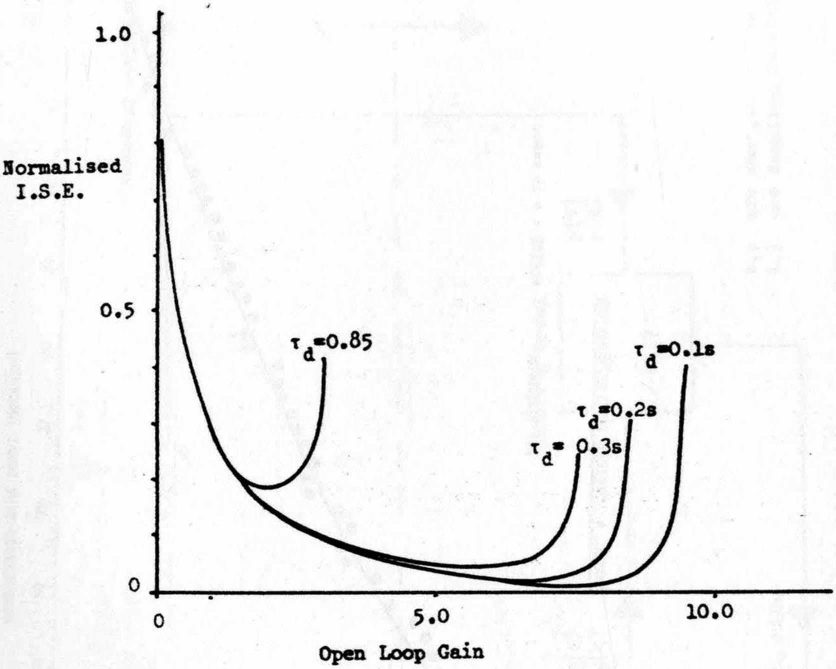


FIGURE 3. THEORETICAL RESULTS SHOWING NORMALISED I.S.E. PLOTTED AGAINST GAIN FOR DIFFERENT VALUES OF DEAD TIME (τ_d) USING A FIRST ORDER PADE APPROXIMATION

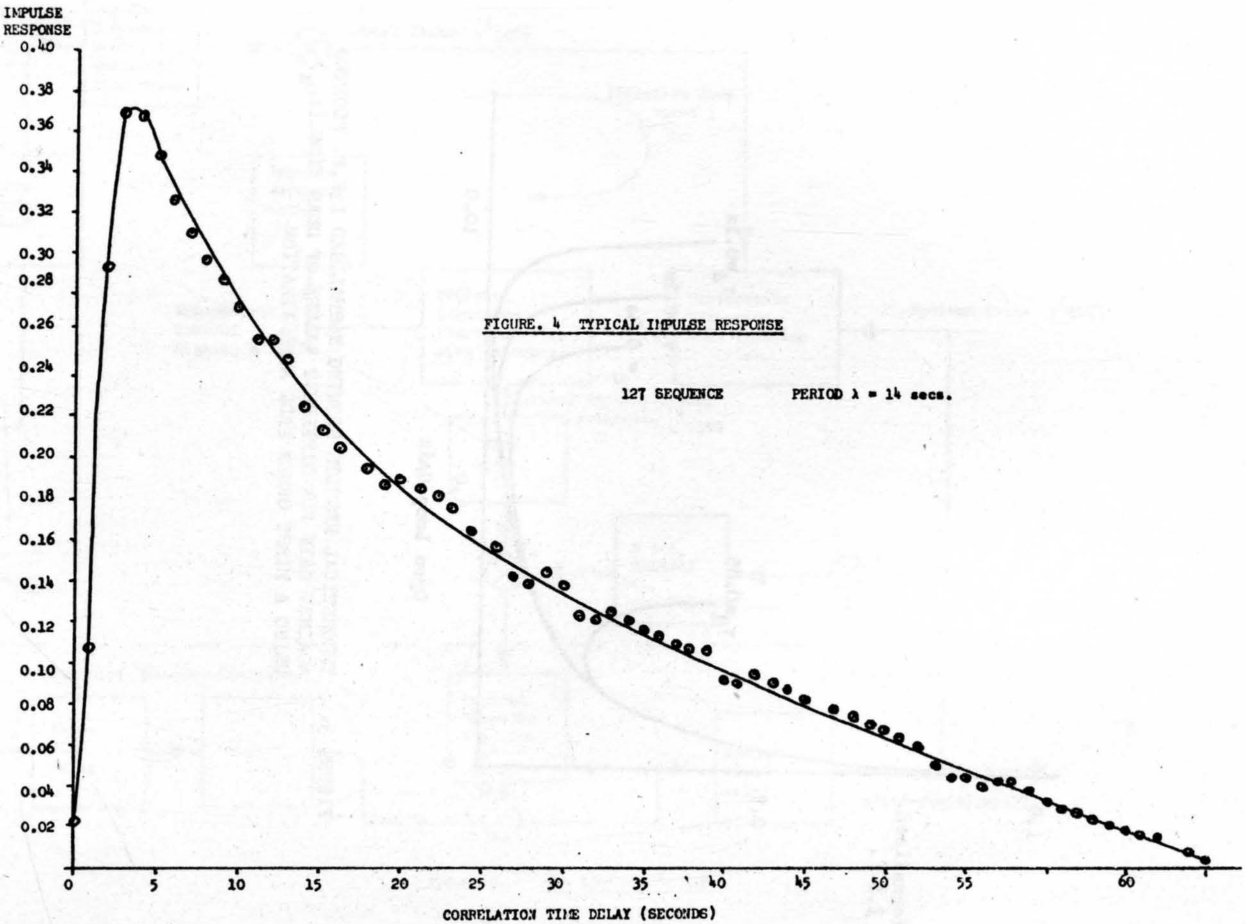


FIGURE. 4 TYPICAL IMPULSE RESPONSE

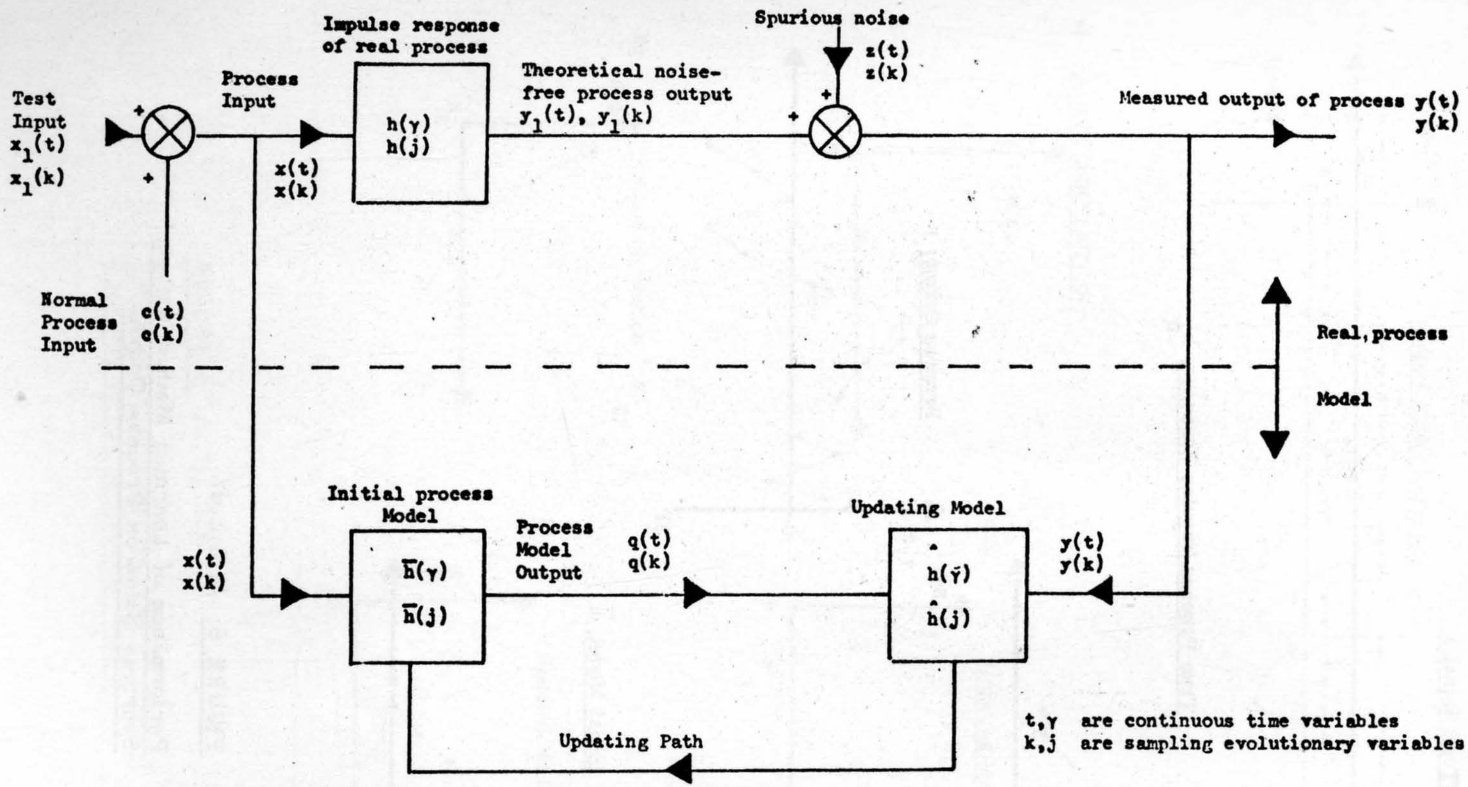


FIGURE 5. THE LEARNING METHOD

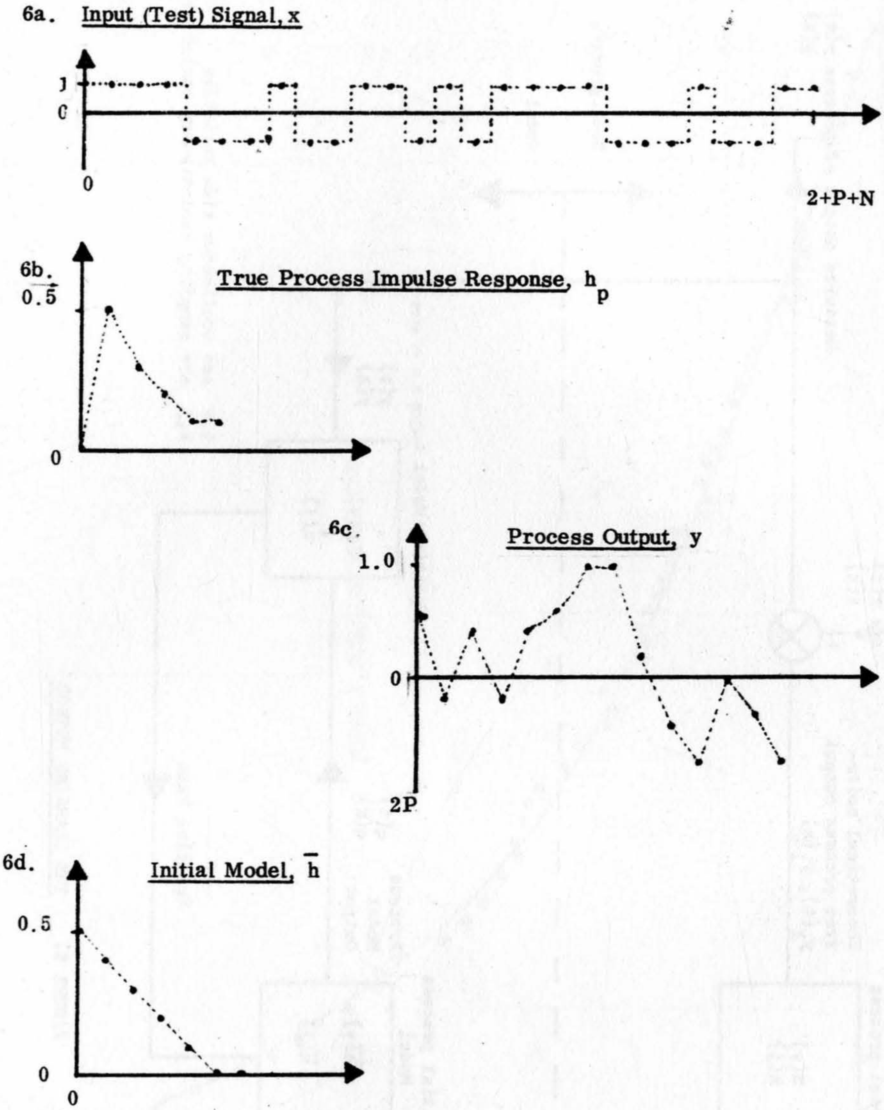


FIGURE 6 /a,b,c,d/

Performance of Learning Method Without Spurious Noise on Process Output

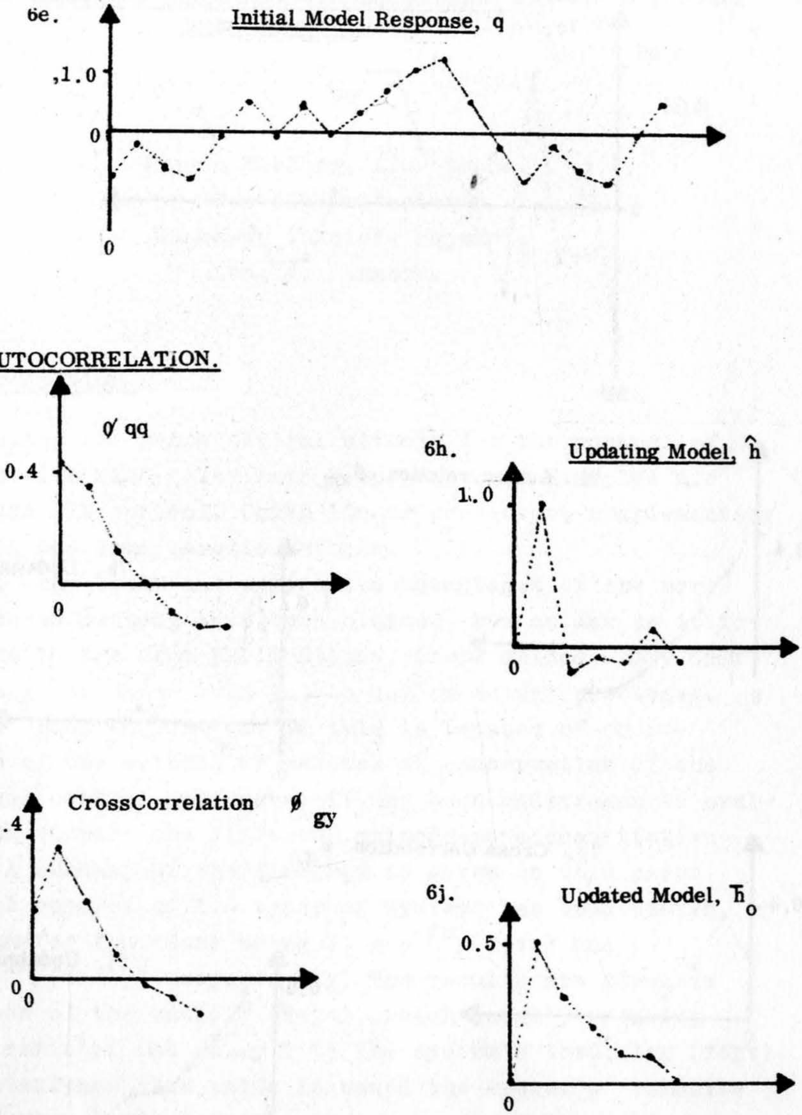


FIGURE 6 /e,f,g,h,j/

Performance of Learning Method Without Spurious Noise on Process Output

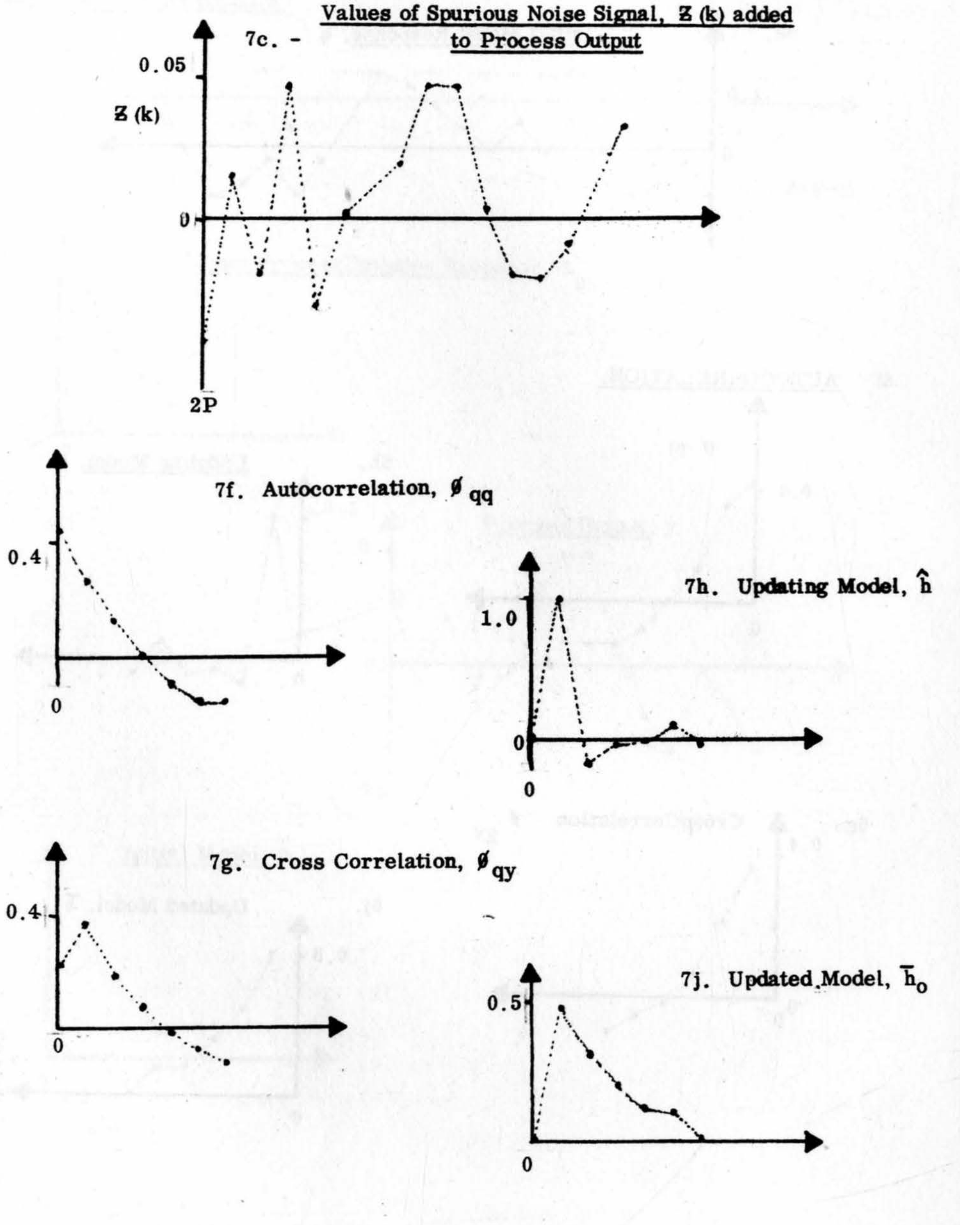


FIGURE 7

Performance of Learning Method with Spurious Noise on Process Output

CONTROL OF SYSTEMS WITH TIME DELAY

by

Gunnar Nielsen, lic. techn.
Servolaboratoriet
Danmarks tekniske Højskole
Lyngby, Danmark.

1: Introduction.

During the years several methods for the control of systems with time delay have been proposed. Examples are I, PI and PID control, Smith linear predictor, complementary feedback and some sampled methods.

In many books and papers the advantages of the more complicated methods have been claimed, but as far as it is possible to see from publications, these methods have only in a very few cases been put to use on actual processes.

In order to find out if this is because of shortcomings of the methods or because of conservatism of the practical control engineers, it has been undertaken to evaluate and compare the different methods on a quantitative basis.[7] A summary of the findings is given in this paper.

The control of two types of systems has been tested, the transfer functions being $f_1 = e^{-Ts}/(1+\tau s)$ and $f_2 = e^{-Ts}/(1+\tau s)^2$ respectively. The results are given as functions of the ratio $T/(T+\Sigma\tau)$, which roughly speaking is the ratio of the delay T to the system's total lag $(T+\Sigma\tau)$. For convenience this ratio is named the system's 'relative delay'. When it is zero the system has no delay, and when it is one it has delay only.

In order to make the comparisons quantitative, a performance criterion is used. As will be seen, there are certain conceptual advantages in choosing the IAE-index, and besides, it mostly results in good controller settings.

If the system has the transfer function

$$f = e^{-sT} / \prod_{i=1}^{n} (1+s\tau_i) \tag{1.1}$$

it is possible to evaluate the value I of the IAE-index for the uncontrolled system with a unit step input. All poles are real, and the error will then have no zero crossings:

$$I = \int_0^\infty |e| dt = \int_0^T 1 \, dt + \int_T^\infty e(t) dt \tag{1.2}$$

$$= T + \int_0^\infty e(t_1) dt_1 \qquad t_1 = t - T$$

The Laplace transform of $e(t_1)$ is

$$e(s) = \frac{\Pi(1+s\tau_i)-1}{s\Pi(1+s\tau_i)} \tag{1.3}$$

and by using the final value theorem:

$$I = T + \int_0^\infty e \, dt_1 = T + \lim_{s \to 0} s \times \frac{1}{s} \frac{\Pi(1+s\tau_i)-1}{s\Pi(1+s\tau_i)} = T + \Sigma\tau_i \tag{1.4}$$

One of the purposes of introducing feedback is to make the response faster, that is the index value lower. To measure how well this is obtained, the performance index is normalized by dividing by the index value of the uncontrolled system i.e. by $T+\Sigma\tau_i$. Normalized index values greater than one then means that the feedback actually makes the system slower.

2: Ordinary I, PI and PID-control.

a) Reference step changes.

For the systems the controller settings are adjusted to obtain minimum value of the performance index. Fig. 2.1 shows the values of the normalized index as a function of the relative delay for the two systems, and the three different controllers. The response is well damped except in some cases for small values of the relative delay.

b) Load step changes.

