COMMUNICATION SYSTEMS RESEARCH GROUP

THE UNIVERSITY OF TEXAS

ELECTRICAL ENGINEERING DEPARTMENT

AUSTIN

507-53486

COMMUNICATIONS SYSTEMS RESEARCH GROUP
DEPARTMENT OF ELECTRICAL ENGINEERING
THE UNIVERSITY OF TEXAS

OPTIMAL ADAPTIVE FILTER REALIZATIONS FOR SAMPLED STOCHASTIC PROCESSES WITH AN UNKNOWN PARAMETER

by

C. G. Hilborn, Jr. and D. G. Lainiotis

July, 1967

CSRG Technical Report No. 67-3

Prepared for*

National Aeronautics and Space Administration
Manned Spacecraft Center
Under Grant NGR-44-012-066

*This work was also partially supported by the Joint Services Electronics Program under Grant AF-AFOSR-766-66.

ABSTRACT

In this report techniques are derived for realizing Bayes optimal learning systems in analogue delay-feedback form for filtering a sampled stochastic process in the presence of an unknown or random parameter when the optimal filter for known parameter value can be realized. The unknown parameter may take on a bounded but continuous range of real values, and be either an unknown and random constant or a Markov random process. Examples of the techniques are given for both constant and time-varying parameters.

CONTENTS

FIGURES

OPTIMAL ADAPTIVE FILTER REALIZATIONS FOR SAMPLED STOCHASTIC PROCESSES WITH AN UNKNOWN PARAMETER*

I. INTRODUCTION

In this paper, the problem is treated of realizing the Bayes optimal adaptive filter for a sampled stochastic process, where the relevant probability densities are not completely known. Specifically the functional forms are known and would be fully specified by knowledge of a single parameter, such that given the parameter, an optimal (conditional mean) filter could be constructed. Such problems can arise, for example, in the observation and control of systems subject to modeling uncertainty or random parameter variations or in communication or sounding over random or partially unknown channels.

* This work was also partially supported by the Joint Services Electronics Program under Grant AF-AFOSR-766-66.

Magill [1] investigated the Bayes optimal adaptive estimation of a sampled Gauss-Markov process with a constant unknown parameter vector, restricted to take on a known finite set of possible values. The resulting estimation system was shown to be realizable as a linear combination of linear-recursive (i.e., Kalman [2] filters (one for each parameter value), whose outputs are each weighted by the corresponding a posteriori probability of the parameter value, given all data.

If the finite parameter space is actually an approximation to a continuous range of values, the complexity and storage requirements for digital computer implementation increase with the fineness of the quantization. The purpose of this paper is to show how in some situations, the optimal adaptive filter can be directly implemented by simple analogue processing even though the parameter space is continuous. This is done for processes where only a single (scalar) parameter is unknown, and may be either a constant or a realization of a Markov sequence, satisfying a first-order difference equation.

II. BAYES OPTIMAL ADAPTIVE FILTER

Let $X_k \triangleq X(t_k)$, $k = 1, 2, \ldots$ be a sequence of state or signal (column) vectors which are to be sequentially estimated from statistically related observation vectors $Z_k \triangleq Z(t_k)$, $k = 1, 2, \ldots$. Let λ_k denote the collection of data $\{Z_1, Z_2, \ldots Z_k\}$. If $\overline{X}_k$ is an estimate of X_k, the optimal estimate $\hat{X}_k$ is defined as that $\overline{X}_k$ which minimizes the quadratic performance index

$$\varsigma_k = E\{(X_k - \overline{X}_k)' Q (X_k - \overline{X}_k) | \lambda_k\}, \tag{1}$$

where $E\{\ \mid \lambda_k\}$ is the conditional expectation operation, given λ_k, Q is any positive definite symmetric matrix, and prime denotes transpose. It is well known that the optimal estimate $\hat{X}_k$ in this generalized least mean square error sense is the conditional mean $E\{X_k \mid \lambda_k\}$:

$$\hat{X}_k = \int X_k \, p(X_k \mid \lambda_k) \, dX_k. \tag{2}$$

If the density $p(X_k \mid \lambda_k)$ is not exactly known, but actually depends on some parameter θ, i.e., $p(X_k \mid \theta, \lambda_k)$ is known, then $p(X_k \mid \lambda_k)$ may be found by writing

$$p(X_k \mid \lambda_k) = \int p(X_k \mid \theta, \lambda_k) \, p(\theta \mid \lambda_k) \, d\theta, \tag{3}$$

where to find the conditional density $p(\theta \mid \lambda_k)$ of θ, all prior knowledge of θ, even if it is in some cases not "random", is expressed as a probability density $p(\theta)$ (Bayes estimation). Assuming the interchangeability of order of integration, (2) can now be rewritten as

$$\hat{X}_k = \int \hat{X}_k(\theta) \, p(\theta \mid \lambda_k) \, d\theta, \tag{4}$$

where

$$\begin{aligned} \hat{X}_k(\theta) &= \int X_k \, p(X_k \mid \theta, \lambda_k) \, dX_k \\ &= E\{X_k \mid \theta, \lambda_k\}. \end{aligned} \tag{5}$$

