State Estimation in the Presence of non-Gaussian Noise

K.N. Plataniotis
Department of ECE
University of Toronto
Toronto, ON, M5S 3G4, Canada
e-mail: kostas@dsp.toronto.edu

A.N. Venetsanopoulos
Department of ECE
University of Toronto
Toronto, ON, M5S 3G4, Canada
e-mail: anv@dsp.toronto.edu

Abstract

The problem of nonlinear filtering with a non-Gaussian model of measurement errors is considered in this paper. Based on Bayes classification of the observations an approximate solution is introduced. The Bayesian estimator can be applied to any discrete time, linear, or nonlinear system which is observed in additive non-Gaussian measurement noise. The problem of narrowband interference suppression in additive noise is considered as an important example of non-Gaussian noise filtering. It is shown that the approximate filter outperforms currently used approaches and at the same time offers simplicity in the design.

1.0 Introduction

Estimation (filtering) theory has received considerable attention in the past four decades, primarily due to its practical significance in solving engineering and scientific problems. Estimation algorithms can be classified into two major categories. Namely, linear and nonlinear filtering algorithms, corresponding to linear (or linearized) physical dynamic models with Gaussian noise statistics and to nonlinear or non-Gaussian physical models.

The linear estimation problem in particular, has attracted considerable attention as can be seen in books and surveys of the subject [2]. The discrete linear, state estimation problem is described by the following model equations and statement of objective:

$$x(k + 1) = \Phi(k + 1, k)x(k) + w(k) \qquad (1)$$

$$z(k + 1) = H(k + 1)x(k + 1) + v(k + 1) \qquad (2)$$

where $x(k)$ is the n-dimensional state process, $z(k)$ is the m-dimensional measurement process, w(k) is the associated state noise, modeled as an uncorrelated white Gaussian sequence with covariance $Q(k)$ and v(k) is the observation (measurement) noise, modeled as an uncorrelated white Gaussian sequence with covariance $R(k)$. The initial state vector $x(0)$ is Gaussian with mean value $\hat{x}(0)$ and covariance $P(0)$. It is also assumed independent of the noise processes $\forall k > 0$. Given the set $Z^k = [z(1), z(2), ..., z(k - 1), z(k)]$, the *mean-squared-error* optimal filtered estimate $\hat{x}(k|k)$ of $x(k) = \hat{x}(k|k) = E(x(k)|Z^k)$ is to be obtained.

The above problem was first solved by Kalman who obtained its well known filter [2]. The so called Kalman filter is the optimal recursive estimator for the above problem. It must be emphasized that the Kalman filter is the optimal estimator if the processes $w(k)$ and $v(k)$ are Gaussian. The performance of a linear estimator developed under the assumption of additive Gaussian noise can be degraded by the non-Gaussian noise due to the deviation from normality in the tails of the noise distribution [11]. Thus, if either of the noise processes in (1)-(2) are non-Gaussian, the measurement $z(k)$ will be non-Gaussian, and the degradation in the performance of the Kalman filter is rather dramatic. In most cases, $w(k)$, which is in fact a tuning parameter that greatly depends on the filter designer, can be modeled as a Gaussian process. However, this is not the case for the observation noise v(k). Many physical environments can be modeled more accurately as non-Gaussian rather than Gaussian observation channels and are characterized by heavy-tailed non-Gaussian distributions. Thus, the development of an efficient and robust filter which can provide accurate and reliable estimates in the presence of non-Gaussian measurement noise is of paramount importance [5]. In this paper, we discuss approximate nonlinear filters for state estimation when the observation noise is non-Gaussian and the actual signal has a linear state space representation.