In the same way we adjust the settings to minimum index value and obtain the curves fig. 2.2. The open loop index value is here infinite, so we normalize by dividing by $T+\Sigma\tau$ as for the reference step change. The load is entering the process in the front; if it does enter later, the results will fall somewhere between figs. 2.1 and 2.2.

It is seen that only minor amounts of delay increases the index value very much, and for large values of the relative delay, feedback control actually slows the system down.

It is also seen that I-control is poor, and that the difference between PI and PID decreases much as the relative delay increases. If for instance the relative delay is above 75% a PID-controller will only decrease the index by 18% or less from the value obtained by a PI-controller. This is true for both reference and load step changes, and it means that for these processes it is seldom worth while taking the trouble to tune one more controller mode.

A rather interesting result can be seen from fig. 2.2. It relates to the curves 1-2, 1-3 and 2-3; it is seen that for relative delays lower than a certain value, the index decreases faster than proportional to the relative delay. This means that a lower index value may be obtained by increasing the system's time constant(s). The effect is easier to see when redrawing the curves in another diagram fig. 2.3. Here the index is not normalized, and as abscissa is used $n\tau(T=1)$. In this way the effect on the system of increasing the lags is immediately seen.

The large improvement obtainable in this way has to be paid for by a slower control for setpoint changes. As load changes often are much more important than setpoint changes, an improvement along these lines may well be worth considering. It should be noted however, that this method is not a special control method, but a modification of the plant to make it easier to control.

3: Smith predictor.

This was proposed by O. J. Smith[8] in the late 50's. The method has later been much discussed but few applications have been published[5].

The method may best be described with the help of fig. 3.1.

C(s) is a common controller. It is seen that the scheme includes a minor feedback loop around the controller. The transfer function becomes:

$$y = \frac{r(CHe^{-sT})+1He^{-sT}(1+CH(1-e^{-sT}))}{1+CH} \tag{3.1}$$

It follows that the predictor removes the term e^{-sT} from the denominator, thus making faster control possible.

We will first discuss the method for setpoint changes. The transfer function is then

$$y = r\,\frac{CH}{1+CH}\,e^{-sT} \tag{3.2}$$

Apparantly the delay is moved outside the closed loop. If the controller C is very good the closed loop can be made infinitely fast, and the value of the index is T as the error e=r-y will be 1 until the time T and then drop to zero.

This result is shown as a curve in fig. 2.1 marked s. It is seen that for a relative delay of 1 the improvement (28% better than PI, 18% better than PID) is useful but not startling. For lower values of the relative delay the improvement is better. For a value of .5 the improvement is of order 42-57% compared to PI and 31-44% compared to PID depending on the plant transfer functions. This is contradictory to the general recommendation, that the Smith predictor is for systems with high relative delays. However, we have supposed ideal control of the CH-loop, and since this cannot be achieved in practice, the improvements will be smaller, especially for lower values of the relative delay.

For load changes things are more complicated as may be seen from the transfer function

$$y = lHe^{-sT}(1 - \frac{CHe^{-sT}}{1+CH}) \qquad (3.3)$$

However, it can be shown that the minimum index value, when using an infinitely good C-controller will be T as for step changes. This value is shown in fig. 2.2. Whereas this value is the lowest possible for setpoint changes, it is not so for load changes. This is also seen from fig. 2.2. For low values of the relative delay, PID, and sometimes even PI-control, are better, and as we have seen above, if the plant itself may be modified by adding a large lag, then PID or PI-control can be made to perform better than the Smith predictor for all values of the relative delay of the original system.

It must be concluded that for load changes, the Smith predictor is of little value. Furthermore it is complicated to construct, so it is not surprising that it has not been used much for process control.

If the control loop mainly is exposed to reference changes, then the predictor may be of interest. In this case it is necessary to discuss the realisation of the delay which is used in the minor loop. If a control computer is used, the input signal can be stored during the time T, but this will require 20-50 memory locations to obtain reasonable accuracy. Another possibility which is acceptable also when using analog hardware, is to use some finite order transfer function as an approximation. If a first order Padé approximation is used, the control will deteriorate somewhat. Generally it has been found that the control, when adding a predictor with this approximation, did not improve more than when adding one more controller mode, which is of course much easier to do.

The second order Padé approximation worked as well as the exact delay and should therefore be used.

At the same time as Smith in U.S. proposed the Smith predictor, Wolman and Giloi in Germany came up with a control scheme which they named 'Complementary feedback'[2,3]. It is, however, easy to transform one of the methods into the other, and what has been said about the predictor is therefore applicaple to the complementary feedback too.

4: Sampled data control.

In several places[1,4,6,9,10] it has been stated, that sampled data control may perform better for plants with dead time than does continuous control. As the introduction of a sampler generally means that information is lost, this is surprising, and has to be investigated carefully, before it is accepted.

The general recommendation is to sample at such a rate, that the dead time T becomes a multiple of the sampling time T_1, $T = mT_1$. A controller which gives dead beat response to a step reference input is then constructed. There are now two approaches. One of them, interrupting control[4,9,10] uses no hold after the sampler which is assumed to close during the time ΔT. The plant is here taken to be of first order. By using correct settings of a PI-controller and a sampling time equal to T, it is now possible to obtain a response, where the error is 1 until time T, and then during the time ΔT, changes to zero. The index value should then be a little more than T, or about the same as for the Smith predictor.

It is, however, easy to see the disadvantages of this method. The control is obtained by short pulses, and accordingly the peak values are very high. The input has to be a step and to come immediately before the sampling instant. If it comes later, the controller does not react until the next sampling instant, and the index value will be a little more than 2T, which is not very good. For reference inputs of other types, and for load inputs the control will generally be poor. For these reasons the method is not recommended.

The other approach is to use a zero-order hold and then construct a conventional dead-beat controller[6]. In this case the error is 1 until time T and then changes to zero during the time T_1. The index value will then take on values from about $T + \frac{1}{2}T_1$ to about $2T + \frac{1}{2}T_1$ depending on when the step comes in relation to the sampling instants.

This method can be used for plants of higher order than one, and the index may be decreased by increasing the value of m. The method is, however, much dependent on an exact transfer function of the plant being known. As for the other method it is constructed for one type of input and performs badly for other reference inputs and for load inputs in general.

We must conclude that sampled data controllers only under very special circumstances give better control than conventional controllers and even then the improvement is small. They are more complicated and much more sensitive to plant parameter variations. Sampling should therefore not deliberately be introduced in process control equipment, and where it is unavoidable the sampling rate should be as high as possible.

5: Feedforward control.

As we can see from the figures 2.1 and 2.2 feedback control is in many cases worse than no control as far as the value of the index is concerned. If the load can be measured it might therefore be advantageous to control mainly by feedforward techniques, and then use only a slow and simple feedback control to take care of drift, inaccuracies in the feedforward components etc. Even for setpoint changes a sort of feedforward control may have advantages.

In the former chapters the load has been assumed to enter the plant at the input. This is too specific in this case, and the block diagram is chosen as shown in fig. 5.1.

For reference input signals the transfer function is

$$y = r\,\frac{(FC+G)He^{-sT}}{1+HCe^{-sT}} \qquad (5.1)$$

We want this to be equal to the theoretical best possible $y = re^{-sT}$. This may be obtained in several ways by fulfilling the equation

$$(FC+G)H = 1+HCe^{-sT} \qquad (5.2)$$

or

$$FC+G = \frac{1}{H} + Ce^{-sT} \qquad (5.3)$$

A simple choice is $F=e^{-sT}$ and $G=1/H$.

A sufficiently good approximation of F is relatively easy to obtain, and a lead element may be used as an approximation to 1/H. For systems with a high relative delay this should perform well, compared to any other method, and not be too complex to instrument.

For load changes we use the feedforward compensator B. In order to cancel the load input completely, it has to be chosen as $B = e^{sT_a}/H_a$ which is not realisable, but must be approximated by some kind of lead element. If the dynamics of $H_a e^{-sT_a}$ are fast compared to that of $H_b e^{-sT_b}$ i.e. if the load enters in the front end of the plant, this approximation will give a good result, better than any feedback configuration. If it is the other way around i.e. the load enters near the end of the plant, it will be of little value, and a feedback scheme may be better.

6: Conclusion.

Several schemes for the control of dead time plants have been discussed, and to a certain extent, their properties have been compared. As so many points of view are possible, it is of course not to be expected, that a list can be given of the different controllers with a single number measuring the performance. Only in very broad lines it is possible to give information concerning the choice of controller for a given application.

Observing these limitations, several conclusions may, however, be drawn.

The superiority of sampled data controllers has often been claimed, but as it is shown here, this holds only under very special circumstances with little relevance to practical problems. In general it can be said that sampling should be avoided, and if this is impossible, then the sampling rate should be as high as possible.

The advantages of control schemes like the Smith predictor or conditional feedforward are also smaller than generally stated. These schemes are designed for setpoint changes and do not behave well for load changes.

The performance of conventional controllers is good. I-control of course is slow, but it will seldom be worth while using more complicated control schemes than PI-control. It should be remembered, that for plants without delay the improvement in index value when adding a D-term to the controller will often be 2 to 3 times. For delay plants it may be 20%, and a more complicated control scheme such as f. inst. the Smith predictor may give another 20%.

For the control of load changes the advantage of adding a lag to the process should be noted.

Feedforward control relies on a completely different principle than does feedback control, and comparison is therefore difficult. When the conditions are right, however, a very good control may be obtained by this method for both load and reference inputs. This type of control should be seriously considered whenever control of plants with high relative delay is undertaken.

References.

1. P. S. Buckley: Techniques of process control, Wiley 1964 p. 88 - 96.

2. W. Giloi: Optimized feedback control of dead time plants by complementary feedback. IEEE-transactions AC, May 64.

3. W. Giloi: Zur theorie verwirklichung einer regelung für laufzeitstrecken nach dem prinzip der ergänzenden rückführung. Dr.ing.-dissertation Technischen Hochschule, Stuttgart. 1960.

4. M.J.Hausner: Get the benefits of sampled data control from modified process controllers. Control Eng. Mar. 61.

5. Lupfer, D.E. and Oglesby, M.W: Applying dead-time compensation for linear predictor process control ISA Journal. Nov. 61. Vol. 8 No. 11.

6. M. Mori: Discrete compensator controls dead time process. Control Engineering. Jan. 62.

7. G. Nielsen: Control of systems with time delay. Servolaboratoriet, Danmarks tekniske Højskole, Lyngby, 1967.

8. O.J.M.Smith: A controller to overcome dead time. ISA Journal Feb. 59. Vol. 6 No. 2.

9. W.H.Wheather: An examination of techniques for the automatic control of processes with dead times. Systems Res.Cent.Case 1963 No. SRC 33-C-63-12.

10. C. S. Zalkind: Sample data control of a predominantly dead time process. Internal report. The Foxboro Co. 1965.

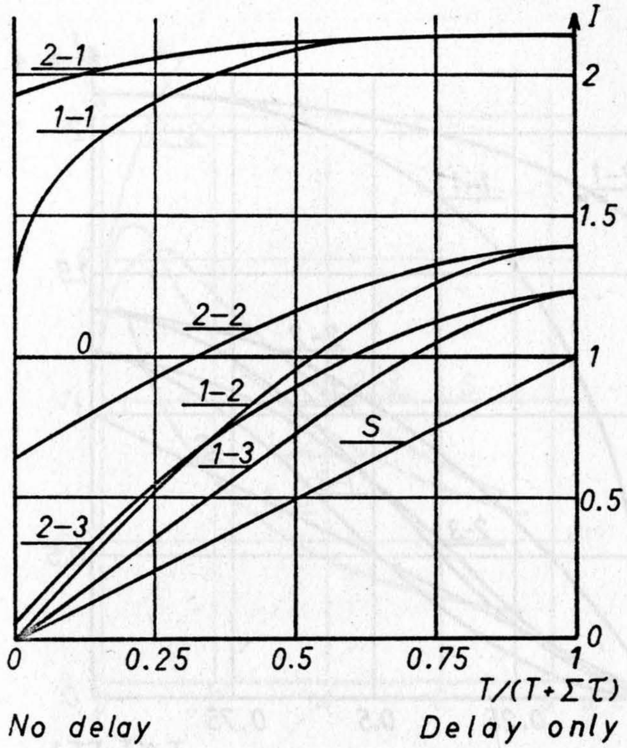


Fig. 2.1. Setpoint change.
First index: Number of lags.
Second index: Number of controller modes:
0: No control, 1: I-control, 2: PI-control, 3: PID-control.
S: Smith predictor, ideal

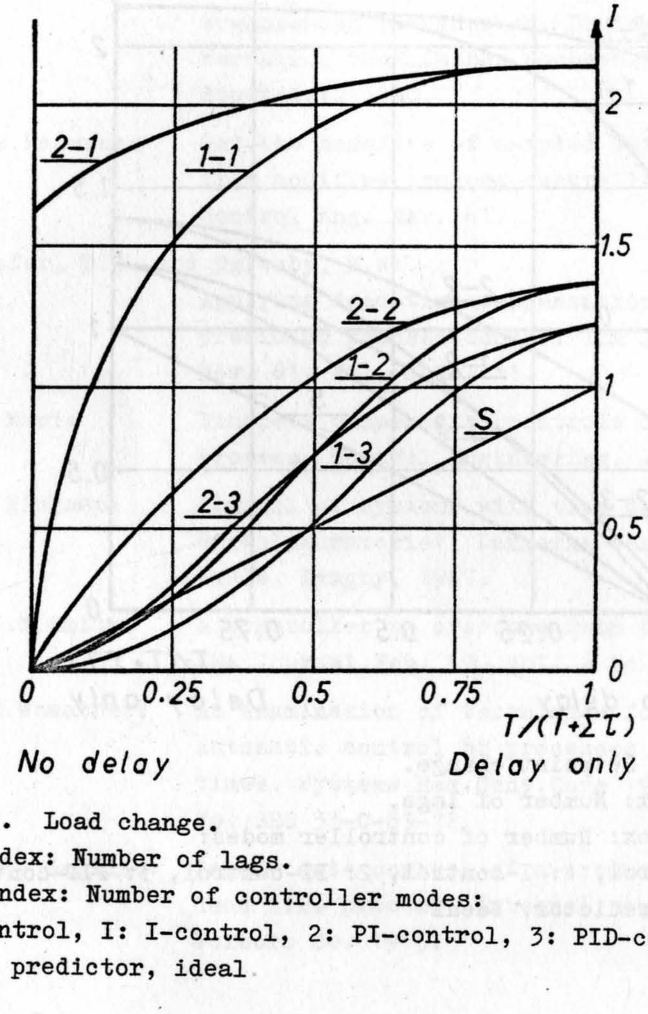


Fig. 2.2. Load change.
First index: Number of lags.
Second index: Number of controller modes:
0: No control, I: I-control, 2: PI-control, 3: PID-control.
S: Smith predictor, ideal

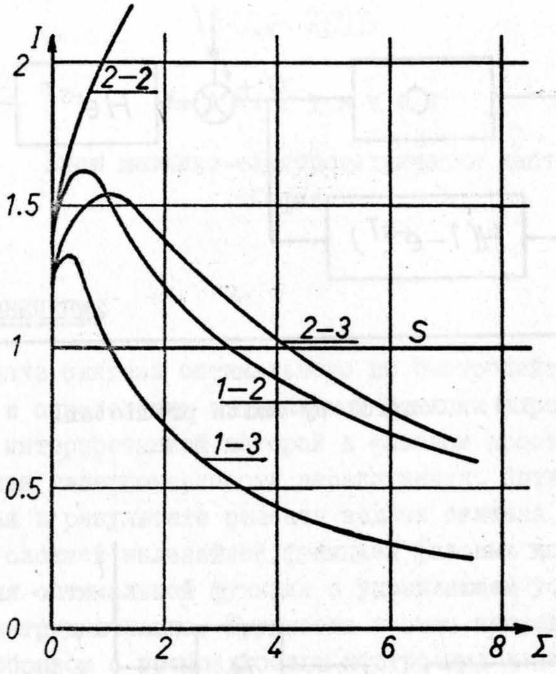


Fig. 2.3. Load change.
First index: Number of lags.
Second index: Number of controller modes:
2: PI-control, 3: PID-control.
S: Smith predictor, ideal.

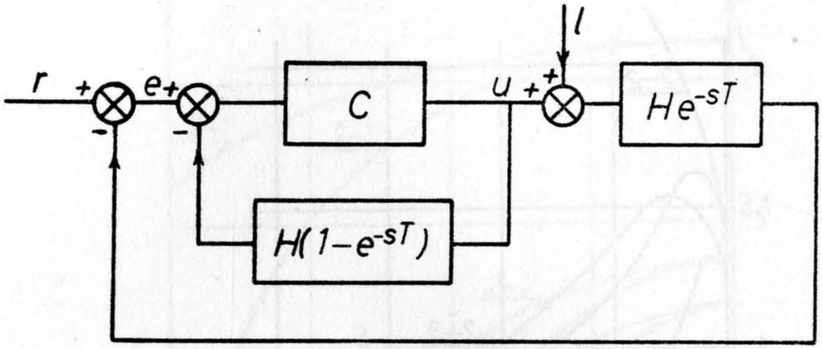


Fig. 3.1. Control by Smith predictor.

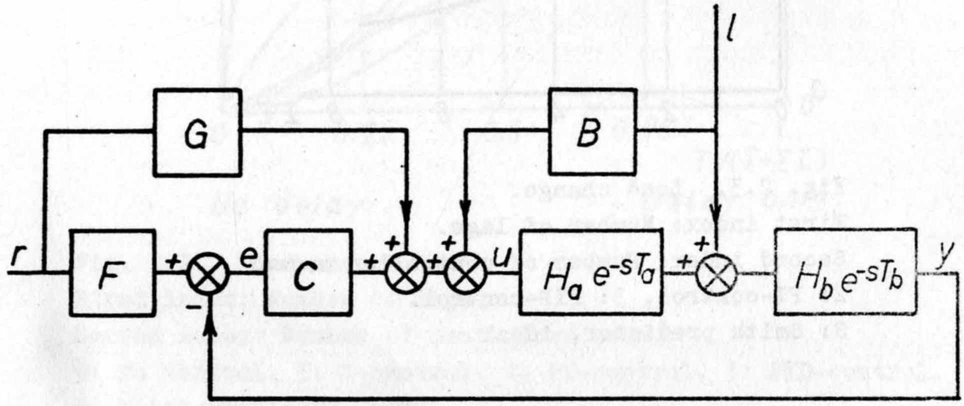


Fig. 5.1. Combined feedforward and feedback control

АППРОКСИМИРУЮЩИЕ СИГНУМ-ФУНКЦИИ ПРИ ПОСТРОЕНИИ КВАЗИОПТИМАЛЬНЫХ ПО БЫСТРОДЕЙСТВИЮ УПРАВЛЯЮЩИХ УСТРОЙСТВ

Л. Ат. Гунчев

Висш машинно-електротехнически институт
София

I. ВВЕДЕНИЕ

Задача синтеза оптимального по быстродействию процесса сводится к определению оптимальной функции управления, геометрической интерпретацией которой в фазовом пространстве является оптимальная гиперповерхность переключения. Оптимальная функция, полученная в результате решения задачи синтеза, является, как правило, сложной нелинейной функцией фазовых координат. Точная реализация оптимальной функции в управляющем устройстве / УУ / системы затруднительно. Трудности в этом направлении связаны главным образом с громоздкостью построения нелинейных функциональных преобразователей от нескольких независимых переменных, а также с большим числом преобразователей и множительных звеньев необходимых для точной реализации строго оптимальной гиперповерхности переключения.

Основным подходом к решению задачи построения квазиоптимального УУ является нахождение аппроксимирующей гиперповерхности переключения в фазовом пространстве, близкой к строго оптимальной. При этом подходе аппроксимирующие функции должны относиться к классу удобных для технической реализации функций. Конечно, можно искать решение этой задачи в классе всех сравнительно удобных для технического построения нелинейных функций фазовых координат [1,2], например квадратичных функций, некоторых параболических функций, нелинейных функций одной независимой переменной и т.д.

Теория оптимальных процессов возникла еще в начале 50-х годов, но ее практическое применение к построению систем управления с неизменной частью третьего порядка и выше оказалось довольно затруднительным. Публикации [3], примыкающие к вопросам аппроксимации сложной нелинейной функции управления, начали появляться также в 50-е годы. Для получения требуемой точности приближения к оптимальной гиперповерхности переключения оказалось необходимым расширить класс аппроксимирующих функций и усложнить их, но тогда практическое построение УУ становится связанным со значительными трудностями и иногда выходит за пределы целесообразной технической реализации.

Представляет несомненный интерес ввести сильное ограничение принятых для технического построения функций и искать решение задачи аппроксимации оптимального по быстродействию управления в классе линейных функций и сигнум-функций линейных комбинаций фазовых координат, которые в наибольшей степени отвечают требованию простоты технического выполнения.

2. ЭКВИВАЛЕНТНЫЕ СИГНУМ-ФУНКЦИИ

Оптимальная функция $U_0(x_1,\dots,x_m)$, полученная в результате решения задачи синтеза, относится к определенному классу эквивалентных синтезирующих функций $U_{cэ}(x_1,\dots,x_m)$, которые эквивалентны по знаку оптимальной функции, т.е.

$$\mathrm{sign}\left[U_{cэ}(x_1,\dots,x_m)\right]=\mathrm{sign}\left[U_0(x_1,\dots,x_m)\right], \qquad /1/$$

где $x_1,\dots,x_m$ – фазовые координаты системы.

Формирование оптимального по быстродействию управляющего воздействия $V_0(t)$, которое поступает на вход объекта, можно производить на базе любой эквивалентной синтезирующей функции, так как

$$V_0(t)=\mu\,\mathrm{sign}\left[U_{cэ}(x_1,\dots,x_m)=\mu\,\mathrm{sign}\left[U_0(x_1,\dots x_m)\right], \qquad /2/$$

где μ – модуль входного управляющего воздействия.

