Journal of Intelligent and Robotic Systems **19:** 207–231, 1997.

Nonlinear Filtering of Non-Gaussian Noise

K. N. PLATANIOTIS, D. ANDROUTSOS and A. N. VENETSANOPOULOS
Department of Electrical and Computer Engineering, University of Toronto, Toronto, M5S 3G4, Canada. e-mail:kostas@dsp.toronto.edu

(Received: 15 February 1996; accepted: 29 September 1996)

Abstract. This paper introduces a new nonlinear filter for a discrete time, linear system which is observed in additive non-Gaussian measurement noise. The new filter is recursive, computationally efficient and has significantly improved performance over other linear and nonlinear schemes. 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 new filter outperforms currently used approaches and at the same time offers simplicity in the design.

Key words: adaptive algorithms, Gaussian mixtures, Kalman filter, narrowband interference, non-Gaussian filtering.

List of Symbols

$x(k+1)$ = system state
$\Phi(k+1,k)$ = state transition matrix
$w(k)$ = state noise
$z(k)$ = measurement vector
$H(k)$ = measurement matrix
$v(k)$ = measurement noise
$Q(k)$ = state noise covariance
$R(k)$ = measurement noise covariance
$\widehat{x}(0)$ = mean value of the initial state vector
$P(0)$ = covariance of the initial state vector
$\widehat{x}(k \mid k)$ = estimate of the system state $x(k)$ utilizing measurements up to time instant k
$P(k \mid k)$ = estimation covariance
$K(k)$ = filter gain at time index k
$\widehat{x}(k \mid k-1)$ = one-step ahead state prediction
$\widehat{z}(k \mid k-1)$ = innovation sequence
$P_z(k \mid k-1)$ = innovation covariance
$f(\cdot)$ = probability density
$N(\mu, B)$ = Gaussian density with mean value $\mu$and covariance B

1. 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. As a result of the combined research efforts of many scientists in the field, numerous estimation algorithms have been developed. These 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 [1]. The discrete linear, state estimation problem is described by the following model equations and statement of objective:

$$x(k+1) = \Phi(k+1,k)x(k) + w(k), \tag{1}$$

$$z(k+1) = H(k+1)x(k+1) + v(k+1), \tag{2}$$

where $x(k)$ is the n-dimensional state process, $\Phi(k+1,k)$ is the $(n \times n)$ state transition matrix describing the transition of the state from time-step $(k+1)$ to time-step (k), $z(k)$ is the m-dimensional measurement process and $H(k+1)$ is an $(m \times n)$ matrix which relates the observations (measurements) z to the state x. The associated state noise $w(k)$, is modeled as white Gaussian sequence with covariance $Q(k)$.

The observation noise $v(k)$ is considered white Gaussian with covariance $R(k)$ and uncorrelated to the state noise process, that is $E(v(k)w(j)) = 0$, $\forall k, j$. The matrices Φ, H, Q and R are assumed known. The initial state vector $x(0)$, which is in general unknown, is modeled as a random variable, Gaussian distributed with mean value $\widehat{x}(0)$ and covariance $P(0)$, and is considered uncorrelated to the noise processes $\forall k > 0$.

Given the set of measurements $Z^k = [z(1), z(2), \ldots, z(k-1), z(k)]$, we desire the *mean-squared-error* optimal filtered estimate $\widehat{x}(k \mid k)$ of $x(k)$:

$$\widehat{x}(k \mid k) = E\bigl(x(k) \mid Z^k\bigr). \tag{3}$$

The above problem was first solved by Kalman with his well known filter [1]. The 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. Hence, if either of 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 will be 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. On the other hand, this is not the case for the observation noise $v(k)$. Despite the fact that most engineering systems in control, communication and intelligent signal processing are developed under the assumption that the interfering noise is Gaussian, 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. Examples include, among others, natural phenomena, such as atmospheric noise, lightning spikes and ice cracking, and a great variety of man-made noise sources, such as electronic devices, neon lights, relay switching noise in telephone channels and automatic ignition systems [2, 3].

In such an environment, the Kalman filter cannot provide the optimal solution due to the Gaussian assumption in which it is based. Since non-Gaussian measurement noise is usually non-stationary, very dependent on the physical environment, and may be of infrequent occurrence, an intelligent adaptive estimator which can continuously adjust its structure to match changing noise characteristics is of paramount importance.

Such a filter is introduced and analyzed in this paper. The new filter utilizes an intelligent adjustment mechanism which monitors the changes in the noise source and adaptively adjusts its parameters to maintain the required performance.

The rest of the paper is organized as follows: in Section 2 mathematical models for non-Gaussian noise environments are discussed. In Section 3 we briefly review previous approaches to the problem. The new estimator is introduced in this section. Motivation and implementation issues are also discussed there. In Section 4 simulation results are presented and the performance of the proposed adaptive filter is described. Finally, Section 5 summarizes our conclusions.

