Recursive Filtering with Non-Gaussian Noises

Wen-Rong Wu and Amlan Kundu

Abstract— **The Kalman filter is the optimal recursive filter, although its optimality can only be claimed under the Gaussian noise environment. In this paper, we consider the problem of recursive filtering with non-Gaussian noises. One of the most promising schemes, which was proposed by Masreliez, uses the nonlinear score function as the correction term in the state estimate. Unfortunately, the score function cannot be easily implemented except for simple cases. In this paper, a new method for efficient evaluation of the score function is developed. The method employs an adaptive normal expansion to expand the score function followed by truncation of the higher order terms. Consequently, the score function can be approximated by a few central moments. The normal expansion is made adaptive by using the concept of conjugate recentering and the saddle point method. It is shown that the approximation is satisfactory, and the method is simple and practically feasible. Experimental results are reported to demonstrate the effectiveness of the new algorithm.**

I. Introduction

The problem of estimating the state of a linear stochastic system when the plant and observation noise are non-Gaussian is a difficult problem. This problem is the main focus of this paper. Consider a linear system described as follows:

$$x_{k+1} = \phi_k x_k + w_k \tag{1}$$

$$z_k = H_k x_k + v_k \tag{2}$$

where x_k is the state vector, and w_k and v_k represent white noise sequences and are assumed to be mutually independent. The basic problem is to estimate the state x_k from the noisy observation z_k. The probability density of the state conditioned on all the available observation data is called the *a posteriori* density. If this density is known, an estimation for any type of performance criterion can be easily found. Denote $f(.)$ as a density and $Z^k = \{z_0, z_1, \cdots, z_k\}$. Then, the *a posteriori* density can be recursively determined as follows [1]:

$$f(x_k|Z^k) = \frac{f(x_k|Z^{k-1})f(z_k|x_k)}{f(z_k|Z^{k-1})} \tag{3}$$

$$f(x_k|Z^{k-1}) = \int f(x_{k-1}|Z^{k-1})f(x_k|x_{k-1})\,dx_{k-1} \tag{4}$$

where the normalizing constant $f(z_k|Z^{k-1})$ is given by

$$f(z_k|Z^{k-1}) = \int f(x_k|Z^{k-1})f(z_k|x_k)\,dx_k. \tag{5}$$

Manuscript received September 26, 1994; revised October 27, 1995. The associate editor coordinating the review of this paper and approving it for publication was Dr. Thomas F. Quatieri.

W.-R. Wu is with the Department of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.

A. Kundu is with US WEST Advanced Technologies, 4001 Discovery Dr., Suite 280, Boulder, CO 80303 USA.

Publisher Item Identifier S 1053-587X(96)03943-8.

The $f(z_k|x_k)$ in (3) is determined by the observation noise density $f(v_k)$ and (2). Similarly, $f(x_k|x_{k-1})$ in (4) is determined by the state noise density $f(w_k)$ and (1). Theoretically, knowing these densities, we can determine the *a posteriori* density $f(x_k|Z^k)$. However, it is often impossible to carry out the integration in (4). Consequently, the *a posteriori* density cannot be determined for most applications. The only exception is when the initial state and all the noise sequences are Gaussian. In this case, (3) and (4) lead to Kalman filter equations.

In earlier approaches, attention was focused on the approximation of density functions. Sorenson and Stubberud [1] employed the Edgeworth expansion to approximate the *a posteriori* density. The Edgeworth expansions are characterized by the central moments of the approximated density. Recursive relations were established for a finite number of these moments [1]. Although this approach has some advantages, the approximated densities tend to be negative for some values. To overcome the problem, Sorenson and Alspach [2] developed the Gaussian sum method. In this approach, all the densities are approximated by a mixture of Gaussian densities. Utilizing the properties of Gaussian densities, they were able to keep track of the evolution of the *a posteriori* density. However, the number of Gaussian components will grow exponentially as the filter propagates. The large number of Gaussian components often makes the whole algorithm unmanageable. By assuming that $f(x_k|Z^{k-1})$ is Gaussian, Masreliez derived the recursive estimation equations based on the concept of score function [3], [4]. This algorithm is conceptually attractive, and results are often nearly optimal. However, Masreliez had found difficulties in implementing score functions except for simple cases.

To make the score function based filter applicable to wider class of problems, West [6] had tried to modify Masreliez's method by using an ad hoc scheme. He expanded the *a priori* density $f(z_k|x_k)$ around the state prediction by means of Taylor series and truncated it at the second-order term. By doing that, one can approximate the density $f(z_k|x_k)$ by a Gaussian density. Although this approach does give better results, there is no guarantee that the appropriate Gaussian density should exist. Some of the most common distributions such as uniform and Laplacian distributions cannot be approximated by this method.

Another method, which uses the nonlinear filtering theory, was developed by Makowski *et al.* [16], [17]. The idea here is to employ some probability measure transformation such that the transformed noise is Gaussian, and the standard Kalman filter can be applied. After that, the results are mapped back to the original probability measure. This approach is mathematically involved and computationally expensive. It is shown that when the initial condition is the only non-Gaussian component, the filtering scheme is finite dimensional [18]. However, when the system is driven by non-Gaussian noise, the computation grows linearly with time. Thus, this method may not be suitable for practical implementation in many situations. Some other related nonlinear techniques are also described in the literature [5], [7].

From the above discussions, it is clear that the score function-based filter could be useful in many situations if an efficient method to evaluate the score function is readily available. The focus of this paper is to make Masreliez' filter more applicable by simplifying the evaluation of score function using an approximation. To get this approximation, we have used a distribution expansion scheme. In this process, we need to transform the original distribution to a new one so that the new distribution has its mean at the point where the score function is to be evaluated. Then, we apply the normal expansion technique to this transformed distribution and its derivative. This transformed distribution is called the *conjugate distribution* [10]. The transformation procedure is called *conjugate recentering* [13]. The approximated score function is then obtained by truncating the expansion. It is shown that the score function can be effectively approximated using only a few moments of the distribution under consideration. The main cost of the computation is to find the conjugate distribution that can be easily found by searching for the so-called *saddle point* [11]. A simple Newton method can be used to carry out this process. The approximation technique is objectively verified by detailed simulations.

It is relevant to note here that in some special problems, the score function can be computed in a straightforward manner using the analytic expression. For instance, the Masreliez filter has been used in suppressing narrowband interference in direct sequence spread-spectrum signal [19]. In this case, the observation noise is highly non-Gaussian due to the presence of a chipped spread-spectrum signal. By assuming that the additive channel noise is Gaussian, analytical expressions of the score function are derived in [19]. Under non-Gaussian channel noise, the analytical expressions may not be easy to obtain. An approximation technique could be more effective and, often, the only alternative. The score function also appears in another important practical problem—locally optimum detection of a known signal in additive noise [20], [21]. The locally optimum detector test statistic is obtained as a linear combination of the score function of the noise distribution evaluated at various observations. In this problem, analytically tractable score functions are generally considered. Here again, the technique described in this paper could be useful when the score function under consideration is difficult to handle analytically.

Our paper is structured as follows. A brief description of the Masreliez filter is given in Section II. In Section III, the evaluation of the 1-D score function and its derivative, as applied to the Masreliez filter, is described. The evaluation of the multidimensional score function and its derivative is described in Section IV. Section V contains a detailed description of simulations. Conclusions are also drawn in this section.

II. The Score Function Approach

In this section, we briefly review Masreliez's algorithm [4]. Consider the linear system described in (1) and (2). Denote $f(z_k|Z^{k-1})$ as the density of z_k conditioned on the previous observations. We name $f(z_k|Z^{k-1})$ the observation prediction density (OPD) and assume it is twice differentiable. Similarly, $f(x_k|Z^{k-1})$ is the density of x_k conditioned on the previous observations and is named the state prediction density (SPD). The filtering problem is to estimate the state vector x_k from all the noisy observations Z^k up to the present time k. Let $E\{w_k w_j^t\} = Q_k \delta_{kj}$. Assume that $f(x_k|Z^{k-1})$ is a Gaussian density with mean $\overline{x}_k$ and covariance matrix M_k. Then, as Masreliez has shown, the minimum variance state estimation $\hat{x}_k$ and its covariance matrix $P_k = E\{(x_k - \hat{x}_k)(x_k - \hat{x}_k)^t|Z^k\}$ can be recursively calculated as follows:

$$\hat{x}_k = \overline{x}_k + M_k H_k^t g_k(z_k) \tag{6}$$

$$P_k = M_k - M_k H_k^t G_k(z_k) H_k M_k \tag{7}$$

$$\overline{x}_{k+1} = \phi_k \hat{x}_k \tag{8}$$

$$M_{k+1} = \phi_k P_k \phi_k^t + Q_k \tag{9}$$

where $g_k(.)$ is a column vector with components:

$$\{g_k(z_k)\}_i = -\left[\frac{\partial f(z_k|Z^{k-1})}{\partial (z_k)_i}\right] [f(z_k|Z^{k-1})]^{-1} \tag{10}$$

and $G_k(z_k)$ is a matrix with elements

$$\{G_k(z_k)\}_{ij} = \frac{\partial \{g_k(z_k)\}_i}{\partial (z_k)_j}. \tag{11}$$

The function $g_k(.)$ is the so-called score function of $f(z_k|Z^{k-1})$. Although the above result is an approximate solution, it suggests a way to modify the Kalman filter under a non-Gaussian environment. Assume that w_k is Gaussian. The state estimate of the Kalman filter is given by

$$\hat{x}_k = \overline{x}_k + M_k H_k^t (H_k M_k H_k^t + R_k)^{-1}(z_k - H_k \overline{x}_k) \tag{12}$$

where $R_k = E\{v_k v_k^t\}$. We can see that $\hat{x}_k$ is a linear function of the residual $z_k - H_k \overline{x}_k$. However, the score function $g(.)$ is a nonlinear function of the residual. This nonlinear function de-emphasizes the influence of large residuals when the observation noise density is long tailed and, on the other hand, emphasizes the large residuals when the observation noise density is short tailed. It is easy to check that the Masreliez filter is reduced to the standard Kalman filter if x_0, w_k, and v_k (for all k's) are all Gaussian.

The following procedure summarizes the implementation of the filter.

- Step 0) Assume that at stage $k-1$, $\hat{x}_{k-1}$ and P_{k-1} are known.
- Step 1) Calculate $\overline{x}_k = \phi_{k-1}\hat{x}_{k-1}$ and $M_k = \phi_{k-1} P_{k-1} \phi_{k-1}^t + Q_{k-1}$.
- Step 2) Approximate the SPD $f(x_k|Z^{k-1})$ by a Gaussian distribution with mean $\overline{x}_k$ and covariance matrix M_k.

Step 3) Find the OPD $f(z_k|Z^{k-1})$ by convolving $f(H_k x_k|Z^{k-1})$ with $f_{v_k}(.)$.
Step 4) Find $g_k(z_k)$ and $G_k(z_k)$.
Step 5) Apply (6) and (7) to find $\hat{x}_k$ and P_k.
Step 6) Let $k \to k+1$, and start all over from Step 1.

The procedure outlined above is straightforward in principle. However, the convolution operation in Step 3 is difficult to implement except for simple cases. In addition, in Step 4, the differentiation operations involved in the evaluation of the score function and its derivative are not trivial.