Процесс аппроксимации оптимального управляющего воздействия можно производить при помощи эквивалентных аппроксимирующих функций $U_{aэ}(x_1,\dots,x_m)$, которые эквивалентны по знаку вибранным для реализации функциям $U_{pэ}(x_1,\dots,x_m)$, т.е.

$$\mathrm{sign}\left[U_{aэ}(x_1,\dots,x_m)\right]=\mathrm{sign}\left[U_{pэ}(x_1,\dots,x_m)\right]. \qquad /3/$$

Если в процессе приближения аппроксимирующая функция выбрана в виде эквивалентной функции $U_{aэ}(x_1,\dots,x_m)$, то достаточно построить в УУ эквивалентную по знаку реализующую функцию, чтобы обеспечить идентичность входного управляющего воздействия, так как

$$V_a(t) = \mu \operatorname{sign}\left[U_{рэ}(x_1,\dots,x_m)\right] = \mu \operatorname{sign}\left[U_{aэ}(x_1,\dots,x_m)\right]. \qquad /4/$$

Эквивалентные аппроксимирующие функции необходимы для процесса приближения, но они не реализуются. УУ квазиоптимальной системы реализуется при помощи простых и удобных для технического построения функций, посредством которых синтезирована эквивалентная реализующая функция $U_{рэ}(x_1,\dots,x_m)$. В качестве функции, принятых для технического выполнения, рассмотрим линейные функции фазовых координат

$$w_i(x_1,\dots,x_m) = C_{i1}x_1 + C_{i2}x_2 + \dots + C_{im}x_m + C_{i0} \qquad /5/$$
$$(i = 1,2,\dots,n)$$

и сигнум-функции от линейных комбинаций координат. При синтезе реализующей функции можно производить операции, которые тоже отвечают требованию простоты технического выполнения, например сумма сигнум-функций, произведение сигнум-функций, сумма линейной функции и сигнум-функции, произведение линейной функции на сигнум-функцию.

Управление, определенное посредством сигнум-функции от суммы сигнум-функции линейной комбинации координат w_1 и линейной функции w_2, т.е.

$$\overset{2}{U}(x_1,\dots,x_m) = \mu \operatorname{sign}\left[\overset{2}{U}_{рэ}(x_1,\dots,x_m)\right] = \mu \operatorname{sign}\left[\operatorname{sign} w_1(x_1,\dots,x_m) + w_2(x_1,\dots,x_m)\right] \qquad /6/$$

является эквивалентным "квадратичным" управлением, а именно:

$$\overset{2}{U}_э(x_1,\dots,x_m) = \mu \operatorname{sign}\left[w_1(x_1,\dots,x_m) + \left|w_1(x_1,\dots,x_m)\right| w_2(x_1,\dots,x_m)\right] = \mu \operatorname{sign}\left[\overset{2}{U}_{aэ}(x_1,\dots,x_m)\right]. \qquad /7/$$

Действительно, легко показать, что

$$\begin{aligned}\operatorname{sign} \overset{2}{U}_{рэ}(x_1,\dots,x_m) &= \operatorname{sign}\left[\operatorname{sign} w_1(x_1,\dots,x_m) + w_2(x_1,\dots,x_m)\right] = \\ &= \operatorname{sign}\left[w_1(x_1,\dots,x_m) + \left|w_1(x_1,\dots,x_m)\right| w_2(x_1,\dots,x_m)\right] = \\ &= \operatorname{sign} \overset{2}{U}_{aэ}(x_1,\dots,x_m),\end{aligned} \qquad /8/$$

т.е. эквивалентная аппроксимирующая функция $\overset{2}{U}_{aэ}(x_1,\dots,x_m)$ содержит произведение линейных функций, а следовательно и квадратичные члены фазовых координат. Далее также можно определить другие формы эквивалентного "квадратичного" управления при помощи суммы двух сигнум-функций или произведение линейной функции на сигнум-функцию.

3. АППРОКСИМИРУЮЩИЕ ЭКВИВАЛЕНТНЫЕ СИГНУМ-ФУНКЦИИ

Аппроксимирующая функция $\mathcal{U}^{2}_{aэ}(x_1,\dots,x_m)$ в форме /8/является одной из возможных форм аппроксимирующих эквивалентных сигнум-функций второго порядка. При помощи последовательности трех сигнум-функций от линейных комбинаций координат можно получить эквивалентную аппроксимирующую функцию $\mathcal{U}^{3}_{aэ}(x_1,\dots,x_m)$ в виде полинома третьего порядка. Также можно определить управляющее воздействие в виде сигнум-функции от суммы трех сигнум-функций и т.д. Аппроксимирующая эквивалентная функция $\mathcal{U}^{3}_{aэ}(x_1,\dots,x_m)$ содержит произведения трех линейных функций и таким образом управление определенное этой функцией является эквивалентным "кубичным" управлением. Полученные результаты посредством нескольких сигнум-функций от линейных функций координат можно обобщить с целью получения выражений аппроксимирующих функций, которые эквивалентны по знаку реализующим функциям в классе линейных функций и сигнум-функций линейных комбинаций координат.

Если аппроксимирующая функция $\mathcal{U}^{n}_{aэ}(x_1,\dots,x_m)$, определена в виде аппроксимирующего сигнум-полинома n -го порядка

$$\overset{n}{G}(x_1,\dots,x_m)=\sum_{j=1}^{n} g_j(x_1,\dots,x_m), \qquad /9/$$

где

$$g_j(x_1,\dots,x_m)=w_j(x_1,\dots,x_m)\left|\sum_{i=1}^{j-1} g_i(x_1,\dots,x_m)\right|, \qquad /10/$$

т.е. $g_j(x_1,\dots,x_m)$ - ряд сигнум-функций линейных комбинаций координат $w_j(x_1,\dots,x_m)$, а именно:

$$g_1(x_1,\dots,x_m)=w_1(x_1,\dots,x_m);$$

$$g_2(x_1,\dots,x_m)=w_2(x_1,\dots,x_m)\left|g_1(x_1,\dots,x_m)\right|;$$

$$g_3(x_1,\dots,x_m)=w_3(x_1,\dots,x_m)\left|g_1(x_1,\dots,x_m)+g_2(x_1,\dots,x_m)\right|$$

и т.д. , то существует в классе линейных функций и сигнум-функций линейных комбинаций координат эквивалентная реализующая функция

$$\mathcal{U}_{pэ}(x_1,\dots,x_m)=\operatorname{sign}\Big\{\operatorname{sign}\Big[\operatorname{sign}\dots\operatorname{sign}(\operatorname{sign}w_1(x_1,\dots,x_m)+ \qquad /11/$$

$$+w_2(x_1,\dots,x_m))+\dots+w_{n-2}(x_1,\dots,x_m)\Big]+w_{n-1}(x_1,\dots,x_m)\Big\}+w_n(x_1,\dots,x_m),$$

которая удовлетворяет зависимости

$$\operatorname{sign} U_{aэ}(x_1,\dots,x_m)=\operatorname{sign} G^n(x_1,\dots,x_m)=\operatorname{sign} U_{pэ}(x_1,\dots,x_m). \quad /12/$$

Можно показать справедливость высказанной выше теоремы об эквивалентном управлении, если записать полином $G^n(x_1,\dots,x_m)$ вида /9/ в развернутой форме:

$$G^n(x_1,\dots,x_m)=G^1(x_1,\dots,x_m)+w_2(x_1,\dots,x_m)\left|G^1(x_1,\dots,x_m)\right|+\dots+$$
$$+w_i(x_1,\dots,x_m)\left|G^{i-1}(x_1,\dots,x_m)\right|+\dots+w_n(x_1,\dots,x_m)\left|G^{n-1}(x_1,\dots,x_m)\right|. \quad /13/$$

Полином i -го порядка получается от полинома $(i-1)$ -го порядка путем умножения на кусочно-линейную функцию, а именно:

$$G^i(x_1,\dots,x_m)=G^{i-1}(x_1,\dots,x_m)\left[\operatorname{sign}G^{i-1}(x_1,\dots,x_m)\right]\left[w_i(x_1,\dots,x_m)+\operatorname{sign}G^{i-1}(x_1,\dots,x_m)\right] \quad /14/$$

В соответствии с этим свойством G -полиномов можно записать зависимость для знака i -го полинома:

$$\operatorname{sign}G^i(x_1,\dots,x_m)=\operatorname{sign}\left[\operatorname{sign}G^{i-1}(x_1,\dots,x_m)+w_i(x_1,\dots,x_m)\right]. \quad /15/$$

Применением последнюю рекуррентную зависимость для некоторых значений порядка полинома, например,

$$i=3,\ \operatorname{sign}G^3(x_1,\dots,x_m)=\operatorname{sign}\left[\operatorname{sign}G^2(x_1,\dots,x_m)+w_3(x_1,\dots,x_m)\right]=$$
$$=\operatorname{sign}\left[\operatorname{sign}\left(\operatorname{sign}w_1(x_1,\dots,x_m)+w_2(x_1,\dots,x_m)\right)+w_3(x_1,\dots,x_m)\right]; \quad /16/$$

$$i=4,\ \operatorname{sign}G^4(x_1,\dots,x_m)=\operatorname{sign}\left[\operatorname{sign}G^3(x_1,\dots,x_m)+w_4(x_1,\dots,x_m)\right]$$

или

$$\operatorname{sign}G^4(x_1,\dots,x_m)=\operatorname{sign}\left\{\operatorname{sign}\left[\operatorname{sign}\left(\operatorname{sign}w_1(x_1,\dots,x_m)+\right.\right.\right.$$
$$\left.\left.\left.+w_2(x_1,\dots,x_m)\right)+w_3(x_1,\dots,x_m)\right]+w_4(x_1,\dots,x_m)\right\}$$ и т.д.

Таким образом, путем индикации для $i=n$ получается выражение /11/, при помощи которого определяется знак аппроксимирующего полинома $G^n(x_1,\dots,x_m)$.

Если аппроксимирующую функцию $U_{aэ}(x_1,\dots,x_m)$, определить в виде сигнум-полинома

$$G^n_\Sigma(x_1,\dots,x_m)=\sum_{i=1}^{n}G^i(x_1,\dots,x_m)\left|\prod_{\substack{j=1\\ j\neq i}}^{n}G^j(x_1,\dots,x_m)\right|, \quad /17/$$

где

$$G^j(x_1,\dots,x_m)=\sum_{i=1}^{j}g_i(x_1,\dots,x_m),$$

то существует в классе линейных функций и сигнум-функций линейных комбинаций координат эквивалентная реализующая функция

$$\begin{aligned} U_{рэ}(x_1,\dots,x_m) &= \operatorname{sign} W_1(x_1,\dots,x_m) + \operatorname{sign}\left(\operatorname{sign} W_1(x_1,\dots,x_m) + W_2(x_1,\dots,x_m)\right) + \operatorname{sign}\left[\operatorname{sign}\left(\operatorname{sign} W_1(x_1,\dots,x_m) + W_2(x_1,\dots,x_m)\right) + \right. \\ &\left. + W_3(x_1,\dots,x_m)\right] + \operatorname{sign}\left\{\operatorname{sign}\left[\operatorname{sign}\left(\operatorname{sign} W_1(x_1,\dots,x_m) + W_2(x_1,\dots,x_m)\right) + \right.\right. \\ &\left.\left. + W_3(x_1,\dots,x_m)\right] + W_4(x_1,\dots,x_m)\right\} + \dots + \\ &+ \operatorname{sign}\left\{\operatorname{sign}\left\{\operatorname{sign}\left[\operatorname{sign}\dots\operatorname{sign}\left(\operatorname{sign} W_1(x_1,\dots,x_m) + W_2(x_1,\dots,x_m)\right) + \right.\right.\right. \\ &\left.\left.\left. + \dots + W_{n-2}(x_1,\dots,x_m)\right] + W_{n-1}(x_1,\dots,x_m)\right\} + W_n(x_1,\dots,x_m)\right\}. \end{aligned} \qquad /18/$$

Знак полинома $G^n_\Sigma(x_1,\dots,x_m)$ сохраняется при умножении на определенно положительное выражение и можно записать:

$$\operatorname{sign} G^n_\Sigma(x_1,\dots,x_m) = \operatorname{sign}\left\{\frac{1}{\left|\prod_{i=1}^{n} G^i(x_1,\dots,x_m)\right|} \cdot \left[\sum_{i=1}^{n} G^i(x_1,\dots,x_m)\left|\prod_{\substack{j=1\\ j\neq i}}^{n} G^j(x_1,\dots,x_m)\right|\right]\right\} = \sum_{i=1}^{n} \frac{G^i(x_1,\dots,x_m)}{\left|G^i(x_1,\dots,x_m)\right|}. \qquad /19/$$

Таким образом получается выражение для знака аппроксимирующего полинома $G^n_\Sigma(x_1,\dots,x_m)$:

$$\operatorname{sign} G^n_\Sigma(x_1,\dots,x_m) = \sum_{i=1}^{n} \operatorname{sign} G^i(x_1,\dots,x_m). \qquad /20/$$

На базе рекуррентной зависимости /I5/ получаются реализующие сигнум-функции:

$$\operatorname{sign} G^i(x_1,\dots,x_m) = \operatorname{sign}\left\{\operatorname{sign}\left[\operatorname{sign}\dots\operatorname{sign}\left(\operatorname{sign} W_1(x_1,\dots,x_m) + W_2(x_1,\dots,x_m)\right) + \dots + W_{i-1}(x_1,\dots,x_m)\right] + W_i(x_1,\dots,x_m)\right\}. \qquad /21/$$

$(i = 1, 2, \dots, n)$

После подстановки в выражении /20/ сигнум-функций /2I/ для различных значений i, получается формула эквивалентной реализующей функций /I8/. Таким образом обобщения теорема об эквивалентном управлении доказана.

4. ПОСТРОЕНИЕ НЕКОТОРЫХ КВАЗИОПТИМАЛЬНЫХ УПРАВЛЯЮЩИХ УСТРОЙСТВ ПРИ ПОМОЩИ ЭКВИВАЛЕНТНЫХ СИГНУМ-ФУНКЦИЙ

После того как определены некоторые классы аппроксимирующих функций, эквивалентные по знаку выбранным для реализации функциям, может быть решена задача построения структуры квазиоптимального УУ. Сущность проделанных операций состоит в определении вида аппроксимирующих функций, которые эквивалентны по знаку реализующим функциям, построенным при помощи линейных функций и сигнум-функций координат. Приборная реализация УУ системы на базе выражений типа линейных функций координат достигается в схеме весьма просто; легко также реализовать сигнум-функции линейных комбинаций координат при помощи релейных элементов.

Аппроксимирующим полиномом $\overset{n}{G}(x_1,\ldots,x_m)$ в виде /9/ определяется эквивалентная реализующая функция в форме /II/ и при ее помощи реализуется эквивалентное управляющее воздействие посредством последовательности n сигнум-функций от сумм сигнум-функций и линейных комбинаций фазовых координат системы. Структура УУ, построенного на базе эквивалентной реализующей функции /II/, показана на рис.I.

Возможности преобразования эквивалентных реализующих структур неисчерпаемы, и этим путем можно расширять класс аппроксимирующих функций. Эквивалентными по знаку преобразованиями можно модифицировать эквивалентные аппроксимирующие функции и приближать их к виду эквивалентных синтезирующих функций. Последние функции принадлежат к классу оптимальной функции и является тоже оптимальными функциями переключения, так как, найдя одну из этих функций, можно реализовать при ее помощи строго оптимальное управление. Конечно, оптимальное управление является единственным, но синтезирующие его функции составляют множество. Это множество является бесконечным, поскольку неограниченно можно преобразовать одну функцию в другую,эквивалентную по знаку.

Суть последних операций заключается в том, что при помощи полученных эквивалентных аппроксимирующих функций можно построить эквивалентное УУ системы, которое имеет более простую структуру для цели технического построения. Достаточно любую эквивалентную синтезирующую функцию представить в конечной форме, определяемой рассмотренными аппроксимирующими полиномами, чтобы реализовать оптимальное управление посредством принятых для технического построения функций.

Структура УУ системы определяется выбором типа эквивалентного аппроксимирующего полинома. В зависимости от сложности и типа модели неизменной части системы управления определяется вид аппроксимирующих функций. При помощи аппроксимирующих функций в виде /I7/ получена эквивалентная реализующая структура в форме /I8/, которая состоит из последовательно-параллельного соединения релейных элементов. Это почти не усложнят структуру УУ по сравнению со схемой УУ, показанной на рис. I.

Требуемая точность приближения определяет порядок эквивалентного аппроксимирующего полинома. Согласно утверждениям рассмотренных теорем, число элементов УУ определено порядком этого полинома. Возможны любые приближения путем увеличения членов аппроксимирующего полинома и соответственно добавления новых звеньев в структурной схеме УУ. Таким образом задача синтеза структуры квазиоптимального УУ сводится к задаче выбора наилучшего вида для процесса аппроксимации выражения в классе эквивалентных аппроксимирующих функций и определения его формы в зависимости от требуемой точности приближения к строго оптимальному процессу. Рассматриваемые аппроксимирующие G -полиномы, определенные в виде /9/ и /I7/, выражаются при помощи линейных комбинаций координат $W_i(x_1,\dots,x_m)$. Эквивалентные реализующие функции, определенные соответственно в форме /II/ и /I8/, выражаются также при помощи линейных комбинаций координат. Существенно то обстоятельство, что эквивалентные функции содержат те же самые линейные комбинации $W_i(x_1,\dots,x_m)$, которые имеются в соответствующих им аппроксимирующих полиномах. Последнее свойство рассматриваемых эквивалентных функций управления дает возможность решить задачу об определении параметров УУ квазиоптимальной системы в процессе аппроксимации. Действительно, для определения линейных функций и сигнум-функций, при помощи которых построены эквивалентные реализирующие функции, нужно лишь найти коэфециенты в линейных функциях фазовых координат. Таким образом задача синтеза параметров квазиоптимального УУ сводится к нахождению в процессе аппроксимации, согласно принятому критерию приближения, неизвестных коэффициентов в линейных функциях координат.

При решении задачи определения параметров квазиоптимального УУ, аналитическое выражение оптимальной функции управления раскладывается в ряд. После оценки остаточного члена ряда можно с любой заданной точностью выразить оптимальную функцию конечным числом членов ряда, например n членов:

$$\mathcal{U}(x_1,\dots,x_m) = \mathcal{U}_{(0)} + \frac{\partial \mathcal{U}_{(0)}}{\partial x_1} x_1^0 + \frac{\partial \mathcal{U}_{(0)}}{\partial x_2} x_2^0 + \frac{\partial \mathcal{U}_{(0)}}{\partial x_3} x_3^0 + \cdots \qquad /22/$$

$$+\sum_{i=1}^{n} \frac{1}{i!}\left\{\frac{\partial \mathcal{U}_{(0)}}{\partial x_1} x_1^0 + \frac{\partial \mathcal{U}_{(0)}}{\partial x_2} x_2^0 + \frac{\partial \mathcal{U}_{(0)}}{\partial x_3} x_3^0 + \cdots\right\}^{(i)}, \qquad /22/$$

где $x_i^0 = x_i - x_{i0}, \quad (i = 1, 2, \ldots, m);$

$$\mathcal{U}_{(0)} = \mathcal{U}(x_1, x_2, x_3, \ldots)\Bigg|_{\substack{x_1 = x_{10} \\ x_2 = x_{20} \\ x_3 = x_{30} \\ \cdots}} ; \quad \frac{\partial \mathcal{U}_{(0)}}{\partial x_1} = \frac{\partial \mathcal{U}(x_1, x_2, x_3, \ldots)}{\partial x_1}\Bigg|_{\substack{x_1 = x_{10} \\ x_2 = x_{20} \\ x_3 = x_{30} \\ \cdots}} \text{ и т.д.},$$

а показатели степени, в которую возводятся выражения, стоящие в скобках, имеют символический смысл для определения порядка производных. Далее при помощи любой системы вспомогательных функций, определенной рассмотренными теоремами, можно записать вид аппроксимирующего полинома n -го порядка. Например, в качестве такого полинома посредством функций $g_j(x_1, \ldots, x_m)$ в виде /I0/ можно записать аппроксимирующий сигнум-полином n -го порядка в следующей форме:

$$G^n(x_1, \ldots, x_m) = w_1(x_1, \ldots, x_m) + w_2(x_1, \ldots, x_m)\left|w_1(x_1, \ldots, x_m)\right| + \\ + \cdots + w_n(x_1, \ldots, x_m)\left|\sum_{j=1}^{n} g_j(x_1, \ldots, x_m)\right|. \qquad /23/$$

Неизвестные коэффициенты C_{ij} в линейных комбинациях $w_i(x_1, \ldots, x_m)$ входящие в /23/, можно определить любыми существующими методами. В частности, это определение можно сделать путем приравнивания коэффициентов при одинаковых степенях в выражении /22/ и в выражении /23/ после написания его в развернутой форме по степеням фазовых координат.

Для илюстрации применения этого метода синтеза квазиоптимального УУ рассмотрен пример системы третьего порядка, неизменная часть которой состоит из последовательного соединения двух апериодических звеньев /постоянные времени: $T_1 = 0{,}0625, T_2 = 0{,}I345$; передаточные коэффициенты: $k_1 = k_2 = I$ / и одного интегрирующего звена / $T_3 = 0{,}0075$ /. В качестве аппроксимирующего полинома принят сигнум-полином

$$G^2(x_1, x_2, x_3) = \mu_1\left[C_{11}x_1 + C_{12}x_2 + C_{13}x_3\right] + \\ + \sigma_1\left[C_{11}x_1 + C_{12}x_2 + C_{13}x_3\right]\left[C_{21}x_1 + C_{22}x_2 + C_{23}x_3\right], \qquad /24/$$

где $\sigma_1 = \operatorname{sign} \mathcal{U}^1(x_1, x_2, x_3) = \operatorname{sign}\left[C_{11}x_1 + C_{12}x_2 + C_{13}x_3\right].$

Методом эквивалентных сигнум-функций определена самая простая структура квазиоптимального УУ в виде последовательного соеди -

нения двух релейных элементов, показанная на рис.2.

Аналитическим путем получено время строго оптимального процесса при подаче на вход системы единичной функции. После реализации указанной структуры квазиоптимального УУ на модели получено время переходного процесса, которое отклоняется на 22% от времени строго оптимального процесса. Конечно, можно построить УУ при помощи n релейных элементов и этим повысить порядок аппроксимирующего G -полинома, а также и точность приближения к оптимальному процессу.

