Automatica, Vol. 7, pp. 465–479. Pergamon Press, 1971. Printed in Great Britain.

Recursive Bayesian Estimation Using Gaussian Sums*

Estimation recurrente bayesienne utilisant les sommes gaussiennes

Rekursive Bayes-Schätzung unter Benutzung Gauss'scher Summen

Рекурсивная байесовская оценка с помощью гауссовских сумм

H. W. SORENSON† and D. L. ALSPACH‡

The approximation of the probability density $p(\mathbf{x}_k|\mathbf{Z}_k)$ of the state of a noisy dynamic system conditioned on available measurement data using a convex combination of gaussian densities is proposed as a practical means for accomplishing non-linear filtering.

Summary—The Bayesian recursion relations which describe the behavior of the *a posteriori* probability density function of the state of a time-discrete stochastic system conditioned on available measurement data cannot generally be solved in closed-form when the system is either non-linear or non-gaussian. In this paper a density approximation involving convex combinations of gaussian density functions is introduced and proposed as a meaningful way of circumventing the difficulties encountered in evaluating these relations and in using the resulting densities to determine specific estimation policies. It is seen that as the number of terms in the gaussian sum increases without bound, the approximation converges uniformly to any density function in a large class. Further, any finite sum is itself a valid density function unlike many other approximations that have been investigated.

The problem of determining the *a posteriori* density and minimum variance estimates for linear systems with non-gaussian noise is treated using the gaussian sum approximation. This problem is considered because it can be dealt with in a relatively straightforward manner using the approximation but still contains most of the difficulties that one encounters in considering non-linear systems since the *a posteriori* density is nongaussian. After discussing the general problem from the point-of-view of applying gaussian sums, a numerical example is presented in which the actual statistics of the *a posteriori* density are compared with the values predicted by the gaussian sum and by the Kalman filter approximations.

1. INTRODUCTION

THE PROBLEM of estimating the state of a non-linear stochastic system from noisy measurement data is considered. This problem has been the subject of considerable research interest during the past few years, and JAZWINSKI [1] gives a thorough discussion of the subject. 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 [2, 3], which are valid only for linear, gaussian systems, continue to be widely used for non-linear, non-gaussian systems. Of course continued application has resulted in the development of *ad hoc* techniques [e.g. Refs. 4, 5] that have improved the performance of the Kalman filter and which give it some of the characteristics of non-linear filters.

Central to the non-linear estimation and stochastic control problems is the determination of the probability density function of the state conditioned on the available measurement data. If this *a posteriori* density function were known, an estimate of the state for 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 differential, or difference, equations [6–8] 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. It is to this approximation problem that this discussion is directed.

1.1. *The general problem*

The approximation that is discussed below is introduced as a means of dealing with the following system and filtering problem. Suppose that the state $\mathbf{x}$ evolves according to

$$\mathbf{x}_{k+1}=\mathbf{f}_k(\mathbf{x}_k, \mathbf{w}_k) \tag{1}$$

* Received 24 August 1970; revised 18 December 1970. The original version of this paper was not presented at any IFAC meeting. It was recommended for publication in revised form by associate editor L. Meier.

† Department of the Aerospace and Mechanical Engineering Sciences, University of California at San Diego, La Jolla, California 92037, U.S.A.

‡ Department of Electrical Engineering, Colorado State University, Fort Collins, Colorada, U.S.A.

and that the behavior of the state is observed imperfectly through the measurement data

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{v}_k). \tag{2}$$

The $\mathbf{w}_k$ and $\mathbf{v}_k$ represent white noise sequences and are assumed to be mutually independent.

The basic problem that is considered is that of estimating the state $\mathbf{x}_k$ from the measurement data* $\mathbf{Z}_k$ for each k, that is, the filtering problem. Generally, one attempts to determine a "best" estimate by choosing the estimate to extremize some performance criterion. For example, the estimate could be selected to minimize the mean-square error. Regardless of the performance criterion, given the *a posteriori* density function $p(\mathbf{x}_k|\mathbf{Z}_k)$, any type of estimate can be determined. Thus, the estimation problem can first be approached as the problem of determining the *a posteriori* density. This is generally referred to as the Bayesian approach [9].

1.2. *The Bayesian approach*

As has been demonstrated by the great interest in the Kalman filter, it is frequently desirable to determine the estimates recursively. That is, an estimate of the current state is up-dated as a function of a previous estimate and the most recent or new measurement data. In the Bayesian case, the *a posteriori* density can be determined recursively according to the following relations.

$$p(\mathbf{x}_k|\mathbf{Z}_k) = \frac{p(\mathbf{x}_k|\mathbf{Z}_{k-1})p(\mathbf{z}_k|\mathbf{x}_k)}{p(\mathbf{z}_k|\mathbf{Z}_{k-1})} \tag{3}$$

$$\eta(\mathbf{x}_k|\mathbf{Z}_{k-1}) = \int p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1})p(\mathbf{x}_k|\mathbf{x}_{k-1})d\mathbf{x}_{k-1} \tag{4}$$

where the normalizing constant $p(\mathbf{z}_k|\mathbf{Z}_{k-1})$ in equation (3) is given by

$$p(\mathbf{z}_k|\mathbf{Z}_{k-1}) = \int p(\mathbf{x}_k|\mathbf{Z}_{k-1})p(\mathbf{z}_k|\mathbf{x}_k)d\mathbf{x}_k. \tag{5}$$

The initial density $p(\mathbf{x}_0|\mathbf{z}_0)$ is given by

$$p(\mathbf{x}_0|\mathbf{z}_0) = \frac{p(\mathbf{z}_0|\mathbf{x}_0)p(\mathbf{x}_0)}{p(\mathbf{z}_0)}. \tag{6}$$

The density $p(\mathbf{z}_k|\mathbf{x}_k)$ in equation (3) is determined by the *a priori* measurement noise density $p(\mathbf{v}_k)$ and the measurement equation (2). Similarly, the $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ in equation (4) is determined by $p(\mathbf{w}_{k-1})$ and equation (1). Knowledge of these densities and $p(\mathbf{x}_0)$ determines the $p(\mathbf{x}_k|\mathbf{Z}_k)$ for all k. However, it is generally impossible to accomplish the integration indicated in equation (4) in closed form so that the density cannot actually be determined for most applications. The principal exception occurs when the plant and measurement equations are linear and the initial state and the noise sequences are gaussian. Then, equations (3–6) can be evaluated and the *a posteriori* density $p(\mathbf{x}_k|\mathbf{Z}_k)$ is gaussian for all k. The mean and covariances for this system are known as the Kalman filter equations.

When the system is non-linear and/or the *a priori* distributions are non-gaussian, two problems are encountered. First the integration in equation (4) cannot be accomplished in closed form and, second the moments are not easily obtained from equation (3). These problems lead to the investigation of density approximations for which the required operations can be accomplished in a straightforward manner. A particularly promising approximation is considered here. Emphasis in this discussion is on the approximation itself rather than the non-linear filtering problem but the latter has been stated because it provides the motivation for considering an approximation. The approximation of probability density functions is discussed in section 2. The use of the gaussian sum approximation for a linear system with non-gaussian *a priori* density functions for the initial state and for the plant and measurement noise sequences is discussed in section 3.

2. APPROXIMATION OF DENSITIES

The problem of approximating density functions in order to facilitate the determination of the *a posteriori* density and its moments has been considered previously [10] using Edgeworth and Gram–Charlier expansions. While this approach has several advantages and its utility has been demonstrated in some applications, it has the distinct disadvantage that when these series are truncated, the resulting density approximation is not positive for all values of the independent variable. Thus, the approximation is not a valid density function itself. To avoid or to at least reduce the negativity of the density approximation, it is sometimes necessary to retain a large number of terms in the series which can make the approximation computationally unattractive. Further, it is well known that the series converges to a given density only under somewhat restrictive conditions. Thus, although the Edgeworth expansion is a convenient choice in many ways, it has proven desirable to seek an approximation which eliminates these difficulties. The *gaussian sum* approximation that is described here exhibits the basic advantages of the Edgeworth expansion but has none of the disadvantages already noted. That is, the approximation is always a valid density function and, further, converges uniformly to any density of practical concern. An approximation of this type was

* The upper case letter $\mathbf{Z}_k$ denotes the set $(\mathbf{z}_0, \mathbf{z}_1, \ldots \mathbf{z}_k)$.

suggested briefly by AOKI [11]. More recently, CAMERON [12] and LO [13] have assumed that the *a priori* density function for the initial state of linear systems with gaussian noise sequences has the form of a gaussian sum.

2.1. *Theoretical foundations of the approximation*

Consider a probability density function p which is assumed to have the following properties.