2.0 The Problem of Nonlinear Filtering

Several models have been used to date to model non-Gaussian noise environments. Some of these models have been developed directly from the underlying physical phenomenon, most notably the *Middleton* Class A, B and C model [8], [9] while others are empirically devised. A fi-

nite sum of known Gaussian distributions can be used in the development of approximate empirical distributions which relate to many physical non-Gaussian phenomena. A special case of the Gaussian sum approach, the ($\epsilon - mixture$), is of particular interest. This model is particular useful when a nominal Gaussian noise environment with an impulsive or 'heavy-tailed' component is assumed. The probability density function for such a model can be written as $f(x) = (1 - \epsilon)f_G(x) + \epsilon f_{nG}(x)$ where $\epsilon \in (0, 1)$, f_G is the pdf of the nominal or background Gaussian density function, and f_{nG} is the pdf of the dominant non-Gaussian noise often taken to be a heavy-tailed density, such as a Laplacian density or a Gaussian density with a large variance (covariance). The mixing parameter ϵ regulates the contribution of the non-Gaussian component. Usually it varies between 0.01 to 0.25 [14]. When a Gaussian density with large variance is used to emulate the non-Gaussian dominant component, the ratio of the dominant to nominal density variances K is on the order of 10 to 10,000 [14], [4]. Due to its flexibility, many different naturally occurring noise distribution shapes can be approximated using the $\epsilon - mixture$ approach. The Gaussian sum approximation is applied to the problem of state estimation in non-Gaussian measurement noise. The system model (1)-(2) is used, however, it is assumed that the density function of the measurement noise $v(k)$ is not Gaussian but a linear combination of Gaussian terms, such as $f(v(k)) = \sum_{i=1}^{l} \epsilon_i N(\mu_i, R_i)$. Such an assumption about the noise statistics results in a predictive measurement density $f(z(k)|x(k), Z^{k-1})$ which has a similar form [13]. A number of filtering techniques have been derived utilizing such empirical noise modeling schemes to tackle the problem of estimating the state of a linear system in a non-Gaussian noise environment. In particular, Sorenson and Alspach [13], [1] considered the case where both the state and the measurement noise sequences are non-Gaussian. They have assumed that the noise sequences have a uniformly convergent series expression in terms of known Gaussian distributions. They used a fixed number of Gaussian terms with known moments in a truncated approximation series to develop an optimal, under these assumptions, minimum mean square error filtering algorithm. The output of the Gaussian sum filter is formed by combining the outputs of a bank of Kalman filters, each one matched to a specific term of the Gaussian sum. The major disadvantage of their approach however, is its computational complexity since the numerical computations in the filter increase almost exponentially in time. In the Gaussian case, the number of terms involved in the derivation of the optimal Kalman filter remain constant. If however, for any time index (k) the density $f(x(k)|Z^{k-1})$ is a Gaussian mixture of (l_1) components and the distribution $v(k)$ contains (l_2) components, then the combined density needed

for the next step calculations is a mixture with $(l_1 x l_2)$ components. Thus, the computational burden at each stage becomes larger as the number of terms increases in the mixture used in the calculations. Thus, the Gaussian sum method is not attractive and in many cases is not feasible in practice due to its computational complexity. In an attempt to alleviate the computational burden associated with the Gaussian sum approach, Masreliez proposed a different filtering methodology based on the so called 'score-function'. His approximate conditional mean filter is computationally more efficient than that of Alspach and Sorenson, but requires the use of problem dependent nonlinearities in the filter gain and a rather demanding convolution operation in the development of the nonlinear score function [6], [7]. An approximate filter called adaptive Gaussian sum filter (AGSF) which is computationally attractive and does not require any problem dependent nonlinearities can be used instead [12]. The approximate filter utilizes the same methodology with that of the Gaussian sum filter of Sorenson and Alspach. However, in order for the procedure to be practical the number of terms in the mixture is controlled at each step. A Bayesian learning technique is employed to collapse the resulting non-Gaussian sum mixture to an equivalent Gaussian term. Thus, at the end of the current cycle of the filter, the resulting Gaussian mixture $f(z(k)|x(k), Z^{k-1})$ is collapsed and approximately represented with only one equivalent Gaussian term. In the next filtering cycle, the calculations involve only the l terms used in the representation of the measurement noise, resulting in fixed complexity. In this way the new filter resolves the computational burden of the Gaussian sum filter without the use of any problem dependent nonlinearities, such as those required by the Masreliez filter. The main points of our strategy are: (i) for each Gaussian term in $f(v(k)) = \sum_{i=1}^{l} \epsilon_i N(\mu_i, R_i)$ a dedicated Kalman filter is employed, (ii) based on the interim results from these dedicated Kalman filters we obtain a Bayesian a-posteriori estimate of the Gaussian mixture $f(z(k)|x(k), Z^{k-1})$ needed in the filtering process, and (iii) the optimal (in the minimum mse error sense) Gaussian approximation for the above mixture is derived. Then the first two moments of this equivalent Gaussian term are used to complete the filtering cycle of a recursive, Kalman-like filter. The form of the estimator is as follows:

$$\hat{x}(k|k-1) = \Phi(k, k-1)\hat{x}(k-1|k-1) \qquad (3)$$

$$P(k|k-1) = \Phi(k, k-1)P(k-1|k-1)\Phi(k, k-1)^{\tau} + Q(k-1) \qquad (4)$$

with initial conditions $\hat{x}(0|0) = x(0)$ and $P(0|0) = P(0)$.

$$\hat{z}(k|k-1) = \sum_{i=1}^{l} w_i(k)\hat{z}_i(k|k-1) \qquad (5)$$

$$\hat{z}_i(k|k-1) = H(k)\hat{x}(k|k-1) + \mu_i \qquad (6)$$

$$P_z(k|k-1) =$$

$$\sum_{i=1}^{l} (P_{zi}(k|k-1) + \tilde{z}(k|k-1)\tilde{z}(k|k-1)^{\tau})w_i(k) \quad (7)$$

$$P_{zi}(k|k-1) = H(k)P(k|k-1)H^{\tau}(k) + R_i \quad (8)$$

$$c(k) =$$

$$((2\pi)^{-m}|P_{zi}|^{-1}exp(-0.5(||\tilde{z}(k|k-1)||^2_{P_{zi}^{-1}(k|k-1)})))\epsilon_i \quad (9)$$

$$w_i(k) = \frac{c(k)}{\sum_{i=1}^{l}c(i)} \quad (10)$$

$$K(k) = P(k|k-1)H^{\tau}(k|k-1)P_z^{-1}(k|k-1) \quad (11)$$

$$P(k|k) = (I - K(k)H(k))P(k|k-1) \quad (12)$$

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K(k)(z(k) - \hat{z}(k|k-1)) \quad (13)$$

where $|.|$ denotes the determinant of the matrix, $||.||$ the inner product in the Euclidean space R^m.

The collapsed density which is used in the approximation has only one Gaussian term, thus it can be incorporated in the recursive form of the usual Kalman filter. The performance of the nonlinear filter depends on the approximation of the Gaussian mixture by the single Gaussian term. The closer the Gaussian density approximates the mixture, the better the filter estimates are. The rationale of this approximation lies in the fact that some of the members in the original density carry small mixing weights at a particular time instant and hence the information that they carry can safely be ignored for practical purposes. As a measure of the goodness of the approximation the *Bhattacharyya coefficient* between two densities $f_i(x)$ and $f_j(x)$ defined as $\rho_{ij} = \int [f_i(x)f_j(x)]^{0.5} dx$ with $0 \leq \rho_{ij} \leq 1$ and $\rho_{ij} = 1$ if $f_i(x) = f_j(x)$ [3] may be utilized to measure the distance between the mixture density and the Gaussian term resulting after the collapse of the mixture. It can be seen in the theorem that the density $f(z(k)|x(k), Z^{k-1})$ is represented by a finite number of parameters which are obtained using recursive Kalman filters. Thus, the Gaussian sum density is formed as the combination of the output of a number of linear filters operating in parallel, resulting in a nonlinear filter with computational complexity similar to that of the linear Kalman filter.

