

Gaussian Sum Approximations in Nonlinear Filtering and Control*

D. L. ALSPACH

*Department of Applied Mechanics and Engineering Sciences,
University of California, San Diego,
La Jolla, California 92037*

ABSTRACT

This paper presents a survey of some recent results in the fields of nonlinear filtering and stochastic control resulting from the use of Gaussian sum approximations to certain density functions. When trying to estimate and control the state of a nonlinear stochastic dynamic system from noisy measurement data, the *a posteriori* density of the state conditioned on the measurement data contains all of the available information about the system state. If these densities were available, it should be possible to develop "optimal" estimation and control policies. The well-known Bayesian recursion relations describe the evolution of this *a posteriori* density in terms of given *a priori* densities, the system dynamics and measurement equations and the available data. Unfortunately, it is seldom possible to solve these closed forms for the required *a posteriori* densities. Over the last few years it has been shown that when the *a priori* densities are approximated by a sum of Gaussian densities with positive weighting coefficients that linear non-Gaussian problems can be solved analytically and that new natural solutions to nonlinear problems are available. Recently this method has been extended to dual deterministic optimal control problems.

1. INTRODUCTION

It has been known for some time that in order to model many physical, chemical, biological, and other systems adequately one must use mathematical models which take into account the stochastic nature of the systems under consideration. Such systems can seldom be observed in such a way that the total system behavior or state is known precisely at all times. Rather, the system is observed through measurement of variables which yield incomplete information and which contain unknown or random errors. In order to study the systems behavior or to control that behavior, it is generally necessary to estimate, from the contaminated measurement data and the stochastic dynamical system equations, the state of the system and certain unknown parameters of the system. In addition it is frequently necessary to take some type of control action based on these estimates.

The problem of finding an optimal control policy based on some performance criteria involves, except for certain simple cases, a complex combination of the

*This work was partially supported by the National Science Foundation under Grant NSF-GK-27869.

estimation problem and the control performance criterion. However, the estimation problem can be considered separately from the control problem and, in this sense, is the simpler of the two.

The problem of estimating the state of a nonlinear stochastic system from noisy measurement data has been the subject of considerable research interest during the past few years, but, although a great deal has been published on the subject, the basic objective of obtaining a solution that can be implemented in a straightforward manner for specific applications has not been satisfactorily realized. This is manifested by the fact that the Kalman filter equations [1, 2], which are valid only for linear, Gaussian systems, continue to be widely used for nonlinear, non-Gaussian systems. Of course, continued application has resulted in the development of *ad hoc* techniques [3-5] that have improved the performance of the Kalman filter and which give it some of the characteristics of nonlinear filters.

However, as more and more emphasis is placed on optimum system performance, the available estimation and control policies for inherently stochastic systems must be improved. As the system models improve and computers with greater capability become available, it will be necessary to develop techniques which utilize all pertinent existing information to obtain such policies. The ability to utilize all of the information contained in non-Gaussian densities and to propagate this through nonlinear plants will be essential for this purpose.

Central to the nonlinear estimation and stochastic control problems is the determination of the probability density function of the state conditioned on the available measurement data. This density contains all information which is theoretically available concerning the state of the system. Thus, if this *a posteriori* density function were known, an estimate of the state of any performance criterion could be determined. Unfortunately, although the manner in which the density evolves with time and additional measurement data can be described in terms of difference (or differential) equations, these relations are generally very difficult to solve, either in closed-form or numerically, so that it is usually impossible to determine the *a posteriori* density for specific applications. Because of this difficulty, it is natural to investigate the possibility of approximating the density with some tractable form. One approach to the solution of this problem and some results which have been obtained are discussed below.

2. RECURSIVE NONLINEAR FILTERING: THE BAYESIAN APPROACH

This paper briefly summarizes the results obtained in the field of nonlinear filtering and stochastic control by use of the Gaussian sum approximation to certain density functions. This work is all based on the Bayesian [6] approach and is therefore limited by the assumption of the existence of the probability

density functions used below. Except for the case of linear systems with Gaussian noise but possibly non-Gaussian *a priori* densities discussed in Sec. 3, all work on nonlinear filtering based on this method has been concerned with the discrete time model given below.

Suppose that the state vector $\mathbf{x}_k$ of the system evolves according to the nonlinear stochastic difference equation:

$$\mathbf{x}_{k+1} = \mathbf{f}_{k+1}(\mathbf{x}_k) + \mathbf{w}_k \quad k = 0, 1, \dots, N-1, \quad (1)$$

where $\mathbf{x}_k$ is the n -dimensional state of the system at the time t_k ; $\mathbf{w}_k$ is the n -dimensional plant random noise vector and where the initial state $\mathbf{x}_0$ is a random variable with a known density function $p(\mathbf{x}_0)$. In addition the white noise sequence $\mathbf{w}^k \triangleq (\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_k)$ has a known probability density function $p(\mathbf{w}^k) = p(\mathbf{w}_0)p(\mathbf{w}_1)\dots p(\mathbf{w}_k)$, which is independent of the initial state.

The behavior of the plant is observed imperfectly through measurement quantities $\mathbf{z}_k$ that are related in a prescribed fashion to the state but which contain random errors.

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k; \quad k = 0, 1, \dots, N, \quad (2)$$

where $\mathbf{z}_k$ is the m -dimensional measurement vector of known measurement data at the time t_k , and $\mathbf{v}_k$ is the m -dimensional measurement noise vector contaminating the data at t_k . The noise sequence $\mathbf{v}^k \triangleq (\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_k)$ is assumed to be a white-noise sequence with a known distribution $p(\mathbf{v}^k)$ and to be independent of the initial state and all plant noise. The restriction to additive noise terms is not necessary but reduces the complexity of two of the densities required for the Bayesian recursion relations given below.