Рассмотрен также пример синтеза другой системы третьего порядка, неизменная часть которой состоит из последовательного соединения трех интегрирующих звеньев / $T_1 = T_2 = T_3 = 1S$ /. Структура УУ рассматриваемой системы совпадает со структурой изображенной на рис. 2, где отмечены коэффициенты C_{ij} линейных комбинаций координат w_i / x_1, x_2, x_3 /. Аналитическим путем определены коэффициенты C_{ij} линейных комбинаций координат, а именно:

$$C_{11} = \frac{\partial \mathcal{U}(0)}{\partial x_1}; \quad C_{21} = C'_{21}\sigma_1; \quad C'_{21} = \frac{1}{2}\,\frac{\dfrac{\partial^2 \mathcal{U}(0)}{\partial x_1^2}}{\dfrac{\partial \mathcal{U}(0)}{\partial x_1}};$$

$$C_{12} = \frac{\partial \mathcal{U}(0)}{\partial x_2}; \quad C_{22} = C'_{22}\sigma_1; \quad C'_{22} = \frac{1}{2}\,\frac{\dfrac{\partial^2 \mathcal{U}(0)}{\partial x_2^2}}{\dfrac{\partial \mathcal{U}(0)}{\partial x_2}}; \qquad /25/$$

$$C_{13} = \frac{\partial \mathcal{U}(0)}{\partial x_3}; \quad C_{23} = C'_{23}\sigma_1; \quad C'_{23} = \frac{1}{2}\,\frac{\dfrac{\partial^2 \mathcal{U}(0)}{\partial x_3^2}}{\dfrac{\partial \mathcal{U}(0)}{\partial x_3}}.$$

Оптимальный переходный процесс при подаче на вход системы единичной функции показан на рис. 3а. Опыты произведенные на модели, построенной по схеме УУ изображенной на рис.2 и коэффициентами C_{ij}, определенными в первом приближении формулами /25/, дали успешные результаты. Для того чтобы добиться более хорошей формы переходного процесса было произведено некоторое изменение коэффициента C'_{22}. Квазиоптимальный процесс полученый путем аппроксимации при помощи полинома $G^2(x_1, x_2, x_3)$ и реализации на модели эквивалентной сигнум-функции, показан на рис. 3б. Видно, что квазиоптимальный процесс по быстродействию практически не отличается от оптимального процесса. Аналитическим способом определено время оптимального процесса при различных начальных отклонениях по ошибке. При двукратном изменении начального условия от 100% до 200% отклонение от оптимального времени не больше 60%. Система работает удовлетворительно и при десятикратном изменении начальных условий, а именно в диапазоне от 25% до 250% номинального скачкообразного сигнала, и тогда отклонение времени квазиоптимального процесса от оптимального при соответствующих начальных условиях не превосходит 80%. Переходный процесс квазиоптимальной системы при линейном входном сигнале показан на рис.3в. Видно, что квазиоптимальный процесс почти не отличается от оптимального процесса. При входном сигнале в виде параболы второй степени переходный процесс в квазиоптимальной системе также удовлетворительно приближается к оптимальному процессу. Все эти экспериментальные результаты, при различных классах входных сигналов и в широком диапазоне изменения начальных условий, получены при неизменных значениях коэффициентов C_{ij}, определенных указанным выше способом.

В общем случае при высоком порядке неизменной части системы и сложном характере ограничений математическое описание строго оптимальной гиперповерхности переключения может оказаться затруднительным. В этом случае определение параметров УУ нельзя производить аналитическим путем. Здесь существует возможность применять любые известные методы поиска значений коэффициентов линейных комбинаций координат $w_i(x_1,\dots,x_m)$, которые входят в эквивалентные реализующие функции. Например, если имеется модель объекта управления, то реализуется структура УУ на модели и экспериментальным путем определяются неизвестные коэффициенты.

Рассмотрим пример построения УУ системы высокого порядка. Для цели илюстрации приведем результаты исследования приближенно-оптимальной следящей системы управления копировально-фрезерного станка. В предыдущих работах [4,5] определена предельно простая в реализации структура УУ следящей системы, при помощи предложенного автором метода эквивалентных функций, как показано на рис. 4. Неизменная часть этой системы рассматривалась в виде последовательного соединения двух динамических блоков: а/ первый блок $C1$ состоит из второго каскада электронного усилителя и электромашинного усилителя /апериодическое звено первого порядка/; б/ второй блок $C2$ – исполнительный двигатель и редуктор /звено второго порядка/.

В качестве выходной координаты блока $C2$ рассматривается ошибка следящей системы δ, а для блока $C1$ - ток I якорной цепи электромашинного усилителя - двигателя. В УУ поступают сигналы этих координат и их производных в виде линейных комбинаций, а именно:

$$w_2(\delta,\dot{\delta}) = k_\delta \delta + k_{\dot{\delta}} \dot{\delta} = k_{x2} x_2 + k_{\dot{x}2} \dot{x}_2 ;$$

$$w_1(I,\dot{I}) = k_I I + k_{\dot{I}} \dot{I} = k_{x1} x_1 + k_{\dot{x}1} \dot{x}_1 . \qquad /26/$$

В качестве аппроксимирующей функции принят полином второго порядка

$$G^2(\delta,\dot{\delta},I,\dot{I}) = k_\delta \delta + k_{\dot{\delta}} \dot{\delta} - \sigma \left[k_\delta k_I \delta I + k_\delta k_{\dot{I}} \delta \dot{I} + k_{\dot{\delta}} k_I \dot{\delta} I + k_{\dot{\delta}} k_{\dot{I}} \dot{\delta} \dot{I} \right], \qquad /27/$$

где $\sigma = \mathrm{sign}\left[k_\delta \delta + k_{\dot{\delta}} \dot{\delta} \right]$.

Осцилограмма переходного процесса регулируемой величины /перемещение фрезы/ $S = f(t)$ в линейной системе приведена на рис.5а. Переходный процесс исследован при скачкообразном входном сигнале, соответствующем мгновенному максимальному рассогласованию между копировальным прибором и фрезой. После подбора значений неизвестных коэффициентов в линейных комбинациях координат $w_1(I,\dot{I})$ и $w_2(\delta,\dot{\delta})$ получается минимальное время переходного процесса t_n = 0,13 сек. Осцилограмма процесса при максимальном рассогласовании дана на рис. 5б. Исследования на модели показали, что уменьшение амплитуды входного сигнала на 40% от максимального сигнала почти не влияет на время переходного процесса. Дальнейшее уменьшение входного сигнала, соответствующего рассогласованию фрезы, приводит к уменьшению времени переходного процесса по сравнению с временем при максимальном рассогласовании фрезы. Таким образом быстродействие системы значительно улучшается /примерно в 7 раз/ по сравнению с переходным процессом в линейной системе, и полностью устраняется перерегулирование выходной величины.

5. ЗАКЛЮЧЕНИЕ

I. В работе показана возможность решения задачи синтеза квазиоптимальной системы методом эквивалентных функций управления. Предложенным методом определяются эквивалентные по знаку управляющие функции в классе наиболее удобных для технической реализации функций.

2. При помощи полученных в работе теорем показано, что знак аппроксимирующей функции совпадает со знаком выбранной для технического построения функции, которая синтезируется посредством линейных функций и сигнум-функций фазовых координат системы.

3. Если процесс аппроксимации оптимальной функции производится при помощи эквивалентных функций, то реализация полученного аппроксимирующего выражения не является необходимой. При аппроксимации согласно этому методу достаточно реализовать эквивалентную функцию, определенную в классе выбранных для технической реализации функций.

4. На основе предложенной методики можно определить структуру квазиоптимального управляющего устройства системы, которая обусловлена выбором вида эквивалентных аппроксимирующих функций. Требуемая точность приближения к строго оптимальной системе предопределяет число элементов структуры.

5. В случае, когда известно математическое описание оптимальной функции управления, полученные выражения эквивалентных функций дают возможность определить параметры квазиоптимального управляющего устройства аналитическим путем. Рассматриваемый метод может быть применен и в случае, когда неизвестно математическое описание оптимальной функции управления.

А.

Л и т е р а т у р а:

I. Фельдбаум А.А., Вычислительные устройства в автоматических системах, Госуд.изд.физико-матем.литературы, М. 1959.

2. Лернер А.Я., Улучшение динамических свойств автоматических компенсаторов, гл.I, "Автоматика и телемеханика",1952,№2, гл.II,1952, № 4.

3. Pike E.W., Silverberg T.R., Designing Mechanical Computers, Machine Design, 1952,v.24, N 7, p.p.131 - 137(part I);N 8,p.p.159--163 (part II).

4. Гунчев Л.Ат.,Димитров Л.Б.,Электронное моделирование следящей системы с характеристиками,близкими к оптимальным, Годишник МЕИ,София, т.XII, кн.II,1962.

5. Гунчев Л.Ат., О построении приближенно-оптимальных систем управления высокого порядка при помощи простых эквивалентных функций управления,Известия ВУЗ,"Электромеханика",№ I, 1967.

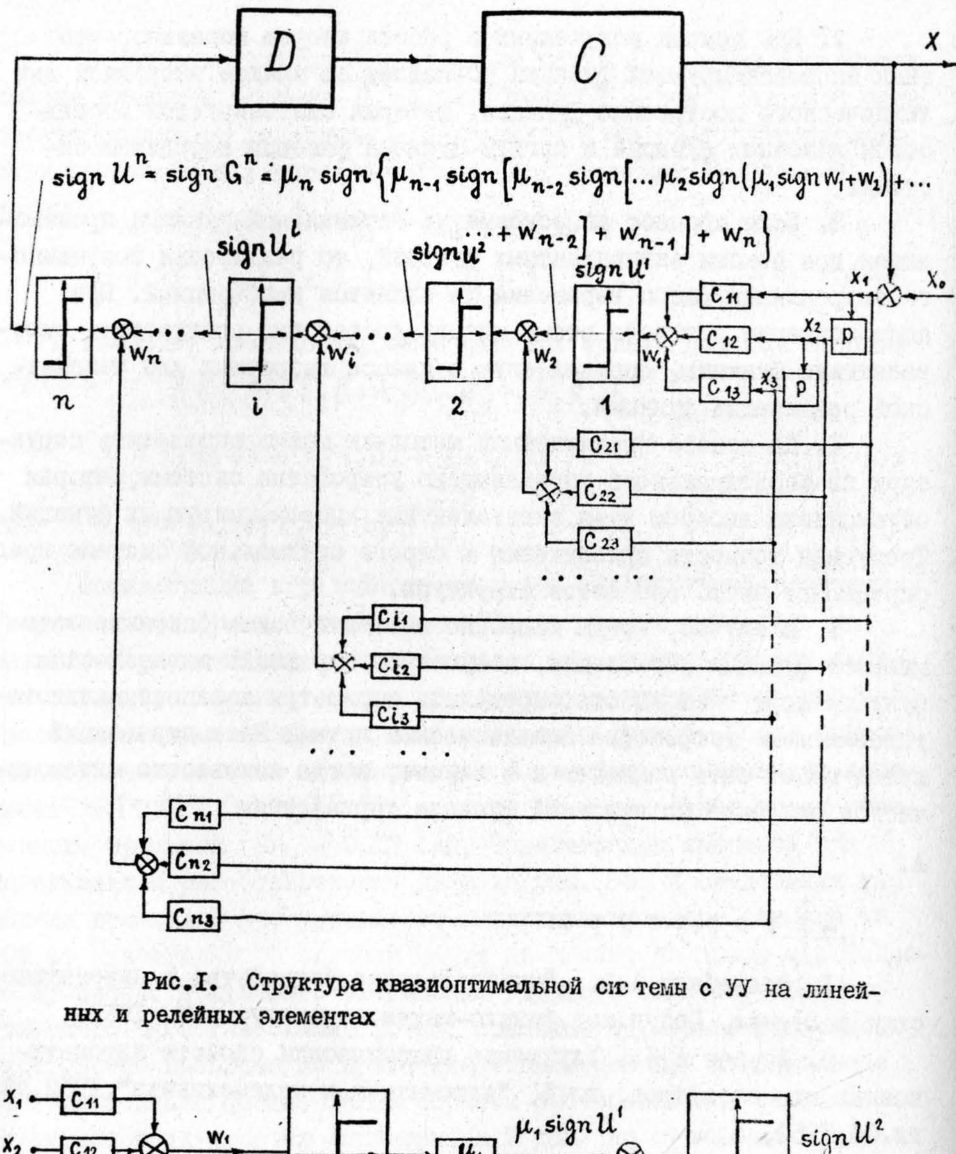


Рис.I. Структура квазиоптимальной системы с УУ на линейных и релейных элементах

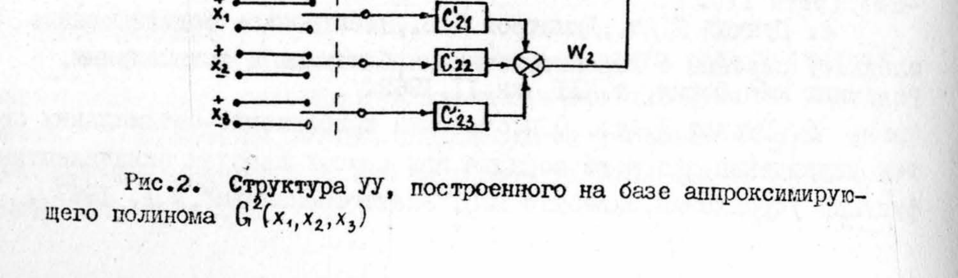


Рис.2. Структура УУ, построенного на базе аппроксимирующего полинома $G^2(x_1, x_2, x_3)$

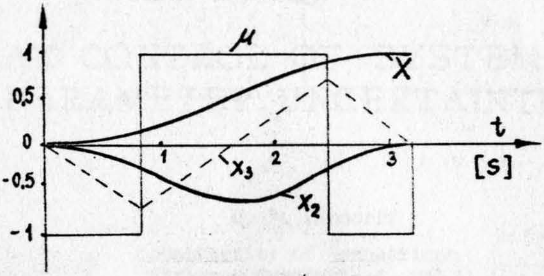


Рис.3. а/

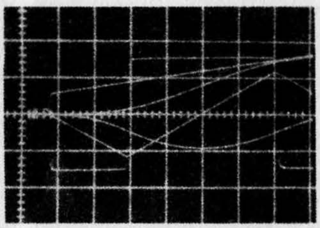


Рис. 3. б/

Рис. 3. в/

Рис.3. Переходные процесы: а/ оптимальной системы; б/квазиоптимальной системы при единичном входном сигнале; в/квазиоптимальной системы при линейном входном сигнале; X_0 – входной сигнал, X – выходной сигнал, x_2 и x_3 – первая и вторая производные ошибки $x_1 = X_0 - X$; μ – входное воздействие объекта

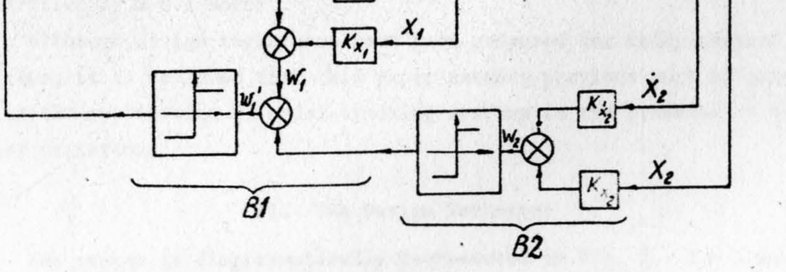


Рис.4. Простейшая структура квазиоптимальной системы высокого порядка

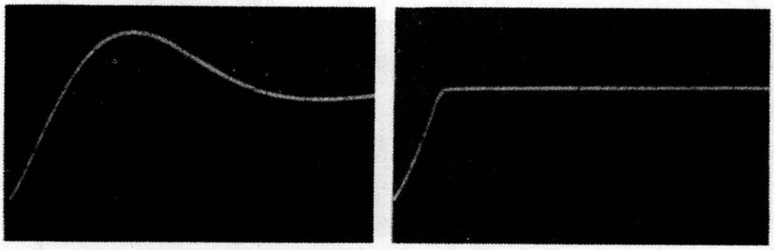


Рис. 5. а/ Рис. 5. б/

Рис.5. Переходный процесс перемещения фрезы $S = f(t)$ при единичном входном сигнале: а/ в линейной системе; б/ в квазиоптимальной системе

RELAY CONTROL OF SYSTEMS WITH PARAMETER UNCERTAINTIES

D. P. Lindorff

University of Connecticut
Storrs, Connecticut, USA

I. Introduction

The application of Liapunov theory to the design of model-tracking systems, taking into account bounds on parameter uncertainties, has been treated by several authors[1,2,3] under the assumption that the state variables of the plant have been defined as phase variables, i.e. for an n^{th} order plant, the state has been defined by the output and its n-1 derivatives. However, in some applications it is impractical to derive the required number of derivatives of the output variable because of the noise which enters the measurements. In fact, in some cases it is impossible to derive all of the state variables from a given output. This situation arises if one output is unobservable with respect to another. For these reasons it may be necessary, or at least desirable, to develop a design procedure which allows freedom in choosing the state variables.

It has been suggested[4] that, by the use of a semidefinite Liapunov function, the rigid requirement on the form of the state variables can be relaxed. The purpose of this paper is to develop this concept, with application to a model-tracking system using a relay controller. Although the theory developed in this paper is applicable to a wider class of problems, the specific plant to be controlled is pictured in Fig. 1, and consists of an inverted pendulum mounted on a cart whose position is controlled by a d-c motor.

Although design techniques have been proposed for relay control systems, it is believed that this paper extends previous work by generalizing the applicaticn of model-tracking systems in the presence of parameter variations.

II. The Design Technique

The system is diagrammatically represented in Fig. 2. The plant is defined according to the linear differential equation

$$\dot{\underline{x}} + A\,\underline{x} + \underline{b}\,u \qquad (2.1)$$

where $\underline{x}$ is a n-dimensional state vector with $\dot{\underline{x}} = dx/dt$, and u is a scalar forcing function. The elements of A and $\underline{b}$ can be slowly varying within known bounds. It is assumed that (2.1) represents a completely controllable system.

The relay is assumed to be ideal, and hence is characterized by the sign function

$$u = \mu L \operatorname{sgn} \gamma(\underline{e}) \qquad (2.2)$$

where $\underline{e} = \underline{s} - \underline{x}$, $\underline{s}$ being the response of the model. L and γ are to be determined. μ equals +1 or -1 depending upon the specific problem involved. $\gamma(\underline{e})$, which represents a linear function of the error-state variables, is defined by

$$\gamma(\underline{e}) = \underline{k}'\,\underline{e} \qquad (2.3)$$

where $\underline{k}' = [k_1 k_2 \ldots k_n]$ denotes the transpose of $\underline{k}$. The model is represented by the equation

$$\dot{\underline{s}} = \mathfrak{a}\,\underline{s} + \underline{\beta}\, m(r,\underline{s}) \qquad (2.4)$$

with $\mathfrak{a}$ and $\underline{\beta}$ defined as nominal values of A and $\underline{b}$. m is a signal which produces the desired response, $\underline{s}$, of the model. r is the input to the model.

The deviations in A and $\underline{b}$ are denoted by

$$\Delta = A - \mathfrak{a}, \qquad (2.5)$$

$$\underline{\delta} = \underline{b} - \underline{\beta}. \qquad (2.6)$$

It is assumed that elements of Δ and $\underline{\delta}$ can be slowly varying within known bounds. By permitting m to include feedback signals of the model, it is possible to cause an unstable plant to track a stable model using relay control.

The design problem consists of determining values for L and $\underline{k}$ so that the plant will track the model with zero error when $\Delta = 0$, $\underline{\delta} = \underline{0}$, and with a bounded error when the elements of Δ and $\underline{\delta}$ are within the stated bounds. Hence, it may be said that the plant is to be controlled with limited force so that, in the presence of bounded parameter variations, the error vector will remain bounded.

The design approach is to define a positive semidefinite quadratic form

$$V = 1/2\ \gamma^2(\underline{e}). \tag{2.7}$$

Now if the control law can be found which causes the time derivative

$$\dot{V} = \gamma\ \dot{\gamma} \tag{2.8}$$

to be negative semidefinite so that $V > 0$, $\dot{V} < 0$ except on the hyperplane (switching plane) defined by $\gamma = 0$, then if at some time $\underline{e}$ lies on the switching plane, its motion will be confined to the switching plane thereafter. The conditions must then be derived which assure stable (bounded) motion on the switching plane.

Since stability of motion off the switching plane is not assured by use of a semidefinite V function, it is pertinent to question why V cannot be made a positive definite function. If V were taken to be positive definite as has been previously assumed[1-3], then in the presence of parameter uncertainties $\dot{V}$ could generally be assured of being negative definite only if the vector $\underline{b}$ in (2.1) were to contain but one non-zero element as would be the case if (2.1) were in phase-variable form. In this paper such a restriction on $\underline{b}$ is not allowed.

In the following discussion, it is to be assumed that V is semidefinite as given by (2.7), and that the error vector is initially on the switching plane. Although no quantitative statement can be made as to the stability of motion off the switching plane, it is reasonable to assume that the system will be stable for small perturbations about the switching plane. A quantitative treatment of the convergence problem for large deviations about the switching plane lies outside the scope of this paper.

In formulating the control law, it is necessary to obtain a differential equation for the error vector. Thus, using (2.1), (2.4), and (2.5), it follows that

$$\dot{\underline{e}} = A\ \underline{e} - \Delta\ \underline{s} + \underline{\beta}\ m - \underline{b}\ u, \tag{2.9}$$

or in scalar form

$$\dot{e}_i = f_i(\underline{e}) + g_i(\underline{s},m) - b_i u,\ i = 1,\ \ldots,\ n. \tag{2.10}$$

To insure that chatter motion will take place, i.e. that motion will be constrained to lie on the switching plane, it is sufficient, as mentioned above, to require that $V > 0$, $\dot{V} < 0$, when $\gamma \neq 0$. This condition will be satisfied if, for $\gamma \neq 0$,

$$\text{sgn}\,\dot{\gamma} = -\,\text{sgn}\,\gamma. \tag{2.11}$$

Eq. (2.11) can be satisfied if the magnitude of u is large enough to control the sign of $\dot{\gamma}$. Consider the explicit form for γ and $\dot{\gamma}$. Thus, upon expanding (2.3),

$$\gamma = k_1 e_1 + k_2 e_2 + \ldots + k_n e_n, \tag{2.12}$$

whereupon

$$\dot{\gamma} = k_1 \dot{e}_1 + k_2 \dot{e}_2 + \ldots + k_n \dot{e}_n. \tag{2.13}$$

From (2.10) and (2.13) it is seen that $\dot{\gamma}$ can be expressed in the form

$$\dot{\gamma} = \phi(\underline{e}, \underline{s}, m) - \sum_{i=1}^{n} k_i b_i u. \tag{2.14}$$

Now if u is made to satisfy the following two conditions,

$$L = |u| \geq \left| \frac{\phi(\underline{e}, \underline{s}, m)}{\sum_{i=1}^{n} k_i b_i} \right| \tag{2.15}$$

$$\text{sgn}\left(\sum_{i=1}^{n} k_i b_i u\right) = \text{sgn}\,\gamma, \tag{2.16}$$

then it follows that (2.11) will be satisfied.

