Adaptive Sampled-data Systems

B. WIDROW

Introduction

Adaptive or self-optimizing systems have the capability of automatically modifying their own structures in order to achieve performance optimization. An adaptive capability is particularly useful in cases where the nature of system input signals is not known, even statistically. In other cases, the nature of the input might be known to be changeable; for example, input statistics can be non-stationary. An adaptive system that continually searches for the optimum within its allowed class of possibilities by an orderly trial-and-error process would give performance vastly superior to that of a fixed system in many of these instances.

Several ways of classifying adaptation schemes have been proposed in the literature. The author finds it convenient to think merely in terms of closed loop and open loop adaptation processes.

The open loop adaptation process involves making measurements of input or environmental characteristics, applying this information to a formula or a computational algorism, and using the results to set the adjustments of the adaptive system. Closed loop adaptation, on the other hand, involves automatic experimentation with these adjustments to optimize a measured system performance. Open loop adaptation is usually simpler to implement where it is applicable. Closed loop adaptation is more fundamental and more generally applicable.

The purpose of this paper is to study adaptation, particularly closed loop adaptation. The objective is to gain an understanding of how automatic system synthesis can be achieved by making use of 'performance feedback'.

Feedback and Trial-and-error Processes

Iterative or trial-and-error processes are integral parts of adaptive systems. They provide the mechanism of adaptation. It is often convenient to represent such processes as feedback systems; the error of trial and error is analogous to the 'error' of feedback control. Many of the relaxation and iterative methods employed by numerical analysts appear to be linear feedback systems when represented in this manner. An example of importance in this discussion is that of surface exploration for stationary points.

Many of the commonly used gradient methods search the surfaces by making changes in the independent variables (starting with an initial guess) in proportion to measured partial derivatives to obtain the next guess, and so forth. These methods give rise to geometric (exponential) decays in the independent variables as they approach a stationary point for second-degree or quadratic surfaces. One-dimensional surface-searching is illustrated in *Figure* 1.

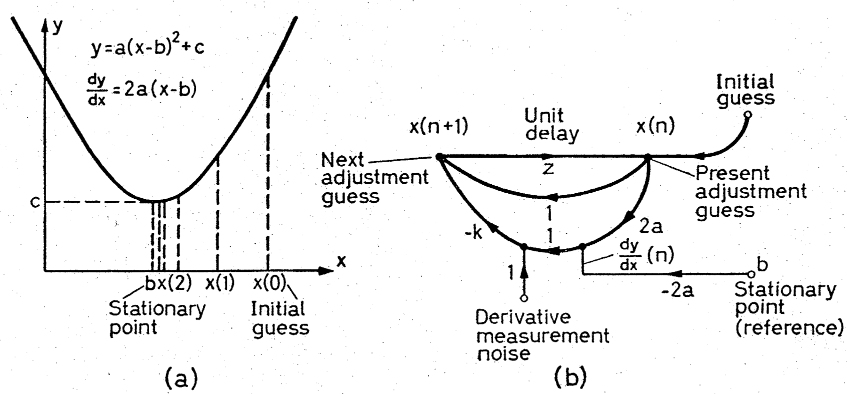

Figure 1. *One-dimensional surface searching*

The surface being explored in *Figure* 1 is given by equation 1. The first and second derivatives are given by equations 2 and 3.

$$y = a(x - b)^2 + c \tag{1}$$

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 2a(x - b) \tag{2}$$

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = 2a \tag{3}$$

A sampled-data feedback model of the iterative process is shown in *Figure* 1 (*b*)[1–3]. The flow graph can be reduced, and the transfer function from any point to any other point can thus be found. The resulting characteristic equation is

$$(2ak - 1)z + 1 = 0 \tag{4}$$

In order to choose the 'loop gain' k to get a specific transient decay rate, one would have to measure the second derivative (2a) at some point on the curve.