2. Non-Gaussian Noise Modeling

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 [4, 5]. On the other hand, empirically devised noise models have been used over the years to approximate many non-Gaussian noise distributions. Based on the Wiener approximation theorem, any non-Gaussian noise distribution can be expressed as, or approximated sufficiently well by, a finite sum of known Gaussian distributions.

This so-called 'Gaussian sum' approach is summarized in the following lemma [6, 7]:

LEMMA. *Any density $f(x)$ associated with an n-dimensional vector x can be approximated as closely as desired by a density of the form*

$$f_A(x) = \sum_{i=1}^{l} a_i N(\mu_i, B_i) \tag{4}$$

for some integer l, and positive scalars a_i with $\sum_{i=1}^{l} a_i = 1$, where $N(\cdot)$ is a Gaussian density with mean value μ_i and covariance matrix B_i:

$$N(\mu_i, B_i) = \frac{1}{(2\pi)^n} |B_i|^{-0.5} \exp\left(-0.5\|x - \mu_i\|^2_{B_i^{-1}}\right). \tag{5}$$

It can be shown that the density $f_A(x)$ can converge uniformly to any density function of practical interest by letting the number of terms increase and each elemental covariance approach the zero matrix [6].

This approximation procedure has been used to develop empirical distributions which relate to many physical non-Gaussian phenomena. A special case of the Gaussian sum approach, the *ε-mixture* is of particular interest. The *ε-mixture* noise model has been extensively used to describe a non-Gaussian noise environment in many communication and control systems, such as spread-spectrum communication systems, target tracking in the presence of glint noise, jamming or clutter suppression, outlier rejection in image processing applications and intelligent processing in interferometric and multi-range measurement systems.

The probability density function for such a model is of the following form:

$$f(x) = (1 - \varepsilon) f_G(x) + \varepsilon f_{nG}(x), \tag{6}$$

where $\varepsilon \in (0, 1)$, 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 and 0.25 [8]. 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 λ is on the order of 10 to 10 000 [8, 9].

3. Filtering in Non-Gaussian Noise

3.1. RELATED PREVIOUS WORK

A number of filtering techniques utilize empirical models, such as the Gaussian-sum or the ε-mixture noise model to tackle the problem of estimating the state of a linear system in a non-Gaussian environment.

For a state space model, such as the one described in (1)–(2) a Gaussian-sum additive measurement noise of the form:

$$f(v(k)) = \sum_{i=1}^{l} a_i N(\mu_i, R_i) \tag{7}$$

results in a predictive measurement density $f(z(k) \mid x(k), Z^{k-1})$ which has a similar form [10].

Based on this observation Sorenson and Alspach had developed a mean-squared state estimator for a non-Gaussian noise environment [10, 11]. They assumed that the noise sequences have a uniformly convergent series expression in terms of known Gaussian distributions. A fixed number of Gaussian terms with known moments is then used to develop an optimal (under these assumptions), minimum-mean-square-error filtering algorithm. The output of their so-called

'Gaussian sum' filter is formed by combining elemental estimates from a bank of Kalman filters, each one matched to a specific term of the Gaussian sum.

Their methodology has a major drawback, namely its computational complexity since the numerical computations in the filter increase almost exponentially in time. For the case of additive Gaussian noise, the number of terms involved in the derivation of the optimal Kalman filter remain constant. In their filter, however, for any time index (k) the predictive measurement density $f(x(k) \mid Z^{k-1})$ is a Gaussian mixture of (l_1) components with $(l_1 > l)$, and since the noise distribution $v(k)$ of (7) contains (l) components, the combined density needed for the next step calculations is a mixture with $(l_1 x l)$ components. Thus, the computational burden at each stage becomes larger as the number of terms in the mixtures increases. Hence, due to its computational complexity the 'Gaussian sum' filter is not practical and in many applications not feasible.

In an attempt to alleviate the computational burden associated with the 'Gaussian sum' filters, Masreliez introduced a new filter which is more robust than the Kalman filter [12, 14]. His methodology is based on the so-called 'score-function'. This function is a custom-tailored nonlinearity, used in a Kalman-like recursive filter to de-emphasize the effect of large noise residuals on the state estimate. The filter designer has to decide on the form of the nonlinearity based knowledge on about the noise characteristics which must be available *a priori*. It is obvious, that since its 'score function' is custom-tailored to fit a specific noise form, the Masreliez filter cannot operate in a changing noise environment and cannot adjust its characteristics on-line to match changing noise sources. Thus, although is more robust than the linear Kalman filter in a fixed nonlinear environment it cannot be classified as an 'intelligent' (adaptive) filtering algorithm [12, 13].