That is, since $\hat{X}_k(\theta)$ is the optimal estimate of X_k given θ, $\hat{X}_k$ is obtained by averaging the output of a filter designed for a known parameter over the conditional or "learning" distribution $p(\theta \mid \lambda_k)$ of the parameter. These facts are well known - in principle.

If, for example, the processes given θ are normal, $\hat{X}_i(\theta)$ is linear even though the overall estimate $\hat{X}_k$ is nonlinear. Cases where $\hat{X}_k(\theta)$ is nonlinear, but easily realizable should not, however, be ruled out.

In the following, the symbols α and β_k are used in place of θ for constant and time-varying parameters, respectively. Even though the constant parameter is a special case of the time-varying parameter, the two cases will be treated separately because of the greater importance and applicability of the constant unknown parameter case.

III. CONSTANT UNKNOWN PARAMETER

In the following, it is assumed that (i) the conditional filter $\hat{X}_k(\alpha)$ can be constructed so that its output can be varied as a function of α, (ii) the value of the conditional prediction density $p(Z_k|\alpha, \lambda_{k-1})$ can similarly be generated as a function of α (λ_k is, in a given realization, fixed) and (iii) all a priori knowledge of α is expressed by a density function $p(\alpha)$ defined over some finite interval $[\alpha_1, \alpha_2]$.

With these assumptions, $p(\alpha|\lambda_k)$ can be written in the recursive form

$$p(\alpha|\lambda_k) = \frac{p(Z_k|\alpha, \lambda_{k-1})\, p(\alpha|\lambda_{k-1})}{\int_{\alpha_1}^{\alpha_2} p(Z_k|\alpha, \lambda_{k-1})\, p(\alpha|\lambda_{k-1})\, d\alpha}, \tag{6}$$

where $p(\alpha|\lambda_0) \triangleq p(\alpha)$. Substituting (6) into (4) with α replacing θ gives

$$X_k = \frac{\int_{\alpha_1}^{\alpha_2} \hat{X}_k(\alpha)\, p(Z_k|\alpha, \lambda_{k-1})\, p(\alpha|\lambda_{k-1})\, d\alpha}{\int_{\alpha_1}^{\alpha_2} p(Z_k|\alpha, \lambda_{k-1})\, p(\alpha|\lambda_{k-1})\, d\alpha} \tag{7}$$

If some portion (say all) of the (constant) sample interval $T = t_{k+1} - t_k$ is used for processing, then equations (6) and (7) suggest the real-time implementation of Figure 1, where α is periodically (with period T) "swept" linearly from α_1 to α_2, and the input Z_k is assumed to be constant over the sample interval. The symbols $\times$ and $\div$ denote zero-memory multiplier and divider devices, with numerator and denominators of the dividers indicated by "n" and "d" respectively. The samplers are operated at the end of each sweep cycle, and the integrators are then reset to zero for the next cycle.