The first and second derivatives are given by equations 5 and

6. These relations are precise for parabolas, and are approximate for higher degree curves (see *Figure* 2).

$$\left.\frac{dy}{dx}\right|_{x_B} = \frac{1}{2\delta}[C - A] \tag{5}$$

$$\left.\frac{d^2y}{dx^2}\right|_{x_B} = \frac{1}{\delta^2}[C - 2B + A] \tag{6}$$

A two-dimensional parabolic surface is described by equation 7, the partial derivatives by equations 8, and the second partial derivatives by equations 9.

$$y = ax_1^2 + bx_2^2 + cx_1 + dx_2 + ex_1x_2 + f \tag{7}$$

$$\frac{\partial y}{\partial x_1} = 2ax_1 + c + ex_2 \quad \frac{\partial y}{\partial x_2} = 2bx_2 + d + ex_1 \tag{8}$$

$$\frac{\partial^2 y}{\partial x_1^2} = 2a \quad \frac{\partial^2 y}{\partial x_1 \partial x_2} = e \quad \frac{\partial^2 y}{\partial x_2^2} = 2b \tag{9}$$

A vector flow-graph model of a two-dimensional iterative surface searching process is given in *Figure* 3 (*a*). The branches

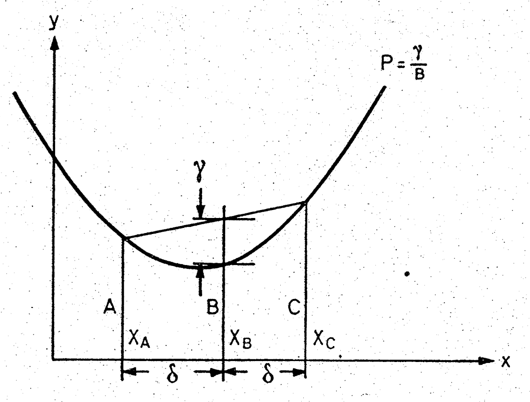

Figure 2. *Measurement of derivatives; definition of the perturbation*

in this graph are capable of carrying two-dimensional samples, indicated by column matrices. This flow graph can be reduced straightforwardly by making use of the rules of matrix algebra. There are as many natural frequencies (decay rates) as there are independent coordinates. The multidimensional loop gain in this case is determined by choice of the matrix of k's.

There are many surface searching methods in common use. Among these are the method of steepest descent, Newton's method, and the Southwell relaxation method.

The flow graph of *Figure* 3 (*a*) can represent Newton's method, wherein the matrix of k's is the inverse of the matrix of second partials. Multidimensional transients die out completely in one step. A modified Newton's method has the same matrix of k's, only scaled by a factor less than unity. Transients die out geometrically, not in one step, and are of a single time constant. Successive adjustments proceed along a straight line in multidimensional space from the initial guess to the stationary point. Cross-coupling among the coordinates is eliminated.

The flow graph of *Figure* 3 (*a*) can also represent the method of steepest descent. Here, the matrix of k's is a diagonal one, with identical elements on the main diagonal. This corresponds to vector changes in adjustment being proportional to the successive local gradient vectors. Cross-coupling is present.

The flow graph of *Figure* 3 (*b*) represents surface searching by the Southwell procedure. Adjustment along each coordinate each time is set to minimize y. This corresponds to the matrix of k's being a diagonal one, with

$$k_{11} = \frac{1}{2a} \quad \text{and} \quad k_{22} = \frac{1}{2b}$$

Cross-coupling is present, but transients are of a single time constant.