In addition, despite the fact that is computationally more efficient than the Alspach and Sorenson filter, it still requires a rather demanding convolution operation in the evaluation of the nonlinear score function. In summary, the need of an 'appropriate' nonlinearity and the calculations required for its evaluation often limits the practicality of this method in many engineering applications.

3.2. THE NEW FILTER

Thus, a new adaptive filter is needed that is computationally attractive and does not require any problem-dependent nonlinearities in its design. To this end, such a filter is introduced here. The new filter utilizes the same methodology with that of the 'Gaussian sum filter' [10, 11]. However, in order for the procedure to be practical, the number of terms in the Gaussian mixture is controlled adaptively at each step. A Bayesian learning technique is utilized to collapse, in an intelligent way, 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 $f(z(k) \mid x(k), Z^{k-1})$ 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' approach without the use of problem dependent nonlinearities, such as those required by the Masreliez filter.

The main points of our strategy can be summarized as follows:

1. For each Gaussian term in (7), which describes the observation noise, a dedicated Kalman filter is employed. These filters can operate in parallel to reduce the processing time.
2. Based on the interim results from these dedicated Kalman filters we obtain a Bayesian *a posteriori* approximation of the Gaussian mixture $f(z(k) \mid x(k), Z^{k-1})$ required in the filtering process. It should be noted at this point that the filter adaptively approximates the predictive measurement density at every filter cycle.
3. Thus, through the Bayesian adaptation, the optimal (in the minimum mean square error sense) Gaussian approximation for the above mixture can be obtained. Then, the first two moments of this equivalent Gaussian term are used to complete the filtering cycle of a recursive, Kalman-like filter.

The equations of the new filter are summarized in the following theorem.

THEOREM (The adaptive filter). *For the linear dynamic system described in* (1)–(2), *if the additive measurement noise is modeled by* (7), *the estimate* $\widehat{x}(k \mid k)$ *of the system state* $x(k)$ *at time step* k *can be computed recursively as follows:*

$$\widehat{x}(k \mid k) = \widehat{x}(k \mid k-1) + K(k)(z(k) - \widehat{z}(k \mid k-1)), \tag{8}$$

$$P(k \mid k) = (I - K(k)H(k))P(k \mid k-1), \tag{9}$$

$$\widehat{x}(k \mid k-1) = \Phi(k, k-1)\widehat{x}(k-1 \mid k-1), \tag{10}$$

$$P(k \mid k-1) = \Phi(k, k-1)P(k-1 \mid k-1)\Phi(k, k-1)^{\tau} + Q(k-1), \tag{11}$$

with initial conditions $\widehat{x}(0 \mid 0) = \widehat{x}(0)$ *and* $P(0 \mid 0) = P(0)$.

$$K(k) = P(k \mid k-1)H^{\tau}(k \mid k-1)P_z^{-1}(k \mid k-1), \tag{12}$$

$$\widehat{z}(k \mid k-1) = \sum_{i=1}^{l} w_i(k)\widehat{z}_i(k \mid k-1), \tag{13}$$

$$\widehat{z}_i(k \mid k-1) = H(k)\widehat{x}(k \mid k-1) + \mu_i, \tag{14}$$

$$P_{zi}(k \mid k-1) = H(k)P(k \mid k-1)H^{\tau}(k) + R_i, \tag{15}$$

with the corresponding innovation covariance and the a posteriori *weights used in the Bayesian approximation defined as*:

$$P_z(k \mid k-1) = \sum_{i=1}^{l} \Big(P_{zi}(k \mid k-1) + \big(\widehat{z}(k \mid k-1) - \widehat{z}_i(k \mid k-1)\big) \times \times \big(\widehat{z}(k \mid k-1) - \widehat{z}_i(k \mid k-1)\big)^{\tau} \Big) w_i(k), \tag{16}$$

$$w_i(k) = \frac{\Big((2\pi)^{-m}|P_{zi}|^{-1} \exp\big(-0.5(\|z(k) - \widehat{z}_i(k \mid k-1)\|^2_{P_{zi}^{-1}(k|k-1)})\big)\Big) a_i}{c(k)}, \tag{17}$$

where $|\cdot|$ *denotes the determinant of the matrix,* $\|\cdot\|$ *is the inner product in the Euclidean space* R^m, a_i *are the initial weighting coefficients used in* (7) *and* $c(k)$ *is a normalization factor defined as follows*:

$$c(k) = \sum_{i=1}^{l} \Big((2\pi)^{-m} |P_{zi}|^{-1} \times \times \exp\big(-0.5(\|z(k) - \widehat{z}_i(k \mid k-1)\|^2_{P_{zi}^{-1}(k|k-1)})\big) \Big) a_i. \tag{18}$$

Proof. The proof is given in the appendix.