Since the state vector $\mathbf{x}_k$ is a random variable, in the probabilistic context of this discussion, the *a posteriori* density $p(\mathbf{x}_k|\mathbf{z}^k)$ provides the most complete description possible of $\mathbf{x}_k$. This density is determined recursively from the Bayesian recursion relations given below. The filtering density is given first and the prediction density is next.

$$p(\mathbf{x}_k|\mathbf{z}^k) = c_k p(\mathbf{x}_k|\mathbf{z}^{k-1})p(\mathbf{z}_k|\mathbf{x}_k), \quad (3)$$

$$p(\mathbf{x}_k|\mathbf{z}^{k-1}) = \int_{R^n} p(\mathbf{x}_{k-1}|\mathbf{z}^{k-1})p(\mathbf{x}_k|\mathbf{x}_{k-1})d\mathbf{x}_{k-1}, \quad (4)$$

where the normalizing constant c_k is given by

$$1/c_k \triangleq p(\mathbf{z}_k|\mathbf{z}^{k-1}) = \int_{R^n} p(\mathbf{x}_k|\mathbf{z}^{k-1})p(\mathbf{z}_k|\mathbf{x}_k)d\mathbf{x}_k, \quad (5)$$

and the initial condition for (3) is

$$p(\mathbf{x}_0|\mathbf{z}^{-1}) \triangleq p(\mathbf{x}_0).$$

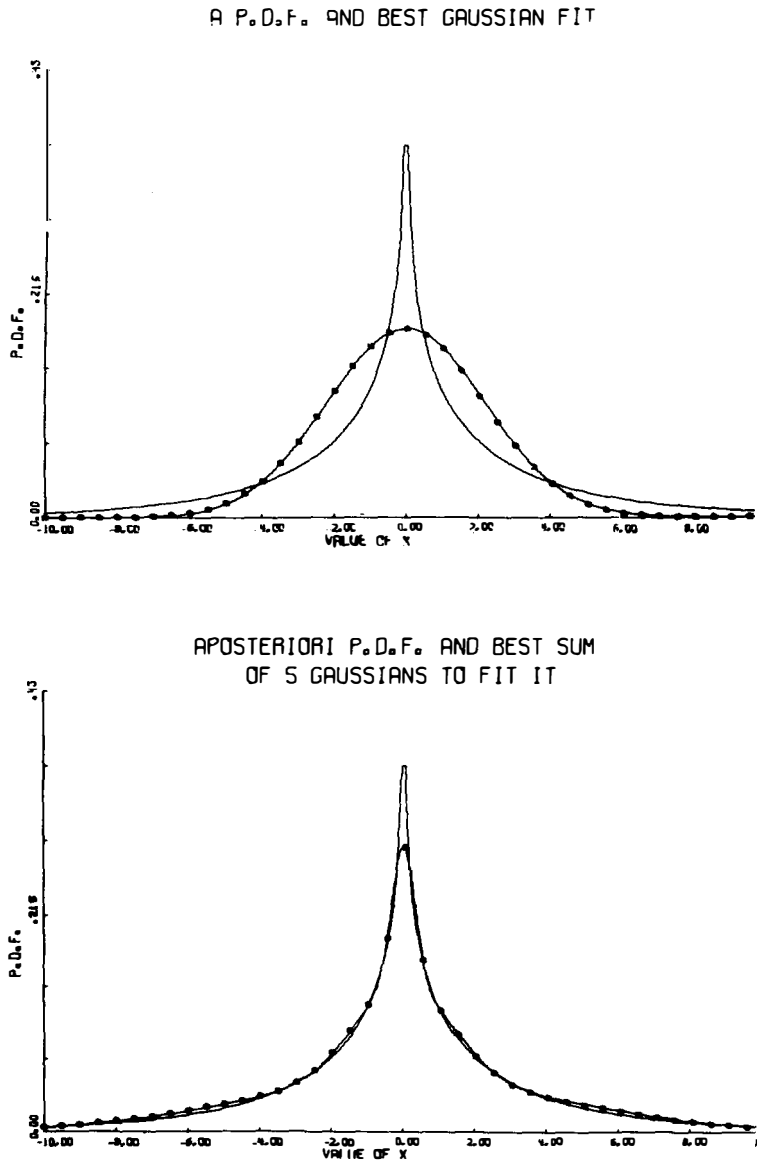
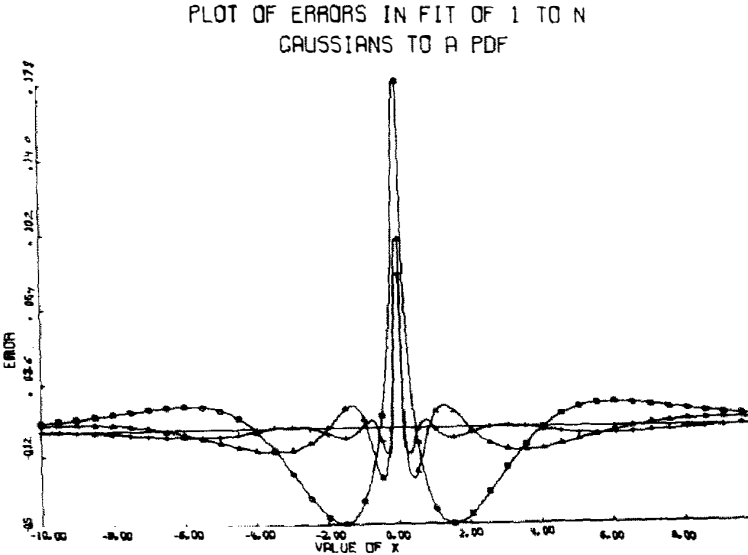
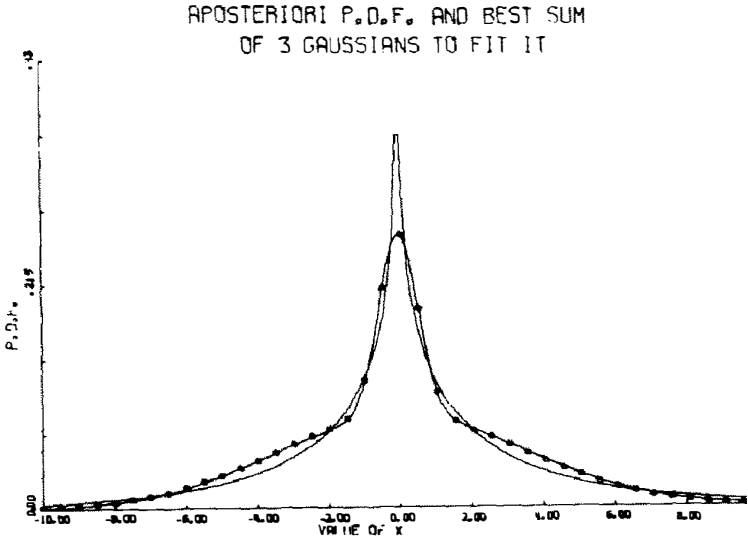


Fig. 1. Above and opposite. 1, 3, and 5 L^2 search fit approximations to product density.