1. p is defined and continuous at all but a finite number of locations
2. $\int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1$
3. $p(x) \geq 0$ for all x

It is convenient although not necessary to consider only scalar random variables. The generalization to the vector case is not difficult but complicates the presentation and, it is felt, unnecessarily detracts from the basic ideas.

The problem of approximating p can be conveniently considered within the context of delta families of positive type [14]. Basically, these are families of functions which converge to a delta, or impulse, function as a parameter characterizing the family converges to a limit value. More precisely, let $\{\delta_\lambda\}$ be a family of functions on $(-\infty, \infty)$ which are integrable over every interval. This is called a *delta family of positive* type if the following conditions are satisfied.

(i) $\int_{-a}^{a} \delta_\lambda(x)\mathrm{d}x$ tends to one as λ tends to some limit value λ_0 for some a.
(ii) For every constant $\gamma > 0$, no matter how small, δ_λ tends to zero uniformly for $\gamma \leq |x| < \infty$ as λ tends to λ_0.
(iii) $\delta_\lambda(x) \geq 0$ for all x and λ.

Using the delta families, the following result can be used for the approximation of a density function p.

Theorem 2.1. The sequence $p_\lambda(x)$ which is formed by the convolution of δ_λ and p

$$p_\lambda(x) = \int_{-\infty}^{\infty} \delta_\lambda(x-u)p(u)\mathrm{d}u \tag{7}$$

converges uniformly to $p(x)$ on every interior subinterval of $(-\infty, \infty)$.

For a proof of this result, see KOREVAAR [14]. When p has a finite number of discontinuities, the theorem is still valid except at the points of discontinuity. It should be noted that essentially the same result is given by Theorem 2.1 in FELLER [15]. If $\{\delta_\lambda\}$ is required to satisfy the condition that

$$\int_{-\infty}^{\infty} \delta_\lambda(x)\mathrm{d}x = 1$$

it follows from equation (7) that p_λ is a probability density function for all λ.

It is basically the presence of the gaussian weighting function that has made the Edgeworth expansion attractive for use in the Bayesian recursion relations. The operations defined by equations (3–6) are simplified when the *a priori* densities are gaussian or closely related to the gaussian. Bearing this in mind, the following delta family is a natural choice for density approximations. Let

$$\delta_\lambda(x) \triangleq N_\lambda(x) = (2\pi\lambda^2)^{\frac{1}{2}} \exp[-\tfrac{1}{2}x^2/\lambda^2]. \tag{8}$$

It is shown without difficulty that $N_\lambda(x)$ forms a delta family of positive type as $\lambda \to 0$. That is, as the variance tends to zero, the gaussian density tends to the delta function.

Using equations (7) and (8), the density approximation p_λ is written as

$$P_\lambda(x) = \int_{-\infty}^{\infty} p(u)N_\lambda(x-u)\mathrm{d}u\,. \tag{9}$$

It is this form that provides the basis for the gaussian sum approximation that is the subject of this discussion.

While equation (9) is an interesting result, it does not immediately provide the approximation that can be used for specific application. However, it is clear that $p(u)N_\lambda(x-u)$ is integrable on $(-\infty, \infty)$ and is at least piecewise continuous. Thus, (9) can itself be approximated on any finite interval by a Riemann sum. In particular, consider an approximation of p_λ over some bounded interval (a, b) given by

$$p_{n,\lambda}(x) = \frac{1}{k}\sum_{i=1}^{n} p(x_i)N_\lambda(x-x_i)[\xi_i - \xi_{i-1}] \tag{10}$$

where the interval (a, b) is divided into n subintervals by selecting points ξ_i such that

$$a = \xi_0 < \xi_1 < \ldots \xi_n = b.$$

In each subinterval, choose the point x_i such that

$$p(x_i)[\xi_i - \xi_{i-1}] = \int_{\xi_{i-1}}^{\xi_i} p(x)\mathrm{d}x$$

which is possible by the mean-value theorem. The constant k is a normalizing constant equal to

$$k = \int_a^b p(x)\mathrm{d}x$$

and insures that $p_{n,\lambda}$ is a density function. Clearly, for $(b-a)$ sufficiently large, k can be made arbitrarily close to 1. Note that it follows that

$$\frac{1}{k}\sum_{i=1}^{n} p(x_i)[\xi_i-\xi_{i-1}]=1 \tag{11}$$

so that $p_{n,\lambda}$ essentially is a convex combination of gaussian density functions N_λ. It is basically this form that will be used in all future discussion and which is referred to hereafter as the *gaussian sum approximation.* It is important to recognize the $p_{n,\lambda}$ that are formed in this manner are valid probability density functions for all n, λ.

2.2. *Implementation of the gaussian sum approximation*

The preceding discussion has indicated that a probability density function p that has a finite number of discontinuities can be approximated arbitrarily closely outside of a region of arbitrarily small measure around each point of discontinuity by the gaussian sum $p_{n,\lambda}$ as defined in equation (10). These asymptotic properties are certainly necessary for the continued investigation of the gaussian sum. However, for practical purposes, it is desirable, in fact imperative, that p can be approximated to within an acceptable accuracy by a relatively small number of terms of the series. This requirement furnishes an additional facet to the problem that is considered in this section.

For the subsequent discussion, it is convenient to write the gaussian sum approximation as

$$p_n(x)=\sum_{i=1}^{n} \alpha_i N_{\sigma_i}(x-\mu_i) \tag{12}$$

where

$$\sum_{i=1}^{n} \alpha_i=1\,; \qquad \alpha_i\geq 0 \text{ for all } i\,.$$

The relation of equation (12) to equation (10) is obvious by inspection. Unlike equation (10) in which the variance λ^2 is common to every term of the gaussian sum, it has been assumed that the variance σ_i^2 can vary from one term to another. This has been done to obtain greater flexibility for approximations using a finite number of terms. Certainly, as the number of terms increase, it is necessary to require that the σ_i tend to become equal and vanish.

The problem of choosing the parameters α_i, μ_i, σ_i to obtain the "best" approximation p_n to some density function p can be considered. To define this more precisely, consider the L^k norm. The distance between p and p_n can be defined as

$$\left\|p-p_n\right\|^k=\int_{-\infty}^{\infty}\left|p(x)-\sum_{i=1}^{n}\alpha_i N_{\sigma_i}(x-\mu_i)\right|^k dx\,. \tag{13}$$

Thus, one can attempt to choose

$$\alpha_i,\ \mu_i\ \sigma_i\ (i=1,\ 2,\ \ldots,\ n)$$

so that the distance $\|p-p_n\|$ is minimized. As the number of terms n increases and as the variance decreases to zero, the distance must vanish. However, for finite n and nonzero variance, it is reasonable to attempt to minimize the distance in a manner such as this. In doing this, the stochastic estimation problem has been recast at this point as a deterministic curve fitting problem.

There are other norms and problem formulations that could be considered. In many problems, it may be desirable to cause the approximation to match some of the moments, for example, the mean and variance, of the true density exactly. If this were a requirement, then one could consider the moments as constraints on the minimization problem and proceed appropriately. For example, if the mean associated with p is μ, then the constraint that p_n have mean value μ would be

$$\mu=\int_{-\infty}^{\infty} x p_n(x)dx=\int_{-\infty}^{\infty} x\left[\sum_{n=1}^{n}\alpha_i N_{\sigma_i}(x-\mu_i)\right]dx$$

$$\mu=\sum_{i=1}^{n}\alpha_i\mu_i\,. \tag{14}$$

Thus, equation (14) would be considered in addition to the constraints on the α_i stated after equation (12).

Studies related to the problem of approximating p with a small number of terms have been conducted for a large number of density functions. These investigations have indicated, not surprisingly, that densities which have discontinuities generally are more difficult to approximate than are continuous functions. The results for two density functions, the uniform and the gamma, are discussed below. The uniform density is discontinuous and is of interest from that point of view. The gamma function is nonzero only for positive values of x so is an example of a nonsymmetric density that extends over a semi-infinite range.

Consider the following uniform density function

$$p(x)=\begin{cases}\frac{1}{4} \text{ for } -2\leq x\leq 2\\ 0 \text{ elsewhere}\end{cases} \tag{15}$$

This distribution has a mean value of zero and variance of 1·333.

Two different methods of fitting equation (15) have been considered. First, consider an approximation that is suggested directly by (10) and referred to subsequently as a *Theorem Fit.* The parameters of the approximation are chosen in the following general manner.

(1) Select the mean value μ_i of each gaussian so that the densities are equally spaced on $(-2, 2)$. By an appropriate location of the densities the mean value constraint (14) can be satisfied immediately.