3.3. COMMENTS

• The new filter is easy to implement, requires no special information and can adapt to changes in the noise environment. Through the intelligent re-evaluation of its *a posteriori* weights in (16)–(18), the filter can follow the true underlying noise conditions coping with uncertainties which may occur due to large variations in the noise signal. As such, our approach constitutes a form of intelligent signal processing since we can identify in it characteristics unique on intelligent mechanisms, namely:

1. ‘self-learning’: the filter utilizes a self-learning mechanism (the Bayesian adaptation of (16)–(18)) in order to track changes in a non-stationary noise environment.
2. ‘robust performance’: its performance is insensitive in extreme values (outliers).

• The collapsed density, which is used to approximate $f(z(k) \mid x(k), Z^{k-1})$, has only one Gaussian term, thus it can be incorporated in the recursive form of the usual Kalman filter. The nonlinear weights ensure that the collapsed equivalent density captures any skewness or bimodality existing in the original nonlinear mixture.

• The performance of the nonlinear filter depends on the approximation of the Gaussian mixture by the single Gaussian term. The rationale of this approximation lies in the fact that some of the members in the original density have small mixing weights at a particular time instant and hence the information that they carry can safely be ignored for practical purposes. The *Bhattacharyya coefficient*, which is defined as:

$$\rho_{ij} = \int [f_i(x) f_j(x)]^{0.5}\, \mathrm{d}x \tag{19}$$

with $0 \leqslant \rho_{ij} \leqslant 1$ and $\rho_{ij} = 1$ if $f_i(x) = f_j(x)$ [15], can be used to measure the validity of the approximation by calculating the distance between the actual Gaussian mixture and the Gaussian term resulting after the collapse of the mixture.

• It can be seen in the theorem above that the density $f(z(k) \mid x(k), Z^{k-1})$ is represented by a finite number of parameters ($\widehat{z}_i(k \mid k-1)$, $P_{zi}(k \mid k-1)$) which are obtained using recursive Kalman filters, each matched to a specific set of initial conditions (μ_i, R_i), $i = 1, 2, \ldots, l$. 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 considerable implementation advantages due to its partitioned structure. Namely, its naturally decoupled parallel structure lends to parallel processing, since its parts consist of the set or recursive linear (Kalman) filters which are easily implementable, and the set of *a posteriori* coefficients used in the adaptive density approximation which are easily obtained by the recursive Bayes algorithm of (16)–(18). This decoupled parallel structure of the new algorithm lends it well to easy implementation in term of array processors. With the advent of inexpensive processors, the decoupled, parallel-processing nature of the new algorithm is of substantial practical importance since it affords inexpensive realizations of an nonlinear estimator. In practical terms, due to its parallel, decoupled structure the new filter has computational complexity similar to that of the simple Kalman (linear) filter (Figure 1).

By contrast, other nonlinear filters, such as the Masreliez filter or filters with pre-processing nonlinearities are computationally expensive. Such designs do not allow for off-line gain computations and require expensive numerical evaluations of convolutions in the realization of the specialized nonlinear functions which they use. To emphasize the difference between the proposed here approach and the Masreliez filter we consider a system similar to the one described in (1)–(2) with state and measurement dimension $(n \times 1)$ when the observation noise given as a Gaussian mixture. For such a system the Kalman filter, and thus our approach through its parallel implementation, requires $2n^3 + 3n^2 + 2n + 3$ multiplications and one division per recursion. The Masreliez filter requires $2n^3 + 3n^2 + 2n + 9$ multiplications, seven divisions, one exponential and one square root operations in addition to the numerical method required for the evaluation of the score function. Therefore, we can conclude that the Masreliez filter requires more

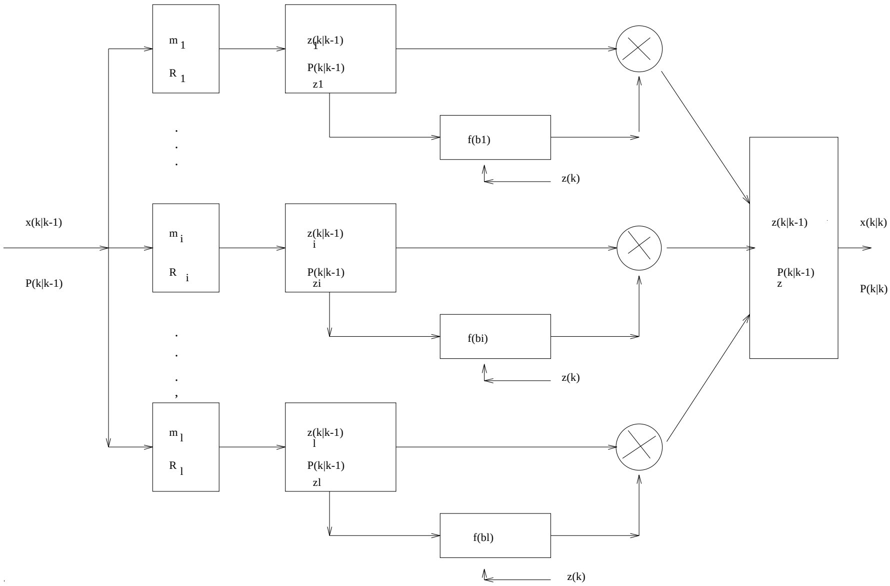

Figure 1. The new nonlinear filter.

computations than our methodology. A detailed analysis of the computations involved in the implementation of recursive filters is provided in [17].