If v_k is Gaussian but w_k is not, we can show that the filter described in (6)–(9) will be reduced to the standard Kalman filter. In order to deal with this problem, Masreliez has derived another filter for the case when w_k is non-Gaussian and v_k is Gaussian. Assume that the system is described as given by (1) and (2), and the observation noise density is Gaussian with zero mean and covariance matrix R_k. Let the matrix $H_k^t R_k^{-1} H_k$ be nonsingular for all k. Then, the minimum variance estimation of x_k and its covariance matrix can be calculated recursively as follows:

$$T_k^{-1} = H_k^t R_k^{-1} H_k \tag{13}$$

$$\hat{x}_k = T_k H_k^t [R_k^{-1} z_k - g_k(z_k)] \tag{14}$$

$$P_k = T_k - T_k H_k^t G_k(z_k) H_k T_k \tag{15}$$

$$\overline{x}_{k+1} = \phi_k \hat{x}_k \tag{16}$$

$$M_{k+1} = \phi_k P_k \phi_k^t + Q_k. \tag{17}$$

Since we do not make the Gaussian assumption for the SPD, the score function can no longer be obtained from the convolution of the SPD and the observation noise density. Denote the convolution operation as "*." We have shown elsewhere that if the density

$$f(H_k \phi_k x_{k-1}|Z^{k-1}) * f(v_k) \tag{18}$$

is assumed to be Gaussian, the score function can be obtained easily. For a detailed discussion, see [15].

III. Evaluation of One-Dimensional Score Functions

In this section, we describe the derivation of the score function approximation scheme. The heart of this approximation procedure is a distribution approximation scheme. By using the concept of adaptive normal approximation of distributions, we find both $f(.)$ and $f'(.)$. From that, we find the expansion of $f'(.)/f(.)$ and truncate it appropriately to obtain the approximation of the score function.

A. Normal Approximation

Let ξ be a random variable with mean m and variance σ^2 and $g(x)$ be the density of $(\xi - m)/\sigma$. It has been shown that $g(x)$ can be expanded as follows [8]:

$$g(x) = \phi(x)\left\{1 + \frac{\rho_3}{3!} H_3(x) + \frac{\rho_4}{4!} H_4(x) + \frac{\rho_5}{5!} H_5(x) + \frac{(\rho_6 + 10\rho_3^2)}{6!} H_6(x) + \cdots \right\} \tag{19}$$

where $\phi(.)$ is a unit normal density, $H_n(x)$'s are Hermite polynomials, which are defined as

$$\left(\frac{\partial}{\partial x}\right)^n e^{-x^2/2} = (-1)^n H_n(x) e^{-x^2/2} \tag{20}$$

and $\rho_n = \kappa_n/\kappa_2^{n/2}$ is the nth standardized cumulants where κ_n denotes the nth cumulant [8] of $f_\xi(x)$. This is called the normal expansion.

An approximation is made by retaining several terms in the expanded series given by (19). This is called the normal approximation. If the distribution being approximated is Gaussian, we expect the normal approximation to be exact. This can be easily verified. The moment generating function (MGF) of a Gaussian distribution (zero mean) is $e^{\sigma^2 T^2/2}$. As a consequence, $\kappa_n = 0$ for $n \geq 3$. If a distribution is close to the Gaussian distribution, its cumulants must be small for $n \geq 3$, and the error is small when the approximation is made with only a few terms. In order to safely use the approximation, some conditions have to be imposed on the distributions being approximated. For example, the distribution has to be unimodal, continuous, and smooth. It is well known that the normal approximation is good around the mean. In the tail region, the approximation can be quite poor (even become negative). Based on this observation, we employ an adaptive expansion scheme which can greatly improve the approximation in the tail region.

B. Adaptive Normal Approximation of Distributions

Let $f(.)$ be a distribution function whose MGF exists in some neighborhood of the origin. In addition, assume that $f(.)$ is continuous and its first few derivatives exist. The basic idea is to use a low-order approximation at each point of the distribution instead of using a high-order approximation at a single point for the whole distribution. Suppose that we want to approximate the distribution at a point, say, x_0. We first transform the original distribution to a distribution that has its mean at x_0. Then, we apply the normal expansion technique to expand the transformed distribution and evaluate it at x_0. Since the normal expansion is good around the mean, this approach is expected to yield a better result than the straightforward normal expansion. The procedure for the transformation is called *recentering* [13], and the transformed distribution is called the *conjugate density* [10]. We now formally state the definition.

Definition: If there exist constants α and T_0 such that $g(z) = \alpha e^{T_0 z} f(z + x_0)$ is a density with zero mean, $g(z)$ is called the conjugate density of $f(x)$ at the point x_0.

If we assume that the conjugate density is known, $g(z)$ can be normalized and expanded as in (19):

$$g(z) = \frac{1}{\sigma_z} \phi\left(\frac{z}{\sigma_z}\right)\left\{1 + \frac{\rho_{z,3}}{3!} H_3\left(\frac{z}{\sigma_z}\right) + \frac{\rho_{z,4}}{4!} H_4\left(\frac{z}{\sigma_z}\right) + \cdots \right\} \tag{21}$$

where σ_z and $\rho_{z,i}$ are, respectively, the standard deviation and the ith standardized cumulant of $g(z)$. From the definition, we know that $f(z+x_0) = \alpha^{-1}e^{-T_0 z}g(z)$. Then, $f(x_0) = \alpha^{-1}g(0)$.

$$g(0) = \frac{\phi(0)}{\sigma_z}\left\{1 + \frac{1}{8}\rho_{z,4} - \frac{1}{48}\left(\rho_{z,6} + 10\rho_3^2\right) + \cdots\right\}. \quad (22)$$

Note that in (22), the odd terms are all equal to zero. If the distribution is close to a Gaussian one, only the first term in the series would provide a good approximation, i.e.

$$\begin{aligned} f(x_0) &= \alpha^{-1}g(0) \\ &\approx \frac{\phi(0)}{\alpha\sigma_z}. \end{aligned} \quad (23)$$

In order to use (23), one has to find the conjugate density. This is carried out as follows. Let the MGF of $f(x)$ be $M(T)$ and $K(T)$ be $\ln[M(T)]$. $K(T)$ is called the cumulant MGF [9]. Then

$$\begin{aligned} M(T) &= e^{K(T)} \\ &= \int_{-\infty}^{+\infty} e^{Tx} f(x)\, dx. \end{aligned} \quad (24)$$

Let T be real. Multiplying both sides of (24) by e^{-Tx_0}, changing the variable x to s where $s = x - x_0$, and differentiating both sides with respect to T, we have

$$\frac{\partial}{\partial T} e^{K(T)-Tx_0} = \frac{\partial}{\partial T}\int_{-\infty}^{+\infty} e^{Ts} f(s+x_0)\, ds. \quad (25)$$

Let us consider the infinite series expansion of the exponential function inside integral in (25). Since MGF exists for a range of T, within this range, each integral corresponding to one term should also exist. Thus, MGF becomes a power series in T. This power series converges for a specific range of T values. Then, one can differentiate this power series with respect to T. At the same time, one can differentiate e^{Ts} under the integral sign and do term-by-term integration. Both sides turn out to be equal. Therefore, one can interchange integral and differentiation [22] in this case.

$$e^{K(T)-Tx_0}[K'(T) - x_0] = \int_{-\infty}^{+\infty} s e^{Ts} f(s+x_0)\, ds. \quad (26)$$

Now, if we choose T_0 and α such that

$$K'(T_0) - x_0 = 0 \quad (27)$$

and

$$\begin{aligned} \frac{1}{\alpha} &= \int_{-\infty}^{+\infty} e^{T_0 s} f(s+x_0)\, ds \\ &= e^{K(T_0)-T_0 x_0} \end{aligned} \quad (28)$$

then

$$\begin{aligned} &\int_{-\infty}^{+\infty} \alpha e^{T_0 s} f(s+x_0)\, ds = 1 \\ &\int_{-\infty}^{+\infty} s\alpha e^{T_0 s} f(s+x_0)\, ds = 0. \end{aligned} \quad (29)$$

Hence, $g(s) = \alpha e^{T_0 s} f(s+x_0)$ is the conjugate density.

The solution of $K'(T) - x_0 = 0$ is referred to as the saddle point of $e^{-Tx_0}M(T)$. The existence of the saddle point is discussed in [11]. We assume that $f(x) = 0$ for $x < a$, $x > b$ and $M(T)$ converges for $-d_1 < T < d_2$. Note that a, b, d_1 and d_2 can be finite or infinite. If $a < x < b$ and

$$\lim_{T \to -d_1} K'(T) = a$$

and

$$\lim_{T \to d_2} K'(T) = b \quad (30)$$

then a unique real root for $K'(T) = x$ exists, and as T increases from $-d_1$ to d_2, $K'(T)$ increases continuously from $x = a$ to $x = b$. The condition is easily satisfied for a wide class of distributions.

Once the conjugate density is found, its moments can also be found. This is done by differentiating (26) n times and using (27) and (28). Denote the nth moments by $\mu_{z,n}$, and then

$$\mu_{z,n} = K^{(n)}(T_0). \quad (31)$$

The relation of moments and cumulants (denoted by $\kappa_{z,n}$) can be found in [8]. In particular, we have $\kappa_{z,n} = \mu_{z,n}$ for $n = 1, 2, 3$.

The MGF of a distribution is nothing but the Laplace transform of the distribution. One of the most important properties of the Laplace transform is that the convolution in the temporal or spatial domain can be transformed into multiplication in the frequency domain. This property is directly applicable to MGF's. This is also the key concept that allows us to avoid the convolution operation involved in the estimation of the score function as required in Masreliez's approach. This is discussed next.