In the special case in which ϵ takes the values of either 0 or 1, the sequence of $(\epsilon(1), \epsilon(2), ..., \epsilon(k-1), \epsilon(k), ...)$ becomes a Markov sequence of random quantities. In this scenario the measurement errors at the kth time instant belong to one of two models. The standard model $v_0(k)$ where $v_0(k) \sim N(\mu_0, R_0)$ or the alternative model $v_1(k) \sim N(\mu_1, R_1)$, $R_1(k) = (KR_0(k))$ which corresponds to rarely occurring dominant noise.

The optimal estimate $\hat{x}(k|k)$ requires knowledge of the actual value ϵ for that particular time instant. When the value ϵ at time k is fixed the density function $f(\tilde{z}(k|k-1)|\epsilon)$ is Gaussian with covariance $P_{z0})(k|k-1) = H(k)P(k|k-1)H^{\tau}(k) + R_0$ for $\epsilon = 0$, or $P_{z1}(k|k-1) = H(k)P(k|k-1)H^{\tau}(k) + R_1$ for $\epsilon = 1$. The optimal procedure for estimating ϵ reduces to the calculation of the ratio $\lambda(k|k-1) = \frac{f(\tilde{z}(k|k-1)|\epsilon=1)}{f(\tilde{z}(k|k-1)|\epsilon=0)}$ with $\hat{\epsilon} = 1$ if $\lambda(k|k-1) \geq 1$. The state estimate is given by the following set of equations:

$$\hat{x}(k|k-1) = \Phi(k, k-1)\hat{x}(k-1|k-1) \quad (14)$$

$$P(k|k-1) = \Phi(k,k-1)P(k-1|k-1)\Phi(k,k-1)^{\tau} + Q(k-1) \quad (15)$$

with initial conditions $\hat{x}(0|0) = x(0)$ and $P(0|0) = P(0)$.

$$\hat{z}(k|k-1) =$$

$$H(k)\hat{x}(k|k-1) + [(1-\hat{\epsilon})\mu_0(k) + \hat{\epsilon}\mu_1(k)] \quad (16)$$

$$P_{zi}(k|k-1) =$$

$$H(k)P(k|k-1)H^{\tau}(k) + [(1-\hat{\epsilon})R_0(k) + \hat{\epsilon}R_1(k)] \quad (17)$$

$$K(k) = P(k|k-1)H^{\tau}(k|k-1)P_z^{-1}(k|k-1) \quad (18)$$

$$P(k|k) = (I - K(k)H(k))P(k|k-1) \quad (19)$$

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K(k)(z(k) - \hat{z}(k|k-1)) \quad (20)$$

3.0 Application to Narrowband Interference Suppression in Impulsive Channels

The approximate filter described here is studied for its rejection capability of an autoregressive interferer in a non-Gaussian channel. The problem of interest is the suppression of a narrowband interferer in a direct-sequence spread-spectrum (DSSS) system operating as an Nth order autoregressive process of the form $i_k = \sum_{n=1}^{N} \Phi_n i_{k-n} + e_k$ where e_k is a zero mean white Gaussian noise process and $\Phi_1, \Phi_2, ..., \Phi_{N-1}, \Phi_N$ are the autoregressive parameters known to the receiver. The received signal can be modeled using the following state-space representation:

$$x(k) = \Phi x(k-1) + w(k)$$

$$z(k) = Hx(k) + v(k) \quad (21)$$

with $x(k) = [i_k, i_{k-1}, ..., i_{k-N+1}]^{\tau}$, $w(k) = [e_k, 0, ..., 0]^{\tau}$, $H = [1, 0, ..., 0]$, and

$$\Phi = \begin{vmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_N \\ 1. & 0. & \cdots & 0. \\ \cdots\cdots\cdots\cdots\cdots \\ 0. & 0. & \cdots & 1. \end{vmatrix}$$