The interested reader can refer to it for more information on the subject.

• Finally, the filter proposed here has a robust nature. This can be demonstrated by noting that its decoupled (parallel) structure (Figure 1) has a natural 'failure detection' mechanism built into its weighted-sum adaptation mechanism of (16)–(18). If one of the mixture component conditional filter fails, its estimate $\widehat{z}_i(k \mid k-1)$ will be inferior to those of the other filters. This, in turn, will make the corresponding adaptation weight tend to 0, as can been seen from (16)–(18), which will cut off the diverging component from the mixture, in essence, correcting the filter malfunction.

4. Application to Narrowband Interference Suppression in Impulsive Channels

The adaptive nonlinear filter proposed here can successfully be applied for smoothing out non-Gaussian measurement noise in a plethora of engineering applications. The problem of narrowband interference rejection is discussed in this section as an important example [16].

In many communication and control systems the accurate estimation of an unknown narrowband signal is needed. The nonlinear filters discussed in the previous sections can be used to reject autoregressive interference in a non-Gaussian channel. Spread-spectrum techniques provide an effective way to cope with narrowband interference. The basic idea is to spread the bandwidths of transmitting signals so that they are much greater than the information rate. The problem of interest in this paper 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:

$$i_k = \sum_{n=1}^{N} \Phi_n i_{k-n} + e_k, \tag{20}$$

where e_k is a zero mean white Gaussian noise process and $\Phi_1, \Phi_2, \ldots, \Phi_{N-1}, \Phi_N$ are the autoregressive parameters known to the receiver.

The Direct Sequence Spread Spectrum (DSSS) modulation waveform is written as:

$$m(t) = \sum_{k=0}^{N_c-1} c_k q(t - k\tau_c), \tag{21}$$

where N_c is the pseudonoise chip sequence used to spread the transmitted signal and $q(\cdot)$ is a rectangular pulse of duration τ_c. The transmitted signal can be then expressed as:

$$s(t) = \sum_{k} b_k m(t - kT_b), \tag{22}$$

where $b(k)$ is the binary information sequence and $T_b = N_c\tau_c$ is the bit duration. Based on that, the received signal is defined as:

$$z(t) = as(t - \tau) + n(t) + i(t), \tag{23}$$

where a is an attenuation factor, τ is a delay offset, $n(t)$ is wideband Gaussian noise and $i(t)$ is narrowband interference. Assuming that $n(t)$ is band limited and hence white after sampling, with $\tau = 0$ and $a = 1$ for simplicity, if the received signal is chip-matched and sampled at the chip rate of the pseudonoise sequence, the discrete time sequence resulting from (23) can be written as follows:

$$z(k) = s(k) + n(k) + i(k). \tag{24}$$

Given these assumptions, we can safely consider $s(k)$ to be a sequence of independent identically distributed (i.i.d.) samples taking values $+1$ or -1.

Using the model in (24) a state space representation for the received signal and the interference can be constructed as follows:

$$\begin{aligned} x(k) &= \Phi x(k-1) + w(k), \\ z(k) &= Hx(k) + v(k) \end{aligned} \tag{25}$$

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

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

The additive observation noise $v(k)$ in the state space model is defined as:

$$v(k) = n(k) + s(k).$$

Since the first component of the system state $x(k)$ is the interference $i(k)$, an estimate of the state contains an estimate of $i(k)$ which can be subtracted from the received signal in order to increase the system's performance. For the case of $v(k)$ being a Gaussian process the optimal estimator is the Kalman filter. However, in the problem under consideration the measurement noise $v(k)$ is the sum of two independent variables, one is Gaussian distributed and the other takes on values -1 or $+1$. This is a non-Gaussian sequence with probability density:

$$f(v(k)) = (1 - \varepsilon)N(\mu, \sigma_n^2) + \varepsilon N(-\mu, \lambda\sigma_n^2) \tag{26}$$

with $\mu = 1$ for the case under consideration. 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:

$$i_k = 1.98 i_{k-1} - 0.9801 i_{k-2} + e_k, \tag{27}$$

where e_k is zero mean white Gaussian noise with variance 0.01.