C. Adaptive Normal Approximation of Score Functions

The distribution approximation technique discussed above can also be extended to find the approximation of score function. The idea is to find the expansion of $f(.)$ and $f'(.)$ via the conjugate recentering. From that, we find the expansion of $f'(.)/f(.)$ and truncate it to obtain the approximation. Since $f(.)$ is derivable at each point, we can obtain $f'(.)$ through a term-by-term differentiation of the expansion of $f(.)$. First, we construct the conjugate density $g(z)$ at x_0 and express $f(x)$ in terms of $g(z)$.

$$\begin{aligned} f(x) &= f(z+x_0) \\ &= \alpha^{-1}e^{-T_0 z}g(z). \end{aligned} \quad (32)$$

Then

$$f'(x) = -T_0\alpha^{-1}e^{-T_0 z}g(z) + \alpha^{-1}e^{-T_0 z}g'(z) \quad (33)$$

$$\left.\frac{f'(x)}{f(x)}\right|_{x=x_0} = -T_0 + \left.\frac{g'(z)}{g(z)}\right|_{z=0}. \quad (34)$$

The expansion of $g(z)$ using (21) is given by

$$\begin{aligned} g(z) = \frac{1}{\sigma_z}\phi\left(\frac{z}{\sigma_z}\right)&\left\{1 + \frac{\rho_{z,3}}{3!}H_3\left(\frac{z}{\sigma_z}\right)\right. \\ &\left. + \frac{\rho_{z,4}}{4!}H_4\left(\frac{z}{\sigma_z}\right) + \cdots\right\}. \end{aligned} \quad (35)$$

Then

$$g'(z) = -\frac{1}{\sigma_z^2}\phi\left(\frac{z}{\sigma_z}\right)\left\{H_1\left(\frac{z}{\sigma_z}\right) + \frac{\rho_{z,3}}{3!}H_4\left(\frac{z}{\sigma_z}\right) + \frac{\rho_{z,4}}{4!}H_5\left(\frac{z}{\sigma_z}\right) + \cdots\right\} \quad (36)$$

$$g(0) = \frac{\phi(0)}{\sigma_z}\left\{1 + \frac{1}{8}\rho_{z,4} - \frac{1}{48}(\rho_{z,6} + 10\rho_3^2) + \cdots\right\} \quad (37)$$

$$g'(0) = -\frac{\phi(0)}{\sigma_z^2}\left\{\frac{1}{2}\rho_{z,3} - \frac{1}{8}\rho_{z,5} + \cdots\right\}. \quad (38)$$

Retaining the first terms in (37) and (38), we obtain the approximation of the score function

$$\frac{f'(x_0)}{f(x_0)} \approx -T_0 - \frac{\rho_{z,3}}{2\sigma_z}. \quad (39)$$

To implement Masreliez's filter, we also need the derivative of the score function. To do this, we can differentiate (34) and truncate the expanded series [involving $g''(.)$]. Thus, this procedure would be similar to what we have done for the score function. Instead, we propose a simpler alternative that finds the derivative of the approximated score function. This is done in the following manner. The relationship between x and T is established through the equation $K'(T) = x$. Thus

$$K''(T)\frac{dT}{dx} = 1$$
$$\frac{dT}{dx} = \frac{1}{K''(T)}. \quad (40)$$

Take the derivative of the score function in (39) with respect to x. Then

$$\frac{d}{dx}\left(\frac{f'}{f}\right)\Big|_{x=x_0} \approx \left\{-\frac{dT}{dx} - \frac{dT}{dx}\frac{d}{dT}\left[\frac{K^{(3)}(T)}{2K''(T)^2}\right]\right\}\Big|_{T=T_0} \approx -\frac{1}{\sigma_z^2}\left[1 + \frac{\mu_{z,4}}{2\sigma_z^4} - \frac{\mu_{z,3}}{\sigma_z^6}\right]. \quad (41)$$

The approximation results we have obtained here are closely related to some of the asymptotic approximation techniques in statistics. Elsewhere [15], we have shown the equivalence of our distribution and score function approximation schemes with Daniels' saddle-point method [11] and Hampel's small sample asymptotic method when the sample size is one [13]. For detailed discussion, see [15].

D. Error Analysis

The approximation scheme described above usually works very well for distributions not deviating too much from the Gaussian form. Since a general error analysis is very difficult, we show here only some analysis using the results in [11]. As we know, the standard normal approximation becomes poor as x approaches the ends of its admissible range. Therefore, we will investigate the error behavior of our approximations around the end points. From (30), we know that as $x \to a, b$, the saddle point $\to -d_1, d_2$. It is shown in [11] that for a wide class of distributions, their conjugate densities approximate either to the gamma form or to the normal form as the saddle point $\to -d_1, d_2$. In the first case, the standardized cumulant $\rho_{z,n}$ is bounded for a given n; in the second case, $\rho_{z,n} \to 0$. In Appendix A, we define distributions with the gamma and the normal form and show that if conjugate densities belong to these kinds, approximations of score functions and their derivatives are also either bounded or zero.

E. Implementation of the Score Function of the OPD

Next, we apply the score function approximation technique derived above to the Masreliez algorithm. We assume that the MGF of $f(v_k)$ is known. Since $f(x_k|Z^{k-1})$ is Gaussian with known mean and variance, the MGF of $f(H_k x_k|Z^{k-1})$ can be easily found. Let the MGF's of $f(v_k)$ and $f(H_k x_k|Z^{k-1})$ be $M_v(T)$ and $M_x(T)$, respectively.

1) *Find the MGF of* $f(z_k|Z^{k-1})$: Since $f(z_k|Z^{k-1})$ is obtained from the convolution of $f(v_k)$ and $f(H_k x_k|Z^{k-1})$, the MGF of $f(z_k|Z^{k-1})$ is $M_v(T)M_x(T)$.
2) *Find the conjugate density of* $f(z_k|Z^{k-1})$ *at* z_k: Let $K(T) = \ln(M_v(T)M_x(T))$. The conjugate density of $f(z_k|Z^{k-1})$ at z_k is constructed as $g(s) = \alpha_k e^{T_k s} f(s + z_k)$, where T_k is chosen as the saddle point of $\{M_v(T)M_x(T)e^{-Tz_k}\}$, i.e., $K'(T_k) - z_k = 0$, and $1/\alpha_k$ as $\{M_v(T_k)M_x(T_k)e^{-T_k z_k}\}$.
3) *Find the second, the third, and the fourth moment of* $g(s)$: σ_s^2, $\mu_{s,3}$, and $\mu_{s,4}$ can be found by

$$\sigma_s^2 = K''(T_k)$$
$$\mu_{s,3} = K^{(3)}(T_k)$$
$$\mu_{s,4} = K^{(4)}(T_k). \quad (42)$$

4) *Approximate the score function of* $f(z_k|Z^{k-1})$ *and its derivative by*

$$-\frac{f'(z_k|Z^{k-1})}{f(z_k|Z^{k-1})} \approx T_k + \frac{\mu_{s,3}}{2\sigma_s^4} \quad (43)$$

$$\frac{d}{dz_k}\left(-\frac{f'}{f}\right) = \frac{1}{\sigma_s^2}\left[1 + \frac{\mu_{s,4}}{2\sigma_s^4} - \frac{\mu_{s,3}^2}{\sigma_s^6}\right]. \quad (44)$$

The density being approximated is the OPD. As we know, this density is obtained by a convolution of the SPD and the noise density. Since we assume that the SPD is Gaussian, if the noise density is well behaved and unimodal, the general shape of the OPD will not be very different from a Gaussian density. This is the rationale for using the adaptive normal approximation scheme in the filtering problem. If the noise density is not unimodal, we can first approximate it by a weighted sum of some well-behaved unimodal densities and apply the adaptive normal approximation to the individual OPD (SPD convolved with the individual unimodal density). By doing so, the score function can be easily found. This is elaborated upon in the next paragraph.

If the noise distribution is composed of two of more distributions, i.e., a weighted sum of two or more distributions, the calculation of saddle points and moments can be very tedious. Here, we propose a simple method that can overcome this problem. Let the distribution $f(.)$ be a weighted sum of two distributions $f_1(.)$ and $f_2(.)$ with weights w_1 and w_2, where $w_1 + w_2 = 1$. The score function and its derivative for the distribution $f_i(.)$ are denoted by $g_i(.)$ and $G_i(.)$ $(i = 1, 2)$, respectively. Define

$$w_a = \frac{w_1 f_1(x)}{w_1 f_1(x) + w_2 f_2(x)}$$

and

$$w_b = \frac{w_2 f_2(x)}{w_1 f_1(x) + w_2 f_2(x)}. \tag{45}$$

Then, it is simple to show that the score function of $f(.)$, and its derivative can be calculated as

$$g(x) = w_a g_1(x) + w_b g_2(x) \tag{46}$$

$$G(x) = w_a[G_1(x) - g_1(x)^2] + w_b[G_2(x) - g_2(x)^2] + g(x)^2. \tag{47}$$

Note that the method can be easily generalized to find the score function and its derivative for a distribution that is a weighted sum of more than two distributions.

It is possible to include more terms in the expanded series to improve the approximation accuracy. However, to include the next available term in the expanded series, we have to use cumulants up to the sixth order. For a well-behaved and Gaussian-like distribution such as the OPD, the increase in the computation complexity is often not warranted.

IV. Evaluation of Multidimensional Score Functions

In this section, we extend the results obtained in Section III to systems with multidimensional observations. In these kinds of systems, the OPD becomes multivariate, and the filtering problem becomes more complicated. We assume that all the necessary regularity conditions mentioned in Section III also hold here. Using techniques similar to those developed in Section III, we derive the approximation formula for the score function of a multivariate distribution. For the sake of clarity, we only consider the 2-D systems. For systems with more dimensions, the same procedure can be followed in a straightforward manner.

A. The Bivariate Normal Expansion

Using a procedure similar to that described in Section III, we can find the expansion of a multivariate distribution. However, since the expansion involves the multivariate Hermite polynomials, it is not appropriate to use this direct expansion. The reason is that in the multivariate Hermite polynomials, the variables are coupled with each other, and the expansion formula becomes very complex after differentiation. Here, we propose another approach that reduces the complexity of the operation. The main idea is to transform the original distribution to a distribution that has uncorrelated components such that univariate Hermite polynomials can be applied. The procedure can be described as follows:

1) Perform a linear transformation that will make the components of the distribution uncorrelated.
2) Expand the distribution.
3) Perform the inverse transform to obtain the expansion of the original distribution.

Consider a zero mean bivariate distribution $f(x, y)$. If $\sigma_x^2 = 1$, $\sigma_y^2 = 1$, and $E(xy) = 0$, $f(x, y)$ can be expanded as [12]

$$\begin{aligned} f_{x,y}(x, y) &= \\ &\phi(x)\phi(y)\left\{1 + \frac{1}{3!}(H^t\rho)^{[3]}(x, y)\right. \\ &+ \frac{1}{4!}(H^t\rho)^{[4]}(x, y) + \frac{1}{5!}(H^t\rho)^{[5]}(x, y) \\ &\left.+ \frac{1}{6!}[(H^t\rho)^{[6]}(x, y) + 10\{(H^t\rho)^{[3]}\}^2(x, y)] + \cdots\right\} \end{aligned} \tag{48}$$

where $\phi(.)$ is a unit normal density, $(H^t\rho)^{[i]}$ is the expansion of $[H(x)\rho_1 + H(y)\rho_2]^i$ in which $\rho_1^j\rho_2^k$ $(j + k = i)$ is replaced by ρ_{jk} and $H(x)^j H(y)^k$ by $H_j(x)H_k(y)$, and $\{(H^t\rho)^{[i]}\}^l$ is the expansion of $(H^t\rho)^{[i]}$ in which $H_j(x)^m H_k(y)^n$ is replaced by $H_{j\times m}(x)H_{k\times n}(y)$. For example

$$\begin{aligned} (H^t\rho)^{[3]}(x, y) &= \rho_{30}H_3(x) + 3\rho_{21}H_2(x)H_1(y) \\ &\quad + 3\rho_{12}H_1(x)H_2(y) + \rho_{03}H_3(y) \end{aligned} \tag{49}$$

$$\begin{aligned} \{(H^t\rho)^{[3]}\}^2(x, y) &= \rho_{30}^2H_6(x) + 9\rho_{21}^2H_4(x)H_2(y) \\ &\quad + \cdots + \rho_{03}^2H_6(y) \end{aligned} \tag{50}$$

where $H_n(x)$ is the Hermite polynomial with degree n, and ρ_{ij} is the normalized cumulant expressed as $k_{ij}/(k_{20}^{i/2}k_{02}^{j/2})$; k_{ij}'s are the cumulants that are the expanded coefficients of the cumulant MGF.