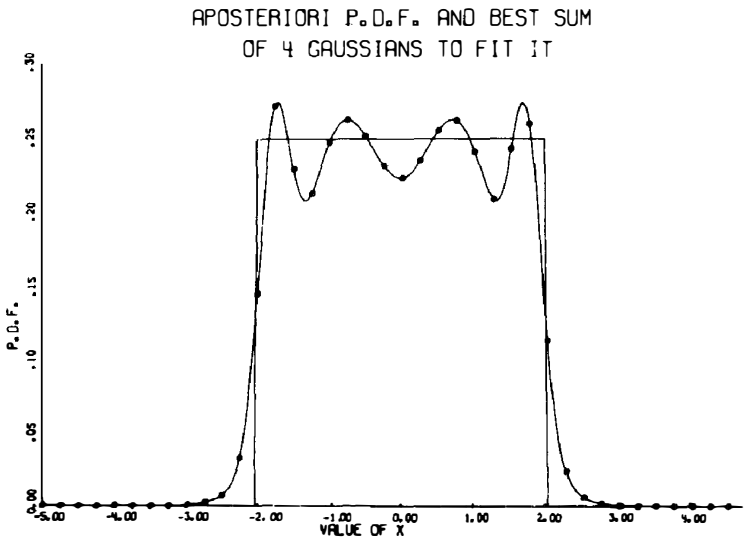
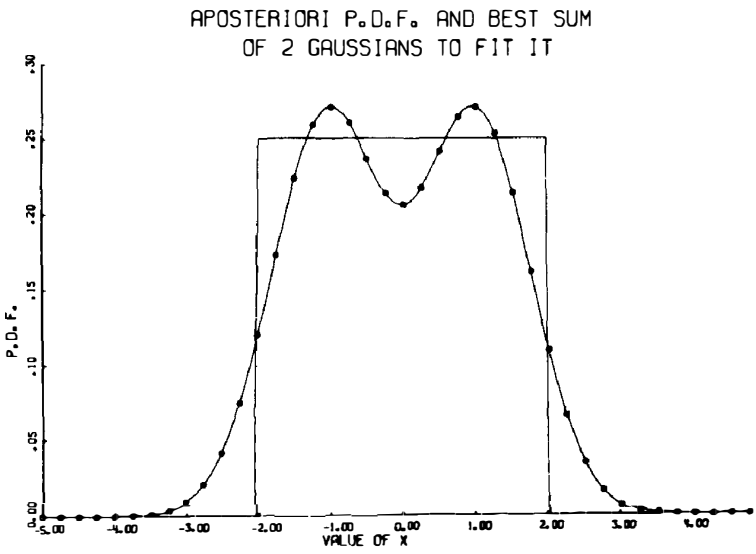
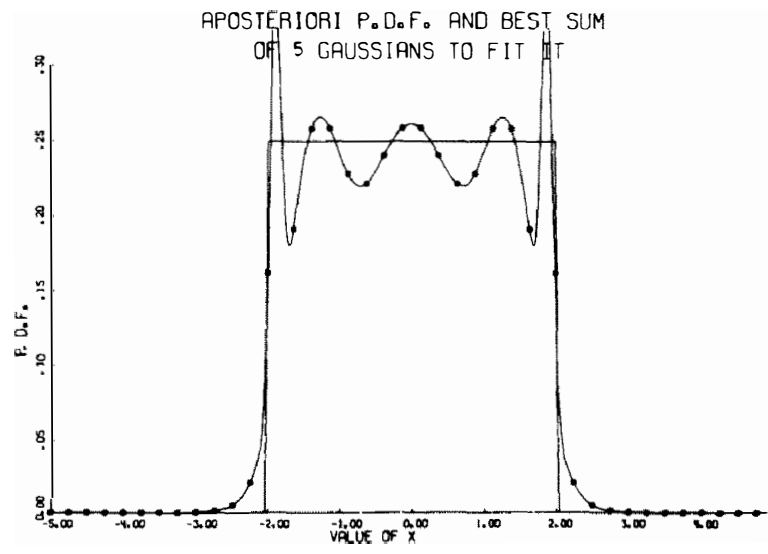
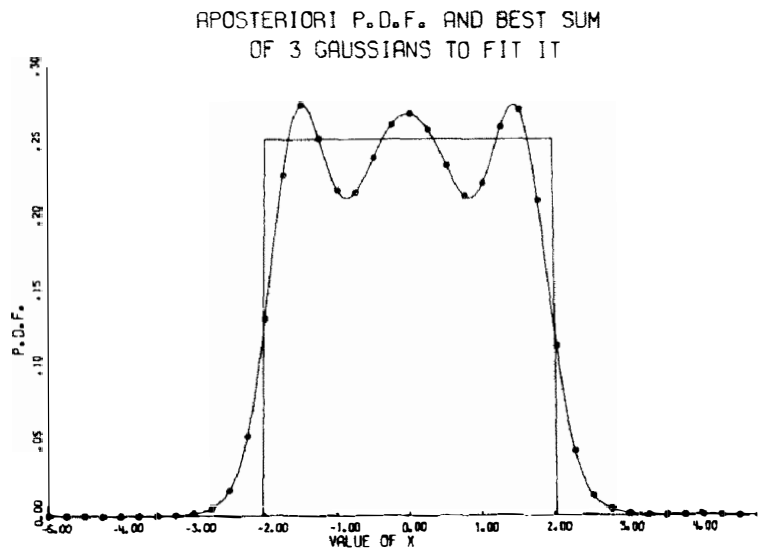


Fig. 2. Above and opposite. Gaussian sum approximation to uniform density L^2 search fit 2 to 5 terms.