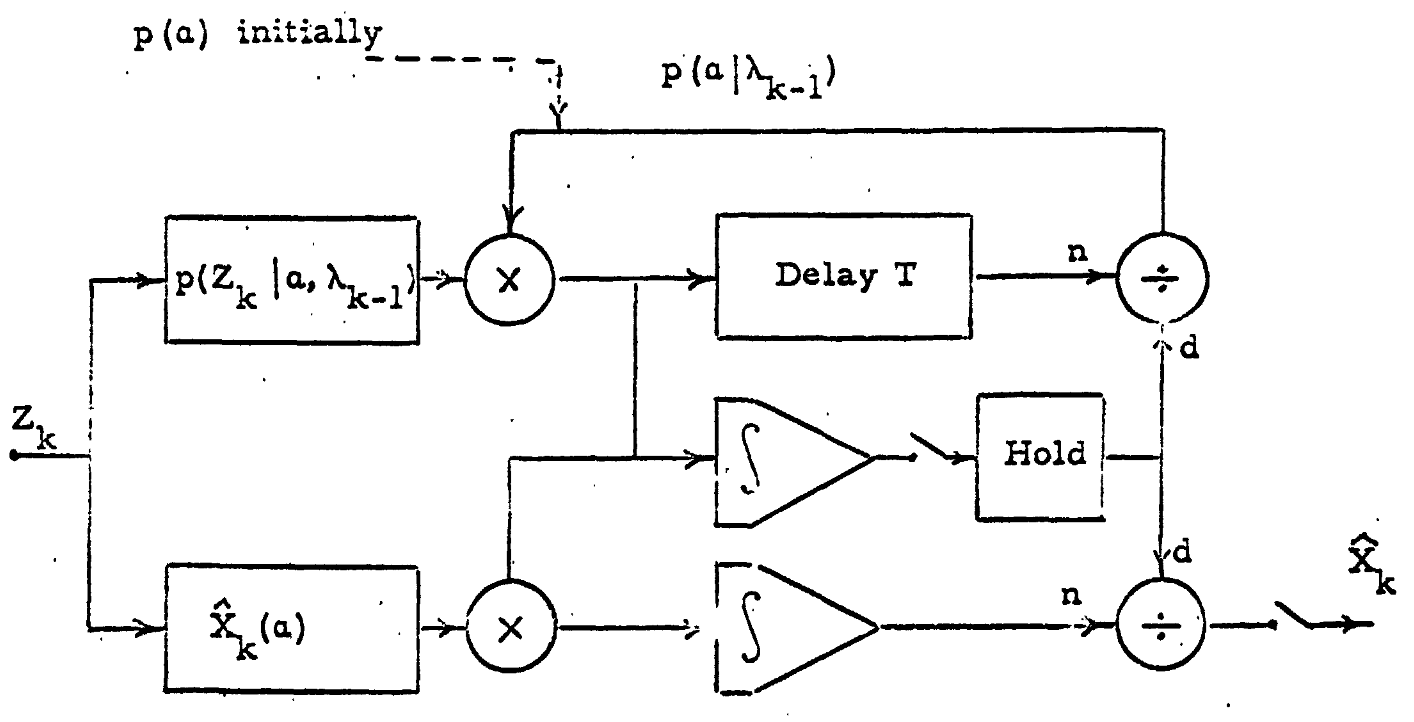

Figure 1. Optimal Adaptive Filter, Constant Parameter Case.

The memory of the learning part of this adaptive system is provided by the delay line with the changing shape of the pulse representing $p(\alpha|\lambda_{k-1})$ characterizing the learning of α. When convergence is possible, the pulse $p(\alpha|\lambda_{k-1})$ ideally narrows down to an impulse "function" $\delta(\alpha-\alpha_o)$, where α_o is the actual value of α. Practically, the bandwidth of the delay line will limit the "sharpness" of the pulse. The bandwidth of the system should therefore be wide enough to provide adequate "resolution" for the problem. It should be noted that the bandwidth of the delay line is roughly equivalent to the fineness of quantization of α needed to use the solution of Magill [1].

Finally, it is well known that from the martingale nature of conditional probability sequences, $p(\alpha|\lambda_k)$ will converge to $\delta(\alpha-\alpha_o)$ with probability one whenever (i) $p(\alpha) > 0$ in a neighborhood of α_o, and (ii) there exists any sequence of statistics $\varphi_1(\lambda_1)$, $\varphi_2(\lambda_2)$, ... converging to α_o with probability one [3]. Then, in this case $\hat{X}_k \rightarrow \hat{X}_k(\alpha_o)$ with probability one, so that the system operates in the limit with exact knowledge of the parameter α.

Example 1: Suppose X_k, $k=1, 2, \ldots$ is a sequence of independent, zero mean, gaussian signals (maximum entropy signals) transmitted over a channel with unknown gain α, where $0 < \alpha_1 \leq \alpha \leq \alpha_2 < \infty$. Let independent, zero mean, gaussian noise be added to each received signal. That is

$$Z_k = \alpha X_k + N_k \quad \text{(all scalar)}.$$

If σ_x^2 and σ_N^2 are the signal and noise powers or variances, then $\hat{X}_k(\alpha)$ is linear in Z_k and given by

$$\hat{X}(\alpha) = \frac{\alpha\sigma_x^2}{\sigma_N^2 + \alpha^2\sigma_x^2} Z_k.$$