If x and y are correlated, we apply the transformation procedure outlined above to obtain the expansion. The transformation can be realized by the Gram–Schmidt orthogonalization procedure. Let $x_t = x$ and $y_t = y - cx_t$. If we choose $c = E(xy)/E(x^2)$, then $E(x_t y_t) = 0$. Then, $f(x_t, y_t)$ can be expanded using (48)

$$\begin{aligned} f_{x_t,y_t}(x, y) &= \\ &\frac{\phi(x)\phi(y)}{\sigma_{x_t}\sigma_{y_t}}\left[1 + \frac{1}{6}(H^t\rho_t)^{[3]}\left(\frac{x}{\sigma_{x_t}}, \frac{y}{\sigma_{y_t}}\right) + \cdots\right] \end{aligned} \tag{51}$$

where ρ_t is the normalized cumulant of the transformed distribution. Now, we can transform back to find the expansion of $f_{x,y}(., .)$. Note that the Jacobian is one. Thus

$$\begin{aligned} f_{x,y}(x, y) = \frac{\phi(x)\phi(y - cx)}{\sigma_{x_t}\sigma_{y_t}}&\left[1 + \frac{1}{6}(H^t\rho_t)^{[3]}\right. \\ &\left.\cdot\left(\frac{x}{\sigma_{x_t}}, \frac{y - cx}{\sigma_{y_t}}\right) + \cdots\right]. \end{aligned} \tag{52}$$

B. Adaptive Bivariate Normal Expansion

The basic idea of applying the conjugate densities in the adaptation scheme can be used directly with a minor modification. Thus, we will omit the derivation whenever it is similar to that of the 1-D case. Let $z = (z_1, z_2)^t$, $a = (a_1, a_2)^t$, and $x = (x_1, x_2)^t$. We now define the bivariate conjugate density.

Definition: If there exist a constant scalar α and a constant vector T_0 such that $g(z) = \alpha e^{T_0^t z} f(z + a)$ is a density with zero mean, $g(z)$ is called the conjugate density of $f(x)$ at a point a.

Let z_1 and z_2 be uncorrelated. We then can express $f(x)$ in terms of $g(z)$ and perform the normal expansion on $g(z)$

using (48). If we only retain the first term in the expansion, we obtain an approximation of $f(a)$, which is similar to (23)

$$f(a) = \alpha^{-1} g(0) \approx \frac{\phi(0)^2}{\alpha \sigma_{z_1} \sigma_{z_2}}. \tag{53}$$

By using the idea of the saddle point, we can find the bivariate conjugate density. Let the cumulant MGF of $f(x)$ be $K(T)$. The saddle point now becomes a vector, and we have to solve for two unknowns in two equations, i.e.

$$\frac{\partial}{\partial T_1} K(T) - a_1 = 0$$
$$\frac{\partial}{\partial T_2} K(T) - a_2 = 0. \tag{54}$$

Thus, T_0 is chosen as the solution of (54), and $1/\alpha = e^{K(T_0) - T_0^t a}$.

Under some regularity conditions, a unique solution to (54) exists. Detailed discussion can be found in [14]. The moments of the conjugate density are computed as in (31). Denote the ijth moment of $g(z)$ by $\mu_{z,ij}$. Then

$$\mu_{z,ij} = \frac{\partial^{(i+j)}}{\partial T_1^i \partial T_2^j} K(T_0). \tag{55}$$

The relationship between bivariate moments and cumulants can be found in [9]. Particularly, we have $\kappa_{z,ij} = \mu_{z,ij}$ for $i + j = 3$.

C. Adaptive Approximation of Bivariate Score Functions

Similar to the 1-D case, the distribution approximation scheme can be used for the score function approximation. Suppose that we want to evaluate the score function of $f(x)$ at a point a. First, we construct the conjugate density of $f(x)$ at the point a. Let $z = x - a$. The conjugate density is $g(z) = \alpha e^{T_0^t z} f(z + a)$. Then

$$f(x) = \alpha^{-1} e^{-T_0^t z} g(z) \tag{56}$$
$$f'(x) = -T_0 \alpha^{-1} e^{-T_0^t z} g(z) + \alpha^{-1} e^{-T_0^t z} g'(z) \tag{57}$$
$$\left.\frac{f'(x)}{f(x)}\right|_{x=a} = -T_0 + \left.\frac{g'(z)}{g(z)}\right|_{z=0}. \tag{58}$$

Next, we uncorrelate z_1 and z_2. Let $\zeta_1 = z_1$ and $\zeta_2 = z_2 - cz_1$, where $c = E(z_1 z_2)/E(z_1^2)$. Using (52), we have

$$g(z_1, z_2) = \frac{\phi(z_1)\phi(z_2 - cz_1)}{\sigma_{\zeta_1}\sigma_{\zeta_2}} \left[1 + \frac{1}{6} (H^t \rho_\zeta)^{[3]} \cdot \left(\frac{z_1}{\sigma_{\zeta_1}}, \frac{z_2 - cz_1}{\sigma_{\zeta_2}}\right) + \cdots \right]. \tag{59}$$

Taking the derivative of (59) and noting that $\phi'(z)|_{z=0} = 0$, we have

$$\left.\frac{\partial}{\partial z_1} g(z_1, z_2)\right|_{z=0} = \frac{\phi(z_1)\phi(z_2 - cz_1)}{\sigma_{\zeta_1}\sigma_{\zeta_2}} \cdot \left.\left\{ \frac{\partial}{\partial z_1} (H^t \rho_\zeta)^{[3]} \left(\frac{z_1}{\sigma_{\zeta_1}}, \frac{z_2 - cz_1}{\sigma_{\zeta_2}}\right) + \cdots \right\}\right|_{z=0} \tag{60}$$

$$\left.\frac{\partial}{\partial z_2} g(z_1, z_2)\right|_{z=0} = \frac{\phi(z_1)\phi(z_2 - cz_1)}{\sigma_{\zeta_1}\sigma_{\zeta_2}} \cdot \left.\left\{ \frac{\partial}{\partial z_2} (H^t \rho_\zeta)^{[3]} \left(\frac{z_1}{\sigma_{\zeta_1}}, \frac{z_2 - cz_1}{\sigma_{\zeta_2}}\right) + \cdots \right\}\right|_{z=0} \tag{61}$$

$$\left.\frac{\partial}{\partial z_1} (H^t \rho_\zeta)^{[3]} \left(\frac{z_1}{\sigma_{\zeta_1}}, \frac{z_2 - cz_1}{\sigma_{\zeta_2}}\right)\right|_{z=0} = -\frac{3}{\sigma_{\zeta_1}} \rho_{\zeta,30} + \frac{3c}{\sigma_{\zeta_2}} \rho_{\zeta,21} - \frac{3}{\sigma_{\zeta_1}} \rho_{\zeta,12} + \frac{3c}{\sigma_{\zeta_2}} \rho_{\zeta,03} \tag{62}$$

$$\left.\frac{\partial}{\partial z_2} (H^t \rho_\zeta)^{[3]} \left(\frac{z_1}{\sigma_{\zeta_1}}, \frac{z_2 - cz_1}{\sigma_{\zeta_2}}\right)\right|_{z=0} = -\frac{3}{\sigma_{\zeta_2}} \rho_{\zeta,03} - \frac{3}{\sigma_{\zeta_2}} \rho_{\zeta,21}. \tag{63}$$

Similar to the 1-D case, we retain the first term in the expansion of $g'(0)/g(0)$. Hence, we obtain the approximation of the score function as

$$\frac{f'(a)}{f(a)} \approx -T_0 - \frac{1}{2} \begin{pmatrix} \dfrac{\rho_{\zeta,30} + \rho_{\zeta,12}}{\sigma_{\zeta_1}} - \dfrac{c(\rho_{\zeta,21} + \rho_{\zeta,03})}{\sigma_{\zeta_2}} \\ \dfrac{1}{\sigma_{\zeta_2}} (\rho_{\zeta,03} + \rho_{\zeta,21}) \end{pmatrix}. \tag{64}$$

The ijth central moment of $g_\zeta(\zeta)$ denoted as $\mu_{\zeta,ij}$ is not directly obtainable, but we can use the ijth central moment of $g(z)$ as computed by (55) to find it. For example

$$\mu_{\zeta,12} = E(\zeta_1 \zeta_2^2) = E[z_1(z_2 - cz_1)^2] = \mu_{z,12} - 2c\mu_{z,21} + c^2 \mu_{z,30}. \tag{65}$$

By the same procedure, we find

$$\mu_{\zeta,20} = \mu_{z,20},$$
$$\mu_{\zeta,02} = \mu_{z,02} - 2c\mu_{z,11} + c^2 \mu_{z,20} \tag{66}$$
$$\mu_{\zeta,30} = \mu_{z,30},$$
$$\mu_{\zeta,03} = \mu_{z,03} - 3c\mu_{z,12} + 3c^2 \mu_{z,21} - c^3 \mu_{z,30} \tag{67}$$
$$\mu_{\zeta,12} = \mu_{z,12} - 2c\mu_{z,21} + c^2 \mu_{z,30},$$
$$\mu_{\zeta,21} = \mu_{z,21} - c\mu_{z,30}. \tag{68}$$

The last entity we have to calculate is the derivative of the score function. In the bivariate case, this gives in a 2×2 matrix. We show the result in Appendix B.

D. Implementation of the Bivariate Score Function of the OPD

Let the MGF of the SPD $f(x_k|Z^{k-1})$ be $M_x(T)$ and the MGF of the observation noise $f(v_k)$ be $M_v(T)$. The implementation of the bivariate score function of the OPD $f(z_k|Z^{k-1})$ can be summarized as follows:

1) *Find the MGF of* $f(z_k|Z^{k-1})$. The MGF of $f(H_k x_k|Z^{k-1})$ is found as $M_x(H_k^t T)$. Therefore, the MGF of $f(z_k|Z^{k-1})$ is $M_x(H_k^t T) M_v(T)$.
2) *Find the conjugate density of* $f(z_k|Z^{k-1})$ *at point* z_k by solving (54).

3) *Find transform coefficient* c *and the* ijth *moments of the conjugate density* using (55).
4) *Find the* ijth *moments of the transformed conjugate density* using (65)–(68).
5) *Find the score function and its derivative* using (64) and the procedure outlined in Appendix B.

In this section, we have discussed the bivariate score function approximation scheme for the sake of simplicity. This scheme can be easily extended to more general multidimensional cases. In general multidimensional cases, however, one has to start with the appropriate equivalent formulae of (48). For the linear transformation and the expansion, the same procedure as used in the bivariate case can be used.

V. Simulations and Conclusions

In the previous sections, we have discussed the new score function approximation technique. In this section, we present some simulation results. For non-Gaussian distributions, we have used a mixed model as shown below.

$$f(x) = pf_i(x - \mu_i) + (1 - p)f_j(x - \mu_j). \tag{69}$$

The subscript of f (i and j) stands for type of distribution and μ for the mean of the distribution. We only consider the two most common types of distribution, namely, "Gaussian" and "Laplacian," which are denoted by g and l, respectively. These distributions are defined as follows:

$$\text{Gaussian} \quad f_g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \tag{70}$$

$$\text{Laplacian} \quad f_l(x) = \frac{1}{2\eta} e^{-|x|/\eta}. \tag{71}$$

We first show two examples of score function approximation. The distributions used are obtained by convolving a Gaussian distribution with non-Gaussian distributions. This type of distribution is what we encounter often in the filtering problem. The non-Gaussian distributions are 1) Gaussian and Laplacian mixed distribution and 2) Laplacian and Laplacian mixed distribution. The Gaussian distribution is assumed to have zero mean and variance σ^2. Using the model in (69), we specify the parameters as follows:

1) $p = 0.99$; $i = g$, $\mu_i = 0$, $\sigma_i^2 = 1$; $j = l$, $\mu_j = 0$, $\eta = 5$; $(\sigma^2 = 1)$.
2) $p = 0.5$; $i = j = l$, $\mu_i = -\mu_j = 3$; $\eta_i = \eta_j = 1$, $(\sigma^2 = 1)$.