The additive observation noise $v(k)$ in the above model is defined as $v(k) = n(k)(+s(k))$: Following the assumptions introduced in [15], the received waveform has been

processed by a matched filter and sampled at the chip rate, to generate the discrete signal $z(k)$ at the receiver. Since in the model above the first component of the state $x(k)$ is the interferer i_k, an estimate of the state contains an estimate of the i_k, which can be subtracted from the received signal in order to increase the system performance. The sequence $v(k)$ is the impulsive noise process derived using the $\epsilon - mixture$ model and $s(k)$ is the DSSS signal which is treated as an i.i.d binary sequence with equal probability to occur $p(s_k = -1) = p(s_k = 1) = 0.5$. The sequences $s(k)$, $n(k)$ and i_k are assumed to be mutually independent. When only the impulsive noise of the channel is assumed present, without the DSSS signal, each v(k) has the following density:

$$f(v(k)) = (1 - \epsilon)N(0, \sigma_n^2) + \epsilon N(0, k\sigma_n^2) \qquad (22)$$

In the simulation studies reported here, the interferer is found by channeling white noise through a second-order infinite-duration impulse response (IIR) filter with two poles at 0.99, therefore $i_k = 1.98i_{k-1} - 0.9801i_{k-2} + e_k$ where e_k is zero mean white Gaussian noise with variance 0.01. To demonstrate the effectiveness of the proposed algorithm in a non-Gaussian environment the regulatory coefficient ϵ is set to be $\epsilon = 0.2$ and the ratio K is taken to be $K = 10$ or $K = 10,000$ with $\sigma_n = 1.0$.

The following algorithms have been tested: two Kalman filters each one matched to a part of the mixture above, the Masreliez filter and the adaptive Gaussian sum filter (AGSF). Filter initialization requires an initial filtered estimate $\hat{x}(0|0)$ and an initial error covariance $P(0|0)$. All four filters are initialized using an initial mean value of 0.01 and a covariance 1.0. The normalized mean square error (*NMSE*) was utilized for filter comparison purposes in all experiments. The data were averaged through Monte Carlo techniques. In the simulation studies, 50 independent runs were processed, each 1000 samples in length. The non-Gaussian noise, for a single run, is depicted in Fig.2 ($K = 10$) and filtering results are listed in Figs. 3-4. Significant findings and corresponding remarks are herein organized in a series of comments, which are supported by the appropriate figures: (i) Due to the independence of the Kalman filter calculations from the actual noise distribution shape, the performance of the optimal linear filter in the presence of non-Gaussian measurement noise is not satisfactory. The outliers generated by the heavy-tails of the non-Gaussian noise distribution result in erroneous estimates of the prediction measurement covariance $P_z(k|k - 1)$. The divergence from the theoretical covariances involved in the calculations of the Kalman filter gain leads to unacceptable estimation results. (ii) The nonlinear Masreliez filter performs relatively well in modest non-Gaussian channels. However, it fails to deliver accurate and reliable estimates in highly impulsive measurement chan-

nels. Due to its complexity, the need for customized, problem dependent nonlinearities in the 'score-function' and the inconsistency in its performance, this filter can not be considered as a general purpose robust recursive estimator. (iii) The approximate filter performed well under all the different noise scenarios selected. From the plots included the improvement accomplished by the utilization of the new filter versus the Kalman filter and the Masreliez filter is immediately evident. The effects have appeared more pronounced at more dense non-Gaussian (impulsive) environments. This trend was also verified during the error analysis utilizing the Monte Carlo error plots. Furthermore, the improvement has been obtained without additional computations. To conclude our study a comprehensive comparison of the filters discussed in this paper must be provided. Taking into consideration all the above mentioned facts, we can decide to use the following crude ranking of performance:

- 0: Poor performance

- 1: Average performance

- 2: Good performance

A similar ranking can be used for the evaluation of the computational complexity:

- 0: High computational complexity

- 1: Average computational complexity

- 2: Low computational performance

Based on the above figures of merit the applicability of the filters to the problem of narrowband interference suppression in impulsive channels is summarized in Table 2. In summary, from the simulation studies reported here we can conclude that the performance of the different linear or nonlinear filter depends on the departure from the normality of the measurement noise and of course the signal-to-noise ratio (SNR). For low SNR and strong non-Gaussian measurement noise the nonlinear filters outperform the linear suboptimal estimators and the the Masreliez filter, exhibiting at the same time significantly less computational complexity.