If (2.15) and (2.16) are satisfied and if at some time $\gamma(\underline{e}) = 0$, then the motion thereafter will remain on the switching plane. It is now necessary to establish the conditions for such motion to be stable. Whereas (2.9) characterizes motion in Euclidian n space (E^n), motion on the switching plane can be described in E^{n-1} space. It follows based on the conditions $\gamma(\underline{e}) = 0$ and $\dot{\gamma}(e) = 0$, that (2.9) can be reduced dimensionally so that, if $\tilde{\underline{e}}$ represents a vector composed of n-1 components of $\underline{e}$, then (2.9) can be expressed as a linear differential equation of the form

$$\dot{\tilde{\underline{e}}} = \underline{h}_1(\underline{k}, A, \tilde{\underline{e}}) + \underline{h}_2(\underline{k}, \underline{b}, \Delta\,\underline{s}, \delta\,\underline{m}). \tag{2.17}$$

It should be noted that motion on the switching plane will not generally call for $u \equiv 0$. Since at $\gamma = 0$ the function sgn γ is not defined, it can be said that, for the idealized relay, $u = L\,\text{sgn}\,\gamma$ will in fact satisfy the requirement for motion on the switching plane. For a discussion of the chatter problem, see Schaefer[5].

Since $\underline{s}$ and m are signals generated by the model, it is found that $\underline{h_2}$ is a bounded forcing function whose magnitude depends directly on parameter deviations, Δ and $\underline{\delta}$. Furthermore, $\underline{h_1} = 0$ if $\Delta = [0]$ and $\underline{\delta} = \underline{0}$. If elements of $\underline{k}$ are chosen so that (2.17) is asymptotically stable with $\underline{h_2} = \underline{0}$, then $\underline{\tilde{e}}$, as well as $\underline{e}$, will be bounded. Finally, with a knowledge of the bounds on $\underline{e}$, $\underline{s}$, and m, the value of L which is required to satisfy (2.15) can be determined. This completes the formal design procedure. The details of the method will be elaborated upon in a specific design application.

It is noted that, in constrast to previous work[2,3], by this design procedure the plant will in general track the model with zero error only if the parameters of the model and the plant are equal.

In effect, a relay controller has been designed to cause the plant to have a desired nominal response characteristic, i.e. the model response. The effect of parameter uncertainty has been related to the response of a linear differential equation (2.17).

III. Design Application

The design technique represented in the preceding section will now be applied to a model-tracking system using a relay controller. The plant consists of a cart supporting an inverted pendulum. As may be seen in Fig. 1, the cart, which is mounted on a track, is connected through a pulley and gearing to a d-c motor. The purpose is to cause the plant to track a linearized model of the nominal plant. This particular plant is of interest because the state variables are most naturally defined other than as phase variables. Thus, referring to Fig. 3, if velocity of the cart is to be controlled, then the state variables logically become

$$x_1 = v \text{ m/sec}$$
$$x_2 = \theta \text{ rad}$$
$$x_3 = \dot{\theta} \text{ rad/sec}$$

and

$$u = e_a \text{ volts.}$$

Relative to (2.1) and based on the linearizing assumption ($\sin\theta \simeq \theta$, $\cos\theta \simeq 1$), the equations of the plant can be written in the form

$$\dot{x}_1 = a_{11}\, x_1 + a_{12}\, x_2 + b_1\, u$$

$$\dot{x}_2 = x_3 \tag{3.1}$$

$$\dot{x}_3 = a_{31}\, x_1 + a_{32}\, x_2 + b_3\, u.$$

In order to study the effect of parameter variations, the distance ℓ from the pivot point to the center of gravity of the pendulum was made to be adjustable by repositioning m_1 on the rod. The nominal values of the parameters, as well as their dependencies on ℓ, are shown in Table 1.

	nominal	ℓ dependence
a_{11}	-800	none
a_{12}	-3	none
a_{31}	800	$1/\ell$
a_{32}	13	$1/\ell$
b_1	15	none
b_3	-15	$1/\ell$

Table I

According to the design procedure, the nominal values in Table I should be assigned to the corresponding terms in (2.4) pertaining to the model. Thus, (2.4) becomes

$$\dot{s}_1 = \alpha_{11}\, s_1 + \alpha_{12}\, s_2 + \beta_1\, m$$

$$\dot{s}_2 = s_3 \tag{3.2}$$

$$\dot{s}_3 = \alpha_{31}\, s_1 + \alpha_{32}\, s_2 + \beta_3\, m$$

where each α_{ij} and β_i has the nominal value of the corresponding a_{ij} and b_i in Table I.

It is noted that the plant is open-loop unstable. In this problem linear feedback was used so that the signal m(t) stablized the model and produced a desired response ($\underline{s}$) to a reference input (r).

From (2.10), (3.1), (3.2), the error equations become

$$\dot{e}_1 = a_{11}\, e_1 + a_{12}\, e_2 + g_1 - b_1\, u$$

$$\dot{e}_2 = e_3 \tag{3.3}$$

$$\dot{e}_3 = a_{31}\, e_1 + a_{32}\, e_2 + g_3 - b_3\, u$$

where

$$g_i = -\Delta_{i1} s_1 - \Delta_{i2} s_2 + \beta_i m, i = 1,3$$

and

$$\Delta_{ij} = a_{ij} - \alpha_{ij}, j = 1,2.$$

Eq. (2.7) in turn becomes

$$V = 1/2\,(\gamma^2)$$

$$= 1/2\,(e_1 + k_2 e_2 + k_3 e_3)^2 \quad (3.4)$$

wherein it is arbitrarily assumed that $k_1 = 1$ and $k_2 > 0$, $k_3 > 0$. It follows, upon substituting (3.3) into the expression for $\dot{\gamma}$, that

$$\dot{\gamma} = \dot{e}_1 + k_2 \dot{e}_2 + k_3 \dot{e}_3$$

$$= a_{11} e_1 + a_{12} e_2 + g_1 - b_1 u + k_2 e_3 + k_3 (a_{31} e_1 + a_{32} e_2 + g_3 - b_3 u), \quad (3.5)$$

whereupon (2.15) becomes

$$L = |u| \geq \left| \frac{\Psi_1(\underline{e}) + (\beta_1 + k_3 \beta_3) m - \Psi_2(\underline{s})}{b_1 + k_3 b_3} \right| \quad (3.6)$$

with

$$\Psi_1 = (a_{11} + k_3 a_{31})\, e_1 + (a_{12} + k_3 a_{32})\, e_2 + k_2 e_3$$

and

$$\Psi_2 = (\Delta_{11} + \Delta_{31})\, s_1 + (\Delta_{12} + \Delta_{32})\, s_2.$$

Eq. (2.16) in turn becomes

$$\operatorname{sgn}(b_1 + k_3 b_3)\, u = \operatorname{sgn} \gamma. \quad (3.7)$$

The equations (3.6) and (3.7) form the basis for a control law. However, before discussing these equations further, it will be necessary to obtain the reduced equation (2.17) which describes motion on the switching plane, assuming that (3.6) and (3.7) will be satisfied.

Using the equations $\gamma(\underline{e}) = 0$, $\dot{\gamma}(\underline{e}) = 0$ which are valid on the switching plane, it is found that (2.17) becomes

$$\left(1 + \frac{b_1}{b_3} r_1\right) \dot{e}_1 = \left[a_{11} + \frac{b_1}{b_3}(r_1 r_2 - a_{31})\right] e_1 + g_1 - \frac{b_1}{b_3} g_3 + \left[a_{12} + \frac{b_1}{b_3}(r_2^2 - a_{32})\right] e_2 \quad (3.8a)$$

$$\dot{e}_2 = - r_1 e_1 - r_2 e_2 \tag{3.8b}$$

where

$$r_1 = \frac{1}{k_3}, \quad r_2 = \frac{k_2}{k_3},$$

and

$$g_i = -\Delta_{i1} s_1 - \Delta_{i2} s_2 + \beta_i m, \quad i = 1,3.$$

It is required at this point that k_2 and k_3 be chosen so that (3.8) will be stable. If (3.8) is rewritten in the homogeneous form

$$\dot{e}_1 = \phi_{11} e_1 + \phi_{12} e_2 \tag{3.9}$$

$$\dot{e}_2 = \phi_{21} e_1 + \phi_{22} e_2,$$

then the characteristic equation is seen to be

$$\lambda^2 - (\phi_{11} + \phi_{22}) \lambda + (\phi_{11} \phi_{22} - \phi_{21} \phi_{12}) = 0. \tag{3.10}$$

Stability requires that $\phi_{11} + \phi_{22} < 0$, $\phi_{11} \phi_{22} > \phi_{21} \phi_{12}$, which can be shown to be equivalent to requiring for $k_1 = 1$ that

$$k_3 > 1, \quad k_2 > \sqrt{5}\, k_3. \tag{3.11}$$

Returning to (3.6) and (3.7), it is required that conditions be established which guarantee $\dot{V} < 0$ with $\gamma \neq 0$. On the assumption that (3.6) is satisfied, it follows that u will control the sign of γ. For the nominal values of b_1, b_3 in Table I, and the requirement $k_3 > 1$, it is seen that (3.7) is satisfied if

$$\operatorname{sgn} u = - \operatorname{sgn} \gamma. \tag{3.12}$$

In order that (3.12) can be used for the full range of variations of b_1 and b_3, it is necessary that k_3 be chosen so that $(b_1 + k_3 b_3)$ will always be of one sign. In this case, k_3 can always be made large enough to meet this requirement.

Attention is now directed to (3.6). In order to determine a sufficiently large value of L, it is necessary to know at least the bounds on $\Psi_1(\underline{e})$, $\Psi_2(\underline{s})$ and m. It is informative first to assume that the plant parameters are at their nominal values. In this case it can be seen that $\Psi_2(\underline{s}) = 0$, $\Psi_1(\underline{e}) = 0$. Hence (3.6) reduces to

$$L = |u| > |m|. \tag{3.13}$$

The statement that $\Psi_1(\underline{e}) = 0$ in the absence of parameter deviations follows from the fact that the forcing function in (3.8) then reduces to zero. Thus, for the nominal plant, if $\underline{e} = \underline{0}$ initially, then the tracking error is zero if (3.12) and (3.13) are satisfied. Since the magnitude of m depends upon the size of the input (r), as well as the response time of the model (which is reflected in m(t)), the sizing of L requires a knowledge of the specifications placed on the system's performance.

At this point in the design, the effect of parameter deviations should be considered. More specifically, the value of L in (3.6) must be made large enough to override the terms $\Psi_1(\underline{e})$ and $\Psi_2(\underline{\dot{s}})$. In Section IV results of simulation studies are presented which bear on this aspect of the problem. It is important to note that it is possible to use simulation methods advantageously in the design since the response of the plant in tracking the model is insensitive to L, provided L is large enough. Thus, by making L excessively large in the simulation, the behavior of $\underline{e}(t)$ can be determined. This information can in turn be used in finding the minimum acceptable value of L.

IV. Results by Simulation and Experiment

The design which has been discussed in the preceding section was completed with the aid of analog-computer simulation and tests on the physical system, in response to a square-wave input to the model.

The results of the simulation studies are presented in Figs. 4, 5, and 6. In these data the value of L was made excessively large, so as to guarantee that (2.15) would be satisfied. In Fig. 4 the responses of the state variables of the model and the nominal plant are shown. The tracking error is seen to be essentially zero as predicted. In Figs. 5 and 6 the value of ℓ in the plant was increased by 40% of its nominal value, and the test signal used in Fig. 4 was again applied. The data in Fig. 5 shows the responses of the same state variables as in Fig. 4, whereas the data in Fig. 6 shows the responses of the error state variables (e_1, e_2, e_3). It is interesting to note that the main effect of this particular parameter variation is to alter the initial response of the plant state variables. The spikes which appear in the error responses were used to establish bounds on the error terms appearing in (3.6).

In Fig. 7 response of the physical system in terms of x_1 (velocity of cart) and x_2 (angle of pendulum) is compared with corresponding terms,

s_1 and s_2, of the model, using a value of L determined from the simulation. Since parameter values of the plant could not be determined exactly, some error was to be expected. The high frequency components present in the response of the plant are attributed to imperfect switching and coulomb friction.

V. Conclusions

The design of a relay controller has been formulated in which a linear plant subject to parameter uncertainties is caused to track a model with bounded error. The method can be applied to multi-output systems, or more generally to systems whose state variables are not necessarily available as phase variables. The method is demonstrated by application to a specific design problem involving the stabilization of a cart-supported inverted pendulum.

From the theoretical standpoint, further study should be made of the tracking error which is incurred due to the use of an imperfect relay. In addition, because a semidefinite Liapunov function is used in the design, it has not been possible to determine bounds on the motion which result if the error states are not initially on the switching plane. This problem merits further consideration.

References

1. L. P. Grayson, "Design via Liapunov's Second Method," Preprints of Tech. Papers, 4th J.A.C.C., prepared by the A.I.Ch.E., New York, N.Y., pp. 589-595, 1963.

2. R. V. Monopoli, "Control of Linear Plants with Zeros and Slowly Varying Parameters," I.E.E.E. Trans. AC-12, N. 1, pp. 80-83, Feb., 1967.

3. D. P. Lindorff, "Control of Nonlinear Multivariable Systems," I.E.E.E. Trans. AC-12, No. 11, pp. 506-515, Oct., 1967.

4. T. M. Taylor, "Determination of A Realistic Error Bound for a Class of Imperfect Nonlinear Controllers," Preprints of Tech. Papers, 9th J.A.C.C., prepared by I.E.E.E., New York, N. Y., pp. 522-538, 1968.

5. J. F. Schaefer, "Chatter Motion of Bang-Bang Controllers", Proc. Second Annual Princeton Conf. on Information Science and Systems, pp. 228-231, 1968.

This work has been sponsored by the National Aeronautics and Space Administration Research Grant NsG-309.

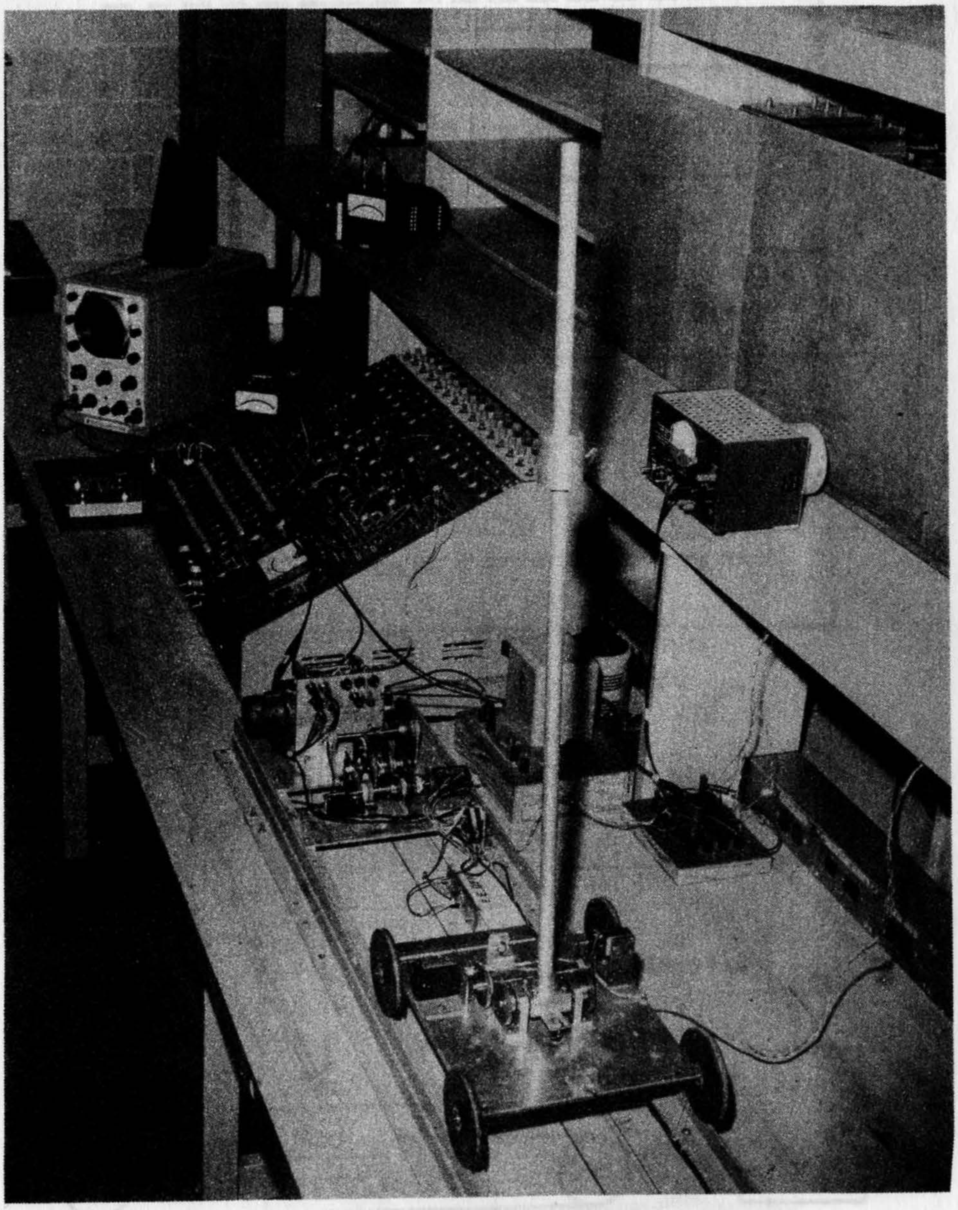


Fig. 1 Cart-driven inverted pendulum.

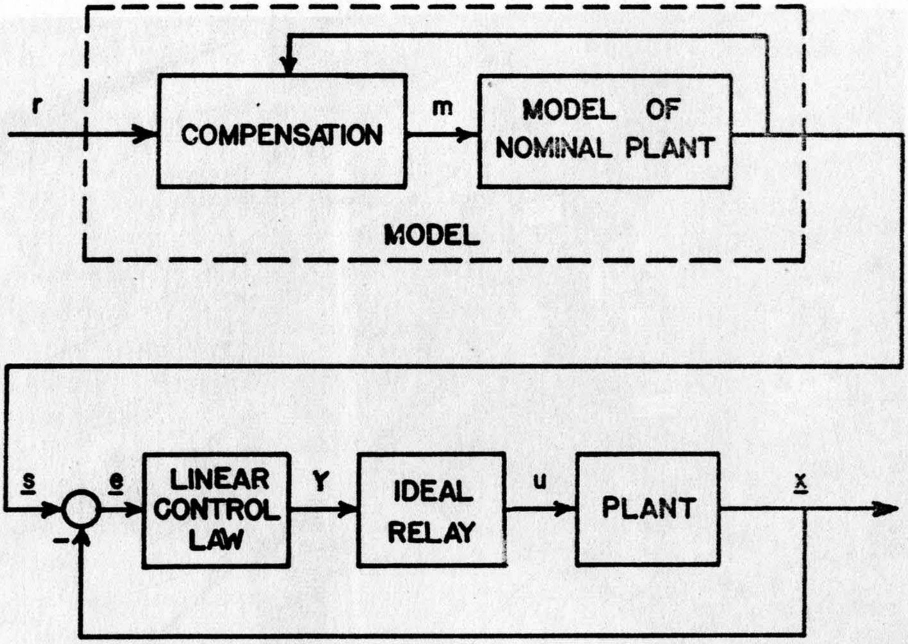


Fig. 2 Block diagram of relay control system.

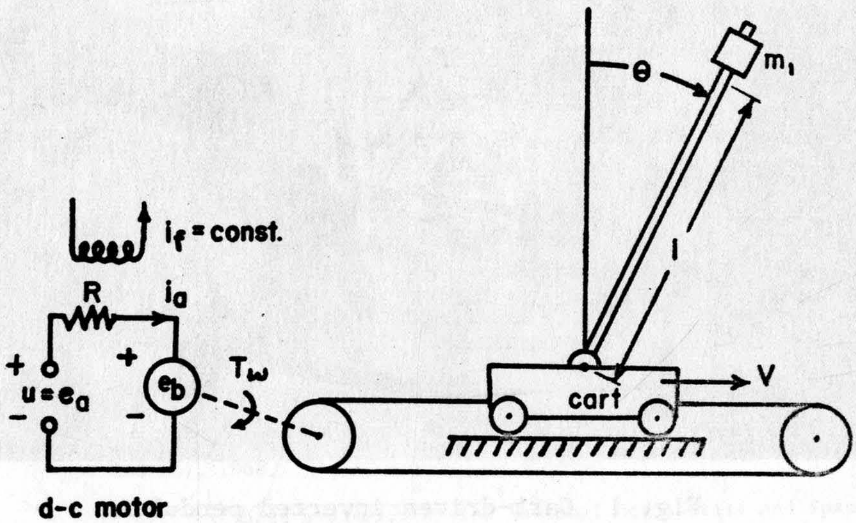


Fig. 3 Schematic representation of unstable plant.