The densities $p(\mathbf{z}_k|\mathbf{x}_k)$ and $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ are determined from (1) and (2), and the *a priori* densities for $\mathbf{v}_k$ and $\mathbf{w}_{k-1}$.

The nonlinear filtering problem is solved when the density $p(\mathbf{x}_k|\mathbf{z}_k)$ can be obtained for all k , however, except when the system (1)–(2) is linear and the *a priori* distributions are Gaussian, it is generally impossible to determine $p(\mathbf{x}_k|\mathbf{z}^k)$ in a closed form using (3)–(5). In the linear Gaussian case, the relations describing the conditional mean and covariance of $p(\mathbf{x}_k|\mathbf{z}^k)$ are the well-known Kalman filter equations [2]. The difficulties associated with the explicit determination of the *a posteriori* density have led to the development of approximate procedures for estimating the state of nonlinear stochastic systems. The most commonly used approximation involves the assumptions that the *a priori* distributions are Gaussian and that the nonlinear system can be linearized relative to latest estimate yielding the “extended” Kalman filter which has seen widespread application to nonlinear systems [2, 5, 7].

The extended Kalman filter has performed well in many applications but there are numerous examples in which unsatisfactory results have been obtained. This has spurred the development of other procedures. Most of these either implicitly or explicitly, retain the assumption that $p(\mathbf{x}_k|\mathbf{z}^k)$ is Gaussian and essentially provide a means for modifying the mean and covariance of the density. However, the Gaussian assumption greatly reduces the generality or richness of possible *a posteriori* densities throwing out all possibility of more realistic multimodal densities.

For this reason, the Gaussian sum method discussed below was developed allowing the approximation of general *a posteriori* densities for nonlinear non-Gaussian systems.

3. GAUSSIAN SUM APPROXIMATION TO A PRIORI DENSITIES

Since it is not generally possible to solve Eqs. (3–5), the Gaussian sum approximation is introduced to allow at least approximate solution of these equations. To maintain the advantages of the Gaussian densities, it is proposed that certain *a priori* densities $p(\cdot)$ be represented or approximated by a linear combination of Gaussian densities.

$$p_A(\mathbf{x}) = \sum_{i=1}^N \alpha_i N(\mathbf{x} - \mathbf{a}_i, B_i), \quad (6)$$

$$N(\mathbf{x}, B) = \exp \{0.5 \mathbf{x}^t B^{-1} \mathbf{x}\} / (2\pi)^{n/2} |B|^{1/2}, \quad (7)$$

where the weighting factors α_i have the constraints:

$$\sum_{i=1}^N \alpha_i = 1 \quad \alpha_i \geq 0. \quad (8)$$

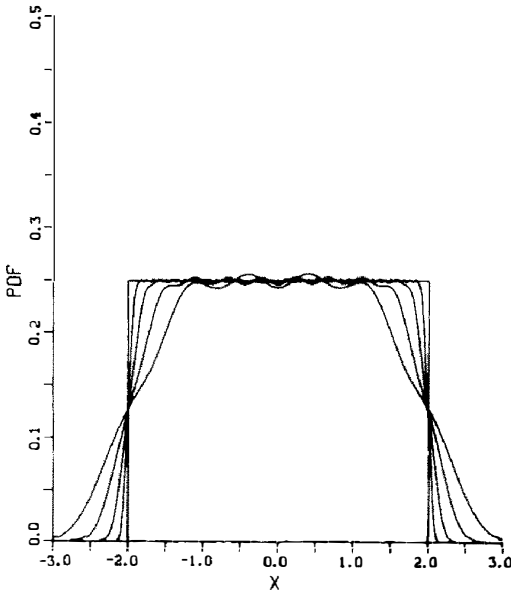


Fig. 3. Best theorem fit, 6, 10, 20, and 49 term approximations.

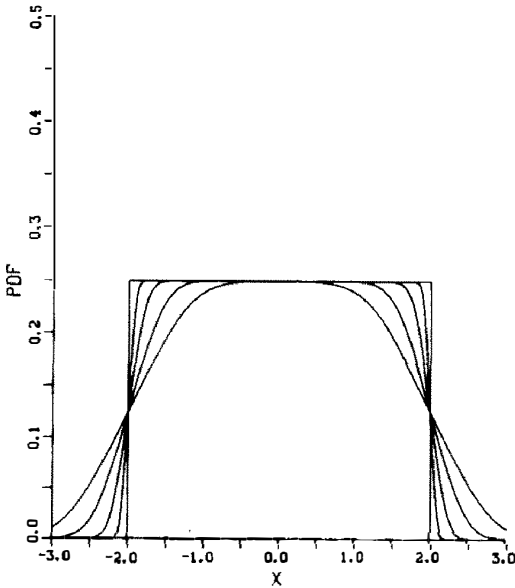


Fig. 4. Smoothed theorem fit, 6, 10, 20, and 49 term approximations.