4.0 Conclusions

The paper has addressed the important problem of state estimation in non-Gaussian observation channels. A robust and computationally efficient approximate filter was introduced. Its application to the problem of narrowband interferer suppression in impulsive noise channels has been considered. Extensive experimentation has been introduced to demonstrate the effectiveness of the new filter. Results indicate that the new filter performs better than the presently available methods. In addition, the approximate filter is

computationally attractive, does not require any problem-dependent nonlinearities in its design, can adapt to varying noise statistics and satisfies the requirements of a general purpose real-time filtering system.

References

[1] D.L. Alspach, H. Sorenson, 'Nonlinear Bayesian estimation using Gaussian sum approximations', IEEE Trans. on Automatic Control, vol. AC-17, no. 4, 1972, pp. 439-448.

[2] J.V. Candy, *Signal Processing: The model-based approach*, Mc-Graw Hill Book Company, N.Y., 1986.

[3] K. Fukunaga, *Introduction to Statistical Pattern Recognition* Academic Press, Second Edition, London, 1990.

[4] L.M. Garth, H.V. Poor, 'Narrowband interference suppression in impulsive channels', IEEE Trans. on Aerospace and Electronic Systems, vol. AES-28, 1992, pp. 15-35.

[5] P.J. Huber, 'Robust estimation of a location parameter', Ann. Math. Statist., vol. 35, 1964, pp.73-101.

[6] C.J. Masreliez, 'Approximate non-Gaussian filtering with linear state and observation relations', IEEE Trans. on Automatic Control, vol. AC-20, 1975, pp. 107-110.

[7] C.J. Masreliez, R.D. Martin, 'Robust Bayesian estimation for the linear model and robustifying the Kalman filter', IEEE Trans. on Automatic Control, vol. AC-22, 1977, pp. 361-371.

[8] D. Middleton, 'Man-made noise in urban environments and transportation systems', IEEE Trans. on Communications, vol. COM-21, 1973, pp. 1232-1241.

[9] D. Middleton, 'Canonical non-Gaussian noise models: Their implication for measurement and for prediction of receiver performance,' IEEE Trans. Electromagn. Compat., vol. EMC-21, 1979, pp. 209-219.

[10] T. Namera, A.L. Stubberud, 'Gaussian sum approximation for non-linear fixed-point prediction', Int. Journal of Control, vol. 38, no. 5, 1983, pp. 1047-1053.

[11] I. Pitas, A.N. Venetsanopoulos, 'Order statistics in digital image processing', Proceedings of IEEE, vol. 80, no. 12, 1992, pp. 1893-1923.

[12] K.N. Plataniotis, D. Androutsos, A.N. Venetsanopoulos, 'Nonlinear filtering of non-Gaussian noise', Journal of Intelligent and Robotic Systems, vol. 19, pp. 207-231, 1997.

[13] H. Sorenson, D.L. Alspach, 'Recursive Bayesian estimation using Gaussian sum' , Automatica, vol. 7, 1971, pp.465-479.

[14] K.S. Vastola, 'Threshold detection in narrowband non-Gaussian noise,' IEEE Trans. on Communication, vol. COM-32, 1984, pp. 134-139.

[15] R. Vijayan, H.V. Poor, 'Nonlinear techniques for interference suppression in spread-spectrum systems', IEEE Trans. on Communication, vol. COM-38, 1990, pp. 1060-1065.

Table 1. Figure of Merit

a	Impulsive channel. No DSSS signal present
b	Intense Impulsive channel. No DSSS signal present
c	Intense Impulsive channel. DSSS signal present
d	Computational complexity