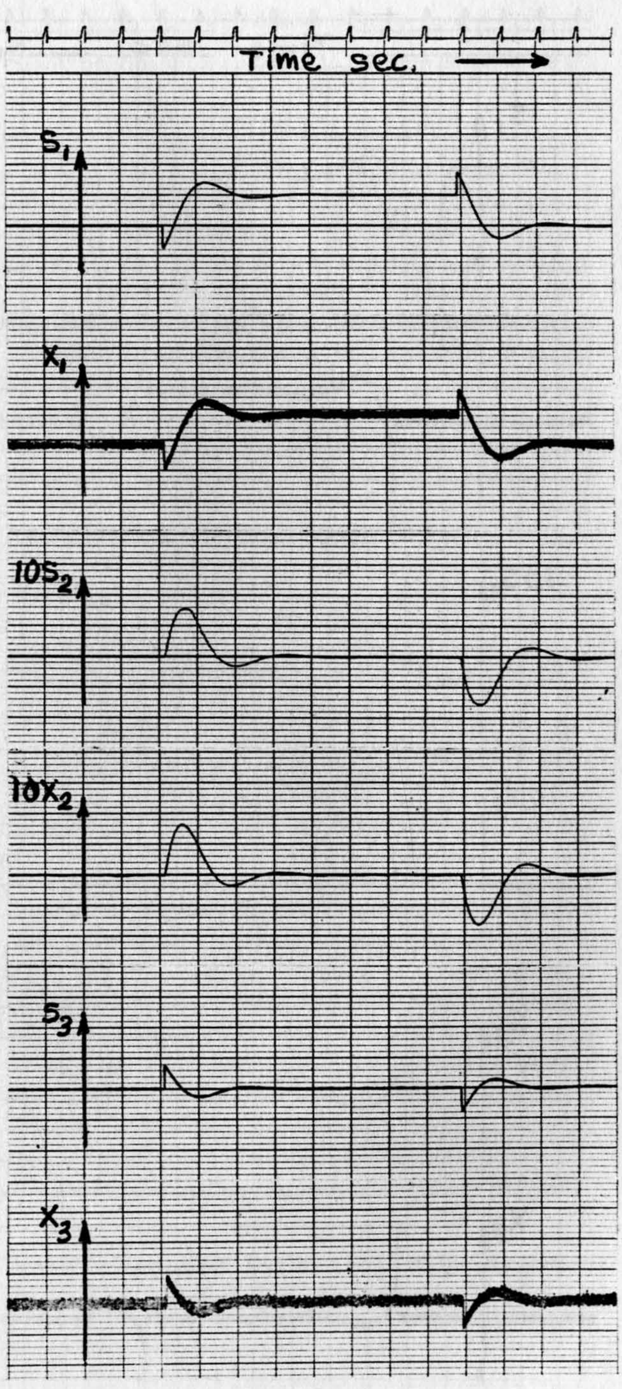


Fig. 4 Response by analog simulation of nominal plant.

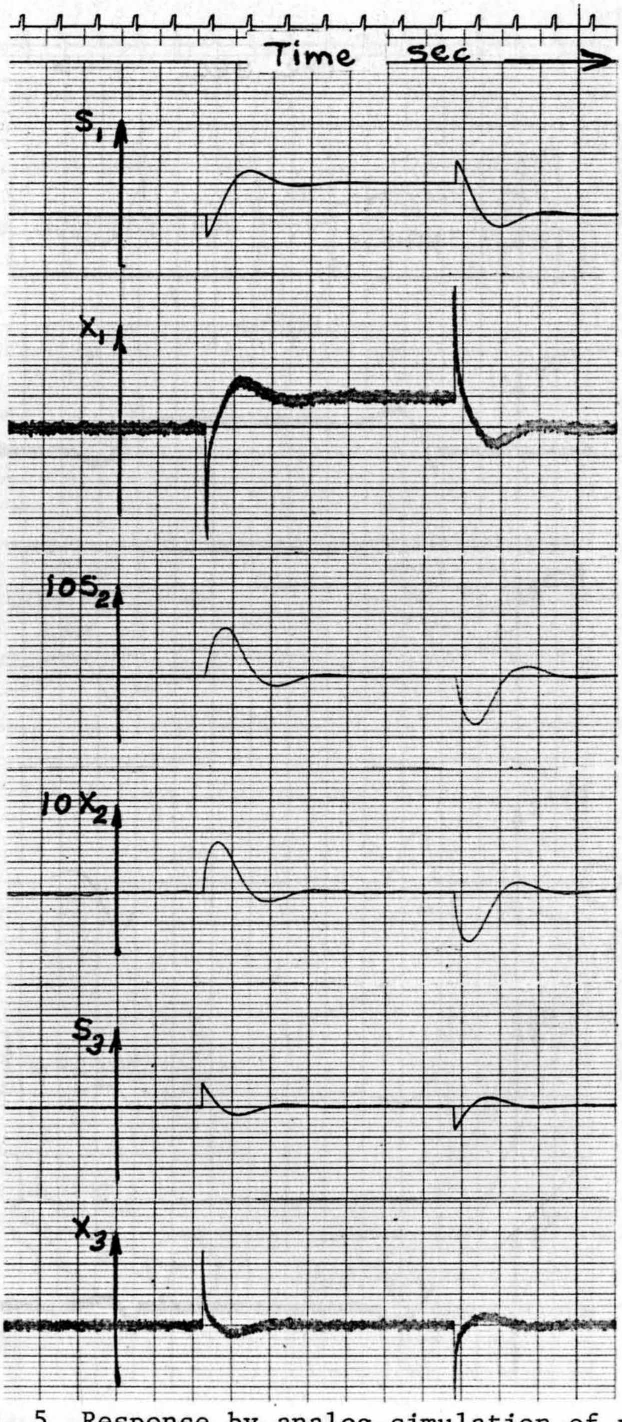


Fig. 5 Response by analog simulation of plant with parameter deviation.

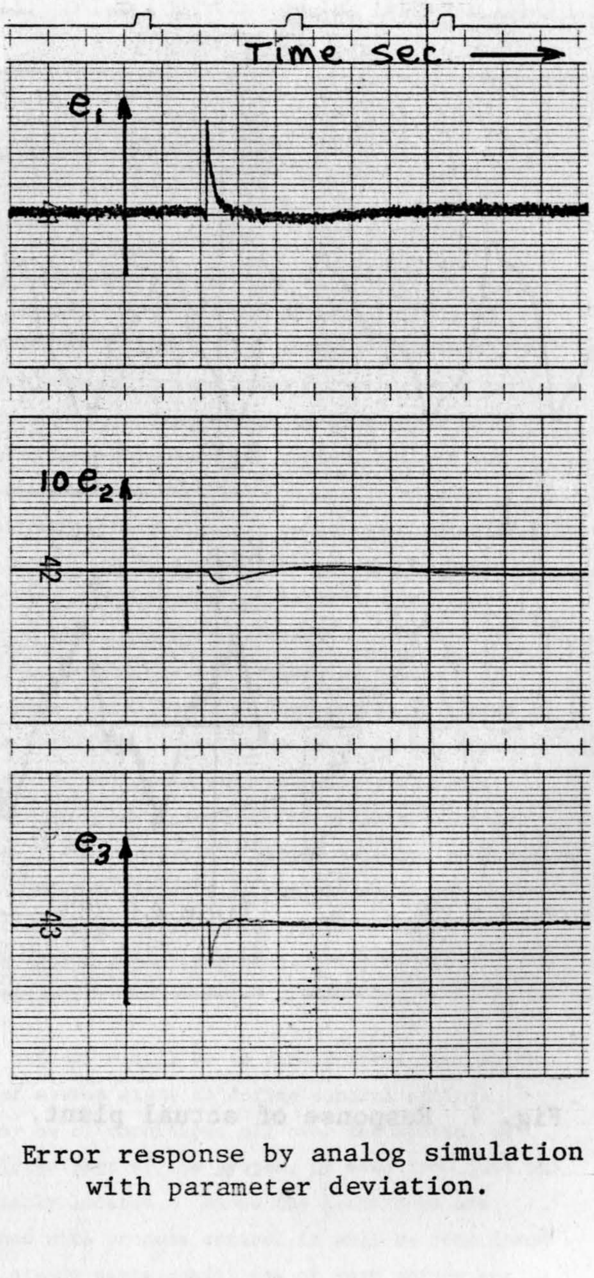


Fig. 6 Error response by analog simulation of plant with parameter deviation.

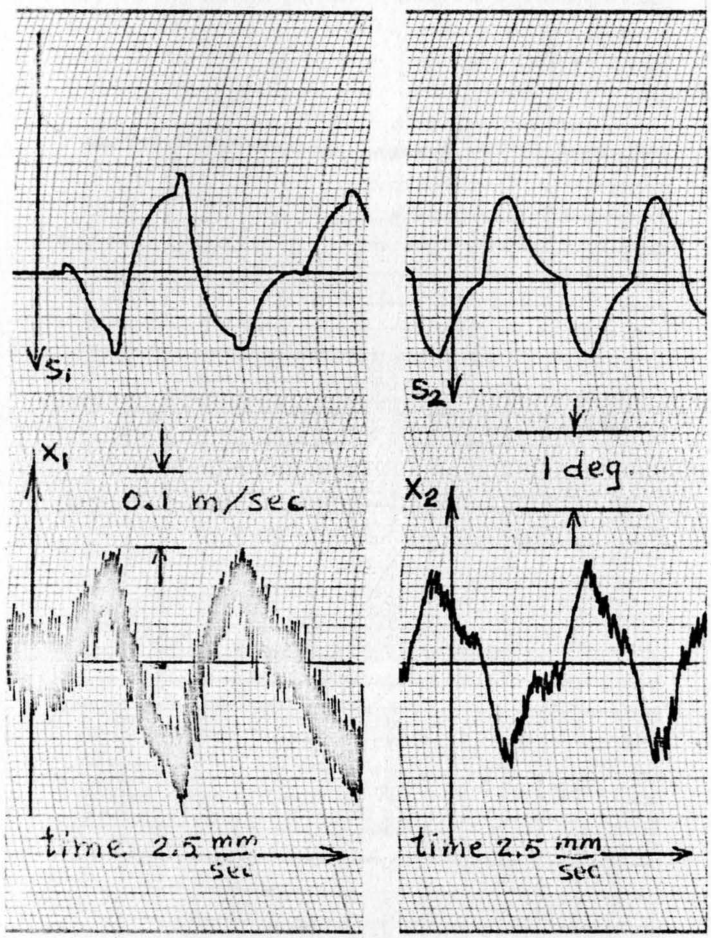


Fig. 7 Response of actual plant.

DESIGN OF LIMITED-INSTRUMENTATION CONTROL SYSTEMS FOR DISTRIBUTED PROCESSES

Dr. M.J. McCann,
Imperial Chemical Industries Ltd.,
Wilmslow, Cheshire, England.

1. Introduction

There are many examples in the literature of modern control theory concerning the optimal control of distributed parameter systems. The design techniques however may often result in very large or intractable computational problems. The purpose of this paper is to demonstrate methods for designing control systems for distributed processes with restrictions on the amounts of instrumentation. The design can be achieved using available modern computational equipment and techniques. A basic concept is that a dynamic system can be controlled best by using information about its current state without dynamic compensations that are influenced by previous values of state. In certain cases these optimal control laws can be developed analytically as functions of system state. In the case of the linear system with integral quadratic performance criteria, the control law is found to be a linear function of system state. Furthermore the change to other performance criteria makes little difference to system design or behaviour as long as the performance criteria makes deviations from design conditions more costly as the deviations increase.

The objective will be to find control laws which use a measure of system state to define control action. This measure may be of conditions all over the system, or from the restricted part of the system, or even from just one instrument suitably located. Since the techniques are largely concerned with process control it will be considered essential that steady state conditions of zero errors are necessarily met.

2. Prototype linear distributed process

To develop the ideas, the process to be considered for control will be somewhat like a single path, shell-and-tube heat exchanger with disturbances entering as fluctuations in inlet temperature to the tube side, and control achieved by variation in shell side temperature. The control objective will be to minimise fluctuations in exit temperature from the tube side using only one probe or instrument to measure the state of the system. The example has the essential character of any process which has a disturbance entering with the moving stream and control action which is distributed along the steam either by changes in environment or by changes in stream velocity.

A representative system with significant wall effects described by the following equations:

$$\left.\begin{aligned} \frac{\partial U_2}{\partial t} &= \frac{G_{12}}{C_2}\left(U_1-U_2\right) + \frac{G_{23}}{C_2}\left(U_3-U_2\right) \\ \frac{\partial U_3}{\partial t} &= \frac{-V_3\,\partial U_3}{\partial x} + \frac{D_3\,\partial^2 U_3}{\partial x^2} + \frac{G_{23}}{C_3}\left(U_2-U_3\right) \end{aligned}\right\} \quad 2.1$$

C_2 and C_3 are the thermal capacities of wall and fluid per unit length of system, G_{12} and G_{23} are the thermal conductivities between shell and wall, and between wall and fluid, U_1 is the control variable (the shell temperature), U_2 the wall temperature and U_3 the fluid temperature, with V_3 and D_3 its velocity and axial diffusivity respectively. Figure 1 shows a diagram of the process and the form of the control system used. This control system involves no dynamic compensation network but includes one probe into the system whose location is to be found. The design can be separated into two parts, the steady state design and the dynamic part as follows.

3. Steady State Design

There are two parts to the steady state design problem: - Finding the required distribution of conditions, in this case, temperature distribution over the system and finding the desired characteristics of the control system to maintain these desired conditions.

In the steady state the equations obtained from Eqn 2.1. are:

$$\left.\begin{aligned} -V_3\frac{\partial U_3}{\partial x} + D_3\frac{\partial^2 U_3}{\partial x^2} - \frac{G_{23}}{C_3} U_3 &= \frac{G_{23}}{C_3} U_2 \\ \text{and } \left(G_{12}+G_{23} \right) U_2 - G_{23}U_3 &= G_{12}U_1 \end{aligned}\right\} \quad 2.2$$

So that provided the required $U_3(X)$ have continuous second derivative etc., U_1 can be found by Equation 2.2. This approach can be extended to non-linear systems and to those with endothermic or exothermic reactions in the fluid. The design of desirable temperature distributions for chemical reactors has been well covered (see Aris: 1)*. Nuclear reactors and heat exchangers have their own substantial literature. For the rest of this discussion it will be assumed that the necessary steady state design requirements have been met, but the variables will represent deviations from these conditions, which deviations can be obtained by linearisation if necessary.

The analysis or measurement of the steady state (deviation) system yields information for control design. Figure 2b shows a steady-state response to a disturbance of size D in U_3 (O) as a function of position along the system. Figure 2c shows the corresponding corrective action which, as shown in figure 2d, is just sufficient to suppress the disturbance at X=L (Beyond X=L over correction results). A probe inserted into the fluid measures $U^*_3(X)$ as in figure 2d, and so to get the corrective action, C, the gain in the controller (K in figure 1) has to be made inversely proportional to $U^*_3(X)$ giving the relationship of figure 2d. The dynamic characteristics of the control system design can then be considered by choosing the best probe location subject to the above choice of the steady state gain requirements.

* Names and associated numbers refer to references given at the end of the paper.

4. Dynamic control (Single probe)

A satisfactory control system must be stable but any control loop around a distributed process can be made unstable by increasing loop gain. In figure 3 the gain K for satisfactory steady-state behaviour (from figure 2) is compared to the gain K_S for critical stability, both plotted as functions of probe location. It can be seen that for $x \geq a$ the system cannot meet the steady state requirements and be stable as well. The probe location has to be in the region $x < a$.

Several approaches are available for finding this probe location. All presuppose that a suitable dynamic simulation of the process is available. It is necessary to point out that a suitable dynamic simulation for a distributed process for these purposes is one which demonstrates in adequate detail, not only the end to end or inlet-outlet behaviour of the system but also shows its internal behaviour as well. The first method to be described is based on step response tests. Referring to figure 4a the response at L due to a step disturbance D is observed. The probe is inserted into the system at a possible location and the loop gain adjusted for zero steady-state error. In principle, this desired steady state gain can be calculated but with analogue simulations particularly it is frequently more simple to adjust until the required conditions are met, rather than to try and set it up by calculation in advance. A response U(L,t) as in figure 4b or 4c probably results.

If the response is like that of (b) then the probe is too close to L, causing the disturbance to be effective at L before the control system has been able to respond. If the probe is moved further away from L towards the source of the disturbance, the control system may act too early giving a response at L like that shown in (c). The ideal arrangement has the disturbance and its corrective action arriving at L virtually simultaneously, as in (d). This gives a useful rule for a preliminary estimate of the best probe location. Suppose the mean transit (residence) time from entry to exit in the fluid is T units, (as indicated on fig. 4a), then the probe should be located so that the time for the disturbance to reach the probe, T_m, should be equal to $T-T_c$, where T_c is the time for a step put into the controller to reach the exit of the system, (L).

In some systems the response (b) may show oscillations because the probe is too close to the 'stability boundary' - or alternatively it may not be possible to adjust for zero error without oscillation occurring. As another alternative it is possible that the controller may be so slow to respond that no really satisfactory location can be found for the probe inside the system, (i.e. for $0 \leq x < a$). This requires either a compensating network in which case the design is that for a "feed forward" control system or a probe located even further upstream.

It makes little difference to the results if step disturbances, white noise or sinusoidal disturbances are used for the testing but the step is the simplest to use. The performance can be appraised by different performance criteria, mean square error, mean modulus or error, etc. with little effect on the conclusions. The choice of an error criterion is just as subjective as a decision based on observation of the response.

5. Sensitivity to Parameter Changes

While the design technique shown above can give good dynamic response it has the serious disadvantage that at least part of the system, that between probe and exit, is outside the control loop and therefore not subject to feedback control. Furthermore it works satisfactorily only for linear processes and also these linear processes must have constant coefficients. It is highly unlikely that these conditions will be met in practical examples, therefore some means must be provided for overcoming this difficulty.

One way of dealing with this is to provide a correction signal based on the integral of error measured at the actual exit point. The system is shown in figure 5. This system has the dynamic advantages of the properly located probe and the long-term accuracy of the integration. Some parameter changes can be handled by modification of the control system. A frequently occurring case is changes in flow rate. The required gain for satisfactory steady state response is a function of flow rate. This gain can be found from steady state tests and incorporated into the controller if the flow rate is measured. A significant change in flow rate (e.g. ± 50%) also makes a difference to the ideal probe location because it changes the response times of the system. However the dynamic responses of the system to both disturbances and to control are changed in the same way, so unless the performance is found to be critically dependent on probe location, the best position for average flow rate should be used. Since the probe is making measurements and deviations from a normal operating condition, any change in the operating condition must be accounted for by a change in the base level relative to which the measurement is made. This can be treated as part of the steady state design problem or by means of an adaptive mechanism as was done by Gaither: 2.

The sensitivity to parameter changes can be significantly reduced by operating several separate but interconnected control regions in sequence along the length of the system, as will be described in the next section.

6. <u>The Sequence of Controllers</u>

The single control region of figure 1 and figure 2 is subdivided into groups of separate control regions each with its own probe into the process stream. This is shown in figure 6. A disturbance entering at D_1 is suppressed by controller C_3 by the time it reaches P_3, but if the action taken by controller C_3 is in any way inappropriate, or if some other disturbance, D_2, enters, of which C_3 is unaware, then C_4 can suppress it and so on down the length of the system. The effect as seen by a particle moving along with the stream is somewhat like a sampled-data control. From the point of view of the static observer each control region acts like a feedforward control if its probe is upstream of the region itself (as for C_3 in fig. 6). As each of the regions is made smaller it becomes more and more likely that the best (single) probe location will be upstream of its control region and therefore under the influence of another controller. If however the probe is influenced by the control region for which it is the instrument (.e.g. if M_3 was within C_3 region) and not by a previous control region then the design procedure is as above for single probe control, for each independent region.

For the system depicted in figure 6 the best probe location and controller gain is first decided for each single control region acting alone, the others present but inactive, with disturbances entering the process upstream of the probe in question. Suppose this results in the probe for region C_3 (in fig 6) being. as shown.

inside region C_2, and likewise region C_4 has its probe in region C_3 etc. Consider a steady state disturbance D_1 entering the system which is detected by M_3 and suppressed by C_3 by the time it reaches P_3. At the measurement point M_4 there will still be an uncorrected deviation which will cause controller C_4 to take unnecessary corrective action. As a result an error will appear at M_5 and be propagated through to C_5 and so on. This propagation of the disturbance through the system in steady state can be suppressed by interconnection between controllers. The analysis follows.

A diagram of a complete system is shown in figure 7, and the parts relevant to the discussion are emphasised. For an individual control region to suppress a disturbance in steady state, the transfer from P_{r-1} at which point a disturbance d_{r-1} may be supposed to appear, to P_{r+1} has to made zero in steady state. Thus, for any particular probe position to be tested the objective is to have:

$$g_{pp}(o)\, g_{pp}(o) + g_{mp}(o)\, g_{cm}\, g_{pc}(o) = 0 \qquad 6.1$$

(where $g_{pp}(s)$, $g_{mp}(s)$ etc are transfer functions) which fixes g_{cm}, a constant gain. The choice of the probe location for m_{r+1} is determined by dynamic requirements as before.

To eliminate the spatial propagation of steady state disturbances the response at C_{r+2} to a disturbance d_{r-1} must be zero so therefore:

$$g_{mp}(o)\, g_{cm}(g_{mc}(o)\, g_{cm} + g_{cc}(o) + g_{pp}(o)\, g_{mp}(o)\, g_{cm} = 0 \qquad 6.2$$

and $g_{cc}(o)$ is fixed. The dynamic non-propagation of disturbances may be improved by making $g_{cc}(s)$ into a compensating network, which will always be physically realizable because it has to match the transfer through the physical components of the controller and process. It is in principle possible to calculate the required gains g_{cm} and $g_{cc}(o)$ from equations 6.1 and 6.2. However the simplest method of design is probably by adjustment on test using a simulation.

A multi-actuator scheme as described above exhibits insensitivity not only to parameter changes but also to errors in design and defects in the control system. Suppose for example that the gain g_{cm} of a particular controller is in error, then the next controller will take corrective action because the signals from its instrument and from its connection with the erroneous controller will not correspond to give zero action. Similarly if a controller saturates then whatever part of its load it cannot handle is passed on to the next section. The cost of this insensitivity is that of extra controller elements, both instruments and actuators.

7. Two-dimensional dynamic systems

Design techniques described above for one dimensional systems can be extended to deal with two and three dimensional systems. The technique will be described by means of an example.

A system of mass or energy transport in two-space dimensions has control applied at a boundary C and a disturbance entering at a boundary D. It is desired that when control is applied, the effect of the disturbances entering at D should not be felt on the desired zero effect line (DZEL). See for example figure 8. Consider first the steady-state problem. A disturbance applied over the boundary at D, most easily considered as a uniform disturbance, will have an influence everywhere in the system, up to, but not including the boundary where C is defined as zero. Similarly a change of C will have an influence everywhere. When C and D are applied together, with opposite polarity there must be a line between them on which the sum total effect is zero. In Figure 9 this is shown as the actual zero effect line (AZEL). If the magnitude of C is increased the line moves towards D and vice versa. If the shape of the distribution of control action with respect to spatial location is changed, the the shape of the AZEL is changed. For satisfactory steady-state performance the shape of AZEL has to be adjusted so that it coincides with the DZEL. However, this may well not be physically possible but some approximate solution can be achieved. If the spatial distribution of disturbance changes, the AZEL changes as well, so a control distribution has to be chosen for only one disturbance distribution. This may be amended by multi-controller systems and its effect on performance minimized by choice of 'worst' distribution. In linear systems with given spatial distribution D and C and with constant magnitude ratio between C and D the actual zero effect line does not move.