(2) The weighting factors α_i are set equal to i/n so

$$\sum_{i=1}^{n} \alpha_i = 1 .$$

(3) The variance σ_i^2 of each gaussian is the same and is selected so that the L^1 distance between p and p_n is minimized.

This approximation procedure requires only a one-dimensional search to determine σ^2.

To investigate the accuracy and the convergence of the approximation, the number of terms in the sum was varied. Figure 1(a) shows the approximation when 6, 10, 20 and 49 terms were included. It is interesting to observe that the approximation retains the general character of the uniform density even for the six-term case. As should be expected, the largest errors appear in the vicinity of the discontinuities at ± 2. These approximations exhibit an apparent oscillation about the true value that is not visually satisfying. This oscillation can be eliminated by using a slightly larger value for the variance of the gaussian terms as is depicted in Fig. 1(b).

The second and fourth moments and the L^1 error are listed in Table 1 for these two sets of approximation. The individual terms were located symmetrically about zero so that the mean value of the gaussian sum agrees with that of the uniform density in all cases. Note that for the best fit the error in the variance is only 1·25 per cent when 20 terms are used. As should be expected, higher order

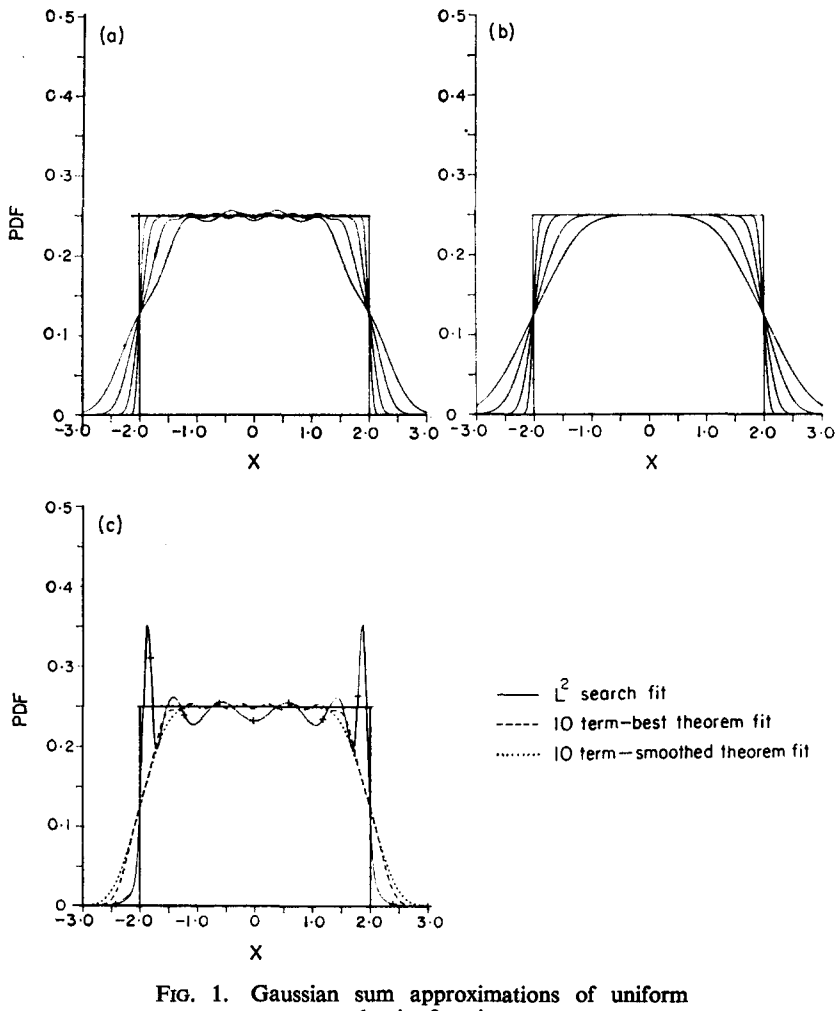

FIG. 1. Gaussian sum approximations of uniform density function.

(a) Best theorem fit—6, 10, 20 and 49 term approximations.

(b) Smoothed theorem fit—6, 10, 20 and 49 term approximations.

(c) L^2 Search fit comparison.

moments converge more slowly since errors farthest away from the mean assume more importance. For example, the fourth central moment has an error of 5 per cent for the 20-term approximation. The errors in the moments of the smoothed fit are aggravated only slightly although the L^1 error increases in a nontrivial manner.

TABLE 1. UNIFORM DENSITY APPROXIMATION

	Variance	Fourth central moment	L^1 error
True values	1·333	3·200	—
Best theorem fit			
6 terms	1·581	5·320	0·2199
10 terms	1·417	3·884	0·1271
20 terms	1·354	3·363	0·0623
49 terms	1·336	3·226	0·0272
Smoothed theorem fit			
6 terms	1·690	6·387	0·2444
10 terms	1·456	4·224	0·1426
20 terms	1·363	3·442	0·0701
49 terms	1·338	3·238	0·0280
L^2 search fit			
6 terms	1·419	3·626	0·0968

As an alternative approach, the parameters α_i μ_i, σ_i^2 were chosen to minimize the L^2 distance which is hereafter referred to as an L^2 search fit. These results are summarized in Fig. 1(c) for $n=6$. Included for comparison in this figure are the 10-term theorem fits from Figs. 1(a) and 1(b). Because of the number of terms involved, obtaining a search fit is significantly more difficult than obtaining a theorem fit for the same number of terms. Thus, the theorem fit may be more desirable from a practical standpoint. The moments and L^1 error for the search fit is also included in Table 1. These values are slightly better than the theorem fit involving ten terms. It is interesting in Fig. 1(c) to note the "spikes" that have appeared at the points of discontinuities. This appears to be analogous to the Gibbs phenomenon of Fourier series.

The second example that is discussed here is the gamma density function. It is defined as

$$p(x)=\begin{cases} 0 & \text{for } x<0 \\ \dfrac{x^3 e^{-x}}{6} & \text{for } x\geq 0. \end{cases} \tag{16}$$

The distribution has a mean value of 4 and second, third, and fourth central moments of 4, 8 and 72 respectively.

First, consider a theorem fit of this density in which the mean values are distributed uniformly on (0, 10). For the uniform density, the uniform placement was natural; for the gamma density it is not as appropriate. For example, for $n=6$ or 10, it is seen in Fig. 2 (a) that the approximating density is not as good, at least visually, as one might hope. The first four central moments are listed in Table 2. Clearly, the higher order moments contain large errors and even the mean value is incorrect in contrast with the uniform density.

TABLE 2. GAMMA DENSITY APPROXIMATION

	Mean	Variance	Third central moment	Fourth central moment	L^1 Error
True values	4	4	8	72	
Theorem fit					
6 terms in (0, 10)	3·94	4·345	5·496	60·78	0·119
10 terms in (0, 10)	3·94	3·861	4·941	47·84	0·053
20 terms in (0, 10)	3·93	3·611	4·653	41·23	0·023
20 terms in (0, 12)	4·00	4·206	7·477	71·41	0·042
L^2 search fit					
1 terms in (0, 10)	3·51	3·510	0	10·53	0·203
2 terms in (0, 10)	3·82	3·427	3·04	34·96	0·078
3 terms in (0, 10)	3·91	3·632	4·711	43·39	0·036
4 terms in (0, 10)	3·95	3·744	5·682	49·57	0·018