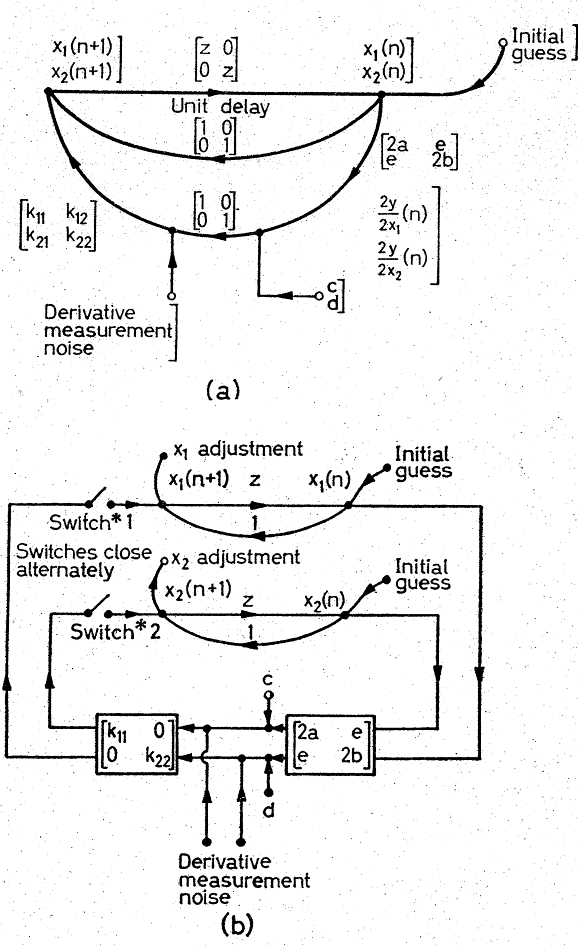

Figure 3. *Two-dimensional surface searching models*

Analysis of an Adaptive Sampled-data Predictor

Consider the general linear sampled-data system formed of a tapped delay line, shown in *Figure* 4. This system is intended to be a statistical predictor. The present output sample $g(n)$ is a linear combination of present and past input samples. The constants in this combination are h_0, h_1, h_2, etc., the predictor impulse–response samples, or the gains associated with the delay-line taps. Their choice constitutes the adjustable part of the predictor design. They may be adjusted in the following manner. Apply a mean square reading meter to $\varepsilon(n)$, the difference between the present input and the delayed prediction. This meter will measure mean square error in prediction. Adjust $h_0, h_1, h_2 \ldots$, until the meter reading is minimized.

The problem of adjusting the h's is not trivial, because their effects upon performance interact. Suppose that the predictor has only two impulses in its impulse response, h_0 and h_1. The mean square error for any setting of h_0 and h_1 can be readily derived:

$$\varepsilon(n) = f(n) - h_0 f(n-1) - h_1 f(n-2)$$

$$\overline{\varepsilon^2}(n) = \phi_{ff}(0)h_0^2 + \phi_{ff}(0)h_1^2 - 2\phi_{ff}(1)h_0 - 2\phi_{ff}(2)h_1 + 2\phi_{ff}(1)h_0h_1 + \phi_{ff}(0) \tag{10}$$

The discrete autocorrelation function of the input is $\phi_{ff}(k)$. The mean square error is a parabolic function of the predictor adjustments h_0 and h_1.

The optimum m impulse predictor can be derived analytically by setting the partial derivatives of $\overline{\varepsilon^2}$ of equation 10 equal to zero. This is the discrete analogue of Wiener's optimization[4] of continuous filters. Finding the optimum system experimentally is the same as finding a minimum of a paraboloid in m dimensions. This could be done manually by having a human operator read the meter and set the adjustment, or it could be done automatically by making use of the iterative gradient methods for surface searching, as described in the previous section. When either of these schemes is employed, an adaptive system results that consists essentially of a 'worker' and a 'supervisor'. The worker in this case predicts, whereas the supervisor has the job of adjusting the worker.

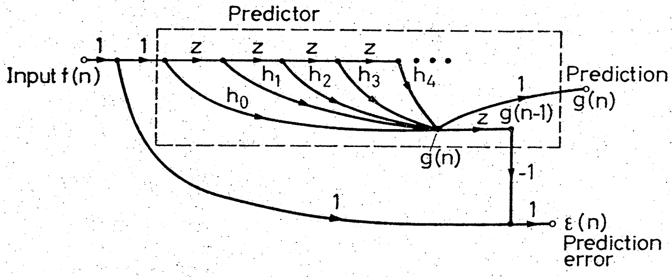

Figure 4. An adjustable sampled-data predictor

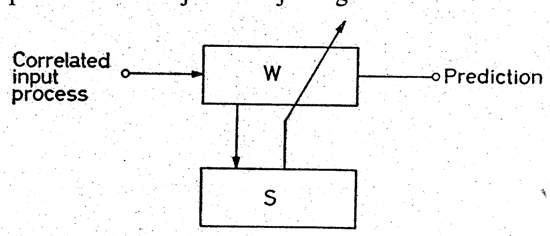

Figure 5. An adaptive predictor

Figure 5 is a block diagram representation of such a basic adaptive unit. The supervisor continually seeks a better worker. Adaptation is a multidimensional feedback process. The 'error' signal is the gradient of mean square error with respect to adjustment.

Noise enters the adaptation feedback system because the input process cannot be continued indefinitely for each measurement of mean square error (A, B, C of *Figure* 2, needed for gradient measurement), and thereby places a basic limitation upon adaptability. It will be shown that the slower the adaptation, the more precise it is. The faster the adaptation, the more noisy (and poor) are the adjustments.