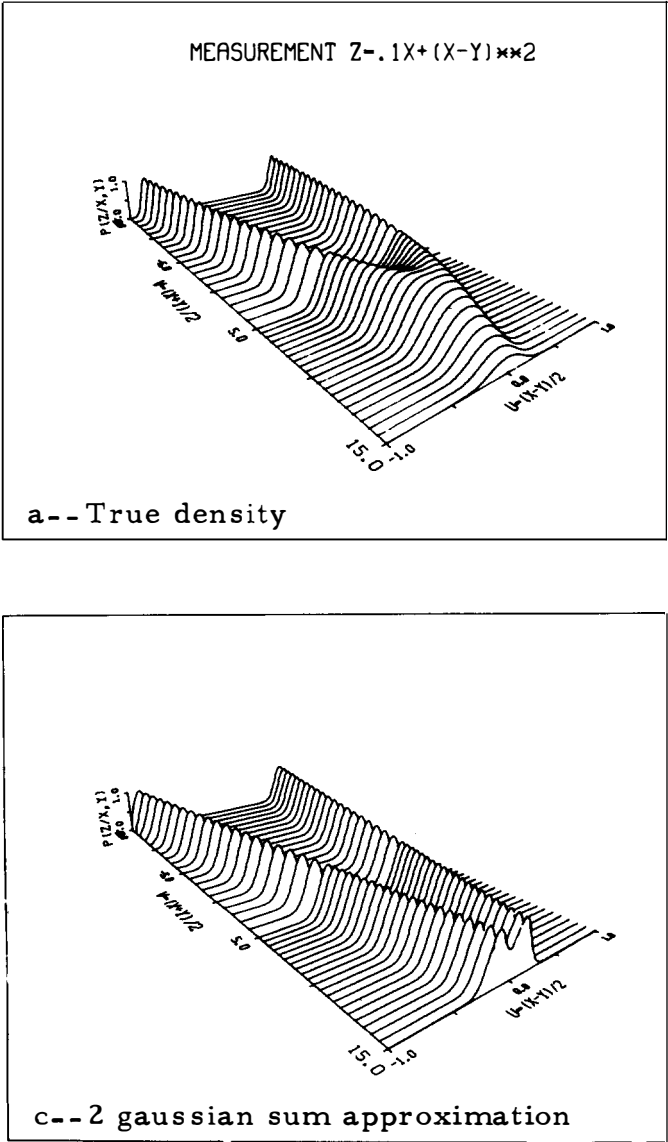
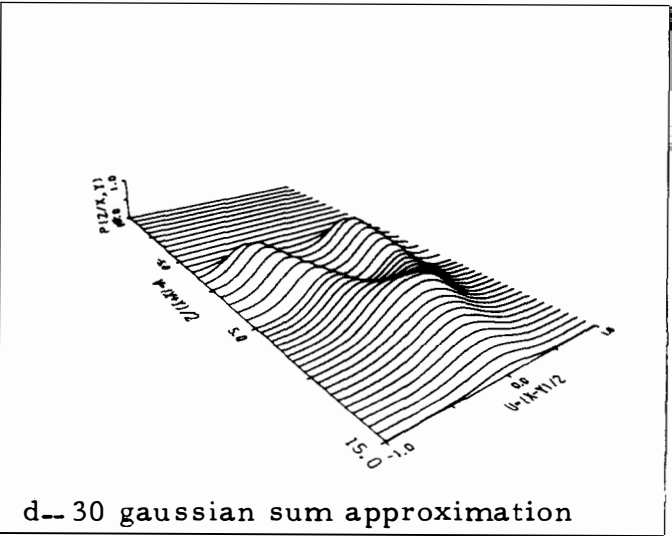
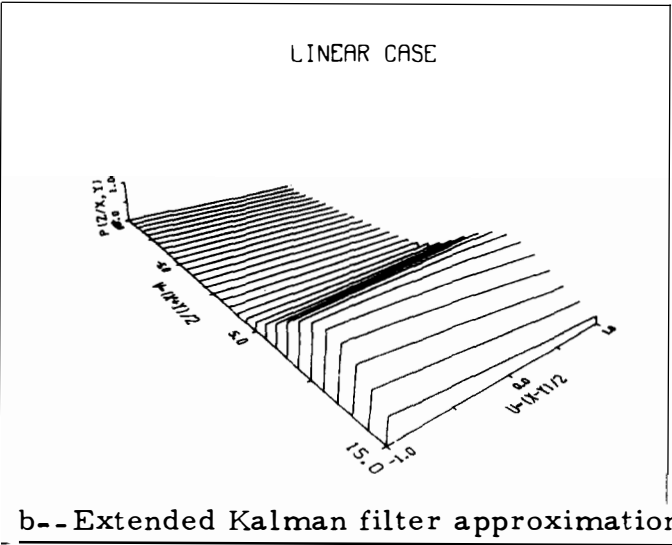


Fig. 5. Above and opposite. Measurement density function and approximation.



These two properties insure that the approximation is always positive and integrates to a value of one. This approximation is thus a true probability density function in its own right, and, when this representation is used in linear problem, the Bayesian recursion relations can be solved exactly in closed form.

There are two questions that must be answered for this approximation to be useful. The first is whether or not there exist approximations of the form for given *a priori* densities, and secondly, can we obtain such approximations in a reasonable manner.

The first question can be answered in the affirmative. In fact, if $g(\cdot)$ is an integrable function on a bounded or unbounded interval J , then given an $\epsilon > 0$, one can find a function $p_A(\cdot)$ of the form of Eq. (6) such that:

$$\int_{R^n} |g(\mathbf{x}) - p_A(\mathbf{x})| d\mathbf{x} < \epsilon.$$

This can be proven by appealing to Wiener's theorem on approximation [8], by direct proof [9] or by considering delta functions of the positive type [10]. Thus this class of density functions is rich enough to approximate all density functions of engineering interest arbitrarily well.

The second problem of obtaining specific approximations to a given *a priori* density has been solved in two ways. First, one may reduce the problem to a deterministic optimization problem and perform a numerical search for the parameters of the Gaussian sum [9, 10]. The results of such approximation are shown in Figs. 1 and 2. A fair amount of computer time is required to solve for each approximate density. A second method based on the concept of delta functions of positive type has proven more useful in practice. It has been found that very adequate approximations can be obtained with no numerical search simply by first spacing the Gaussian means uniformly over the region of state space of interest, and secondly by choosing the variance to give a smooth fit. In this approach, the weighting factors, α_i , are chosen proportional to the magnitude of the value of the function to be approximated at the appropriate points. Examples of Gaussian sum approximation to several densities which were obtained by this method are shown in Figs. 3-5. Other examples are contained in the references.