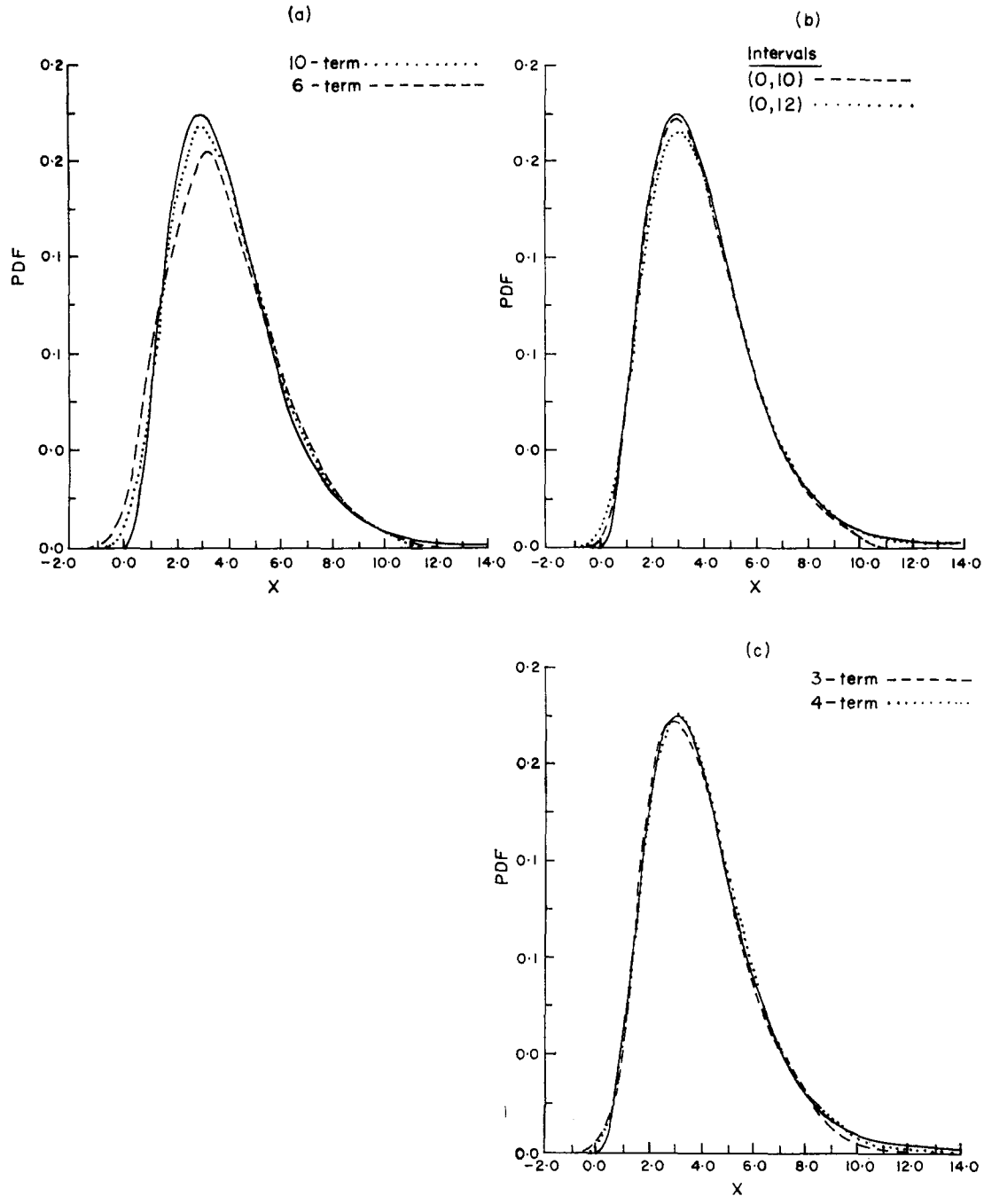

FIG. 2. Gaussian sum approximations of gamma density function.
(a) Best theorem fit—6 and 10 term approximations.
(b) Twenty term approximations over different intervals.
(c) L^2 search fit: 3 and 4 term approximations.

Two different 20-term approximations are depicted in Fig. 2(b). In one the mean values of the gaussian terms are selected in the interval (0, 10), whereas the second approximation is distributed in (0, 12). Note in the first case the gaussian sum tends to zero much more rapidly than the gamma function for $x > 10$. Thus, to improve the approximation and the moments it is necessary to increase the interval over which terms are placed and the second curve indicates the influence of this change. The results of these two cases are included in Fig. 2 and indicate that increasing the interval over which the approximation is valid has significantly improved the moments.

The theorem fit provides a very simple method for obtaining an approximation. However, it is clear that better results could be obtained by choosing at least the mean values of the individual gaussian terms more carefully. Consider now some L^2 search fits. In Fig. 2(c), the search fits for three

and four terms are depicted and the moments are listed in Table 2. Clearly, values of the mean and variance appear to be converging to the true values of 4. Note also that the 4-term approximation is considerably better than the 10-term theorem fit. Thus, the search technique, while more difficult to obtain, points out the desirability of judicious placement of the gaussian terms in order to obtain the most suitable approximation.

3. LINEAR SYSTEMS WITH NONGAUSSIAN NOISE

It is envisioned that the gaussian sum approximation will be very useful in dealing with non-linear stochastic systems. However, many of the properties and concomitant difficulties of the approximation are exhibited by considering linear systems which are influenced by nongaussian noise. As is well-known, the *a posteriori* density $p(x_k/Z_k)$ is gaussian for all k when the system is linear and the initial state and plant and measurement noise sequences are gaussian. The mean and variance of the conditional density are described by the Kalman filter equations. When nongaussian distributions are ascribed to the initial state and/or noise sequences, $p(x_k/Z_k)$ is no longer gaussian and it is generally impossible to determine $p(x_k/Z_k)$ in a closed form. Furthermore, in the linear, gaussian problem, the conditional mean, i.e. the minimum variance estimate, is a linear function of the measurement data and the conditional variance is independent of the measurement data. These characteristics are generally not true for a system which is either non-linear or nongaussian.

For the following discussion, consider a scalar system whose state evolves according to

$$x_k=\Phi_{k,k-1}x_{k-1}+w_{k-1} \tag{17}$$

and whose behavior is observed through measurement data z_k described by

$$z_k=H_kx_k+v_k. \tag{18}$$

Suppose that the density function describing the initial state has the form

$$p(x_0)=\sum_{i=1}^{l_0}\alpha_{i0}N_{\sigma'_{oi}}(x_0-\mu'_{i0}). \tag{19}$$

Assume that the plant and measurement noise sequences (i.e. $\{w_k\}$ and $\{v_k\}$) are mutually independent, white noise sequences with density functions represented by

$$p(w_k)=\sum_{i=1}^{s_k}\beta_{ik}N_{q_{ki}}(w_k-\omega_{ik}) \tag{20}$$

$$p(v_k)=\sum_{i=1}^{m_k}\gamma_{ik}N_{r_{ik}}(v_k-v_{ik}). \tag{21}$$

There are a variety of ways in which the gaussian sum approximation could be introduced. For example, it is natural to proceed in the manner that is to be discussed here in which the *a priori* distributions are represented by gaussian sums as in equations (19), (20) and (21). This approach has the advantage that the approximation can be determined off-line and then used directly in the Bayesian recursion relations. An alternative approach would be to perform the approximation in more of an on-line procedure. Instead of approximating the *a priori* densities, one could deal with $p(x_k/Z_k)$ in equation (3) and the integrand in (4) and derive approximations at each stage. This would be more direct but has the disadvantage that considerable computation may be required during the processing of data. Discussion of the implementation of this approach will not be attempted in this paper.

3.1. *Determination of the* a posteriori *distributions*

Suppose that the *a priori* density functions are given by equations (19, 20 and 21). In using these representations in the Bayesian recursion relations, it is useful to note the following properties of gaussian density functions

Scholium 3.1: For $a\neq 0$,

$$N_\sigma(x-ay)=\frac{1}{a}N_{\sigma/a}(y-x/a). \tag{22}$$

Scholium 3.2:

$$N_{\sigma_i}(x-\mu_i)N_{\sigma_j}(x-\mu_j)=N^2_{[\sigma_i^2+\sigma_j^2]^{1/2}}(\mu_i-\mu_j)N_{\sigma_{ij}}(x-\mu_{ij}) \tag{23}$$

where

$$\mu_{ij}=\frac{\mu_i\sigma_j^2+\mu_j\sigma_i^2}{\sigma_i^2+\sigma_j^2}$$

$$\sigma_{ij}^2=\frac{\sigma_i^2\sigma_j^2}{\sigma_i^2+\sigma_j^2}.$$

The proofs of (22) and (23) are omitted.

Armed with these two results it is a simple matter to prove that the following descriptions of the filtering and prediction densities are true.

Theorem 3.1. Suppose that $p(x_k/Z_{k-1})$ is described by

$$p(x_k/Z_{k-1})=\sum_{i=1}^{l_k}\alpha_{ik}N_{\sigma'_{ik}}(x_k-\mu'_{ik}). \tag{24}$$

Then, $p(x_k/Z_k)$ is given by

$$p(x_k/Z_k)=\sum_{i=1}^{l_k}\sum_{j=1}^{m_k}c_{ij}N_{\sigma_{ij}}(x_k-\varepsilon_{ij}) \tag{25}$$

where

$$c_{ij}=\alpha_{ik}\gamma_{jk}N_{\Sigma}(z_k-\rho_{ij})\Big/\sum_{l=1}^{l_k}\sum_{m=1}^{m_k}\alpha_{ik}\gamma_{mk}N_{\Sigma}(z_k-\rho_{lm}).$$

$$\mu_{ij}=H_k\mu'_{ik}+\nu_{jk}$$

$$\Sigma^2=H_k^2\sigma'^2_{ik}+r_{jk}^2$$

$$\varepsilon_{ij}=\mu'_{ik}+\frac{\sigma'^2_{ik}H_k}{\sigma'^2_{ik}H_k^2+r_{jk}^2}[z_k-\nu_{jk}-H_k\mu'_{ik}]$$

$$\sigma_{ij}^2=\sigma'^2_{ik}-\sigma'^4_{ik}H_k^2/(\sigma'^2_{ik}H_k^2+r_{jk}^2).$$

It is obvious that the $c_{ij}\geq 0$ and that

$$\sum_{i=1}^{l_k}\sum_{j=1}^{m_k}c_{ij}=1.$$

Thus, equation (25) is a gaussian sum and for convenience one can rewrite it as

$$p(x_k/Z_k)=\sum_{i=1}^{n_k}a_{ik}N_{p_{ik}}(x_k-\mu_{ik}) \tag{26}$$

where $n_k=(l_k)(m_k)$ and the a_{ik}, p_{ik} and μ_{ik} are formed in an obvious fashion from the c_{ij}, σ_{ij} and ε_{ij}.