To study the applicability of the proposed algorithm in a non-Gaussian environment the regulatory coefficient ε is set to be $\varepsilon = 0.2$ and the ratio λ is taken to be $\lambda = 10$ or $\lambda = 10\,000$ with $\sigma_n = 1.0$.

The following algorithms are used in the simulation studies reported here:

1. Kalman Filter – I

This linear (Kalman) filter assumes that the measurement noise is Gaussian with mean value $\mu_1 = 1$ and covariance $R_1 = 1.0$. In other words, this filter is matched to the 'nominal' Gaussian component of the noise model in (26).

2. Kalman Filter – II

This linear (Kalman) filter assumes that the measurement noise is Gaussian with mean value $\mu_2 = -1$ and covariance $R_2 = \lambda R_1$. In other words, the filter is matched to the 'dominant' noise component.

3. Adaptive Filter

The proposed here adaptive filter of Theorem 1 is the third filter considered. The filter utilizes two elemental Kalman filters to calculate its adaptive weights.

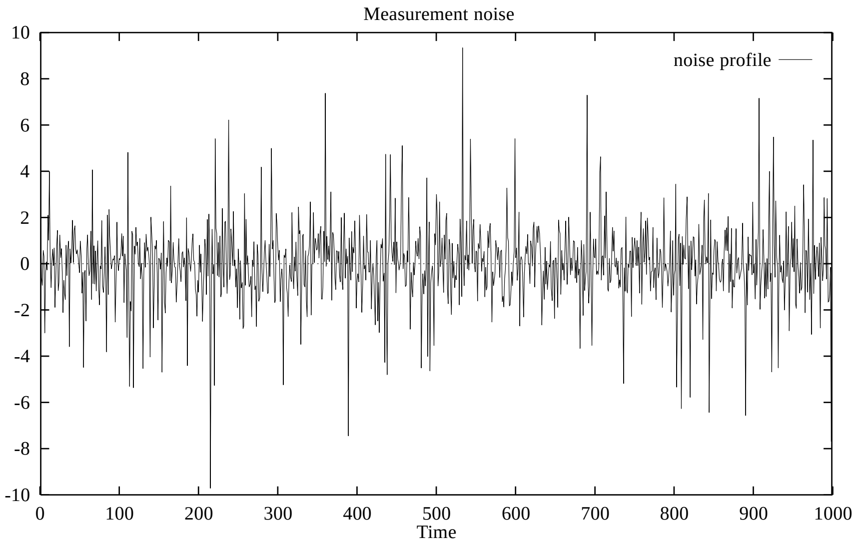

Figure 2. Measurement noise profile ($\varepsilon = 0.2$, $\lambda = 10$).

4. Masreliez Filter

The Masreliez filter with a nonlinear function, defined in [16] to fit the noise model in (26) is the last filter included in the comparison.

All four filters require an initial filtered estimate $\widehat{x}(0 \mid 0)$ and an initial error covariance $P(0 \mid 0)$. The values used in this experiment are; mean value 0.01 and covariance 1.0. In addition, the values of autoregressive parameters of (20) are assumed to be known to the filters.

The normalized mean square error (NMSE) is utilized for filter comparison purposes in all experiments. The data were averaged through Monte Carlo techniques. Given the form of the state vector in (25) the first component of $x(k)$ is used in the evaluation analysis. The NMSE is therefore defined as:

$$\mathrm{NMSE} = \frac{1}{\mathrm{MCRs}} \left(\sum_{k=1}^{\mathrm{MCRs}} \frac{\left(x_{1r}^{k} - \widehat{x}_{1j}^{k}\right)^2}{{x_{1r}^{k}}^{2}} \right), \tag{28}$$

where MCRs is the number of Monte Carlo runs, x_{1r} the actual value and $\widehat{x}_{1j}$ is the outcome of the j-filter under consideration.

In this experiment, 50 independent runs were processed, each 1000 samples in length. The non-Gaussian measurement noise profile, for a single run, is depicted in Figure 2 ($\lambda = 10$) and in Figure 5 for $\lambda = 10\,000$.