4. LINEAR SYSTEMS WITH GAUSSIAN NOISE BUT NON-GAUSSIAN INITIAL STATE

When the system and measurement equations are linear and all the *a priori* densities are Gaussian, all the *a posteriori* densities are also Gaussian with the mean and covariance given by the Kalman filter. Thus the densities are reproducing [11] in the sense that each density has the same form. When all of

the above assumptions hold except that the initial density $p(\mathbf{x}_0)$ is a Gaussian sum instead of Gaussian, the *a posteriori* densities are still truly reproducing, since all the *a posteriori* densities are also Gaussian sums with the same number of terms as the initial density. This use of Gaussian sums for nonlinear filtering was mentioned by Aoki as early as 1965 [12] and has been considered by Cameron [13], Lainiotis [15, 18], and Lo [17], as well as Alspach and Sorenson [9, 10]. This case is particularly interesting since after the initial approximation of the density $p(\mathbf{x}_0)$ by a Gaussian sum density, one obtains the exact solution of the Bayesian recursion relations. All of the terms of the sum propagate forward in time by a set of independent Kalman filter equations, while all nonlinear dependence on the measurement data is contained in the weighting coefficients. These are not modified in propagations through the prediction step of Eq. 4 and given that

$$p(\mathbf{x}_k | \mathbf{z}^{k-1}) = \sum_{i=1}^N \alpha'_i N(\mathbf{x}_k - \hat{\mathbf{x}}'_{k_i}, p'_{k_i})$$

one can easily show that

$$p(\mathbf{x}_k | \mathbf{z}^k) = \sum_{i=1}^N \alpha_i N(\mathbf{x}_k - \hat{\mathbf{x}}_{k_i}, p_{k_i}),$$

where each $\hat{\mathbf{x}}_{k_i}$ and p_{k_i} are related to $\hat{\mathbf{x}}'_{k_i}$ and p'_{k_i} by the usual Kalman filter equations [2] and α_i is given by

$$\alpha_i = \frac{\alpha'_i N(\mathbf{z}_k - H_k \hat{\mathbf{x}}_{k_i}, \Omega_{k_i})}{\sum_{i=1}^N \alpha'_i N(\mathbf{z}_k - H_k \hat{\mathbf{x}}'_{k_i}, \Omega_{k_i})},$$

$$\Omega_{k_i} = R_k + H_k^T p'_{k_i} H_k.$$

One should note that obtaining a best fit to a given *a priori* density $p(\mathbf{x}_0)$ can be performed off-line before the actual filtering begins.

This particular case can be useful in its own right when one has precise knowledge of the statistic of the initial state or when one can transform a more complicated case to this form, but it is more useful as an example of “exact” nonlinear filtering and as an indication of how to attack more general problems.

5. LINEAR SYSTEMS WITH NON-GAUSSIAN NOISE AND INITIAL STATE

If the linear measurements [2] are contaminated by non-Gaussian noise or the linear system [1] is driven by non-Gaussian noise, one must extend the method discussed in the last section. If one simply replaces the non-Gaussian

plant and measurement noise densities by Gaussian densities of the same mean and covariance, a great deal of information may be lost and nonmeaningful estimates obtained. However, if good Gaussian sum approximations of $p(v_k)$ and $p(w_k)$ are found, one may still solve the Bayesian recursion relations (3)–(5) analytically [9, 10]. The resulting recursion relations are similar to those discussed in Sec. 4 involving a number of Kalman filters evolving in parallel with the nonlinear dependence on the measurements contained in the weighting coefficients. The *a posteriori* densities resulting can be made as accurate as desired by choice of the initial approximations [9].

The recursion relations are still reproducing in the sense that all the *a posteriori* densities are still of the Gaussian sum form. However, to be truly reproducing the number of terms in the sum at each stage should remain constant. This is only the case when only the initial density is non-Gaussian. When either the plant or measurement noise density is a Gaussian sum with more than one term, the number of terms in the exactly calculated *a posteriori* densities grows from stage to stage. It has been shown [10] that there are two factors which can lead to the restriction of this growth by the use of natural approximations. The first arising from the form of the weighting coefficients of each term allows one to drop negligible terms. These are terms in which the weighting coefficient has become exponentially small.

THEOREM 1. *If $p(x)$ and $p_A(x)$ are Gaussian sum density functions:*

$$p(x) = \sum_{i=1}^m \alpha_i N(x - a_i, B_i),$$

$$p_A(x) = \frac{\sum_{i=1}^{m-l} \alpha_i N(x - a_i, B_i)}{\sum_{i=1}^{m-l} \alpha_i}, \quad 0 < l \leq m-1,$$

where

$$\alpha_i < \delta, i > m-l.$$

Then

$$\int_{R^n} |p(x) - p_A(x)| \leq 2 \sum_{i=m-l+1}^m \alpha_i \leq 2l\delta$$

The second approximation used results from the tendency of many terms to converge to the same point and the fact that two identical terms can be combined by adding their weighting coefficients. When the terms are only approximately the same, a bound on the L^1 error introduced by both approximations has been found.

THEOREM 2. If $p(\mathbf{x})$ and $p_A(\mathbf{x})$ are Gaussian sum density functions of order m and $m - 1$, respectively, where

$$p(\mathbf{x}) = \sum_{i=1}^m \alpha_i N(\mathbf{x} - \mathbf{a}_i, B_i),$$

with $B_1 = B_2 = B$,

$$p_A(\mathbf{x}) = \sum_{i=3}^m \alpha_i N(\mathbf{x} - \mathbf{a}_i, B_i) + N(\mathbf{x} - \bar{\mathbf{x}}, B)(\alpha_1 + \alpha_2),$$

where

$$\bar{\mathbf{x}} = \frac{\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2}{\alpha_1 + \alpha_2}, (\mathbf{a}_1 - \mathbf{a}_2)^T B^{-1} (\mathbf{a}_1 - \mathbf{a}_2) \leq \epsilon^2,$$

then

$$\int_{R^n} |p(\mathbf{x}) - p_A(\mathbf{x})| d\mathbf{x} \leq \frac{2\alpha_1 \alpha_2}{\alpha_1 + \alpha_2} \epsilon.$$

The proofs of both theorems are straightforward and follow directly from properties of multidimensional Gaussian distributions.

Lastly, experimental evidence shows that specific choices of number of terms in the *a priori* densities and specific thresholds for the two approximations lead to a bound on the number of terms required to describe the *a posteriori* densities for any stage.

6. NONLINEAR SYSTEM

The most frequently used approach for the estimation of the state of a nonlinear system involves the linearization of the system relative to a best estimated value for the state and the subsequent application of the Kalman filter equations. This approach results in approximations of two types. First, the nonlinear physical system is replaced by a linear model, and, second, the non-Gaussian *a posteriori* density function $p(\mathbf{x}_k | \mathbf{z}^k)$ is approximated by a Gaussian density. One other method that may be used on a small class of problems is to transform the nonlinear problem to a linear problem. When this can be done such that both the measurements and system are linear, one usually finds that the Gaussian noise has been transformed to non-Gaussian noise. Under these special conditions the method of Sec. 5 can be applied.