Figs. 1 and 2 show the approximation results. It can be seen clearly that the approximation scheme is quite satisfactory. In particular, for case one, we can barely distinguish the real from the approximated one. This is because the major component of the distribution is Gaussian and the only approximation error is due to a small Laplacian component. In case two, two components are both non-Gaussian. The approximation error is consequently larger. It is also easy to find that the error is mainly concentrated on the areas where the slopes of the score function go through rapid changes. We have also compared the average squared error between the approximated and the exact score function for this case. We have found that it is 1.1642×10^{-3}. This numerical figure shows that the approximation error is small.

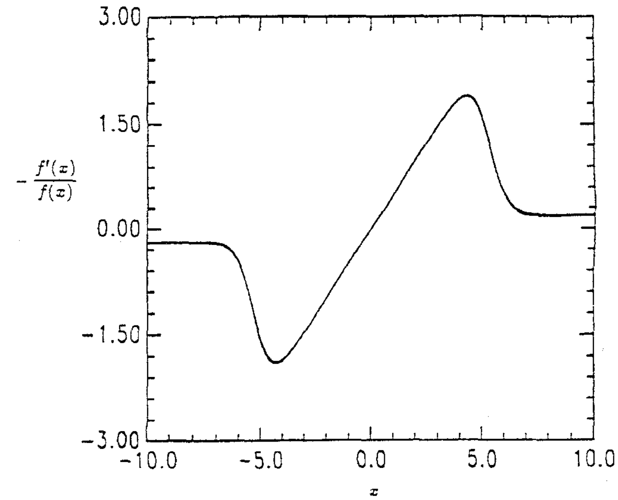


Fig. 1. Approximated and the exact score functions; the mixture of Gaussian and Laplacian convolved with Gaussian distribution.

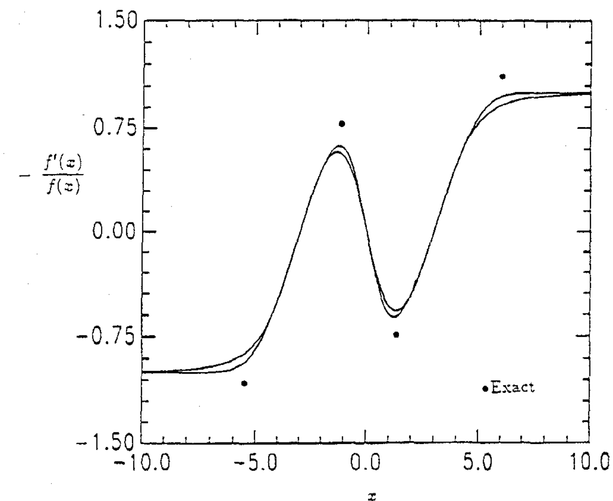


Fig. 2. Approximated and the exact score functions; the mixture of two Laplacian convolved with Gaussian distribution.

Next, we consider the random signal filtering problem. We assume that the random signal is a first-order autoregressive (AR) process and corrupted by additive white noise. This can be described by the following signal model:

$$x_{k+1} = \phi x_k + w_k \tag{72}$$

$$z_k = Hx_k + v_k \tag{73}$$

where

- x_k original signal
- z_k observed signal
- w_k driving noise
- v_k observation noise
- $H = 1$.

Two cases are simulated. We first consider the case where the driving noise is Gaussian and the observation noise is non-

TABLE I
COMPARISON OF KALMAN AND MASRELIEZ FILTERS; GAUSSIAN DRIVING NOISE; GAUSSIAN+LAPLACIAN OBSERVATION NOISE; $\eta = 7.07$; $\phi = 0.8$

$f_v(.)$: Gaussian+Laplacian $\sigma_g^2 = 4, \sigma_l^2 = 100, \eta = 7.07$				$\phi = 0.8, f_w(.)$: Gaussian $\sigma_w^2 = 3.6, \sigma_x^2 = 10, \sigma_{x_0}^2 = 5$	
σ_v^2/σ_x^2	σ_v^2	p	Kalman	Masreliez	Improvement
0.5	5	1.042%	2.514813	2.302512	8.442%
1.0	10	6.250%	3.887884	2.717266	30.11%
1.5	15	11.46%	4.613065	2.954992	35.94%
2.0	20	16.67%	5.374556	3.235764	39.79%
2.5	25	21.88%	5.628642	3.486571	38.06%

TABLE II
COMPARISON OF KALMAN AND MASRELIEZ FILTERS; GAUSSIAN DRIVING NOISE; GAUSSIAN+LAPLACIAN OBSERVATION NOISE; $\eta = 7.07$; $\phi = 0$

$f_v(.)$: Gaussian+Laplcian $\sigma_g^2 = 4, \sigma_l^2 = 100, \eta = 7.07$				$\phi = 0, f_w(.)$: Gaussian $\sigma_w^2 = 10, \sigma_x^2 = 10, \sigma_{x_0}^2 = 5$	
σ_v^2/σ_x^2	σ_v^2	p	Kalman	Masreliez	Improvement
0.5	5	1.042%	3.242469	2.955587	8.839%
1.0	10	6.250%	5.085141	3.721616	26.81%
1.5	15	11.46%	6.139159	4.093556	33.32%
2.0	20	16.67%	6.941551	4.653338	32.96%
2.5	25	21.88%	7.218412	5.001915	30.71%

Gaussian. Then, we consider the case where the driving noise is non-Gaussian and the observation noise is Gaussian. The non-Gaussian noise considered here is Gaussian and Laplacian mixed noise. By using (69), the distributions can be described as $i = g$; $j = l$; $\mu_i = \mu_j = 0$. Two different values of ϕ are used to test the performance of the filter for signals with both high and low correlation. The values of ϕ chosen in our experiments are 0.8 and 0, respectively. A Monte Carlo experiment is then conducted. The signal is generated with 100 points, and the noise sequence is added to the generated signal. The experiment is carried out 100 times. The mean square error is taken as the performance index

$$\text{MSE} = \frac{1}{10000} \sum_{j=1}^{100} \sum_{i=1}^{100} [x_j(i) - \hat{x}_j(i)]^2. \tag{74}$$

To show the effect of nonlinear filtering, we have compared the performance of the Kalman filter vis-à-vis the Masreliez filter. The improvement measure is obtained as follows:

$$\text{Improvement} = \frac{\text{MSE}_{\text{Kalman}} - \text{MSE}_{\text{Masreliez}}}{\text{MSE}_{\text{Kalman}}}. \tag{75}$$

Tables I and II summarize the experimental results for the non-Gaussian observation noise. In terms of MSE, the Masreliez filter provides significant improvement. Figs. 3–5 show a sample run (Table I, $p = 11.46\%$) of the experiment. From these figures, one can see that the standard Kalman filter fails to filter out the high variance Laplacian noise. In addition, the Kalman filter overestimates the variance of Gaussian noise, resulting in too much smoothing of the signal. On the other hand, the score-function based filter efficiently removes both Gaussian and Laplacian noises and restores the signal nicely. Similar procedures are used to test the performance of the filters for non-Gaussian driving noise. We have used the same system and the same type of noises, but the driving noise and the observation noise are interchanged. Tables III and IV summarize the experimental results. Figs. 6–8 show a sample run (Table III, $p = 11.28\%$) of the experiment. We can see that the standard Kalman filter uses a fixed gain and cannot retain the non-Gaussian signal. The score-function-based filter adaptively changes the gain and successfully restores the signal.

To compare the computational complexities of the Masreliez algorithm with the Kalman filter, we consider the system model in (72) and (73). Let the dimension of x_k be $n \times 1$ and H be $1 \times n$. For one cycle, the Kalman filter requires $2n^3 + 3n^2 + 2n + 3$ multiplications and one division. If the distribution of observation noise is a Gaussian mixture (as Masreliez have used in [4]), the Masreliez filter requires $2n^3 + 3n^2 + 2n + 9$ multiplications and seven division, one exponential, and one square root operations. If the distribution of observation noise is a mixed Gaussian/Laplacian distribution (as we have used in previous simulations), then the Masreliez filter requires

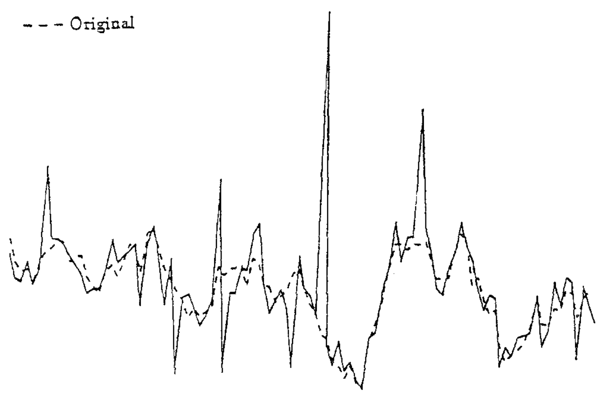


Fig. 3. Corrupted signal; Gaussian+Laplacian observation noise.

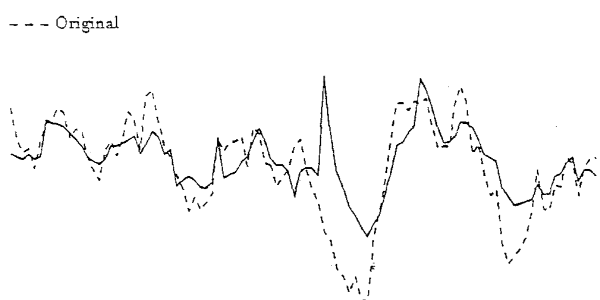


Fig. 4. Kalman filtered signal.

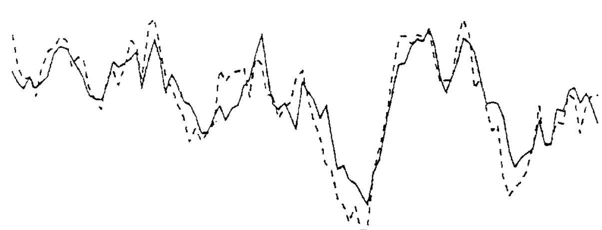


Fig. 5. Masreliez filtered signal.