Consider that the adaptive model has only a single adjustment. A plot of mean square error *versus* h_0 for this simplest system would be a parabola, analogous to the parabola of *Figure* 1. During each cycle of adjustment, the derivative of $y \equiv \overline{\varepsilon^2}$ with respect to $x \equiv h_0$ would have to be measured according to the scheme of *Figure* 2.

Noise in the system adjustment causes loss in steady-state performance. It is useful to define a dimensionless parameter M the 'misadjustment', as the ratio of the mean increase in mean square error to the minimum mean square error. It is a measure of how the system performs on the average, after adapting transients have died out, compared with the fixed optimum system. With regard to the curve of *Figure* 1

$$M = \frac{\bar{y} - c}{c} \tag{11}$$

Consideration of equation 1 shows that $(\bar{y} - c)$, the average increase in y, is equal to the variance in x multiplied by a. This variance is due to derivative measurement noise which propagates by way of the iterative surface-searching process.

The noise propagation path is shown in the flow graph of *Figure* 1 (*b*). Assuming that derivative measurement noises are statistically independent from one iteration cycle to the next, the variance in x equals the variance in derivative noise multiplied by $[1/(8a^2\tau)]$, a conservative approximation to the sum of squares of the impulses of the impulse response from the noise injection point to the adjustment x. The time constant τ is defined so that if $\tau = 1$ adaptation transients decay by a factor $(1/\varepsilon)$ with each iterative cycle.

Equation 5 gives the derivatives as the difference between 'forward' and 'backward' measured values of y multiplied by $1/2\delta$. 'Noise' in the measurements of y (due to finite sample size) causes noisy derivative measurements. A detailed derivation of the variance in derivative measurement is given in reference 5. The result is that

$$(\text{variance in derivative measurement}) = \frac{ac}{NP} \tag{12}$$

The number of forward or backward measurements per cycle is N. The perturbation is P. Relation 12 is based upon several assumptions: that the adjustment x is in the vicinity of the minimum, that the prediction error signal is gaussian distributed (relation 12 is quite insensitive to the shape of this distribution density, however), and that the prediction error samples are uncorrelated (correction for correlation less than 90 per cent is very small).

If the nature of the physical process permits 'data repeating', i.e. if it is possible to apply the same input data to the system for both forward and backward measurements, the variance of the derivative measurement noise does not depend upon the amplitude of the perturbation. Making the same assumptions as were made previously, the expression for the variance with data repeating is

$$(\text{variance in derivative measurement}) = \frac{4ac}{N} \tag{13}$$

It should be noted that in this case N is the total number of error samples per cycle.

The misadjustment equals $(1/(8a\tau c))$ multiplied by the variance in derivative measurement noise. Accordingly

$$M = \frac{1}{8N\tau P} = \frac{1}{4(2N\tau)P} \tag{14}$$

For the data repeating case

$$M = \frac{1}{2(N\tau)} \tag{15}$$

The $(N\tau)$ product is related to the total number of samples 'seen' by the system in adapting to a step transient in input process statistics. Notice that a given effect could be achieved by using many samples per cycle (large N) and few cycles with large steps to adapt (small τ), or by using few samples per cycle (small N) and proceeding towards the optimum with small steps (large τ).

Let the number of samples that elapse in one time constant of adaptation be called the 'adaptation time constant' Γ. Where data repeating is not practised, $\Gamma = 2N\tau$. Where data is repeated, $\Gamma = N\tau$. Expressions 14 and 15 become 16 and 17 respectively.

$$M = \frac{1}{4\Gamma P} \tag{16}$$

$$M = \frac{1}{2\Gamma} \tag{17}$$

These ideas can be applied to multidimensional adaptation by using the flow graphs of *Figure* 3. The misadjustment increases with m^2 when Newton's method is used. The misadjustment increases with m when data is repeated for any single time-constant method. One such method, Southwell's, is easy to implement (no matrix inversion). The misadjustment is given by equation 17 multiplied by m.