The proof of the theorem is straightforward. From the definition of the measurement relation (18) and the measurement noise density (21), one sees that

$$p(z_k/x_k)=\sum_{i=1}^{m_k}\gamma_{ik}N_{r_{ik}}(z_k-H_kx_k-\nu_{ik}).$$

Using this and (24) in (3) and applying the two scholiums, one obtains (25).

The prediction density is determined from (4) and leads to the following result.

Theorem 3.2. Assume that $p(x_k/Z_k)$ is given by (26). Then for the linear system (17) and plant noise (20) the prediction density $p(x_{k+1}/Z_k)$ is

$$p(x_{k+1}/Z_k)=\sum_{i=1}^{n_k}\sum_{j=1}^{s_k}a_{ik}\beta_{jk}N_{\lambda_{ij}}(x_{k+1}-\Phi_{k+1,k}\mu_{ik}-\omega_{jk}) \tag{27}$$

where

$$\lambda_{ij}{}^2=\Phi_{k+1,k}{}^2p_{ik}{}^2+q_{jk}{}^2.$$

It is convenient to redefine terms so that (27) can be written as

$$p(x_{k+1}|Z_k)=\sum_{i=1}^{l_{k+1}}\alpha_{i(k+1)}N_{\sigma_{i(k+1)}}(x_{k+1}-\mu_{i(k+1)}). \tag{28}$$

Clearly, the definition of $p(x_0)$ has the form (28) as does the $p(x_k/Z_{k-1})$ assumed in Theorem 3.1. Thus, it follows that the gaussian sums repeat themselves from one stage to the next and that (26) and (28) can be regarded as the general forms for an arbitrary stage. Thus, the gaussian sum can almost be regarded as a reproducing density [16]. It is important, however, to note that the number of terms in the gaussian sum increases at each stage so that the density is not described by a fixed number of parameters. The density would be truly reproducing if only the initial state were non-gaussian, for example, see Ref. [12]. It is clear in this case that the number of terms in the gaussian sum remains equal to the number used to define $p(x_0)$ so that $p(x_k/Z_k)$ is described by a fixed number of parameters.

There are several aspects that require comment at this point. First, if the gaussian sums for the *a priori* density all contain only one term, that is, they are gaussian, the Kalman filter equations and the gaussian *a posteriori* density are obtained. In fact the ε_{ij} and $\sigma_{ij}{}^2$ in (25) and the means and variances $\lambda_{ik}{}^2$ in (27) each represent the Kalman filter equations for the ijth density combination. Thus, the gaussian sum in a manner of speaking describes a combination of Kalman filters operating in concert. To examine this further, consider the first and second moments associated with the prediction and filtering densities.

Theorem 3.3

$$E[x_k|Z_k]\triangleq\hat{x}_{k/k}=\sum_{i=1}^{l_k}\sum_{j=1}^{m_k}c_{ij}\varepsilon_{ij} \tag{29}$$

$$E[(x_k-\hat{x}_{k/k})^2|Z_k]\triangleq p_{k/k}{}^2=\sum_{i=1}^{l_k}\sum_{j=1}^{m_k}c_{ij}[\sigma_{ij}{}^2+(\hat{x}_{k/k}-\varepsilon_{ij})^2] \tag{30}$$

$$E[x_{k+1}|Z_k]\triangleq\hat{x}_{k+1/k}=\Phi_{k+1,k}\hat{x}_{k/k}+E[w_k] \tag{31}$$

$$E[(x_{k+1}-\hat{x}_{k+1/k})^2|Z_k]=p_{k+1/k}{}^2=\Phi_{k+1,k}{}^2p_{k/k}{}^2+E\{[w_k-E(\omega_k)]^2\} \tag{32}$$

where

$$E(w_k)=\sum_{i=1}^{s_k}\beta_{ik}\omega_{ik}$$

$$E\{[w_k-E(\omega_k)]^2\}=\sum_{i=1}^{s_k}\beta_{ik}(q_{ik}{}^2+\omega_{ik}{}^2)-E^2(w_k).$$

In equation (29), the mean value $\hat{x}_{k/k}$ is formed as the convex combination of the mean values ε_{ij} of the individual terms, or Kalman filters, of the gaussian sum. It is *important to recognize* that the c_{ij}, as is apparent from equation (25), depend upon the measurement data. Thus, the conditional mean is a *non-linear* function of the current measurement data.

The conditional variance $p_{k/k}^2$ described by (30) is more than a convex combination of the variances of the individual terms because of the presence of the term $(\hat{x}_{k/k}-\varepsilon_{ij})^2$. This shows that the variance is increased by the presence of terms whose mean values differ significantly from the conditional mean $\hat{x}_{k/k}$. The influence of these terms is tempered by the weighting factor c_{ij}. Note also that the conditional variance (in contrast to the linear Kalman filter) is a function of the measurement data because of the c_{ij} and the $(\hat{x}_{k/k}-\varepsilon_{ij})$.

The mean and variance of the prediction density are described in an obvious manner. If the gaussian sum is an approximation to the true noise density, these relations suggest the desirability of matching the first two moments exactly in order to obtain an accurate description of the conditional mean $\hat{x}_{k+1/k}$ and variance $P_{k+1/k}^2$.

As discussed earlier, it is convenient to assign the same variance to all terms of the gaussian sum. Thus, if the initial state and the measurement and noise sequences are identically distributed, it is reasonable to consider the variances for all terms to be identical and to determine the consequences of this assumption. Note from Scholium 3.2 that if

$$\sigma_i^2=\sigma_j^2\triangleq\sigma^2$$

then

$$\sigma_{ij}^2=\sigma^2/2 \qquad \text{for all } i,j.$$

Thus, the variance remains the same for all terms in the gaussian sum when $p(x_k/z_k)$ is formed and the concentrations as described by the variance becomes greater. Furthermore, from Scholium 3.2 it follows that the mean value is given by

$$\mu_{ij}=\frac{\mu_i+\mu_j}{2}$$

so the new mean value is the average of the previous means. This suggests the possibility that the mean values of some terms, since they are the average of two other terms, may become equal, or almost equal. If two terms of the sum have equal means and variances. they could be combined by adding their respective weighting factors. This would reduce the total number of terms in the gaussian sum.

The c_{ij} in (25) are essentially determined by a gaussian density. Thus, if $z_k-\rho_{ij}$ becomes very large, then the c_{ij} may become sufficiently small that the entire term is negligible. If terms could be neglected, then the total number of terms in the gaussian sum could be reduced.

The prediction and filtering densities are represented at each stage by a gaussian sum. However, it has been seen that the sums have the characteristic that the number of terms increases at each stage as the product of the number of terms in the two constituent sums from which the densities are formed. This fact could seriously reduce the practicality of this approximation if there were no alleviating circumstances. The discussion above regarding the diminishing of the weighting factors and the combining of terms with nearly equal moments has introduced the mechanisms which significantly reduce the apparent ill effects caused by the increase in the number of terms in the sum. It is an observed fact that the mechanisms whereby terms can be neglected or combined are indeed operative and in fact can sometimes permit the number of terms in the series to be reduced by a substantial amount.

Since weighting factors for individual terms do not vanish identically nor do the mean and variance of more gaussian densities become identical, it is necessary to establish criteria by which one can determine when terms are negligible or are approximately the same. This is accomplished by defining numerical thresholds which are prescribed to maintain the numerical error less than acceptable limits.