TABLE III
COMPARISON OF KALMAN AND MASRELIEZ FILTERS; GAUSSIAN OBSERVATION NOISE; GAUSSIAN+LAPLACIAN DRIVING NOISE; $\eta = 4.472$; $\phi = 0$

$f_w(.)$: Gaussian+Laplacian $\sigma_g^2 = 1, \sigma_l^2 = 40, \eta = 4.472$				$\phi = 0.8, f_v(.)$: Gaussian $\sigma_v^2 = 10$	
σ_x^2/σ_v^2	σ_x^2	p	Kalman	Masreliez	Improvement
0.5	5	2.051%	2.641439	2.120085	19.74%
1.0	10	6.667%	3.856439	2.948058	23.56%
1.5	15	11.28%	4.593211	3.442713	25.05%
2.0	20	15.90%	5.231454	3.978750	23.95%
2.5	25	20.51%	5.713847	4.433931	22.40%

TABLE IV
COMPARISON OF KALMAN AND MASRELIEZ FILTERS; GAUSSIAN DRIVING NOISE; GAUSSIAN+LAPLACIAN OBSERVATION NOISE; $\eta = 4.472$; $\phi = 0.8$

$f_w(.)$: Gaussian+Laplacian $\sigma_g^2 = 1, \sigma_l^2 = 40, \eta = 4.472$				$\phi = 0, f_v(.)$: Gaussian $\sigma_v^2 = 3.6$	
σ_x^2/σ_v^2	σ_x^2	p	Kalman	Masreliez	Improvement
0.5	1.8	2.051%	1.252170	0.864422	30.97%
1.0	3.6	6.667%	1.837671	1.141429	37.89%
1.5	5.4	11.28%	2.165590	1.292514	40.32%
2.0	7.2	15.90%	2.469167	1.476629	40.20%
2.5	9.0	20.51%	2.646290	1.669427	36.92%

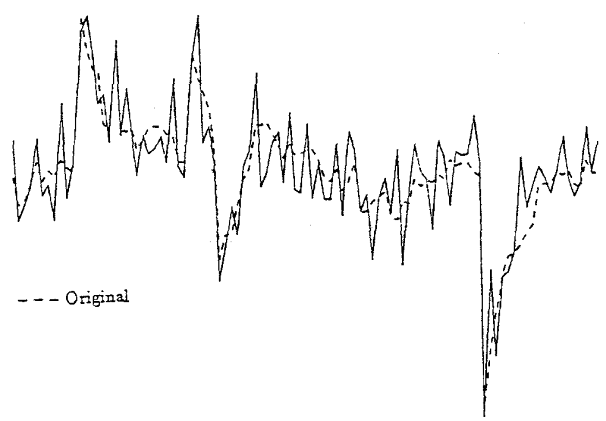


Fig. 6. Corrupted signal; Gaussian observation noise.

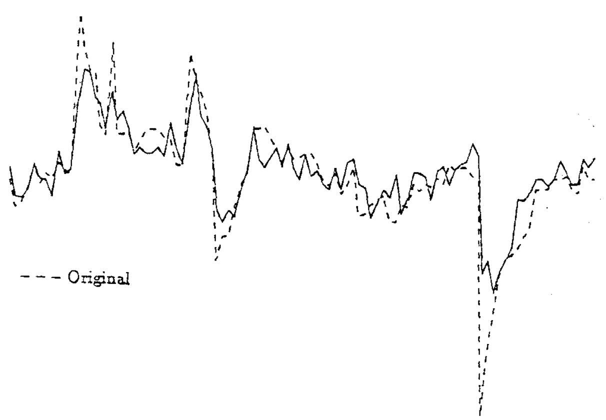


Fig. 7. Kalman filtered signal.

$2n^3 + 3n^2 + 2n + 61$ multiplications and 34 divisions, two exponential, and two square root operations. Here, we use the Newton's method to carry out the saddle point search, and the number of iteration is set to five, which is a typical value for the Newton's method to converge in these experiments. From these figures, we can see that the Masreliez filter requires more computations than the Kalman filter as expected. When the dimension of the system is small, the computational increase is more apparent.

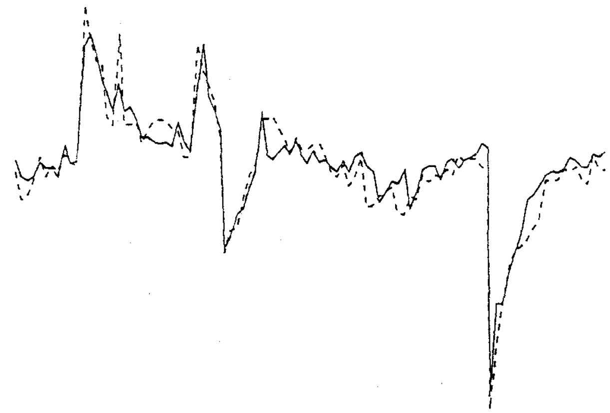


Fig. 8. Masreliez filtered signal.

Finally, we present some simulation results for the 2-D filtering problem. We extend the non-Gaussian noise model in (69) in which $f_i(.)$ and $f_j(.)$ of (69) now are 2-D distributions. To facilitate our simulations, we define a 2-D Laplacian-type distribution as follows: Let v_1 and v_2 be two independent Laplacian random variables with parameters η_1 and η_2. In addition, let $v = (v_1, v_2)^t$ and $\eta = (\eta_1, \eta_2)^t$. Define a vector $x = (x_1, x_2)^T$ and $x = Qv$, where

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{76}$$

is a rotation matrix, and θ is the rotation angle. We then say that $f(x)$ is a 2-D Laplacian-type distribution. To have a better understanding of our 2-D filtering scheme, we first show a 2-D score function and its derivative. Similar to 1-D cases, we have considered the distribution obtained by convolving a Gaussian distribution with a non-Gaussian distribution. The non-Gaussian distribution is a mixed distribution composed of Gaussian and Laplacian-type distributions. Their parameters are specified as follows:

$$\begin{aligned} p &= 0.99 \\ \theta &= 0^\circ \\ \eta &= \begin{pmatrix} 3 \\ 1 \end{pmatrix} \\ \Sigma &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ \Sigma_c &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned} \tag{77}$$

where Σ is the covariance matrix of the Gaussian component in the mixed model, and Σ_c is the covariance matrix of the convolving Gaussian distribution. Figs. 9 and 10 show the second component of the score function ($S_2(x)$ in (94)) and

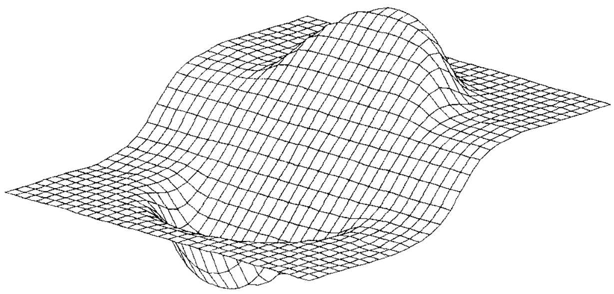


Fig. 9. Approximated 2-D score function.

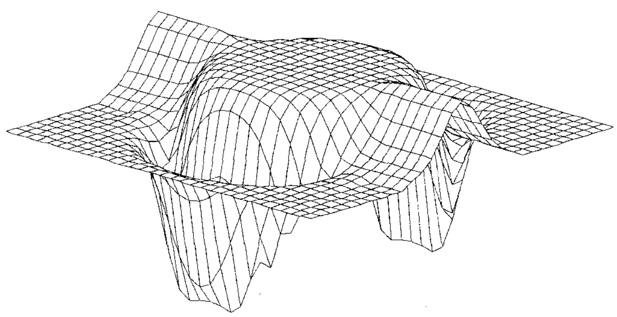


Fig. 10. Derivative of the approximated 2-D score function.

its derivative with respect to x_2 (G_{22} in (102)). From these figures, we can clearly see that the score function limits the influence of large observations. Note that the score functions of all Gaussian distributions are just planes. We have also compared the average squared error between the approximated and the exact score function ($S_2(x)$). We have found that it is only 7.9341×10^{-4}. This figure clearly shows that the approximation error is indeed very small. The last experiment we have carried out is 2-D filtering. We still use the model in (72) and (73). The only change is that the dimension of both x_k and z_k is now 2×1. We have assumed that w_k is Gaussian, and v_k is a mixed distribution composed of Gaussian and Laplacian-type distributions. We arbitrarily chose a 2-D system as follows:

$$\begin{aligned} \phi &= \begin{pmatrix} 0.8 & 0.4 \\ -0.4 & 0.8 \end{pmatrix} \\ H &= \begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix} \\ \sigma_w^2 &= \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \\ p &= 0.9 \\ \eta &= \begin{pmatrix} 9 \\ 5 \end{pmatrix}, \\ \sigma_v^2 &= \begin{pmatrix} 6 \\ 2 \end{pmatrix}. \end{aligned} \tag{78}$$

In (78), we implicitly assume that components of Gaussian distributions are independent; σ_w^2 and σ_v^2 then denote the variances of w_k and the Gaussian part of v_k, respectively. We have used the same criterion as that in (74) and (75). The only difference is that we run a single realization with 2000 data points. To see the filtering results for different observation distribution shapes, we have rotated the Laplacian-type component of our mixed model by five angles, i.e., 0, 45, 90, 135, and 180°. The simulation results are shown in Table V. We can see that the performance improvements are similar to 1-D cases. To test the performance of the Masreliez filter in systems with low correlation, we have performed another experiment where ϕ is set to zero. The results are shown in Table VI. The MSE improvements are less than those observed in highly correlated systems. This behavior is also similar to that in 1-D cases. From these results, the following conclusions are drawn:

1) The score function approximation scheme developed in this paper yields satisfactory results for a wide class of distributions, particularly for the distributions encountered often in the filtering problems. From Figs. 1 and 2, we can see that the approximation error is clearly bounded. It is simple to check that the conjugate densities in our simulations approach the gamma form when the Laplacian densities dominate and approach the normal form when the Gaussian densities dominate. This result is consistent with our error analysis described in Section III.
2) The score-function-based filters outperform the standard Kalman filter in the non-Gaussian noise environment, especially under the mixture-type noise environment.
3) In case of non-Gaussian observation noise, the stronger the observation noise, the greater the improvement. This is because the OPD becomes more non-Gaussian under the influence of stronger observation noise. On the other hand, in case of non-Gaussian driving noise, the stronger the observation noise, the smaller the improvement. This is because the observation noise is Gaussian, and the OPD becomes more Gaussian under the influence of the stronger observation noise.
4) In case of non-Gaussian observation noise, the Masreliez filter can perform better for signals with high correlation. For signals with low correlation, the Gaussian assumption of the SPD becomes less valid. On the other hand, in case of non-Gaussian driving noise, the Masreliez filter can perform better in signals with low correlation. In signals with low correlation, the Gaussian assumption of the density in (18) becomes more tenable.
5) In the score function approach, we assume that the SPD is Gaussian. This assumption may not be valid in some situations. A straightforward extension is to replace this density by a Gaussian sum. We call this the extended Masreliez algorithm. Note that this is different from the approach of Sorenson *et al.* [1]. The Gaussian sum is only used to keep track of the SPD. Preliminary studies indicate that the extended Masreiliez algorithm may outperform the standard Masreliez filter if noises are of mixture-type distributions. Research in this direction is now underway. Finally, we comment that there is a growing interest in the area of non-Gaussian signal processing. Our method is simple and effective and may be used in many applications.