The conditional density $p(Z_k|\alpha, \lambda_{k-1})$ is gaussian, and does not depend on λ_{k-1}:

$$p(Z_k|\alpha, \lambda_{k-1}) = \frac{\exp -\frac{1}{2}\{Z_k^2 / (\sigma_N^2 + \alpha^2\sigma_x^2)\}}{\sqrt{2\pi}\,(\sigma^2 + \alpha^2\sigma_x^2)^{\frac{1}{2}}}.$$

Finally, if all that is known about α is $\alpha_1 < \alpha < \alpha_2$, then

$$p(\alpha) = \frac{1}{\alpha_2 - \alpha_1}, \quad \alpha_1 \leq \alpha \leq \alpha_2,$$

that is, the constant $1/(\alpha_2 - \alpha_1)$ is switched in for $p(\alpha|\lambda_o)$ for $t_1 \leq t \leq t_2$. The part of the system, shown only as the blocks generating $p(Z_k|\alpha, \lambda_{k-1})$ and $\hat{X}_k(\alpha)$(in Figure 1), is shown for this example in Figure 2.

Example 2: Now suppose $Z_k = mX_k + N_k$, where N_k is an independent normal sequence with variance σ_N^2 and X_k is no longer "white" but is a Markov process satisfying the difference equation

$$X_{k+1} = \varphi X_k + U_k,$$

where U_k is a zero mean, variance σ_U^2 normal sequence. (Example 1 is a special case of this example when $\varphi = 0$.) Suppose that any one of the quantities $\{\varphi, \sigma_U^2, m, \sigma_N^2\}$ is the unknown parameter. A recursive linear filter [2] then can be built for $\hat{X}_k(\alpha)$ with the unknown element left as a variable and swept

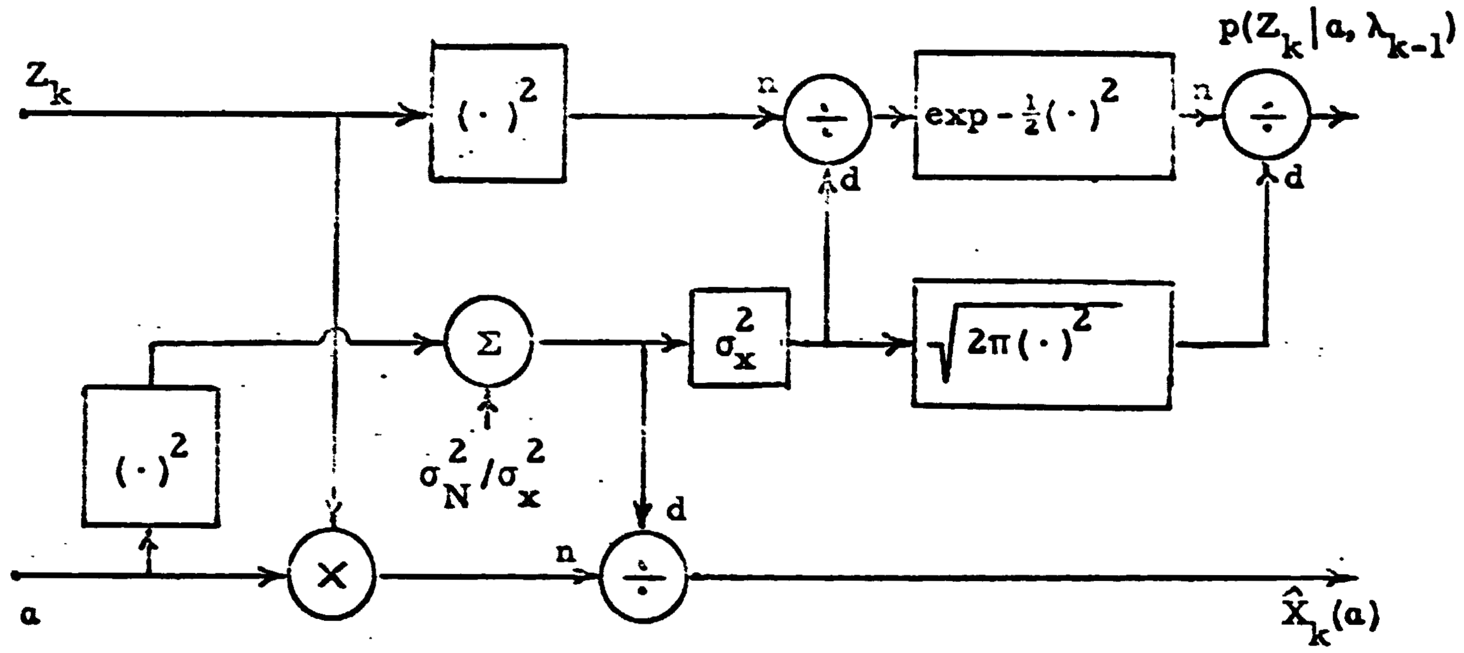

Figure 2. Conditional estimate/prediction density function generator for Example 1.