To quantitatively assess the performance of the different filters we compare their performance against the Kalman filter which is matched to the nominal

Table I. Performance evaluation: $\mu_1 = 1$, $R_1 = 1$, $\mu_2 = 1$, $R_2 = 10$, $\varepsilon = 0.05$

Filter	NMSE	Improvement
KF-I	233.79	–
KF-II	6159.5	No improvement
Adaptive	181.85	22.3%
Masreliez	602.09	No improvement

Table II. Performance evaluation: $\mu_1 = 1$, $R_1 = 1$, $\mu_2 = 1$, $R_2 = 10000$, $\varepsilon = 0.05$

Filter	NMSE	Improvement
KF-I	15786.3	–
KF-II	385191.7	No improvement
Adaptive	208.04	98%
Masreliez	7804.7	50.56%

Table III. Performance evaluation: $\mu_1 = 1$, $R_1 = 1$, $\mu_2 = 1$, $R_2 = 10$, $\varepsilon = 0.1$

Filter	NMSE	Improvement
KF-I	321.7	–
KF-II	4555.24	No improvement
Adaptive	177.67	44.8%
Masreliez	482.96	No improvement

observation noise conditions (KF-I). The *Measure of Improvement* is defined as follows:

$$\mathrm{MI} = \frac{\mathrm{NMSE}_{(\mathrm{KF\text{-}I})} - \mathrm{NMSE}_{(\mathrm{filter})}}{\mathrm{NMSE}_{(\mathrm{KF\text{-}I})}}. \tag{29}$$

Tables I–VI summarize the experimental results obtained for different noise scenarios.

In order to facilitate the performance of the new scheme in a heavy-tailed non-Gaussian noise environment, a second simulation experiment is performed. In this second experiment, the observation noise is assumed to have the following probability density:

$$\begin{aligned} f(v(k)) \;=\; 0.5\Big[&(1-\varepsilon)[N(-\mu,\sigma_n^2) + N(\mu,\sigma_n^2)] + \\ &+ \varepsilon[N(-\mu,k\sigma_n^2) + N(\mu,k\sigma_n^2)]\Big] \end{aligned} \tag{30}$$

Table IV. Performance evaluation: $\mu_1 = 1$, $R_1 = 1$, $\mu_2 = 1$, $R_2 = 10000$, $\varepsilon = 0.1$

Filter	NMSE	Improvement
KF-I	54874.8	–
KF-II	382408.16	No improvement
Adaptive	186.447	99%
Masreliez	2576.46	95%

Table V. Performance evaluation: $\mu_1 = 1$, $R_1 = 1$, $\mu_2 = 1$, $R_2 = 10$, $\varepsilon = 0.2$

Filter	NMSE	Improvement
KF-I	307.14	–
KF-II	5111.49	No improvement
Adaptive	253.013	17.5%
Masreliez	299.17	2.6%

Table VI. Performance evaluation: $\mu_1 = 1$, $R_1 = 1$, $\mu_2 = 1$, $R_2 = 10000$, $\varepsilon = 0.2$

Filter	NMSE	Improvement
KF-I	189938.85	–
KF-II	54239115.3	No improvement
Adaptive	179.8737	99%
Masreliez	2065.71	89%

with $\varepsilon = 0.2$, $\mu = 5$, $\lambda = 100$ and $\sigma_n = 1.0$. In the context of narrowband interference suppression, this noise model corresponds to transmission of binary direct-sequence spread-spectrum over an impulsive channel.

In this experiment, 100 independent runs (Monte Carlo runs), each 1000 samples in length were considered. Due to its high complexity and the unavailability of suitable nonlinear transformation for the 'score function' the Masreliez filter was not included in these simulation studies.

Two different plot types are reported in the paper. First, state estimation plots for single Monte Carlo runs are included to facilitate the performance of the different estimation schemes. In addition, the normalized mean square error plots for all the simulation studies are also reported.

Significant findings and corresponding remarks are here organized in a series of comments, which are supported by the appropriate figures.

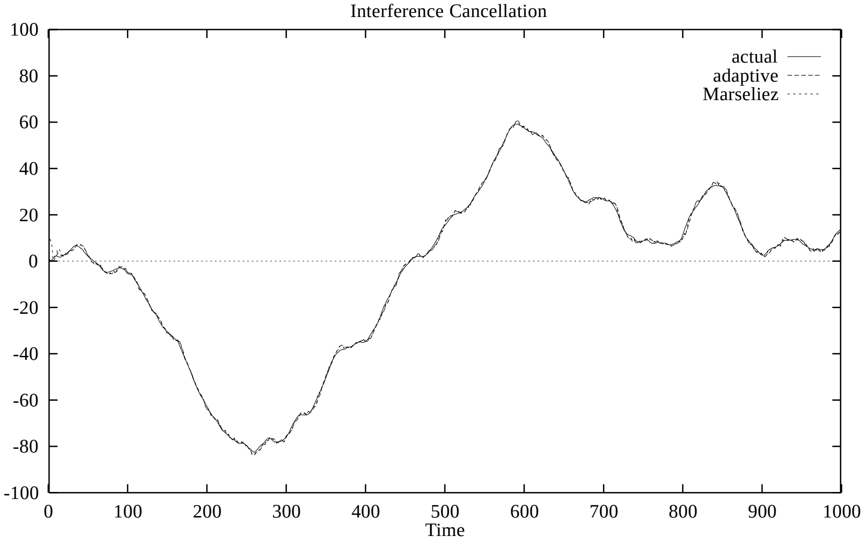

Figure 3. State estimation results ($\varepsilon = 0.2$, $\lambda = 10$).