These principles may be applied in a variety of situations, two of which are illustrated in *Figure* 6. Performance feedback is used in the system of *Figure* 6 (*a*) to achieve imitation of an unknown complex system. The adaptive system 'learns' of the characteristics of the unknown system by imitating its behaviour as best it can. The mean square error is a parabolic function of the adjustments, if the input is stationary and the unknown system is linear. A combination of imitation and prediction enables an adaptive system to predict the output of an unknown dynamic system by making use of both its input and output signals. A conventional predictor would use only the output signal. In *Figure* 6 (*b*), a scheme is shown which combines a low noise, low capacity link for performance feedback with a high capacity, noisy communication link. An adaptive filter is used to separate noise and signal. The mean square error is again a parabolic function of the adjustments, and the rate of adaptation is limited by the low capacity link.

The misadjustment gives a measure of the effectiveness of adaptation. It neither gives information on the magnitude of the minimum mean square error, nor does it give information on the effectiveness of the choice of adjustment variables (are there enough of these and are they the best to use).

An extensive series of simulation studies was undertaken by R. R. Brown with the aid of an IBM 704 digital computer. The results of these experiments have shown that the measured misadjustments rarely differ from their predicted values by as much as 20 or 30 per cent. The results of one of these simulations is given in reference 5.

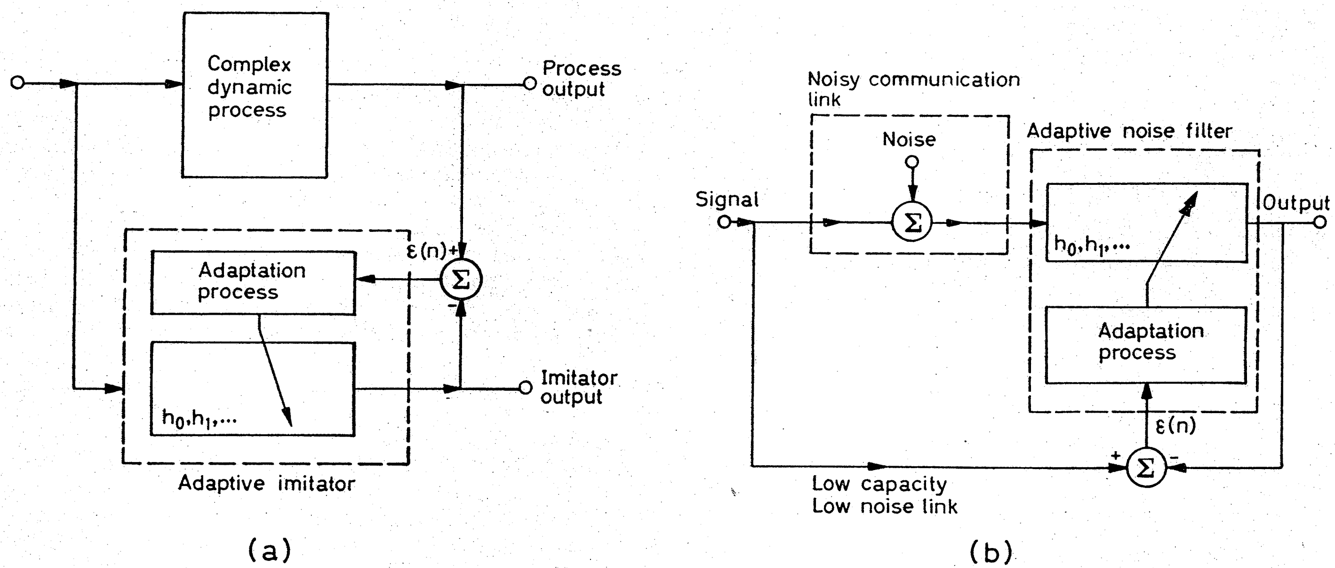

Figure 6. *An adaptive imitator and an adaptive noise filter*

The misadjustment formulae are quite accurate when applied to the situations for which they have been derived. These formulae serve as 'rules of thumb' when other performance criteria than minimization of mean square error are used, and when the worker is non-linear[6]. Solutions to many practical problems in the control and communications fields will be attainable by using these principles. As evidence of this, a ten-impulse filter could adapt to a major change in input process statistics after 'seeing' 200 process samples and would have a steady-state misadjustment of about 10 per cent.

Application of Adaptive Filters to Non-stationary Signals

The mean square error surface is fixed in shape orientation and position when the input to an adaptive system is a stationary process. It is possible to define a quasi-static or instantaneous mean square error surface when the system input is not stationary. The characteristics of this surface vary with the changes in statistical characteristics of the input signal. Performance feedback allows an adaptive system to 'track' continually the changes in process and environment, and to maintain its adjustments at any instant close to the instantaneous mean square error surface minimum.