Consider the effects on the L^1 error of neglecting terms with small weighting factors. Suppose that the density is

$$p(x)=\sum_{i=1}^{n}\alpha_i N_\sigma(x-a_i) \tag{33}$$

and such that $\alpha_1, \alpha_2, \ldots, \alpha_{m-1} (m<n)$ are less than some positive number δ_1. Note that the variance has been assumed to be the same for each term. Consider replacing p by p_A where

$$p_A(x)=\frac{1}{\sum_{i=m}^{n}\alpha_i}\sum_{i=m}^{n}\alpha_i N_\sigma(x-a_i). \tag{34}$$

The following bound is determined without difficulty.

Theorem 3.4.

$$\int_{-\infty}^{\infty}|p(x)-p_A(x)|dx\leq 2\sum_{i=1}^{m-1}\alpha_i \tag{35}$$

$$\leq 2(m-1)\delta_1. \tag{36}$$

The L^1 error caused by neglecting $(m-1)$ terms each of which are less than δ_1 is seen in (35) to be less than twice the sum of the neglected terms. Thus, the threshold δ_1 can be selected by using (36) or (35) to keep the increased L^1 error within acceptable limits.

Consider the situation in which the absolute value of the difference of the mean values of two terms is small. In particular, suppose that a_1 and a_2 are approximately the same and consider the L^1 error that results if the $p(x)$ given in (33) is replaced by

$$p_A(x)=\sum_{i=3}^{n}\alpha_i N_\sigma(x-a_i)+(\alpha_1+\alpha_2)N_\sigma(x-\bar{x}) \tag{37}$$

where

$$\bar{x}=\frac{\alpha_1 a_1+\alpha_2 a_2}{\alpha_1+\alpha_2}.$$

Using (37), one can prove the following bound. For a detailed proof of this and other results, see Ref. [17].

Theorem 3.5.

$$\int_{-\infty}^{\infty}|p(x)-p_A(x)|dx\leq\frac{4\alpha_1\alpha_2 M}{\alpha_1+\alpha_2}|a_2-a_1|. \quad (38)$$

Thus terms can be combined if the right-hand side of (38) is less than some positive number δ_2 which represents the allowable L^1 error. The M in (38) is the maximum value of N_σ and is given by

$$M=\frac{1}{\sqrt{2\pi}\,\sigma}.$$

Observe that as the variance σ decreases, the distance between two terms to be combined must also decrease in order to retain the same error bound.

3.2. *A numerical example*

In this section the results presented in section 3.1 are applied to a specific example. To make (17) and (18) more specific, suppose that the system is

$$x_k=x_{k-1}+w_{k-1} \quad (39)$$

$$z_k=x_k+v_k \quad (40)$$

where the x_0, w_k, $v_k(k=0, 1, \ldots)$ are assumed to be uniformly distributed on $(-2, 2)$ as defined by (15).

The problem that is considered represents something of a worst case for the approximation because the initial state and the noise sequences are assumed to be uniformly distributed with the density discussed in section 2.2. As discussed there, the discontinuities at ± 2 make it difficult to fit this density and necessitates the use of many terms in the gaussian sum. The specific approximation used here contains 10 terms and is shown in Fig. 1(b). It is apparent that this approximation has nontrivial errors in the neighborhood of the discontinuities but nonetheless retains the basic character of the uniform distribution.

The performance of the gaussian sum approximation for this example is described below by comparing the conditional mean and variance provided by the approximation with that predicted by the Kalman filter and with the statistics obtained by considering the true uniform distribution. The latter have been determined for this example after a nontrivial amount of numerical computation. In addition the true *a posteriori* density $p(x_k/Z_k)$ has been computed and is compared with that obtained using the approximation.

The Kalman error variance is independent of the measurement sequence and can cause misleading filter response. For example, for some measurement sequences perfect knowledge of the state is possible, that is the variance is zero, but the Kalman variance still predicts a large uncertainty. For example, suppose that the measurements at each stage are equal to

$$z_k=2k+4, \qquad k=0, 1, \ldots$$

Then, the minimum mean-square estimate for the state based on the uniform distribution is

$$\hat{x}_{k/k}=2k+2, \qquad k=0, 1, \ldots$$

and the variance of this estimate is

$$p_{k/k}{}^2=0 \qquad \text{for all } k.$$

Thus, for this measurement realization the minimum mean-square estimate is error-free. Since the Kalman variance is independent of the measurements, it is necessarily a poor approximation of the actual conditional variance. The square root of the Kalman variance σ_{KAL} and the error in the best linear estimate together with the square root of the variance σ_{GS} and error in the estimate of the state for the ten term gaussian sum for this measurement realization is shown in Fig. 3(a). This shows that a considerable improvement in both the mean and variance is provided by the gaussian sum when compared with the results provided by the Kalman filter.

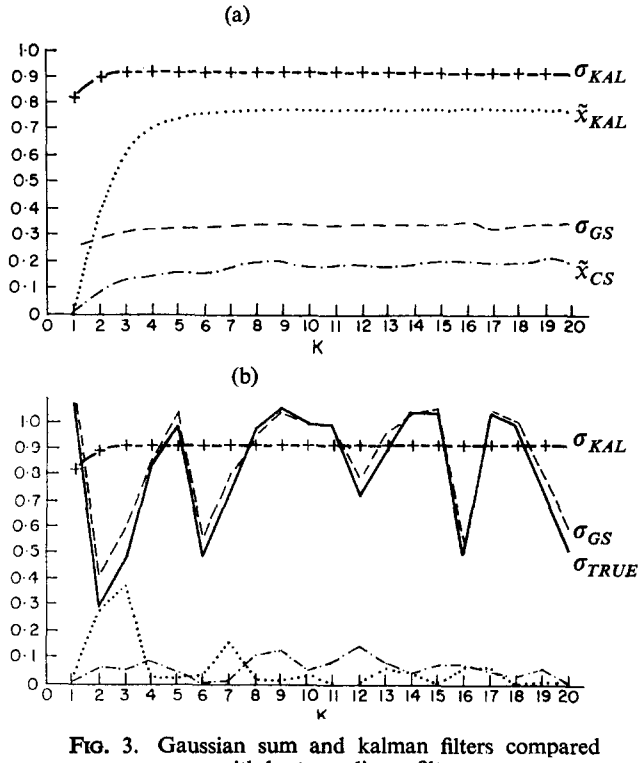

FIG. 3. Gaussian sum and kalman filters compared with best non-linear filter.
(a) Perfect knowledge example.
(b) Random noise example.

The measurement sequence described above is highly improbable. A more representative case is depicted in Fig. 3(b) in which the plant and measurement noise sequences were chosen using a random number generator for the uniform distribution prescribed above. This figure again shows that the gaussian sum approximation to the true variance is considerably improved over the Kalman estimate and, in fact, agrees very closely with the true standard deviation σ_{TRUE} of the *a posteriori* density. However, the Kalman variance is representative enough that the error in the estimate of the mean is not particularly different than that provided by the gaussian sum.

The *a posteriori* density function for the third and fourteenth stages is shown in Figs. 4(a) and 4(b). In these figures, the actual density, the gaussian approximation provided by the Kalman filter and the gaussian sum approximation are all included. At stage 3, the true variance is smaller than the Kalman variance and the Kalman mean has a significant error whereas at stage 14 the true variance is larger than that predicted by the Kalman filter. Note that the error in the *a priori* density approximation at the discontinuities is still evident in the *a posteriori* approximation but that the general character of the density is reproduced by the gaussian sum.