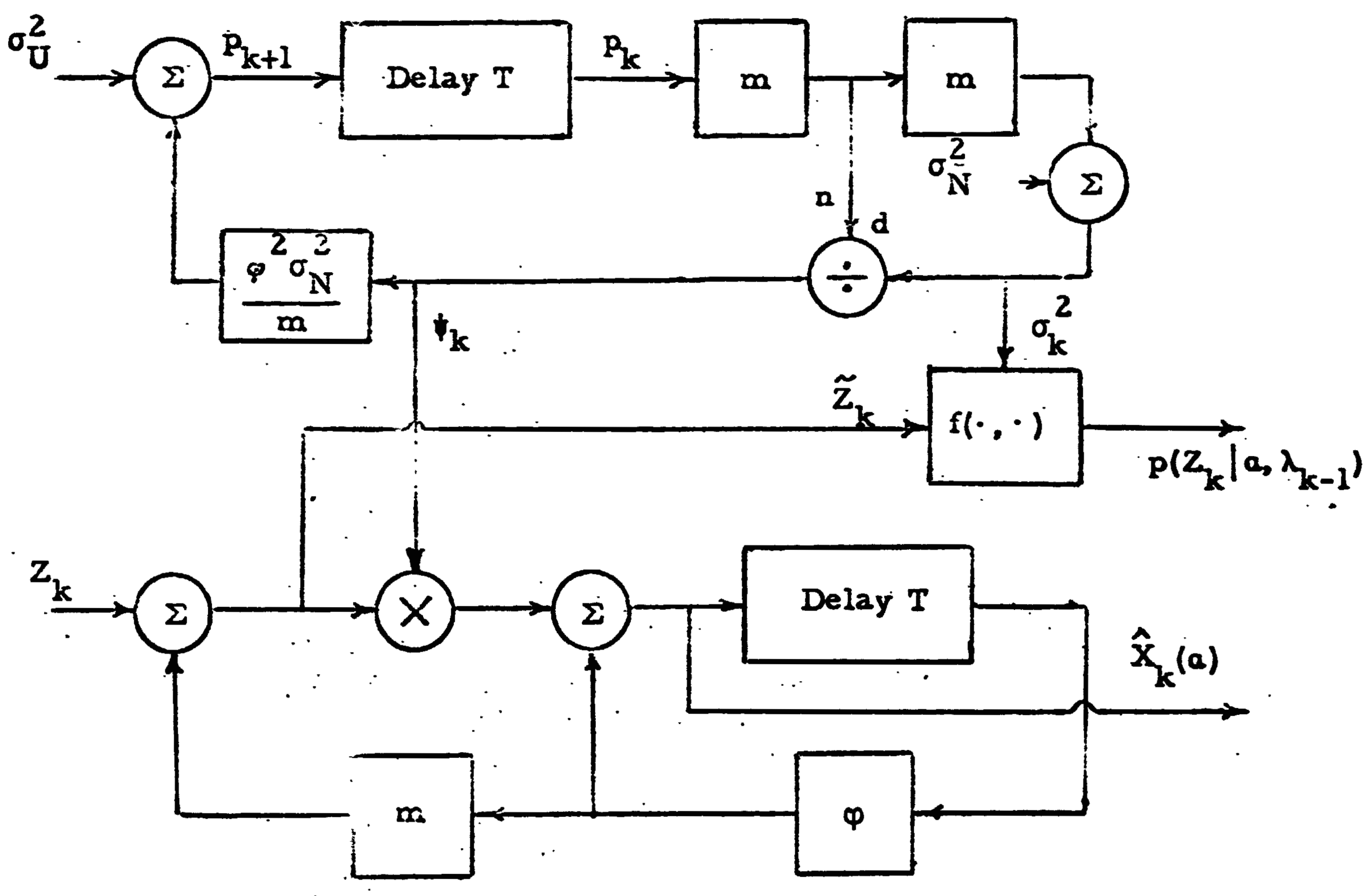

Figure 3. Conditional estimate/prediction density function generator for Example 2.