TABLE V
COMPARISON OF KALMAN AND MASRELIEZ FILTER FOR A 2-D SYSTEM; GAUSSIAN STATE NOISE; GAUSSIAN+LAPLACIAN OBSERVATION NOISE

$f_v(.)$: Gaussian+Laplacian, $f_w(.)$: Gaussian $H = \begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}, \sigma_v^2 = \begin{pmatrix} 6 \\ 2 \end{pmatrix}, \eta = \begin{pmatrix} 9 \\ 5 \end{pmatrix}, p = 0.9, \phi = \begin{pmatrix} 0.8 & 0.4 \\ -0.4 & 0.8 \end{pmatrix}, \sigma_w^2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$				
		Kalman	Masreliez	Improvement
$\theta = 0°$	x_1	3.2261	2.1535	33.25%
	x_2	2.9084	1.6828	42.14%
$\theta = 45°$	x_1	3.7374	2.1610	42.18%
	x_2	2.7530	1.6538	39.93%
$\theta = 90°$	x_1	3.7393	2.1417	36.62%
	x_2	3.5123	1.6442	53.19%
$\theta = 135°$	x_1	3.3403	2.0924	37.36%
	x_2	3.2614	1.7478	46.41%
$\theta = 180°$	x_1	3.5713	2.2008	38.38%
	x_2	2.9494	1.6246	44.92%

TABLE VI
COMPARISON OF KALMAN AND MASRELIEZ FILTER FOR A 2-D SYSTEM; GAUSSIAN STATE NOISE; GAUSSIAN+LAPLACIAN OBSERVATION NOISE; $\phi = 0$

$f_v(.)$: Gaussian+Laplacian, $f_w(.)$: Gaussian $H = \begin{pmatrix} 1 & -0.5 \\ 0.5 & 1 \end{pmatrix}, \sigma_v^2 = \begin{pmatrix} 6 \\ 2 \end{pmatrix}, \eta = \begin{pmatrix} 9 \\ 5 \end{pmatrix}, p = 0.9, \phi = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \sigma_w^2 = \begin{pmatrix} 10 \\ 10 \end{pmatrix}$				
		Kalman	Masreliez	Improvement
$\theta = 0°$	x_1	5.9987	4.2064	29.88%
	x_2	4.2401	2.8014	33.93%
$\theta = 45°$	x_1	6.3976	4.4596	30.29%
	x_2	4.1682	2.7271	34.57%
$\theta = 90°$	x_1	5.1634	4.0313	21.92%
	x_2	5.6822	3.0096	47.03%
$\theta = 135°$	x_1	4.8771	3.9649	18.70%
	x_2	5.7938	3.0484	47.39%
$\theta = 180°$	x_1	6.2548	4.3349	30.70%
	x_2	4.2355	2.7466	35.15%

APPENDIX A

Define the relative approximation error of the density approximation as

$$\epsilon_d = \frac{f(x_0)}{\hat{f}(x_0)} - 1 \tag{79}$$

where $\hat{f}(x_0)$ is the approximation of $f(x_0)$. It is shown in [11] that ϵ_d is bounded for the gamma form and $\epsilon_d \to 0$ for the normal form. The definitions and properties of the gamma and normal forms are described as follows.

1) Gamma Form: Let $g(x) \sim Ax^{\alpha-1}l(x)e^{-cx}$ for $\alpha > 0$ and $c > 0$, where $l(x)$ is continuous, and $l(kx)/l(x) \to 1$ as $x \to \infty$ for every $k > 0$. Then, as $T \to c$

$$M^{(j)}(T) \sim A\frac{\Gamma(j+\alpha)}{(c-T)^{j+\alpha}}\, l\left(\frac{1}{c-T}\right)$$

and

$$\rho_j(T) \sim \alpha^{1-j/2}. \tag{80}$$

2) Normal Form: Let $g(x) \sim e^{-h(x)}$ for large x, where $h(x) > 0$ and $0 < h''(x) < \infty$. Let $v(x)$ and $w(x)$ exist such that

$$\begin{aligned} &\text{i)} \quad [v(x)]^2 h''(x) \to \infty \\ &\text{ii)} \quad e^{-w(x)}h''(x) \to 0 \end{aligned} \tag{81}$$

monotonically as $x \to \infty$, where

$$\begin{aligned} &v(x) > 0 \\ &|v'(x)| \le \alpha < \infty \\ &w(x) = \int \frac{1}{v(x)}\,dx. \end{aligned} \tag{82}$$

Then, $\rho_j(T) \to 0$ as T tends to its upper limiting value.

Since all standardized cumulants are bounded or zero, we find that (35) and (36) have similar expressions. Thus, the relative approximation error of $g'(0)$ is bounded or zero, and we can write

$$g(0) = (1+\epsilon_g)\hat{g}(0) \tag{83}$$

$$g'(0) = (1+\epsilon_{g'})\hat{g}'(0) \tag{84}$$

where ϵ_g and ϵ_g' are some bounded constants or zeros, and $\hat{g}(0)$ and $\hat{g}'(0)$ are approximations of $g(0)$ and $g'(0)$. The relative approximation error of $g'(0)/g(0)$ is then

$$\epsilon_t = \frac{\dfrac{g'(0)}{g(0)}}{\dfrac{\hat{g}'(0)}{\hat{g}(0)}} - 1 = \frac{\epsilon_{g'} - \epsilon_g}{1 + \epsilon_g} \tag{85}$$

which is bounded or zero. Thus, we conclude that the approximation error of the score function $-T_0 + g'(0)/g(0)$ is either bounded or zero.

Next, let us consider the approximation error of the derivative of score function. For the notational simplicity, we let the score function be $h(x)$. From the result above, we have

$$h(x) = (1 + \epsilon_s)\hat{h}(x) \tag{86}$$

$$h(x + dx) = (1 + \epsilon_r)\hat{h}(x + dx) \tag{87}$$

where ϵ_s and ϵ_r are either bounded or zero. The relative approximation error of the derivative of score function is then

$$\epsilon_v = \frac{\dfrac{h(x+dx) - h(x)}{dx}}{\dfrac{\hat{h}(x+dx) - \hat{h}(x)}{dx}} - 1 = \frac{\dfrac{(1+\epsilon_r)\hat{h}(x+dx) - (1+\epsilon_s)\hat{h}(x)}{dx}}{\dfrac{\hat{h}(x+dx) - \hat{h}(x)}{dx}} - 1. \tag{88}$$

When x approaches the end of its admissible range, $\epsilon_r \rightarrow \epsilon_s$. Then, $\epsilon_v \rightarrow \epsilon_s$, which is either bounded or zero. It is interesting to note that using our formulation, the approximation errors of the score function and its derivatives are equal when x is near the end of its admissible range.

APPENDIX B

The relationship between T and x is established through the saddle point. Let the cumulant MGF of $f(x)$ be $K(T)$. Then

$$\frac{\partial}{\partial T_1} K(T) = x_1$$

$$\frac{\partial}{\partial T_2} K(T) = x_2. \tag{89}$$

Taking derivatives of the first and second equation in (89) with respect to x_1 and x_2, we have

$$K_{20}\frac{\partial T_1}{\partial x_1} + K_{11}\frac{\partial T_2}{\partial x_1} = 1$$

$$K_{20}\frac{\partial T_1}{\partial x_2} + K_{11}\frac{\partial T_2}{\partial x_2} = 0 \tag{90}$$

$$K_{11}\frac{\partial T_1}{\partial x_1} + K_{02}\frac{\partial T_2}{\partial x_1} = 1$$

$$K_{11}\frac{\partial T_1}{\partial x_2} + K_{02}\frac{\partial T_2}{\partial x_2} = 0 \tag{91}$$

where $K_{ij} = [\partial^{(i+j)}/\partial T_1^i \partial T_2^j]K(T)$. Solving (90) and (91), we have

$$\frac{\partial T_1}{\partial x_1} = \frac{K_{02}}{K_{20}K_{02} - K_{11}^2}$$

$$\frac{\partial T_2}{\partial x_1} = \frac{-K_{11}}{K_{20}K_{02} - K_{11}^2} \tag{92}$$

$$\frac{\partial T_1}{\partial x_2} = \frac{-K_{11}}{K_{20}K_{02} - K_{11}^2}$$

$$\frac{\partial T_2}{\partial x_2} = \frac{K_{20}}{K_{20}K_{02} - K_{11}^2}. \tag{93}$$

Let the score function be $S(x) = [S_1(x), S_2(x)]^t$ and the MGF of $g_\zeta(.,.)$ be $L(T - T_0)$. Then, from (64), we have

$$S(x) = T + \frac{1}{2}\begin{pmatrix} 1 & -c \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \dfrac{L_{30}}{L_{20}^2} + \dfrac{L_{12}}{L_{20}L_{02}} \\ \dfrac{L_{21}}{L_{20}L_{02}} + \dfrac{L_{03}}{L_{02}^2} \end{pmatrix} \tag{94}$$

where $L_{ij} = [\partial^{(i+j)}/\partial T_1^I \partial T_2^j]L(T)$ and $T = T_0$ for $x = a$. Note that in the case here

$$c = \frac{K_{11}}{K_{20}}. \tag{95}$$

From (94), the derivative of the score function with respect to T_1 can be found as

$$\begin{pmatrix} \dfrac{\partial S_1(x)}{T_1} \\ \dfrac{\partial S_2(x)}{T_1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & -c \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \dfrac{L_{30}^I}{L_{20}^2} - \dfrac{2L_{30}L_{20}^I}{L_{20}^3} + \dfrac{L_{12}^I}{L_{20}L_{02}} - \dfrac{L_{12}(L_{20}^I L_{02} + L_{20}L_{02}^I)}{L_{20}^2 L_{02}^2} \\ \dfrac{L_{21}^I}{L_{20}L_{02}} - \dfrac{L_{21}(L_{20}^I L_{02} + L_{20}L_{02}^I)}{L_{20}^2 L_{02}^2} + \dfrac{L_{03}^I}{L_{02}^2} - \dfrac{2L_{03}L_{02}^I}{L_{02}^3} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & -c^I \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \dfrac{L_{30}}{L_{20}^2} + \dfrac{L_{12}}{L_{20}L_{02}} \\ \dfrac{L_{21}}{L_{20}L_{02}} + \dfrac{L_{03}}{L_{02}^2} \end{pmatrix} \tag{96}$$

where the superscript I denotes the derivative with respect to T_1. For example, $c^I = \partial c/\partial T_1$. By the same way, we can find the derivative of the score function with respect to T_2. The formula is identical to (96), except that the vector $[1, 0]^T$ on the right-hand side is replaced by $[0, 1]^T$, and all superscript I's are replaced with II that denote the derivative with respect to T_2. Thus

$$\begin{pmatrix} \dfrac{\partial S_1(x)}{T_2} \\ \dfrac{\partial S_2(x)}{T_2} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & -c \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \dfrac{L_{30}^{II}}{L_{20}^2} - \dfrac{2L_{30}L_{20}^{II}}{L_{20}^3} + \dfrac{L_{12}^{II}}{L_{20}L_{02}} - \dfrac{L_{12}(L_{20}^{II} L_{02} + L_{20}L_{02}^{II})}{L_{20}^2 L_{02}^2} \\ \dfrac{L_{21}^{II}}{L_{20}L_{02}} - \dfrac{L_{21}(L_{20}^{II} L_{02} + L_{20}L_{02}^{II})}{L_{20}^2 L_{02}^2} + \dfrac{L_{03}^{II}}{L_{02}^2} - \dfrac{2L_{03}L_{02}^{II}}{L_{02}^3} \end{pmatrix}$$

$$+\frac{1}{2}\begin{pmatrix}1 & -c^{II}\\ 0 & 1\end{pmatrix}\begin{pmatrix}\frac{L_{30}}{L_{20}^2}+\frac{L_{12}}{L_{20}L_{02}}\\ \frac{L_{21}}{L_{20}L_{02}}+\frac{L_{03}}{L_{02}^2}\end{pmatrix}. \tag{97}$$