The response of an adaptive filter with a single adjustment to a sudden change in input process is illustrated in *Figure* 7 for two different values of adaptation time constant. The solid lines represent ideal adjustments, and the dotted lines represent the actual noisy geometric adjustments. This kind of response suggests the R–C filter analogue of *Figure* 8.

The 'signal' in *Figure* 8 represents the ideal adjustment. For a stationary process, this is a constant. The variance of the 'noise' that propagates through the R–C filter is inversely proportional to its time constant Γ, and accounts for the misadjustment from small sample size.

When the input process is non-stationary, there are two sources of loss in performance. There is a component of misadjustment that results from small sample size 'noise', and a component due to lag in adaptation or imperfect tracking of the process by the system. In making Γ small, misadjustment due to small sample size becomes large, and in making Γ large, misadjustment due to slow process tracking becomes large.

A stationarily non-stationary worker input signal causes the 'signal' of *Figure* 8 to be a low frequency random process.

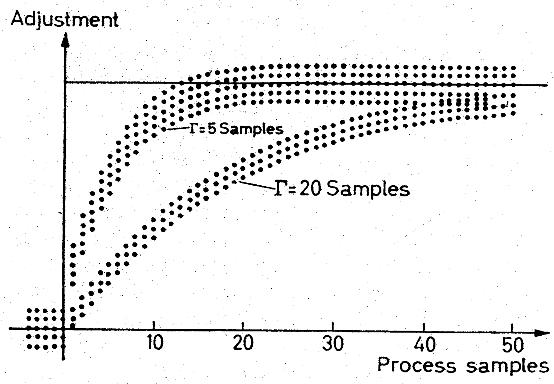

Figure 7. Response of an adaptive system adjustment to a sudden process change

Such an input is generated by applying an uncorrelated random sequence to a sampled-data filter whose impulse response is slowly randomly varied.

The definition of misadjustment is hereby generalized to be the ratio of the difference between the mean square error of an ideal adaptive system and the actual one divided by the mean square error of the ideal system, where averages are taken over many variations of the input process. The ideal system (a non-physical entity) has an instantaneous impulse response which is always optimum for the instantaneous input-process statistics. The misadjustment is proportional to the total variance of the 'tracking error' of *Figure* 8.

The transfer function of the propagation path of the small-sample-size 'noise' (*Figure* 8) is

$$H(S) = \frac{1}{\Gamma s + 1} \tag{18}$$

The transfer function by which signal tracking error develops is

$$1 - H(S) = \frac{\Gamma S}{\Gamma S + 1} \tag{19}$$

The spectral densities of 'signal' and small-sample-size 'noise', as well as the amplitudes of these transfer functions for real frequencies, are sketched in *Figure* 8 (*b*). The behaviour of $|1 - H(S)|$ for $S = j\omega$ and for small ω is approximately like that of $|\Gamma\omega|$. Of importance is $|1 - H(S)|^2$ or $\Gamma^2\omega^2$. If the spectra of 'signal' and 'noise' are roughly as shown, the variance of the 'signal' error for fixed variance of 'signal' is proportional to Γ^2 while the 'noise' variance is proportional to $1/\Gamma$. Therefore, the misadjustment is

$$M = \frac{\alpha}{\Gamma} + \beta\Gamma^2 \tag{20}$$

The constants α and β are properties of the system input process.

Setting the derivative of M to zero gives

$$\frac{\alpha}{\Gamma} = 2\beta\Gamma^2 \tag{21}$$

This means that the adaptation time constant is optimized when the misadjustment component due to small-sample-size 'noise' equals twice the misadjustment component, due to lag in process tracking. The same result can be generalized for any multidimensional iterative process having a single natural frequency.

A Higher Level of Adaptation—the Supervisor's Supervisor

When the nature of the non-stationary input is known, it is possible (but not simple) to calculate the optimum Γ. An alternative that would require much less knowledge of the input process and no elaborate calculation of the value of Γ is to achieve self-optimization of Γ. This can be done by experimentally varying Γ, much as the impulses in the impulse response are varied, with the objective of optimizing long-term performance. A block diagram of such a system, an adaptive predictor, is shown in *Figure* 9. The box W is the worker, S_1 is the supervisor, and S_2 is the supervisor's supervisor. Choice of Γ constitutes the adjustment of S_1.