In Ref. [16], a filtering procedure is defined which is based on the utilization of the Gaussian sum representation of the *a posteriori* density function in conjunction with the linearization procedure that has proven so effective in

Kalman filter applications. This algorithm greatly extends the range of validity of the linearization approximation of the nonlinear system and measurement and eliminates the assumption that $p(x_k | z^k)$ is Gaussian. A general outline of the simplest form of this algorithm follows. Suppose that a stage k , the one stage predicted density $p(x_k | z^{k-1})$, has been calculated and is of the form of a Gaussian sum in Eq. (6). The filtering density $p(x_k | z^k)$ is given from Eq. 3 by

$$p(x_k | z^k) = C_k \sum_{i=1}^N \alpha_i N(x_k - a_i, B_i) p_{v_k}(z_k - h_k(x_k)).$$

As in the extended Kalman filter, linearize h_k relative to a_i so that $p_{v_k}(z_k - h_k(x_k))$ can be approximated by a Gaussian-like function in the region around each a_i . In the choice of Gaussian sum approximations to the *a priori* densities, one makes the region which each term in the sum must cover "small," and then the region in which this linearization must be valid is correspondingly small (obviously much smaller than the region over which it would have to be valid if $p(x_k | z^{k-1})$ were represented by a single Gaussian). Once this is done, the *a posteriori* density $p(x_k | z^k)$ is in the form of a Gaussian sum, and the prediction step of the Bayesian recursion relations Eq. (4) can be performed in the same manner by linearizing $p(x_{k+1} | x_k)$ about each term in the Gaussian sum for $p(x_k | z^k)$. Note that in this algorithm as in Sec. 4, the number of terms in the Gaussian sum does not grow from stage to stage.

Under certain conditions, such as non-Gaussian plant or measurement noise, it becomes necessary to approximate $p(z_k | x_k)$ and $p(x_{k+1} | x_k)$ directly by Gaussian sums. More details about these cases are given in Refs. [9] and [19]; a number of examples of both algorithms are given in the references. These show the possible improvement of nonlinear filters over linearized Gaussian filters.

7. CONTROL THEORY

Often estimating the state of a system is only a stepping stone to allow the control of that system. This would be the case when the system (1) contains a control variable and one is to choose the control at each stage based on all previous measurement data to minimize the expected value of a cost function of the control and the state.

When the system is linear, it is well known that with a number of additional assumptions the principle of optimality can be used to prove that the optimal stochastic control problem can be solved by the solution of two independent problems. First, an optimal estimate of the system state conditioned on all available information is computed, and, secondly, a deterministic optimal con-

trol is calculated for the problem formed by replacing all *a priori* random variables by their mean values. Under the rather fortuitous circumstances when the optimal control is composed of the optimal deterministic control with the state at time k replaced by the best estimate of the state at time k the separation theorem or separation principle is said to be valid. When this theorem does not apply, the control resulting from this combination of an optimal estimate with a deterministic control function is still often used and then the resulting control called the certainty equivalence control [20, 21]. It is partially the separation theorem and partially the extensive use of this suboptimal control that can account for the large interest by control engineers in the inherently secondary problem of estimation theory. In Ref. [22], the proof of the separation theorem is extended using the Gaussian sum method to linear systems with quadratic cost but non-Gaussian initial state and non-Gaussian plant and measurement noise.

In the more common situation where the separation theorem does not apply, the more general dynamic programming approach has not been used partly because of the curse of dimensionality but mostly because it is not generally possible to obtain the required *a posteriori* densities in the analytical form required by the algorithm. However, by using the approximate densities discussed above preliminary investigation into the advantages to be gained from use of optimal and suboptimal dual controllers has been made [23]. This preliminary investigation indicates that there are potentially large benefits in improved performance of stochastic systems from the use of controls different from the certainty equivalence controls and which retain the dual nature of the truly optimal control. The use of approximate *a posteriori* densities of one form or another will be essential for this to be accomplished.

The Gaussian sum approach to deterministic control problems using the duality between optimal control for deterministic systems and estimation for stochastic systems has been investigated by Dajani and Campion [24], following the similar approach [9, 10, 13–18] to the dual linear-Gaussian estimation problem with non-Gaussian initial state-vector. The optimal control problem for nonlinear systems with nonquadratic final-cost function was solved by approximation of the Lyapunov function by a Gaussian sum.

8. DISCUSSION AND POTENTIAL FUTURE RESEARCH

There are several advantages and several limitations to the general approach considered here. First, all of this work is based on the Bayesian approach which must rest on the existence of the appropriate probability density functions. This may pose a conceptual problem, but will hardly affect engineering problems of filtering and control. The nonlinear filters considered here require the

existence of at least a first derivative of the plant and measurement nonlinearities at the points of linearization. For a function that is nowhere differentiable different techniques must be considered. The solution to the Bayesian recursion relations can be implemented in a straightforward manner by the methods considered here for both linear and nonlinear, Gaussian and non-Gaussian problems to give approximate *a posteriori* density functions. These density functions can be made arbitrarily close to the true densities, however the computational requirements can easily become excessive as the number of terms grows. An interesting alternative implementation of this filter is studied in Ref. [25] which considers the use of parallel processing computers to exploit the intrinsic parallel nature of these filters. This is a natural choice, since all the filters discussed here consist of a number of modulated Kalman or extended Kalman filters operating in parallel.

One advantage of this class of filter over some others that have been advanced is the apparent ease with which symmetries in the state transition function and measurement nonlinearities can be utilized to reduce the computational complexity of a problem. A second advantage is the ease with which moments can be obtained once the density approximation is found [10]. An area of future work is the extension of the parallel processing work of Ref. [25]. The Gaussian sum filter should be implemented in detail on an existing parallel processing computer such as the ILLIAC IV and simulation results obtained.