Let the derivative of the score function with respect to x be $G(x)$ and

$$G(a)=\begin{pmatrix}G_{11} & G_{12}\\ G_{21} & G_{22}\end{pmatrix}. \tag{98}$$

Then

$$G_{11}=\frac{\partial S_1}{\partial x_1} = \frac{\partial T_1}{\partial x_1}\frac{\partial S_1}{\partial T_1}+\frac{\partial T_2}{\partial x_1}\frac{\partial S_1}{\partial T_2}\Big|_{T=T_0} \tag{99}$$

$$G_{12}=\frac{\partial S_1}{\partial x_2} = \frac{\partial T_1}{\partial x_2}\frac{\partial S_1}{\partial T_1}+\frac{\partial T_2}{\partial x_2}\frac{\partial S_1}{\partial T_2}\Big|_{T=T_0} \tag{100}$$

$$G_{21}=\frac{\partial S_2}{\partial x_1} = \frac{\partial T_1}{\partial x_1}\frac{\partial S_2}{\partial T_1}+\frac{\partial T_2}{\partial x_1}\frac{\partial S_2}{\partial T_2}\Big|_{T=T_0} \tag{101}$$

$$G_{22}=\frac{\partial S_2}{\partial x_2} = \frac{\partial T_1}{\partial x_2}\frac{\partial S_2}{\partial T_1}+\frac{\partial T_2}{\partial x_2}\frac{\partial S_2}{\partial T_2}\Big|_{T=T_0}. \tag{102}$$

For $T=T_0$, $L_{ij}=\mu_{\zeta,ij}$. The relation of $\mu_{\zeta,ij}$ and $\mu_{z,ij}$ for $i+j=2,3$ has been found in (66)–(68). For $i+j=4$, we have

$$\mu_{\zeta,40}=\mu_{z,04}-4c\mu_{z,31}+6c^2\mu_{z,22} - 4c^3\mu_{z,31}+c^4\mu_{z,40} \tag{103}$$

$$\mu_{\zeta,13}=\mu_{z,13}-3c\mu_{z,22}+3c^2\mu_{z,31}-c^3\mu_{z,40} \tag{104}$$

$$\mu_{\zeta,22}=\mu_{z,22}-2c\mu_{z,22}+2c\mu_{z,31}+c^2\mu_{z,40} \tag{105}$$

$$\mu_{\zeta,31}=\mu_{z,31}-c\mu_{z,40}. \tag{106}$$

Derivatives of c and L_{ij} in (96) and (97) can be obtained from (95), (66)–(68), and (103)–(106). Let $T=T_0$, and denote $\mu^I_{\zeta,ij}=L^I_{ij}$ and $\mu^{II}_{\zeta,ij}=L^{II}_{ij}$. We have the following results:

$$c=\frac{\mu_{z,11}}{\mu_{z,20}} \tag{107}$$

$$c^I=\frac{\mu_{z,21}}{\mu_{z,20}}-c\,\frac{\mu_{z,30}}{\mu_{z,20}} \tag{108}$$

$$\mu^I_{\zeta,20}=\mu_{z,30} \tag{109}$$

$$\mu^I_{\zeta,02}=\mu_{z,12}-c\mu z,\ 21-c^I\mu_{z,11} \tag{110}$$

$$\mu^I_{\zeta,30}=\mu_{z,40} \tag{111}$$

$$\mu^I_{\zeta,03}=\mu_{z,13}-3c\mu_{z,22}+3c^2\mu_{z,31}-c^3\mu_{z,40} - 3c^I(\mu_{z,12}+6c\mu_{z,21}-3c^2\mu_{z,30}) \tag{112}$$

$$\mu^I_{\zeta,12}=\mu_{z,22}-2c\mu_{z,31}+c^2\mu_{z,40} - 2c^I(\mu_{z,21}+c\mu_{z,30}) \tag{113}$$

$$\mu^I_{\zeta,21}=\mu_{z,31}-c\mu_{z,40}-c^I\mu_{z,30} \tag{114}$$

$$c^{II}=\frac{\mu_{z,12}}{\mu_{z,20}}-c\,\frac{\mu_{z,21}}{\mu_{z,20}} \tag{115}$$

$$\mu^{II}_{\zeta,20}=\mu_{z,21} \tag{116}$$

$$\mu^{II}_{\zeta,02}=\mu_{z,03}-c\mu z,\ 21-c^{II}\mu_{z,11} \tag{117}$$

$$\mu^{II}_{\zeta,30}=\mu_{z,31} \tag{118}$$

$$\mu^{II}_{\zeta,03}=\mu_{z,04}-3c\mu_{z,13}+3c^2\mu_{z,22}-c^3\mu_{z,31} - 3c^{II}(\mu_{z,12}+6c\mu_{z,21}-3c^2\mu_{z,30}) \tag{119}$$

$$\mu^{II}_{\zeta,12}=\mu_{z,13}-2c\mu_{z,22}+c^2\mu_{z,31} - 2c^{II}(\mu_{z,21}+c\mu_{z,30}) \tag{120}$$

$$\mu^{II}_{\zeta,21}=\mu_{z,22}-c\mu_{z,31}-c^{II}\mu_{z,30}. \tag{121}$$

We now summarize the evaluation of derivatives of the score function as follows:

1) Substitute (66)–(68) and (103)–(106) into (107)–(121).
2) Substitute the results of Step 1 and (66)–(68) into (96) and (97).
3) Substitute the results of Step 2 into (99)–(102).
4) Use $T=T_0$ in (92) and (93) and substitute the results into (99)–(102).

REFERENCES

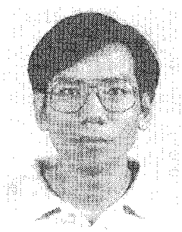
[1] H. W. Sorenson and A. R. Stubberud, "Non-linear filtering by approximation of the *a posteriori* density," *Int. J. Contr.*, vol. 18, pp. 33–51, 1968.
[2] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, pp. 465–479, 1971.
[3] C. J. Masreliez, "Robust recursive estimation and filtering," Ph.D. dissertation, Univ. of Washington, Seattle, 1972.
[4] C. J. Masreliez, "Approximate non-Gaussian filtering with linear state and observation relations," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 107–110, 1975.
[5] A. A. Ershov and R. S. Lipster, "A robust Kalman filter in discrete time," *Automat. Remote Contr.*, vol. 39, pp. 359–367, 1978.
[6] M. West, "Robust sequential approximate Bayesian estimation," *J. R. Statist. Soc. B*, no. 2, pp. 157–166, 1981.
[7] T. Subba Rao and M. Yar, "Linear and nonlinear filters for linear, but not Gaussian processes," *Int. J. Contr.*, pp. 235–246, 1984.
[8] H. Cramer, *Mathematical Methods of Statistics.* Princeton, NJ: Princeton Univ. Press, 1946.
[9] M. G. Kendall and A. Stuart, *Advanced Theory of Statistics.* London: Griffin, 1958, vol. 1.
[10] A. I. Khinchin, *Mathematic Foundation of Statistical Mechanics.* New York: Dover, 1949.
[11] H. E. Daniels, "Saddlepoint approximations in statistics," *Ann. Math. Stat.*, vol. 25, pp. 631–650, 1954.
[12] O. Barndorff-Nielsen and D. R. Cox, "Edgeworth and saddle-point approximations with statistical applications," *J. R. Stat. Soc. B*, vol. 41, pp. 279-312, 1979.
[13] F. R. Hampel, "Some small sample asymptotics," in *Proc. Prague Symp. Asymptotic Statistics*, J. Hajek, Ed. Prague: Charles Univ., 1973, vol. II, pp. 109–126.
[14] I. J. Good, "Saddle-point methods for the multinomial distribution," *Ann. Math. Statist.*, vol. 28, pp. 861–881, 1957.
[15] W.-R. Wu, "Kalman filtering in non-Gaussian environment using efficient score function approximation," Ph.D. dissertation, State Univ. of New York, Buffalo, 1989.
[16] A. M. Makowski, "Results on the filtering problem for linear systems with non-Gaussian initial conditions," in *Proc. 21st IEEE Conf. Decision Contr.*, Dec. 1982.
[17] A. M. Makowski, W. S. Levine, and M. Asher, "The nonlinear MMSE filter for partially observed systems driven by non-Gaussian white noise, with application to failure estimation," in *Proc. 23rd IEEE Conf. Decision Contr.*, Dec. 1984.
[18] C. de Benito and K. A. Loparo, "Nonlinear filtering for discrete time linear systems with non-Gaussian initial data," *IEEE Trans. Automat. Contr.*, vol. 34, pp. 781–783, July 1989.

[19] R. Vijayan and H. V. Poor, "Nonlinear technique for interference suppression," *IEEE Trans. Commun.*, vol. 38, no. 7, pp. 1060–1065, July 1990.

[20] I. Song and S. A. Kassam, "Locally optimum detection of signals in a generalized observation model: The known signal case," *IEEE Trans. Inform. Theory,* vol. 36, no. 3, pp. 502–515, May 1990.

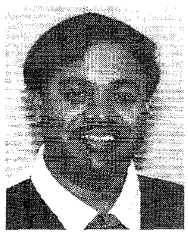
[21] ——, "Locally optimum detection of signals in a generalized observation model: The random signal case," *IEEE Trans. Inform. Theory,* vol. 36, no. 3, pp. 516–530, May 1990.

[22] E. Parzen, *Modern Probability Theory and Its Applications.* New York: Wiley, 1972, p. 216.



Wen-Rong Wu was born in Taiwan, R.O.C., in 1958. He received the B.S. degree in mechanical engineering from Tatung Institute of Technology, Taiwan, in 1980 and the M.S. degrees in mechanical and electrical engineering and the Ph.D. degree in electrical engineering from State University of New York, Buffalo, in 1985, 1986, and 1989, respectively.

Since August 1989, he has been a faculty member with the Department of Communication Engineering of National Chiao Tung University, Taiwan. His research interests include estimation theory, digital signal processing, and image processing.



Amlan Kundu received the B. Tech. degree in electronics from Calcutta University, Calcuta, India, in 1979 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California at Santa Barbara, in 1982 and 1985, respectively.

After graduation, he joined the faculty of the State University of New York (SUNY) at Buffalo. From 1991 to 1992, he was affiliated with Center of Excellence for Document Analysis and Recognition (CEDAR), SUNY at Buffalo. From 1992 to 1994, he was with NCCOSC, San Diego, CA. Currently, he is with US WEST Advanced Technology, Boulder, CO. His research interest includes image/video processing, optimal filtering, pattern recognition, and communications. He is particularly interested in pioneering applications of hidden Markov modeling and hybrid HMM/NN classifiers to nonspeech problems such as handwriting recognition, computer vision, SONAR signal identification etc. He has written more than 60 refereed scientific papers and book chapters in his various fields of research interest.