linearly over the sample period. Suppressing α in the notation, we have

$$\hat{X}_k = \varphi \hat{X}_{k-1} + \psi_k (Z_k - m\varphi \hat{X}_{k-1}),$$

where

$$\psi_k = \frac{m P_k}{m^2 P_k + \sigma_N^2}, \quad P_{k+1} = \frac{\varphi^2 \sigma_N^2}{m} \psi_k + \sigma_U^2,$$

where $P_k = E\{(X_k - \varphi \hat{X}_{k-1})^2\}$. Let $\sigma_k^2 = \mathrm{Var}\{Z_k | \lambda_{k-1}\}$, and let $\tilde{Z}_k = Z_k - m\varphi \hat{X}_{k-1}$, then

$$p(Z_k | \alpha, \lambda_{k-1}) = \frac{e^{-\frac{1}{2}\frac{\tilde{Z}_k^2}{\sigma_k^2}}}{\sqrt{2\pi\sigma_k^2}}$$

$$\triangleq f(\sigma_k^2, \tilde{Z}_k)$$

where

$$\sigma_k^2 = \sigma_N^2 + m^2 P_k .$$

The system for generating $p(Z_k | \alpha, \lambda_{k-1})$ and $\hat{X}_k(\alpha)$ can then be constructed as shown in Figure 3. Then the unknown element of $\{\sigma_U^2, \varphi, \sigma_N^2, m\}$ is swept periodically from α_1 to α_2.

IV. TIME-VARYING PARAMETER

Now suppose the unknown parameter can vary from sample to sample. Specifically, β_k is assumed to be a Markov process satisfying the difference equation

$$\beta_k = \varphi \beta_{k-1} + V_k, \qquad (8)$$

where V_k is an independent sequence with known density $p_v(V_k)$ which is zero outside a finite interval. If this finite range assumption is not strictly true, such an approximation is normally possible with probability arbitrarily close to one. To make the process stable, i.e. bound the range, $|\varphi|$ must be less than one. For definiteness, let φ be non-negative. In addition to assumptions (i - iii) for α, β_k is assumed to affect Z_k "causally", i.e.

$$p(\lambda_k | \beta_j, \ldots, \beta_1) = p(\lambda_k | \beta_k, \ldots, \beta), \quad j \geq k. \qquad (9)$$

The conditional density $p(\beta_k|\lambda_k)$ can then be written recursively as follows:

$$p(\beta_k|\lambda_k) = \frac{p(Z_k|\beta_k, \lambda_{k-1}) p(\beta_k|\lambda_{k-1})}{\int_{b_1}^{b_2} p(Z_k|\beta_k, \lambda_{k-1}) p(\beta_k|\lambda_{k-1}) d\beta_k}, \qquad (10)$$

where

$$p(\beta_k|\lambda_{k-1}) = \int_{b_1}^{b_2} p(\beta_k|\beta_{k-1}) p(\beta_{k-1}|\lambda_{k-1}) d\beta_{k-1}. \qquad (11)$$

and $[b_1, b_2]$ is the range of β_k.

But from the model (8)

$$p(\beta_k|\beta_{k-1}) = p_v(\beta_k - \varphi\beta_{k-1}). \qquad (12)$$

Thus

$$p(\beta_k|\lambda_{k-1}) = \int_{b_1}^{b_2} p_v(\beta_k - \varphi\beta_{k-1})\, p(\beta_{k-1}|\lambda_{k-1})\, d\beta_{k-1}. \tag{13}$$

Equation (13) can be implemented by feeding $p(\beta_{k-1}|\lambda_{k-1})$ through a gain of $\frac{1}{\varphi}$, then a linear time scale compression of φ, and finally through a linear filter with impulse response $p_v(\tau)$. To accomplish the storage for $p(\beta_{k-1}|\lambda_{k-1})$ and make these systems realizable, a delay of $(1-\varphi)$ is added to the time scale change and a delay of φT to the response p_v.

The time scale change can be thought of as a sliding tap delay line of total length $T(1-\varphi)/\varphi$ seconds, with the tap periodically sliding linearly from $T(1-\varphi)/\varphi$ to 0 over the sample intervals, its output being sampled from $k(1-\varphi)T$ to kT. Such a time-varying delay can be realized by a tapped delay line with appropriate switching and smoothing, or in various non-mechanical equivalents. (See [4, 5].)

The total system implementing (13) plus delays is shown in Figure 3 and denoted by the time-varying impulse response $h(\tau, t)$.