There are several other areas requiring more detailed research. First, and most direct, is the application of these methods to a number of additional test problems. Several additional nonlinearities should be investigated as well as problems of higher dimension. Results obtained from these studies should be compared to other possible filters such as the linearized filters [3], the Senne-Bucy digital filters [26], and Spline filters [27]. The classes of problems which require or can justify nonlinear filtering should then be outlined and those classes of problems best suited to each of the recently developed nonlinear filtering methods should be mapped out. Possible combinations of the various methods for optimal results should also be considered. An interesting paper by Center on a generalized least-squares approach to the approximation of conditional density functions has already pointed out many similarities and relationships between the different approaches mentioned above [28]. These relationships mentioned earlier by Lainiotis [29] should be a logical starting place for a study of the relative usefulness of each method.

In addition to the above work, the basic filter algorithms should be extended to include such additional factors as nonadditive noise, state and control dependent noise, unknown plant and measurement covariance matrices, and unknown initial error covariance matrix. Lastly, but perhaps most important, is the continued development and study of stochastic controls which retain the dual nature of the optimal control and can be implemented in a realistic manner.

By using these general methods to obtain stochastic controllers and approximations to optimal nonlinear filters, we will obtain a greater understanding of the nature of the problems, the nature of the truly optimal filter, and the dual nature of the optimal stochastic control. With this greater understanding, it may well be possible to implement an adequate filter or a greatly improved control in a much simpler manner than by the direct use of the methods discussed here or in the references.

REFERENCES

1. R. E. Kalman, A new approach to linear filtering and prediction problems, *J. Basic Eng.* 82D 35-45 (March 1960).
2. H. W. Sorenson, Kalman filtering techniques, *Advances in Control Systems*, Vol. 3, Chap. 5, Academic, New York, 1966.
3. R. P. Wishner, J. A. Tabaczynski, and M. Athans, A comparison of three nonlinear filters, *Automatica* 5 487-496 (1969).
4. H. Cox, Recursive nonlinear filtering, *Proceedings of the National Electronics Conference XXI*, Chicago (1965).
5. C. T. Leondes (Ed.), *Theory and Applications of Kalman Filtering*, NATO AGARDograph 139, No. 629, 7, 058:519. 61:629. 78 (Feb. 1970).
6. Y. C. Ho and R. C. K. Lee, A Bayesian approach to problems in stochastic estimation and control, *Proceedings, Joint Automatic Control Conference* (1964), pp. 382-387.
7. R. S. Bucy, Linear and nonlinear filtering, *Proc. IEEE*, 58, 854-864 (June 1970).
8. N. Wiener, *The Fourier Integral and Certain of its Applications*, Cambridge University Press, Cambridge, 1933.
9. D. L. Alspach, A Bayesian approximation technique for estimation and control of time-discrete stochastic systems, Ph.D. dissertation UCSD, 1970.
10. H. W. Sorenson and D. L. Alspach, Recursive Bayesian estimation using Gaussian sums, *Automatica* 7, No. 4 (July 1971).
11. J. D. Spragins, Reproducing distributions for machine learning, Stanford Elect. Labs, Tech. Report No. 6103-7, 1963.
12. M. Aoki, Optimal Bayesian and min-max control of a class of stochastic and adaptive dynamic systems, *Proceedings IFAC Symposium on Systems Engineering*, Tokyo (1965), pp. 77-84.
13. A. V. Cameron, Control and estimation of linear systems with nongaussian a priori distributions, *6th Allerton Conference* (1968).
14. D. G. Lainiotis, Optimal adaptive estimation: Structure and parameter adaptation, *IEEE Tras. Auto. Control* AC16, 160-170 (1971); also *Proc. IEEE Adaptive Processes Symp.* (1969).
15. D. G. Lainiotis et al., Optimal state-vector estimation for non-Gaussian initial state vector, *IEEE Trans. Autom. Control* AC16, 197-198 (1971).
16. D. G. Lainiotis, Optimal non-linear estimation, *Int. J. Control* 14, 1137-1148 (1971).
17. J. T. H. Lo, Finite-dimensional sensor orbits and optimal nonlinear filtering, *IEEE Trans. Inform. Theory* IT18, No. 5 (September, 1972).
18. S. K. Park and D. G. Lainiotis, Monte-Carlo study of the optimal nonlinear estimator: Linear systems with non-Gaussian initial states, *Int. J. Control* 16, 1029-1040 (1972).

19. D. L. Alspach and H. W. Sorenson, Nonlinear Bayesian estimation using Gaussian sum approximations, *IEEE Trans. Auto. Control* **AC17**, No. 4, 439–448 (August, 1972).
20. S. E. Dreyfus, Some types of optimal control of stochastic systems, *J. Siam Control*, Ser. **A2**, 120–134 (1962).
21. H. A. Simon, Dynamic programming under uncertainty with quadratic criterion function, *Econometrica* **24**, 74–81 (1956).
22. D. L. Alspach and H. W. Sorenson, Stochastic optimal control for linear but non-Gaussian systems, *Int. J. Control* **13**, 1167–1181 (1971).
23. D. L. Alspach, Dual-control based on approximate *a posteriori* density functions, *Proceedings of the 1972 Joint Automatic Control Conference (JACC 72)* and *IEEE Trans. Auto. Control* (October, 1972).
24. M. Z. Dajani and G. Campion, Closed loop control design for nonlinear, nonquadratic systems, *Proc. IEEE Conf. Decision and Control* (December, 1973).
25. Tse Edison, Parallel computation of the conditional mean state estimate for nonlinear systems, *2nd Symp. Nonlinear Est.*, San Diego (1971).
26. R. S. Bucy and K. D. Senne, Digital synthesis of nonlinear filters, *Automatica* **7**, No. 3, 287–298 (May, 1971).
27. R. J. D. de Figueredo and Y. O. Jan, Spline filters, *Proc. 2nd Symp. Nonlinear Estimation*, La Jolla, California (September 13–15, 1971).
28. J. L. Center, Jr., Practical nonlinear filtering of discrete observations by generalized least squares approximation of the conditional probability distribution, *Proc. 2nd Symp. Nonlinear Estimation*, La Jolla, California (September 13–19, 1971).
29. D. G. Lainiotis, Adaptive decomposition of mixtures: a unifying approach, *Proc. Nat. Electron. Conf.* **24**, 104–106 (1968).

Received January, 1974