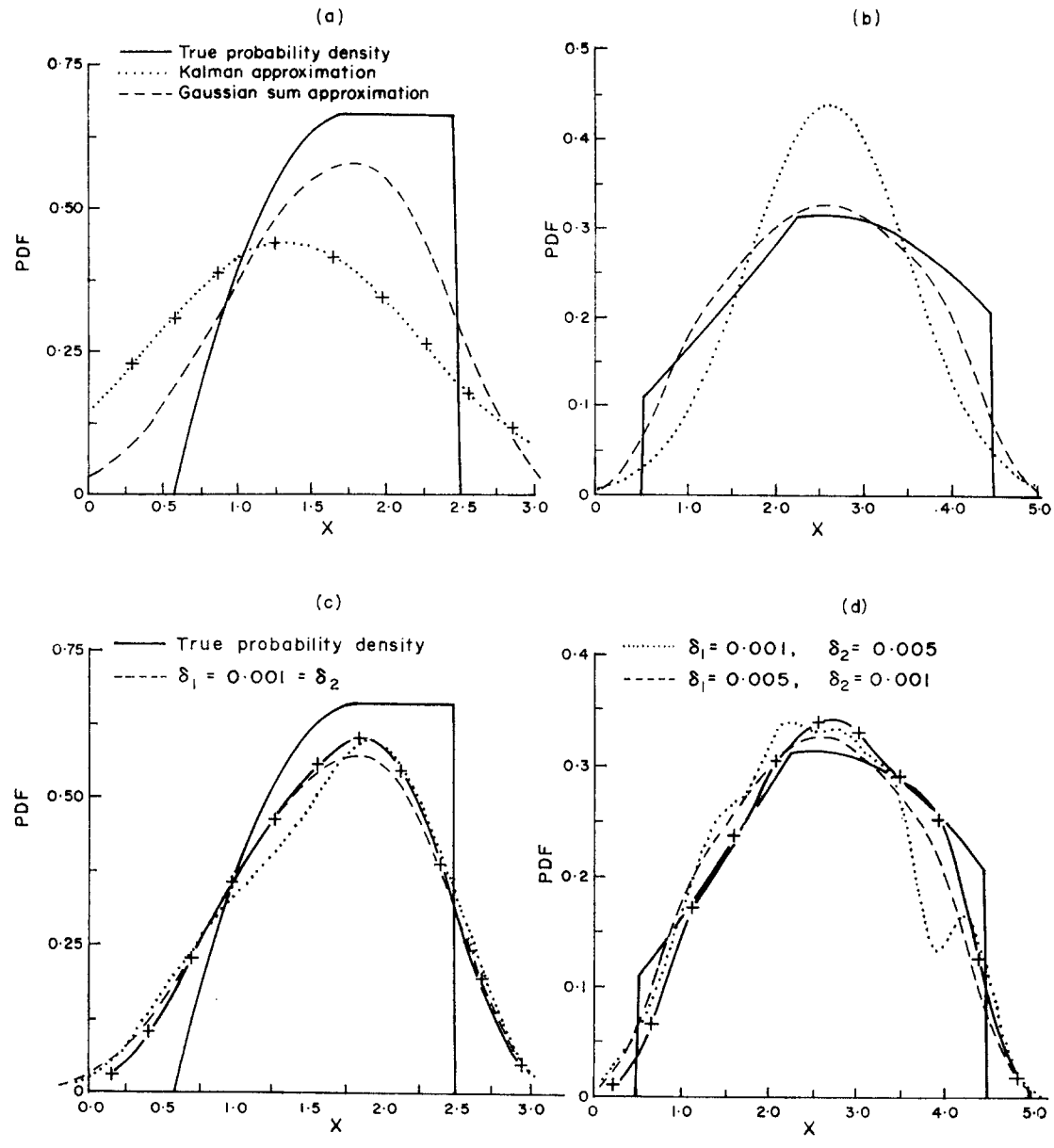

FIG. 4. Influence of δ_1, δ_2 on *a posteriori* density.
(a) Third stage—$\delta_1=\delta_2=0{\cdot}001$.
(b) Fourteenth stage—$\delta_1=\delta_2=0{\cdot}001$.
(c) Third stage—$\delta_1=0{\cdot}001$, $\delta_2=0{\cdot}005$. and $\delta_1=0{\cdot}005$, $\delta_1=0{\cdot}001$.
(d) Fourteenth stage—$\delta_1=0{\cdot}005$, $\delta_2=0{\cdot}001$. and $\delta_1=0{\cdot}001$, $\delta_2=0{\cdot}005$.

In Figs. 4(a) and 4(b) the approximations at each stage are based on terms being eliminated when their weighting functions are less than $\delta_1 = 0{\cdot}001$ and combined when the difference in mean values cause the L^1 error to be less than $\delta_2 = 0{\cdot}001$. The effect of changing these parameters, that is, δ_1 and δ_2, can be seen in Figs. 4(c) and 4(d). In these figures, the effect of changing δ_1 and δ_2 is shown. In one case δ_1 is increased from 0·001 to 0·005 while keeping δ_2 equal to 0·001. Alternatively, δ_2 is increased to 0·005 while the value of δ_1 is maintained equal to 0·001. It is apparent in this example that δ_2 has a larger effect on the density approximation. The increase in this parameter can be seen to introduce a "ripple" into the density function and indicates that the individual terms have become too widely separated to provide a smooth approximation. It is interesting that the effect appears to be cumulative as the ripple is not apparent after three stages but is quite marked at the 14th stage.

The effect of improving the accuracy of the density approximation by including more terms in the *a priori* representations and by retaining more terms for the *a posteriori* representation can be seen in Fig. 5. Twenty terms are included in the *a priori* densities and δ_1 and δ_2 are reduced to 0·0001. The figure presents the actual *a posteriori* density, the gaussian sum approximation, and the gaussian approximation provided by the Kalman filter equations. Comparison of the results for stages 3 and 14 with Fig. 4 indicates the improvements in the approximation that have occurred.

4. CONCLUSIONS

The approximation of density functions by a sum of gaussian densities has been discussed as a reasonable framework within which estimation policies for non-linear and/or nongaussian stochastic systems can be established. It has been shown that a probability density function can be approximated arbitrarily closely except at discontinuities by such a gaussian sum. In contrast with the Edgeworth or Gram–Charlier expansions that have been investigated earlier, this approximation has the advantage of converging to a broader class of density functions. Furthermore, any finite sum of these terms is itself a valid density function.

The gaussian sum approximation is a departure from more classical approximation techniques because the sum is restricted to be positive for all possible values of the independent variable. As a result, the series is not orthogonalizable so that the manner in which parameters appearing in the sum are chosen is not obvious. Two numerical procedures are discussed in which certain parameters are chosen to satisfy constraints or are somewhat arbitrarily selected and others are chosen to minimize the L^k error.

It is anticipated that the gaussian sum approximation will find its greatest application in developing estimation policies for non-linear stochastic systems. However, many of the characteristics exhibited by non-linear systems and some of the difficulties in using the gaussian sum in these cases are exhibited by treating linear systems with nongaussian noise

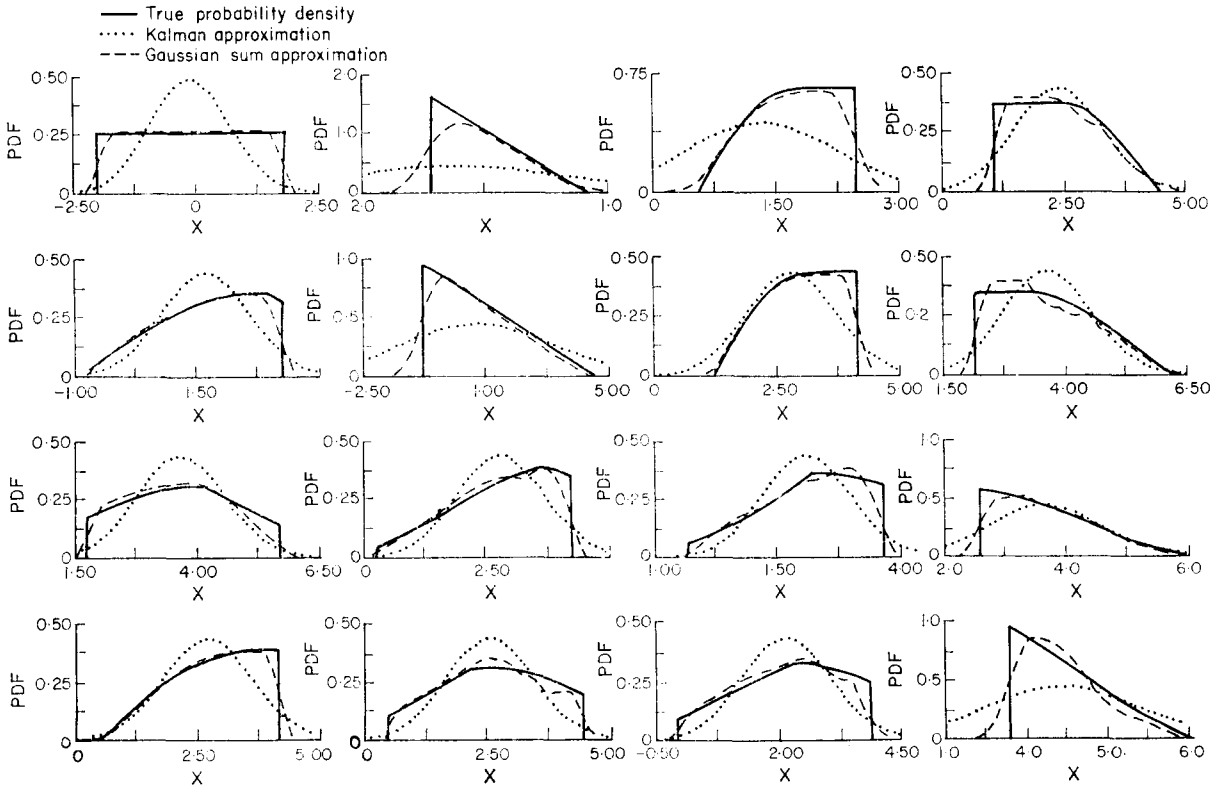

FIG. 5. *A posteriori* density for sixteen stages.

sources. Of course, when the noise is entirely gaussian the problem degenerates and the familiar Kalman filter equations are obtained as the exact solution of the problem.