The optimum Γ minimizes the variance of the random 'tracking error' signal of *Figure* 8. The derivative of the

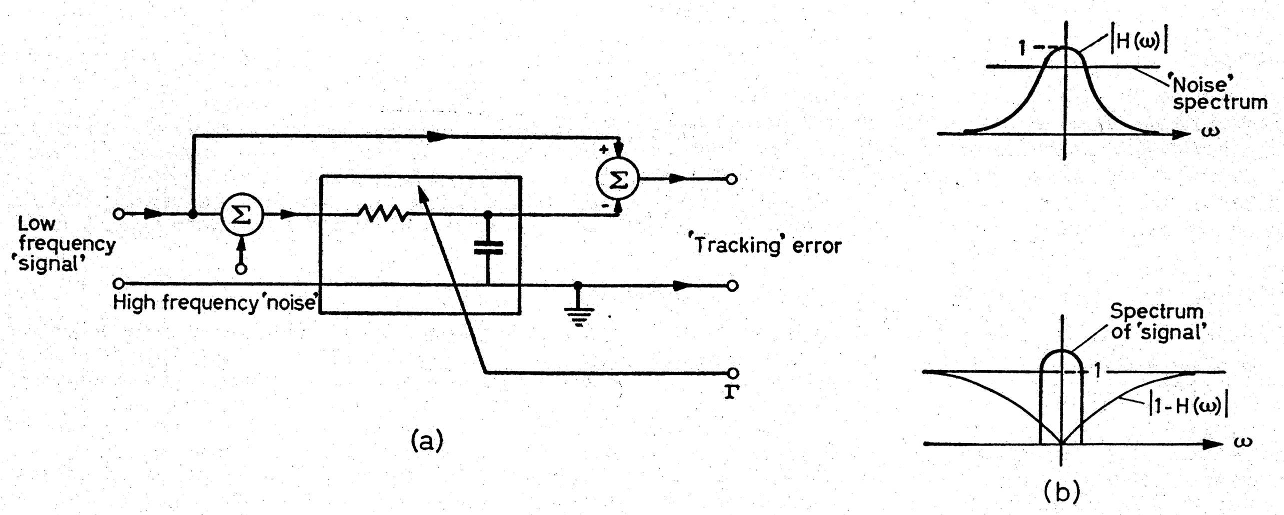

Figure 8. R–C filter analogue of adaptation process

'tracking error' variance with respect to Γ is obtained by measuring the worker's mean square error (long-term averages over many error samples) for 'forward' and 'backward' values of Γ. This may be done with and without data repeating.

The misadjustment M_s of the supervisor S_1 adds directly to the misadjustment M_w of the worker (defined with Γ at optimum) to give the overall misadjustment M.

$$M = M_w + M_s \tag{22}$$

Inspection of *Figure* 8 shows that the random 'tracking error' contains a high frequency component that fluctuates at

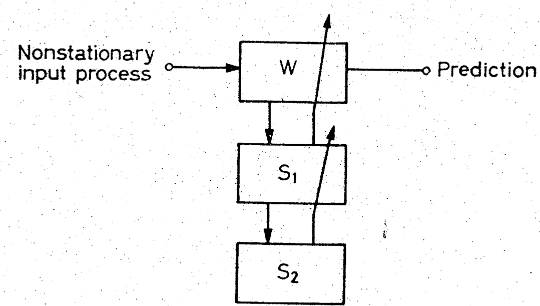

Figure 9. *An adaptive predictor with two levels of adaptation*

the input sampling rate, and a low frequency component that fluctuates with the variations in the input statistics. When the setting of Γ is close to optimum, the variance due to the low frequency component is approximately one-third of the total variance. Errors in the measurement of the variance of the 'tracking error' are due essentially to its low frequency component; the variance in the variance has approximately one-ninth the value that it would have if both the high and low frequency components contributed.

The misadjustment M_s may be computed with the assistance of relations 16 and 17. In the data-repeating case

$$M_s = \frac{1}{9}\left(\frac{1}{2\Gamma_s}\right) \tag{23}$$

where Γ_s is the adaptation time constant of S_1, or the number of samplings of the worker's mean square error per time constant of adaptation of S_1. When S_2 continuously samples the worker's mean square error, the number of samplings is equivalent to half the number of periods of the fastest variation components in the statistics of the non-stationary input process.