Table 2. Filter comparison

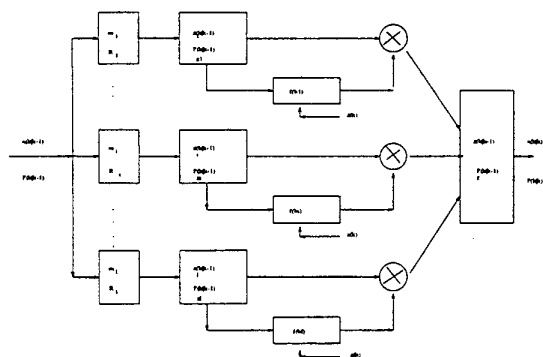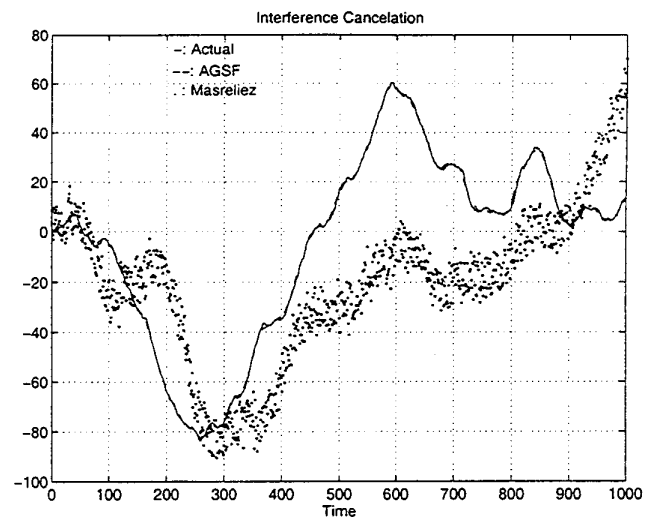| Filter	a	b	c	d
Kalman	0	0	0	2
Masreliez	2	1	-	0
Adaptive Gaussian Sum	2	2	2	0

Figure 1. The Adaptive Filter

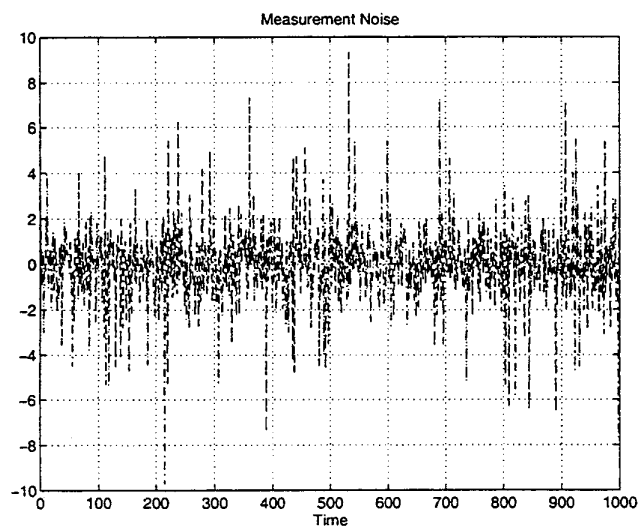Figure 3. Nonlinear filters: Estimation results
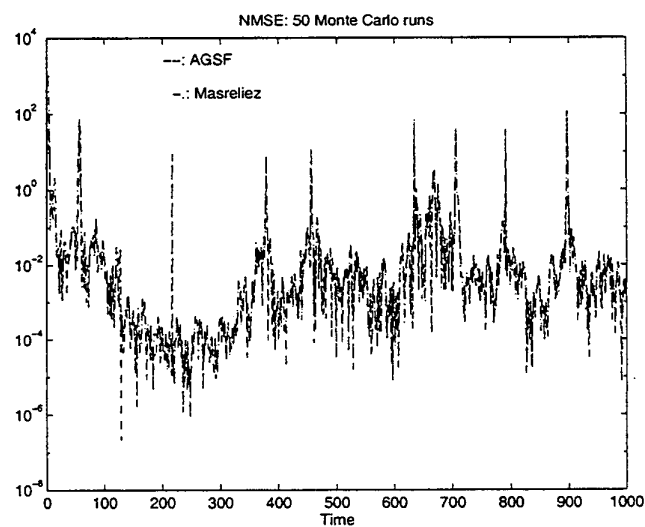

Figure 2. Measurement noise profile $\epsilon = 0.2, K = 10$

Figure 4. State estimation error analysis