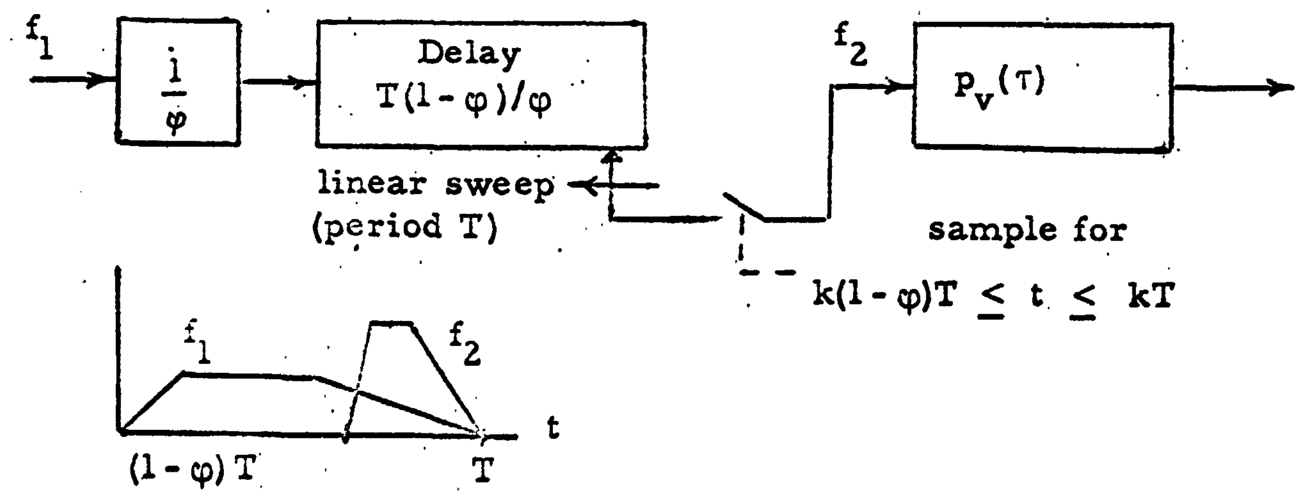

Figure 4. Time-varying Linear System $h(\tau, t)$

Then, using (13) and the fact that

$$\hat{X}_k = \frac{\int \hat{X}_k(\beta_k)\, p(Z_k|\beta_k, \lambda_{k-1})\, p(\beta_k|\lambda_{k-1})\, d\beta_k}{\int p(Z_k|\beta_k, \lambda_{k-1})\, p(\beta_k|\lambda_{k-1})}, \tag{14}$$

the optimal adaptive system can be implemented as shown in Figure 5. The memory of the system is implicit in $h(\tau, t)$. It is interesting to note, that if $V_k \equiv 0$, i.e. $p_v(V_k) = \delta(V_k)$, and $\varphi \to 1$, $h(\tau, t) \to \delta(\tau - t)$, that is a delay of length T and Figure 4 reduces to Figure 1. At the other extreme, if $\varphi = 0$ (β_k an independent sequence) $h(\tau, t) = p_v(T - t)$.