The linear, nongaussian estimation problem is discussed and it is shown that, if the *a priori* density functions are represented as gaussian sums, then the number of terms required to describe the *a posteriori* density is equal to the product of the number of terms of the *a priori* densities used to form it. The apparent disadvantage is seen to cause little difficulty however, because the moments of many individual terms converge to common values which allows them to be combined. Further, the weighting factors associated with many other terms become very small and permit those terms to be neglected without introducing significant error to the approximation.

Numerical results for a specific system are presented which provide a demonstration of some of the effects discussed in the text. These results confirm dramatically that the gaussian sum approximation can provide considerable insight into problems that hitherto have been intractable to analysis.

REFERENCES

[1] A. H. JAZWINSKI: *Stochastic Processes and Filtering Theory*. Academic Press, New York (1970).
[2] R. E. KALMAN: A new approach to linear filtering and prediction problems. *J. bas. Engng* **82D**, 35–45 (1960).
[3] H. W. SORENSON: *Advances in Control Systems*, Vol. 3, Ch. 5. Academic Press, New York (1966).
[4] R. COSAERT and E. GOTTZEIN: A decoupled shifting memory filter method for radio tracking of space vehicles. 18th International Astronautical Congress, Belgrade, Yugoslavia (1967).
[5] A. JAZWINSKI: Adaptive filtering. *Automatica* **5**, 475–485 (1969).
[6] Y. C. Ho and R. C. K. LEE: A Bayesian approach to problems in stochastic estimation and control. *IEEE Trans. Aut. Control* **9**, 333–339 (1964).
[7] H. J. KUSHNER: On the differential equations satisfied by conditional probability densities of Markov processes. *SIAM J. Control* **2**, 106–119 (1964).
[8] J. R. FISHER and E. B. STEAR: Optimal non-linear filtering for independent increment processes, Parts I and II. *IEEE Trans. Inform. Theory* **3**, 558–578 (1967).
[9] M. AOKI: *Optimization of Stochastic Systems, Topics in Discrete-Time Systems*. Academic Press, New York (1967).
[10] H. W. SORENSON and A. R. STUBBERUD: Nonlinear filtering by approximation of the *a posteriori* density. *Int. J. Control* **18**, 33–51 (1968).
[11] M. AOKI: Optimal Bayesian and min-max control of a class of stochastic and adaptive dynamic systems. *Proceedings IFAC Symposium on Systems Engineering for Control System Design*, Tokyo, pp. 77–84 (1965).
[12] A. V. CAMERON: Control and estimation of linear systems with nongaussian *a priori* distributions. *Proceedings of the Third Annual Conference on Circuit and System Science* (1969).
[13] J. T. LO: Finite dimensional sensor orbits and optimal nonlinear filtering. University of Southern California. Report USCAE 114, August 1969.
[14] J. KOREYAAR: *Mathematical Methods*, Vol. 1, pp. 330–333. Academic Press, New York (1968).
[15] W. FELLER: *An Introduction to Probability Theory and Its Applications*, Vol. II, p. 249. John Wiley, New York (1966).
[16] J. D. SPRAGINS: Reproducing distributions for machine learning. Standard Electronics Laboratories, Technical Report No. 6103-7 (November 1963).
[17] D. L. ALSPACH: A Bayesian approximation technique for estimation and control of time-discrete stochastic systems. Ph.D. Dissertation, University of California, San Diego (1970).

Résumé—Les relations recurrentes bayesiennes qui décrivent le comportement de la fonction de densité de probabilité a posteriori de l'état d'un système aléatoire, discret dans le temps, en se basant sur les données des mesures disponibles, ne peuvent être généralement resolues sous une forme fermee lorsque le système est soit non-linéaire, soit non-gaussien. Le present article introduit et propose une approximation de densité, mettant en jeu des combinaisons convexes de fonctions de densité gaussiennes, à titre de méthode efficace pour éviter les difficultes rencontrées dans l'évaluation de ces relations et dans l'utilisation des densités en résultant pour determiner des stratégies d'évaluation particulières. Il est montré que, lorsque le nombre de termes de la somme gaussienne augmente sans limites, l'approximation converge uniformement vers une fonction de densité quelconque dans une catégorie étendue. De plus, toute somme finie est elle-même une fonction de densité valable, d'une manière différente des nombreuses autres approximations étudiées dans le passé.

Le problème de la determination des estimations de la densité a posteriori et de la variance minimale pour des systèmes linéaires avec bruit non-gaussien est traité en utilisant l'approximation des sommes gaussiennes. Ce problème est étudié parce qu'il peut être traité d'une manière relativement simple en utilisant l'approximation mais contient encore la plupart des difficultés rencontrées en considérant des systèmes non-linéaires, puisque la densité a posteriori est non-gaussienne. Après la discussion du problème général du point de vue de l'application des sommes gaussiennes, l'article présente un exemple numérique dans lequel les statistiques réélles de la densité *a posteriori* sont comparées avec les valeurs prédites par les approximations des sommes gaussiennes et du filtre de Kalman.

Zusammenfassung—Die Bayes'schen Rekursionsbeziehungen, die das Verhalten der a priori-Wahrscheinlichkeitsdichtefunktion des Zustandes eines zeitdiskreten stochastischen Systems beschrieben, können nicht allgemein in geschlossener Form gelöst werden, wenn das System entweder nichtlinear oder nicht-gaussisch ist. In dieser Arbeit wird eine Dichteapproximation, die konvexe Kombinationen von Gauss'schen Dichtefunktionen enthält, eingeführt und als bedeutungsvoller Weg zur Umgehung der Schwierigkeiten vorgeschlagen, die sich bei der Auswertung dieser Beziehungen und bei der Benutzung der resultierenden Dichten ergeben, um die spezifische Schätzstrategie zu bestimmen. Ersichtlich konvergiert, da die Zahl der Terme in der Gauss'schen Summe unbegrenzt wächst, die Approximation gleichmäßig zu einer Dichtefunktion in einer umfassenden Klasse. Weiter ist eine endliche Summe selbst eine gültige Dichtefunktion, anders als viele andere schon untersuchte Approximationen.

Das Problem der Bestimmung der Schätzungen der a posteriori-Dichte und des Varianzminimums für lineare Systeme mit nichtgauss'schem Geräusch wird unter Behandlung der Gauss'schen Summenapproximation behandelt.

Dieses Problem wird betrachtet, weil es in einer relativ geraden Art unter Benutzung der Approximation behandelt werden kann, aber immer noch die meisten der Schwierigkeiten enthält, denen man bei der Betrachtung nichtlinearer Systeme begegnet, da die a posteriori-Dichte nichtgaussisch ist. Nach der Diskussion des allgemeinen Problems vom Gesichtspunkt der Anwendung Gauss'scher Summen, wird ein Zahlenbeispiel gebracht, in dem die aktuellen Statistiken der a posteriori-Dichte mit den Werten verglichen werde, die durch die Gauss'sche Summe und durch die Kalman-Filter-Approximationen vorhergesagt wurden.

Резюме—Рекурсивные байесовские выражения описывающие поведение апостериорной фукции плотности вероятности состояния случайной системы, дискретной по времени, основываясь на доступных результатах измерений, не могут быть обычно решены в замкнутой форме когда система либо нелинейна, либо-же не является гауссовской. Настоящая статья вводит и предлагает приближение плотности, прибегающее к сочетаниям гауссовских функций плотности, в качестве эффективного метода для избежания затруднений встречаемых в оценке этих выражений и в использовании вытекающих из них плотностей для опредения данных статегий оценки. Показывается что, когда число членов гауссовской суммы беспредельно возрастает, приближение равномерно стремится к любой функции плотности в широком классе. Сверх этого, всакая конечная сумма сама является приемлемой функцией плотности, в отличии от многочисленных других приближений изучавщихся в прошлом.

Проблема определения оценок апостериорной плотности и минимальной вариантности для линейных систем с не-гауссовским шумом решается с помощью приближения гауссовских сумм. Эта проблема изучается потому что она может быть решена сравнительно просто используя приближение но содержит еще большинство затруднений встречаемых рассматривая нелинейные системы, ибо апостериорная плотность является не-гауссовской. После обсуждения общей проблемы с точки зрения применения гауссовских сумм, статья дает числовой пример в котором действительные статистики апостериорной плотности сравниваются с значениями предсказанными приближениями гауссовских сумм и фильтра калмана.