Conclusion

In this paper an adaptive sampled-data system model that is quasi statistically linear and makes use of performance feedback for self-optimization has been described and evaluated.

Analytical results derived from the model show that misadjustments are inversely proportional to adaptation rates. Adaptation taking place at different 'administrative' levels within the same system may be treated analytically as separate phenomena because of the great disparities in the averaging times. Misadjustments developed at the various levels add.

Closed loop adaptation that makes use of performance feedback permits direct, automatic system synthesis. It has the advantage of being usable where no analytic synthesis procedure exists. More general types of performance feedback were used by Clark and Farley[7] and by Rosenblatt[8] in their computer simulations of adaptive 'nerve nets'. Interconnections among 'neurons' were strengthened or weakened depending upon the success or failure of experimental applications of these nets. A class of adaptive switching circuits was studied by Mattson[9] that is structurally similar to the adaptive sampled-data systems. Again, searching for the 'best' within the space of possibilities was done by means of performance feedback.

It is interesting to note that many natural adaptation processes seem to be based on similar principles. The psychologist trains animals to give desired responses to certain stimuli by 'rewarding' and 'punishing'. The palaeontologist on the other hand studies the effects of various earth and sea environments over millions of years on the structures of living animals. Performance feedback is akin to what evolutionists call 'natural selection', except that in nature, structural perturbations seem to be random in frequency and amplitude.

It can be said that the most important and the most general concept studied in this paper by means of a simple adaptive-system model is that of performance feedback.

References

1 LINVILL, W. K., SITTLER, R. W. and WIDROW, B. *Pulse-Data Systems*. Course 6.54 Notes, Dept. of Electrical Engineering, Mass. Inst. of Technology, 1959

2 RAGAZZINI, F. R. and FRANKLIN, G. F. *Sampled-Data Control Systems*. 1958. New York; McGraw-Hill

3 MASON, S. J. Feedback theory. 1. Some properties of signal flow graphs. *Tech. Rep.* 153, *Res. Lab. of Electronics, Mass. Inst. of Technology*, February 2, 1953. Feedback theory—Further properties of signal flow graphs. *Tech. Rep.* 303, *Res. Lab. of Electronics, Mass. Inst. of Technology*, July 20, 1955

4 WIENER, N. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*. 1949. New York; Wiley

5 WIDROW, B. Adaptive sampled-data systems—A statistical theory of adaptation. *Wescon Conv. Rec., Inst. Radio Engrs, N.Y.* Pt 4 (1959)

6 DELONG, D. F. Analysis of an adaptive sampled-data system. *M.S. Thesis*, Dept. of Electrical Engineering, Mass. Inst. of Technology, January 1959

7 FARLEY, B. G. and CLARK, W. A. Simulation of self-organizing systems by digital computer. *Trans. Inst. Radio Engrs Professional Group of Information Theory* Sept. (1954)

8 ROSENBLATT, F. The perceptron: A theory of statistical separability in cognitive systems. *Cornell Aeronaut. Lab. Rep. No. VG–1196–G–1* Buffalo, New York, January 1958

9 MATTSON, R. A. A self-organizing logical system. 1959 *Eastern Joint Computer Conf. Conv. Rec. Inst. Radio Engrs, N.Y.*

Summary

BP

An adaptive sampled-data system model consisting of an adjustable 'worker' and a 'supervisor' is described and evaluated. 'Performance feedback' is employed to achieve automatic system synthesis.

Iterative gradient methods used in the adjustment of the worker's impulse response are represented by linear sampled feedback models. Small sample size 'noise' causes 'misadjustment' approximately equal to the reciprocal of twice the adaptation time constant multiplied by the number of interacting adjustments (verified by simulation).

For non-stationary inputs, measurement noise misadjustment equals twice that due to process-tracking lag when the adaptation time constant is optimized. Automatic optimization could be accomplished by adaptation on a higher level (supervisor's supervisor). This can be treated analytically as a separate phenomenon, and causes an additive misadjustment component.

Performance feedback makes possible adaptive systems that can cope with non-stationary inputs and can adjust to partial system failure.

Reprinted from

AUTOMATIC AND REMOTE CONTROL

Proceedings of the
First International Congress of the
International Federation of Automatic Control

Moscow, 1960

Published by
BUTTERWORTHS
88 KINGSWAY, LONDON, W.C.2