In heat or material transfer systems the distributions of energy or material will be continuous functions of space, with at worst discontinuous spatial derivatives. Thus, when controlled, the closer the AZEL is approached the smaller will be the observed response.

Any location inside the region dominated by D (see figure 9) can be used for a measurement probe location for proprtional control, but since the smallness of the remaining signal measured near the AZEL necessitates higher gain for steady-state accuracy, there is a "stability" boundary between D and the AZEL. The region between C and AZEL cannot be used for a control probe location because it corresponds to a 'positive' feedback situation requiring physically unrealizable compensation networks.

As with the one-dimensional system the design procedure can be broken into two parts : (1) to adjust the shape of the distribution of control so as to make the AZEL coincide as nearly as possible with DZEL in steady state, and (2) to chose a probe location inside the region dominated by D and on the safe side (nearer D) of the stability boundary to give best dynamic behaviour along DZEL, (or, more easily measured, at apoint on DZEL).

The system shown in figure 10 represents a flowing stream of material in which it is desired to maintain a uniform (zero) temperature profile across the stream at the DZEL as shown. The process is subject to disturbance in incoming stream temperature. Heat is transported through the system by flow, in the direction of positive x and by diffusion through the material. It is possible to control the temperature at the control boundary in the form:

$$c(t).\Psi(x)$$

This means that the distribution of control action, $\Psi(x)$, is chosen in the steady-state design phase and the magnitude, C(t), is determined during operation by the control system.

Such a process can be represented by a suitable analog network as shown in Fig. 11. Alternatively all calculations may be performed digitally.

For steady state design the response due to a (uniformly distributed) disturbance is found at all points in the system. In the analog of fig. 11. D_1, D_2, D_3, D_4, D_5 are set to 10v (say) and $C_1, C_2, C_3, \ldots C_9$ are set equal to zero and the resulting voltages in the process are measured. A plot of such a set of recorded results shown in figure 12 as a perspective drawing of the response surface over the x,y plane of the process. The response due to a uniformly distributed control action ($C_1=C_2=C_3=C_4=\ldots C_g=10_v$, $D_1=D_2=D_3=D_4=D_5=0v$) is shown in figure 13. The effect that it produces on the DZEL is again shown at the right hand end of the diagram.

When these two effects are combined so as to oppose each other the total effect is shown in figure 14. The actual zero effect line (AZEL) could be made to pass through the centre of the DZEL by making the magnitude of the control action 2.4. times that of the disturbance, and this is the condition shown. The response on the DZEL is far from satisfactory, and in fact the magnitude of response near the control boundary is greater than the magnitude of the incoming disturbance!

The influence that the control action has can be improved by shaping its distribution in x. Figure 15 shows the effect of a modified control profile ($\Psi(x)$) which has magnitude 10 units for the first 2/3 of the system and magnitude zero for the remaining 1/3. It can be seen that while not exactly the right shape the effect is has on the DZEL is closer to that of the disturbance so that the cancellation will be better.

When this new control profile is applied to the system to oppose an incoming disturbance, the resulting effect is shown in figure 16. The magnitude of the control action had to be 3.5. times the disturbance magnitude to make the AZEL pass through the middle of the exit stream.

For the purposes of this example, the steady-state design will be taken no further but it is clear that further changes could be made to improve the cancellation. However these changes will tend to make the control action near ω larger and give a smaller or reverse action near the DZEL. At some stage in the design of a real system a practical limit will be reached because of magnitude or gradient (in x) constraints.

8. Choice of instrument position for dynamic behaviour

Since the steady-state magnitude of response when appropriate control is applied is known at every point, (as in figure 8.17), the required gain to give accurate steady-state response from any measure point can be readily calculated as:

$$\frac{\text{Magnitude of Required Control Action}}{\text{Magnitude of Response at Measure Point}}$$

If an analog model is being used for design the constant disturbance D is replaced by a step disturbance and the dynamic response on the DZEL is observed as various measurement probe locations are tried using the appropriate gain in each case. Fig. 17 shows the results of such a test. The control system in this case had no significant dynamic lag so that C(t) was directly proportional to the measurement made by the probe. The best location was found by moving the probe upstream from the centre of the DZEL until the balanced response of Fig 17b was found. Tests of all possible adjacent points showed no improvement

9. Relay Control for two-dimensional systems

A single probe relay control system for a two-dimensional distributed process as described above is extremely simple. The measurement probe has to be very near the AZEL. This is because this is the only locus where having zero average deviation will also produce zero average deviation at the performance measure point, (P in figure 18), and in steady-state the relay controller has to settle for switching about zero. The design procedure is thus to move along the AZEL to find the location which gives best dynamic behaviour at P.

The dynamic response is always a limit cycle because a distributed system is always of high enough (infinite) order to produce 180° phase shaft with finite magnitude response, so that some finite gain will produce a limit cycle. Having the probe near to DZEL and further from the controller makes the limit cycle larger. Having it nearer to the controller makes the limit cycle smaller but reduces accuracy, especially when the distribution of disturbance changes.

The probe has to be located so that the limit-cycle switching of C, which it produces, has the correct ratio of on to off times (mark-space ratio) for balancing the disturbance. Thus as the size of the disturbance changes the required mark space ratio changes. The ideal probe location also shifts slightly so that when the AZEL (and P) has zero mean the mean value of the fluctuating measurement at the probe will be biased enough for the switching about zero to have the correct mark-space ratio. Thus for the more realistic fixed probe location the design has to be carried out for a particular size of disturbance. If the design is carried out for a disturbance magnitude about half the maximum possible which the controller can suppress the design will give a fair performance at maximum disturbance and always better performance for smaller disturbances.

Performance can be improved by the use of two probe control because the frequency of the limit cycle can be increased giving it a smaller magnitude and better overall dynamic behaviours for the complete system. It is still necessary that in steady-state the average sum of the two measures be close to zero. For this to be the case, one probe has to be between the AZEL and the control boundary C and the other between AZEL and disturbance boundary D (see fig. 18). Furthermore, since the steady state mean values of the measurements at all such points (such as M_1 and M_2 in fig 18) are known, the relative weightings to be given to the measures are fixed.

A feasible design procedure is to set up the simulated process without disturbance and adjust the position of the probe near to the control boundary (M_1 in fig. 18) until the limit cycle magnitude and frequency as seen at P is satisfactory. Then put in the disturbance and choose the location of the second probe (M_2) so that good dynamic response to a disturbance results at P. As in the case of the linear control design this can be achieved by looking for early or late control action. It may also be observed that the final system will be slightly better than the linear system in dynamic response, because the controller applies maximum forcing as soon as any disturbance is detected. There is, however always the limit cycle superimposed on this dynamic behaviour.

10. Conclusions

A method has been described for designing the control systems for distributed processes which are linear and for which very good models are available to describe internal behaviour. The method may require a good deal of computational effort but the level of computational difficulty is no greater than that required to produce a working mathematical model. Furthermore, because the work is done with a simulation any real world constraints such as limitations on magnitudes, gradients, numbers of components etc. can readily be incorporated during the design phase.

Alternative design techniques for the same class of distributed processes have been developed by Gaither[2] and Licht.[3] Both these methods used frequency domain concepts. Licht uses spectral analysis to compute the mean square error of a control system as a function of the weighting attached to the information available from each of the possible or allowable probes. He then uses this measure in an optimising technique which results in the choice of the best use (weighting) of the information from the available probes. Gaither's techniques is more closely related to the one described in this paper but his criterion for choosing probe location is based on measurements of phase angle rather than measurements of step responses. The net result is just about the same. Gaither also provides a technique for adaptive adjustment both of the gain of the controller in the linear case and also for the adjustment of the set point or base level relative to which the measurement is made for systems where there are shifts in operating points. There is thus available a group of design techniques for controllers for distributed processes which are no more difficult to carry out than the simulation of the processes themselves.

11. References

1. Aris, R, The Optimal Design of Chemical Reactor, Academic Press 1961.
2. Gaither, P.H. "Modelling and Control of Distributed Parameter Systems". Ph.D. Thesis, Case Institute of Technology, 1965.
3. Licht, B.W., "Statistical Analysis of Distributed Parameter Systems", M.S. Thesis, Case Institute of Technology, 1965.

12. Acknowledgement

The work reported above was performed while the author was a member of the Systems Research Center at Case Institute of Technology, Cleveland, Ohio, U.S.A., and was supported by a number of industrial sponsors.

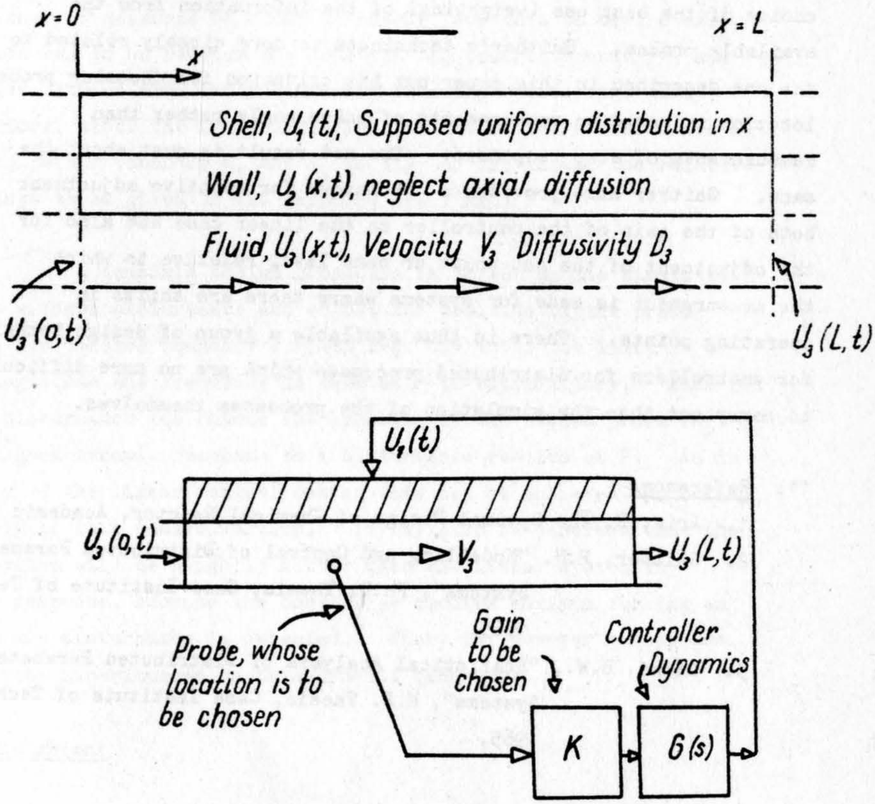


Fig.1. Flow and diffusion system for control design example

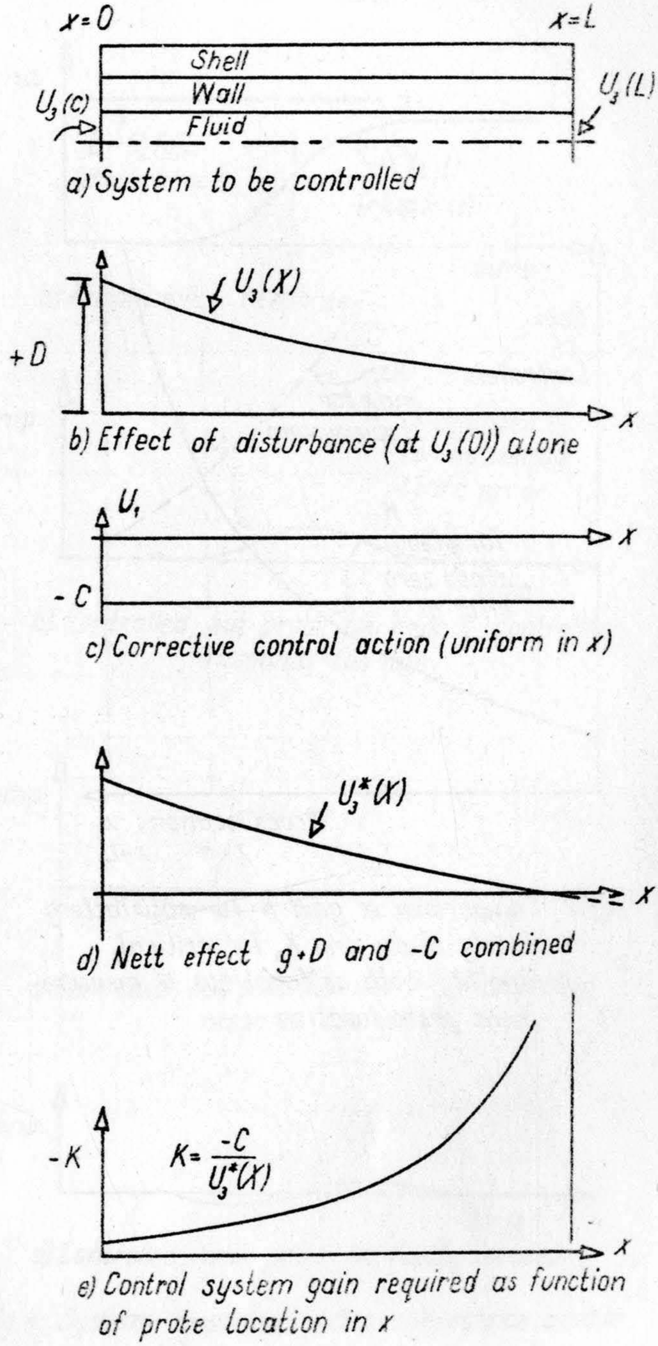


Fig. 2. Development of required steady-state controller gain U_3 probe location

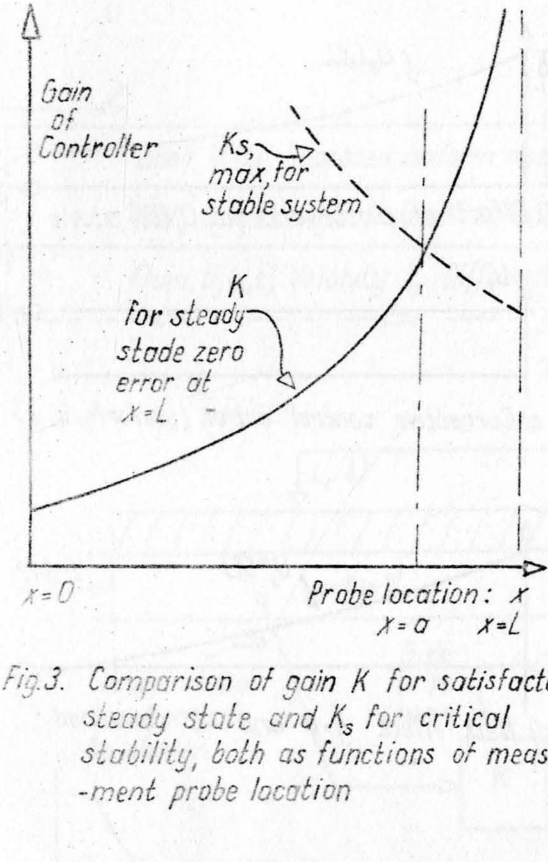


Fig. 3. Comparison of gain K for satisfactory steady state and K_s for critical stability, both as functions of measure-ment probe location

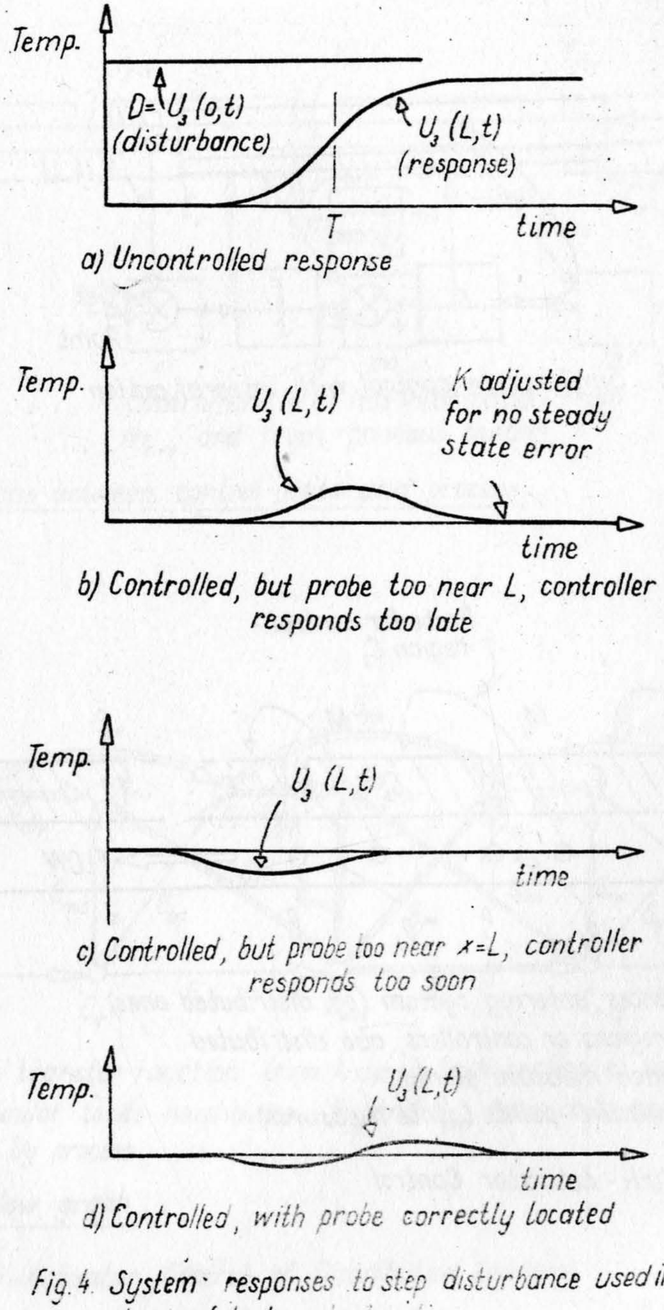


Fig. 4. System responses to step disturbance used in choice of best probe location

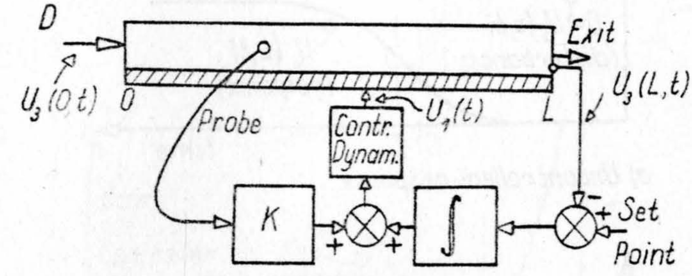
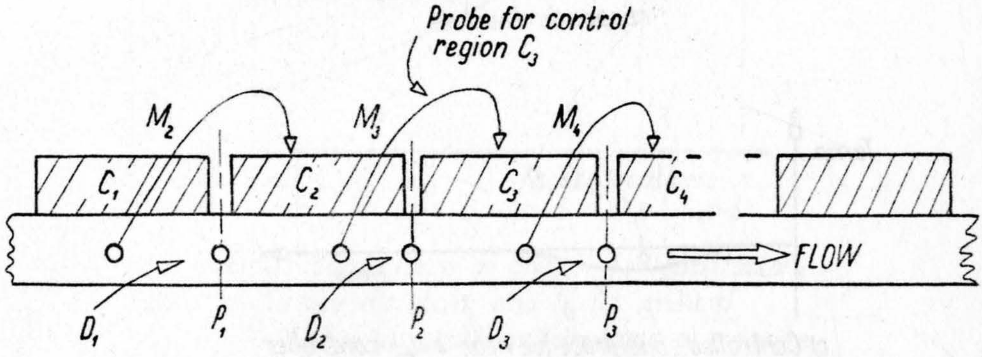
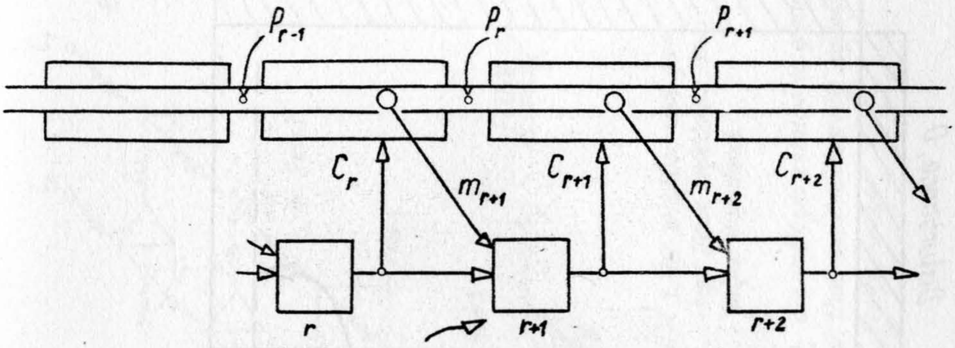


Fig. 5. Single-probe control with integral action

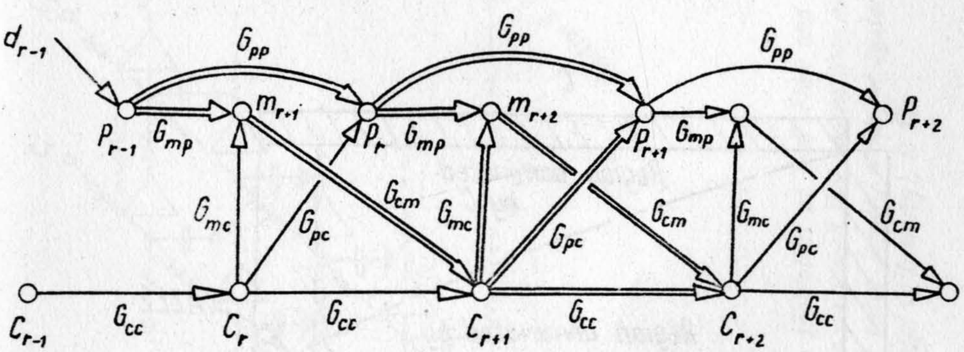


D_i are disturbances entering system (eg. distributed ones)
C_i are control regions or controllers, also distributed
P_i are performance measure points
M_i are control measure points (state measures)

Fig. 6. Multi-Actuator Control



a) Connections between control units and process



Each G_{ik} is transfer function from k to i. Each section of system is here supposed similar to its neighbours. G_{cc} and G_{cm} have to be chosen, rest are determined by process

b) Signal flow graph

Fig. 7. Multi-Actuator Control of Distributed System

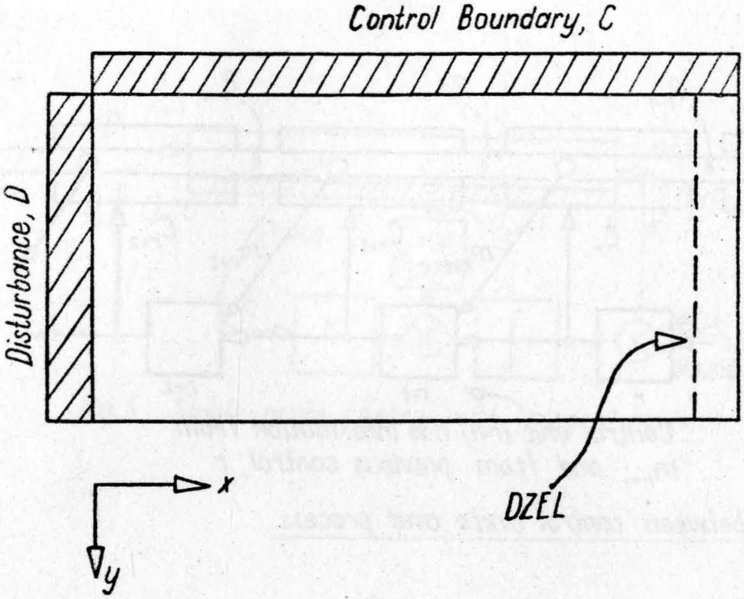


Fig. 8. Disturbance and control and desired zero effect line in a 2-D system

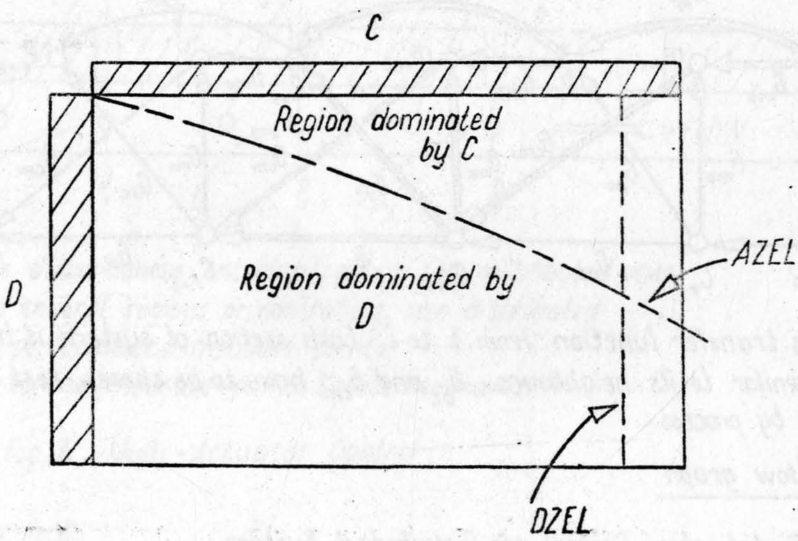


Fig. 9. Actual zero effect line (AZEL) lies between opposing Control (C) and disturbance (D).