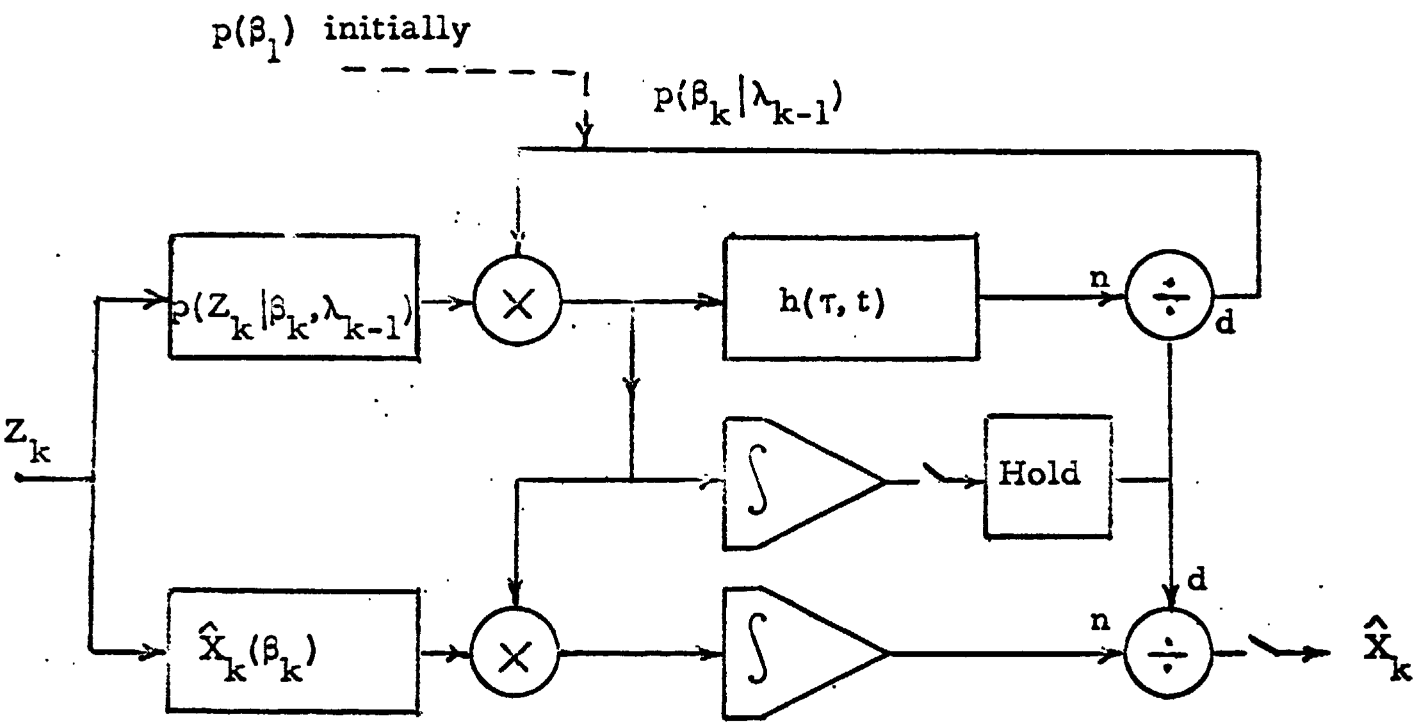

Figure 5. Optimal Adaptive Filter, time-varying parameter case.

Example 3: Suppose $Z_k = \beta_k X_k + N_k$, where X_k and N_k are white normal sequences as in Example 1, and β_k satisfies (8). β_k is a time-varying measurement gain. Suppose the β_k process is in steady state and β_k is normally distributed with mean 0 and variance 1. Then the range of β_k can be very accurately approximated by truncating $p(\beta_k)$ to the [0, 10] interval. Then

$$\hat{X}_k(\beta_k) = \frac{\beta_k \sigma_x^2}{\sigma_y^2 + \beta_k^2 \sigma_x^2} Z_k,$$

and

$$p(Z_k | \beta_k, \lambda_{k-1}) = \frac{e^{-\frac{1}{2} Z_k^2/(\sigma_y^2 + \beta_k^2 \sigma_x^2)}}{\sqrt{2\pi(\sigma_y^2 + \beta_k^2 \sigma_x^2)}}$$

and

$$p(V_k) = \frac{e^{-\frac{1}{2}\frac{[V_k - (1-\varphi)5]^2}{1-\varphi^2}}}{\sqrt{2\pi(1-\varphi^2)}}, \quad -5(1-\varphi^2) \le V_k - (1-\varphi) \le 5(1-\varphi^2).$$

The system for generating $\hat{X}_k(\beta_k)$ and $p(Z_k|\lambda_{k-1})$ is identical to that shown in Figure 2, and the outputs of this system replace the $\hat{X}_k(\beta_k)$ and $p(Z_k|\alpha, \lambda_{k-1})$ blocks of Figure 5. A (truncated) gaussian pulse for $p(\beta_1)$ must be generated for the first cycle.

V. CONCLUSIONS

Real time analogue techniques have been utilized to realize Bayes optimal adaptive filters in the presence of an unknown and random parameter. The systems were found for both constant and time-varying parameters, and shown to be realizable in delay-feedback form whenever the parameter conditional filters and observation densities are realizable as functions of the unknown parameter.

REFERENCES

1. D. T. Magill, "Optimal Adaptive Estimation of Sampled Stochastic Processes," IEEE Trans. on Automatic Control, vol. AC-10 No. 4, pp. 434-439, Oct. 1965. Complete report is Stanford Electronics Labs., Stanford Calif., Report SEL-63-143 (TR No. 6302-3), Dec. 1963.

2. R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," J. Basic Engr., Series D, vol. 82, pp. 35-45, March 1960.

3. M. Aoki, "On Some Convergence Questions in Bayesian Optimization Problems," IEEE Trans. on Automatic Control, Apr. 1965, pp. 180-182.

4. A. Gersho, "Characterization of Time-Varying Linear Systems," Ph. D. dissertation, Cornell University, Ithaca, New York, June 1963, (Report EE 559, Feb. 1963).

5. __________, "Pulse Stretching with Time-Varying Networks," Proceedings of the IEEE (correspondence), Nov. 1966.