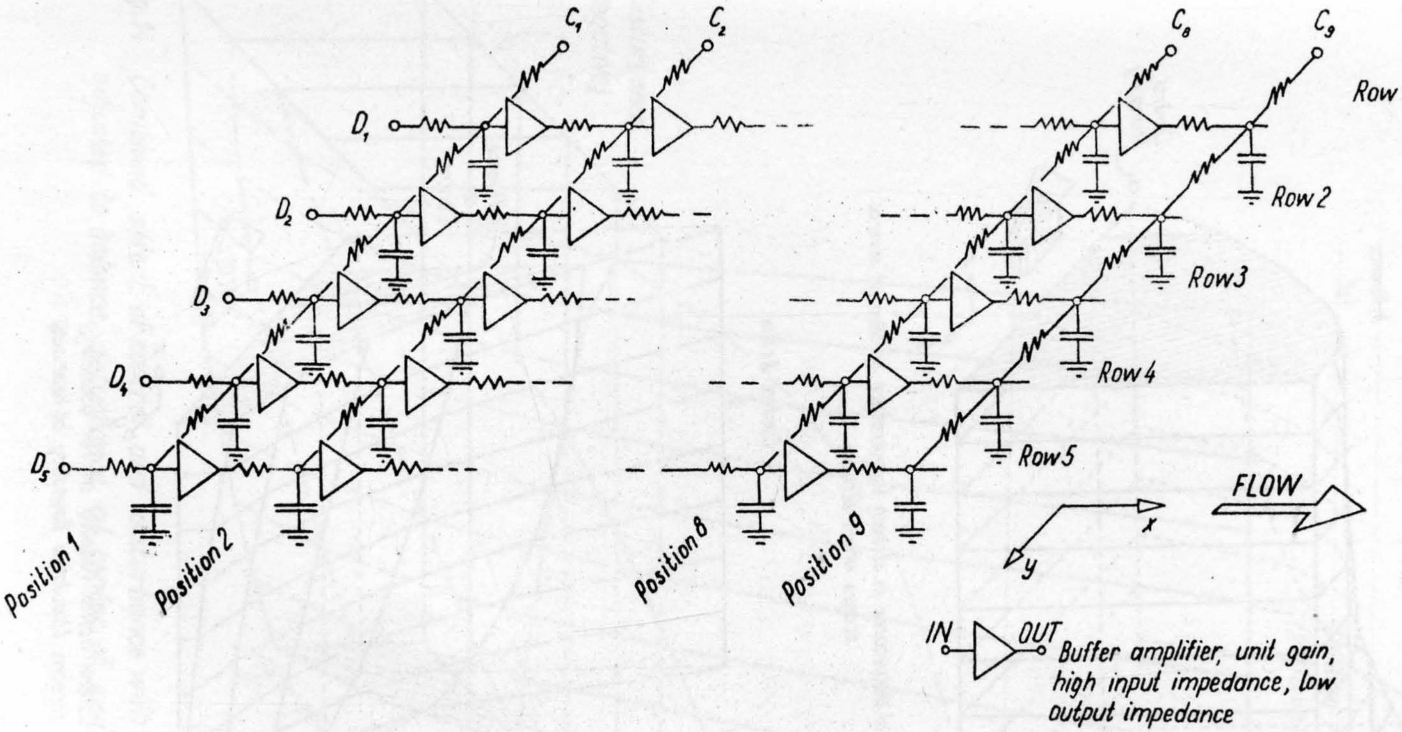


Fig. 11. Special purpose analog simulation of 2-D Flow and - Diffusion System. C_i, D_i are voltages relative to ground. For the case where it is considered that spatial increments Δx and Δy are equal, that axial diffusivity equals lateral diffusivity and ratio: length/width, of system is 9/5 then equivalent Peclet number, (Velocity x length)/(Axial Diffusivity), is 18

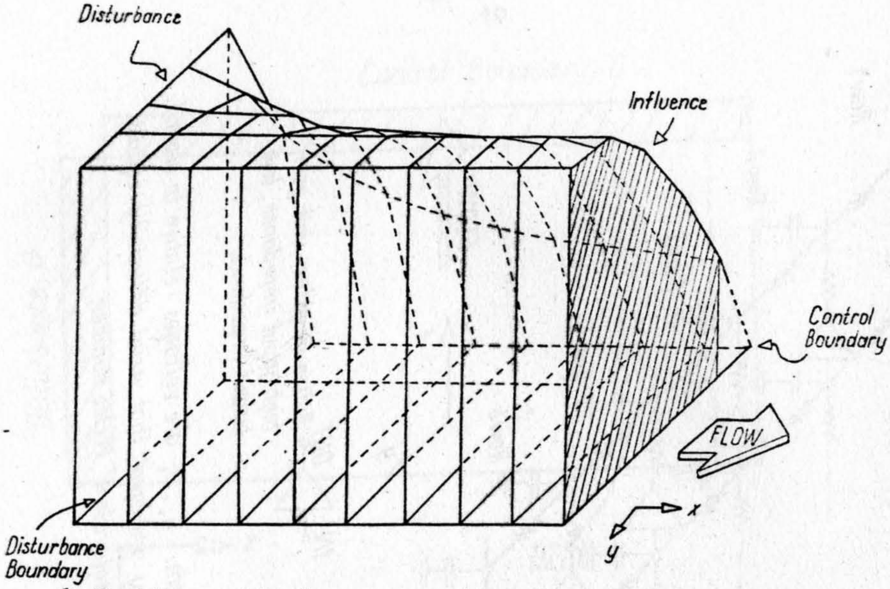


Fig. 12. Influence of disturbance in stream temperature as seen at several stages down-stream

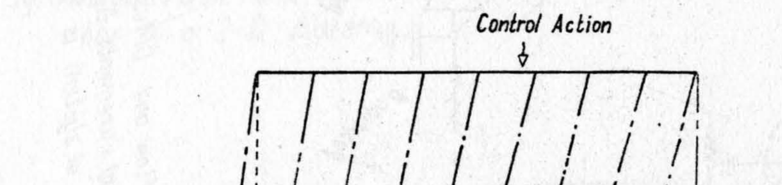


Fig. 13. Influence of Control action (applied on far boundary) on flowing stream. (Insulated boundary at near side)

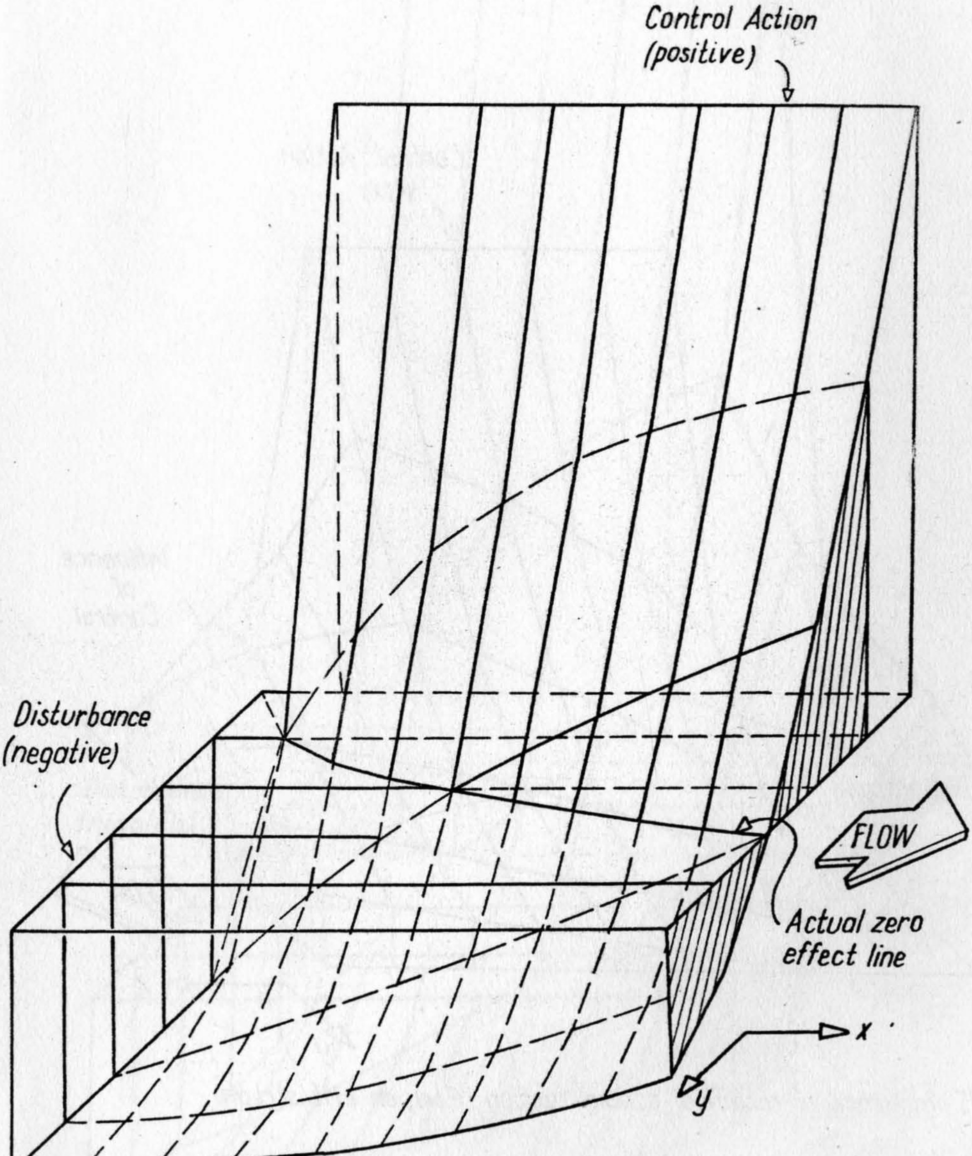


Fig.14. Combined effect of control and disturbance with control adjusted to balance disturbance at center of exit stream

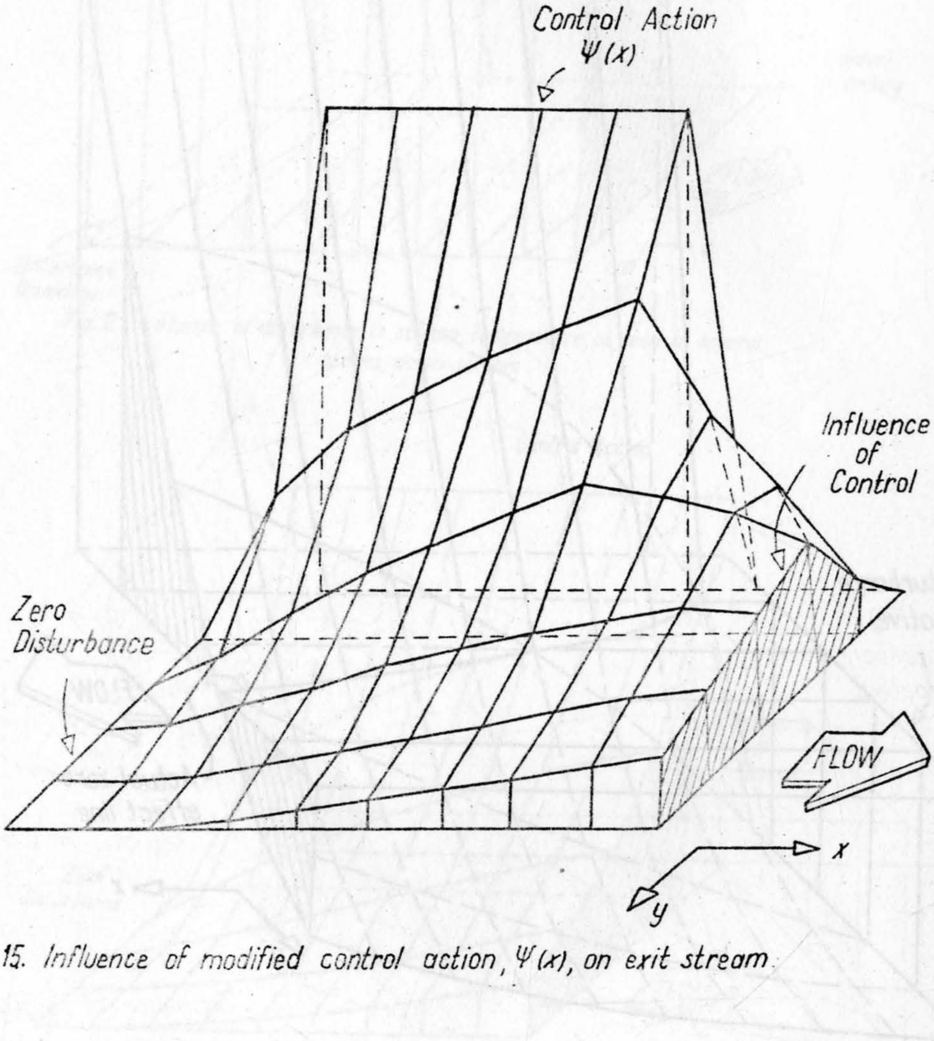


Fig. 15. Influence of modified control action, Ψ(x), on exit stream

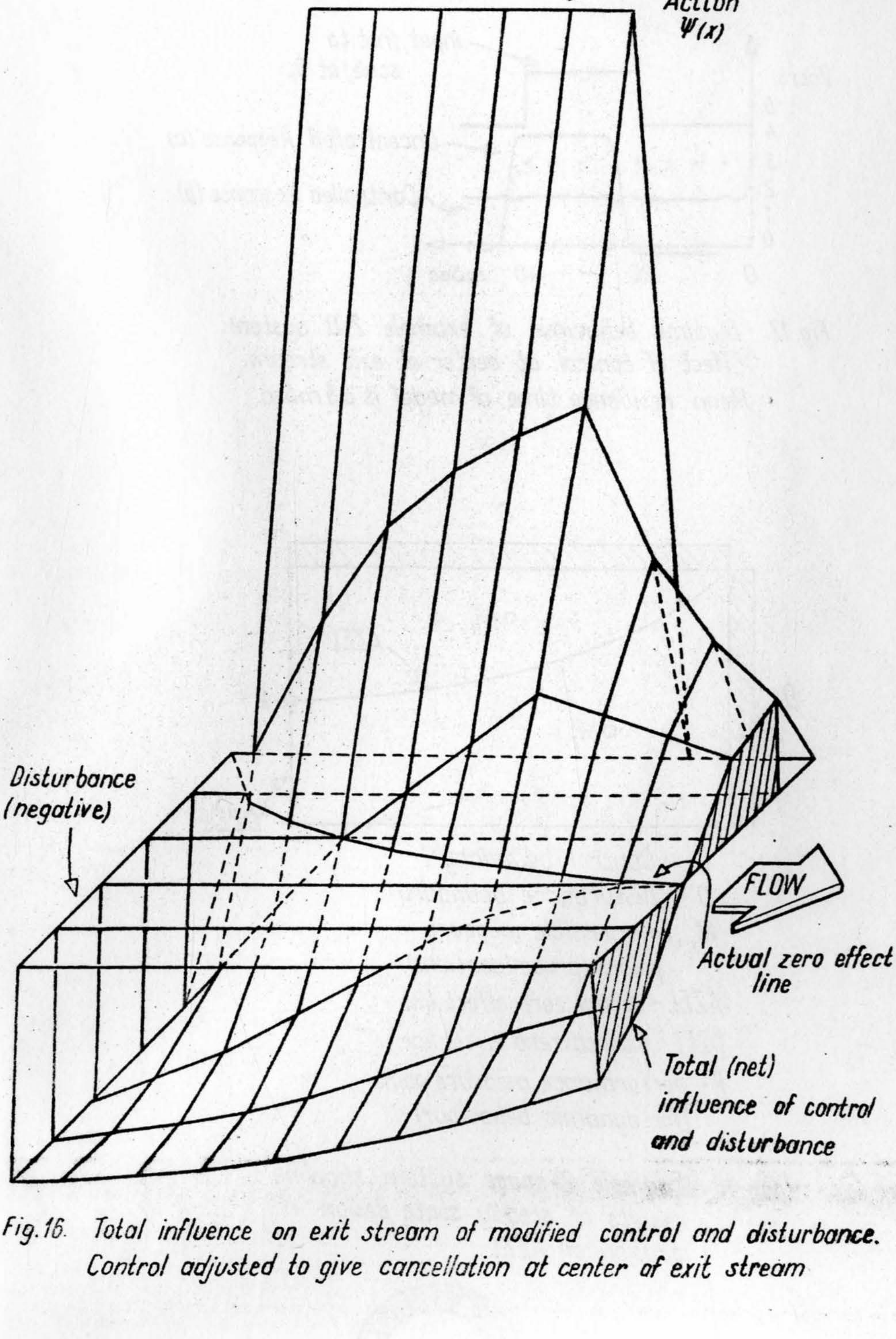


Fig. 16. Total influence on exit stream of modified control and disturbance. Control adjusted to give cancellation at center of exit stream

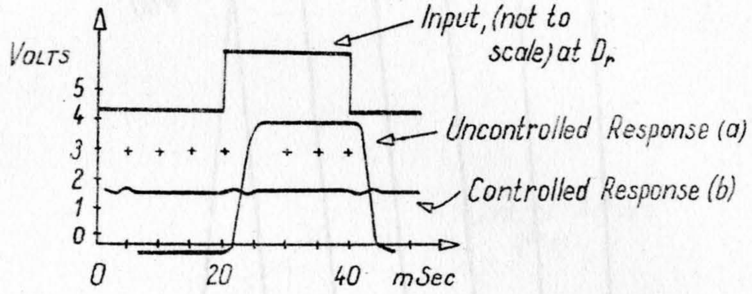
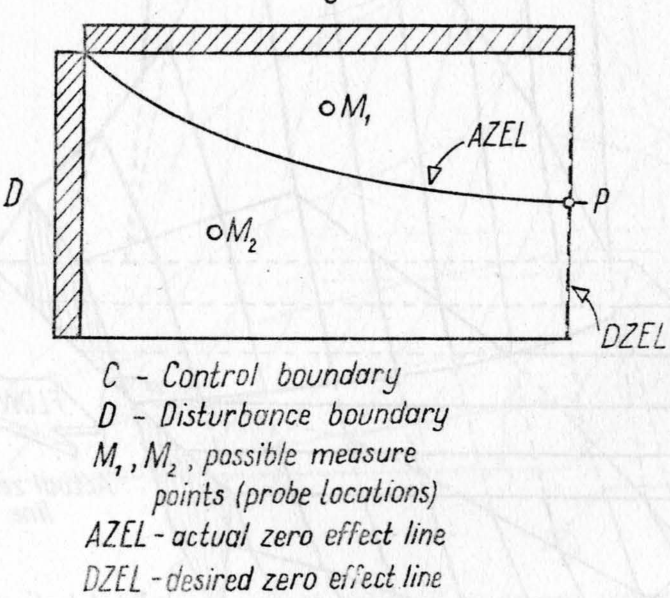


Fig. 17. Dynamic behaviour of example 2-D system Effect of control at center of exit stream. Mean residence time of model is 3.5 mSec



C - Control boundary
D - Disturbance boundary
M_1, M_2, possible measure points (probe locations)
AZEL - actual zero effect line
DZEL - desired zero effect line
P - performance measure point (for dynamic behaviour)

Fig. 18. Example 2-space system showing results of steady state design in location of AZEL