1. 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 $P_z(k \mid k-1)$. The divergence from the theoretical covariances involved in the calculations of the Kalman filter gain leads to unacceptable estimation results.

2. The nonlinear Masreliez filter performs relatively well in some non-Gaussian channels. Given its complexity, the need for customized, problem dependent nonlinearities in the 'score-function' and the inconsistency in its performance (see Tables I–VI) this filter cannot be considered as a general purpose robust recursive estimator.

3. The new filter performed well under all the different noise scenarios selected. From the tables and plots included in the paper we can clearly see the improvement accomplished by the utilization of the new filter versus the Kalman filter and the Masreliez filter (see Tables I–VI). 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.

4. The new filter is computationally more efficient than the Masreliez filter. No *a priori* defined score function is needed in its implementation. The new filter has a natural decoupled structure which makes it suitable for parallel implementation in real time. On the contrary the Masreliez filter requires numerical methods, such as the iterative Newton method for the evaluation of the nonlinearities used. It is widely accepted that such iterative processes have heavy

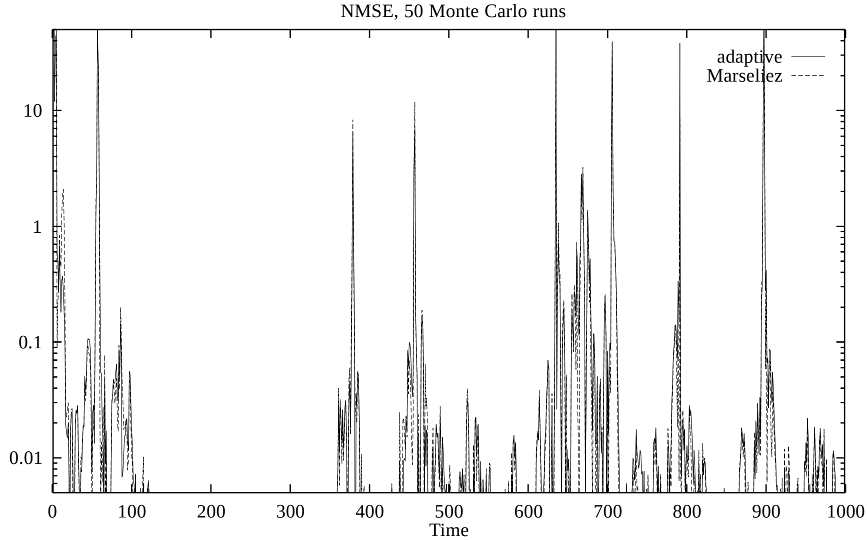

Figure 4. State estimation results ($\varepsilon = 0.2,\ \lambda = 10$).

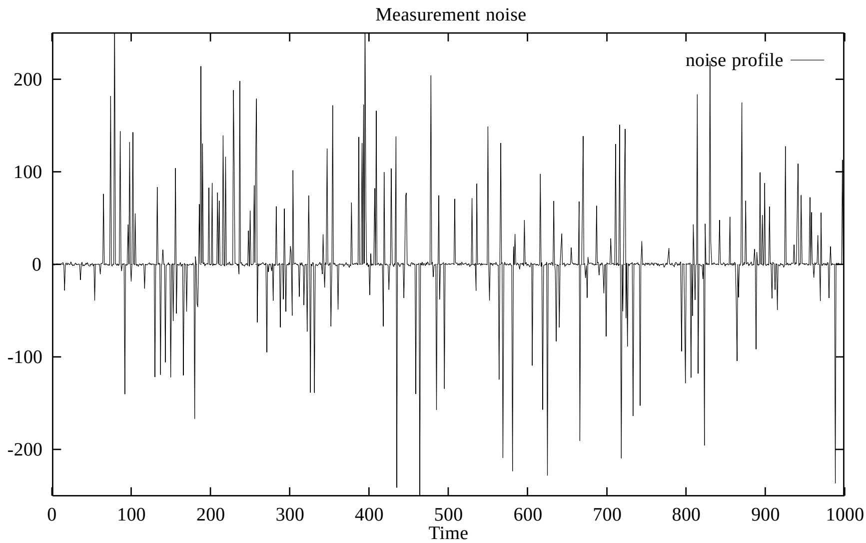

Figure 5. Measurement noise profile ($\varepsilon = 0.2,\ \lambda = 10\,000$).

computational requirements and usually their real-time implementation is not feasible. Furthermore, the development of multidimensional score functions cannot, in general, be handled analytically and often approximation formulas have to be used. On the contrary, the new filter has the same structure in any dimension and no modification is necessary.

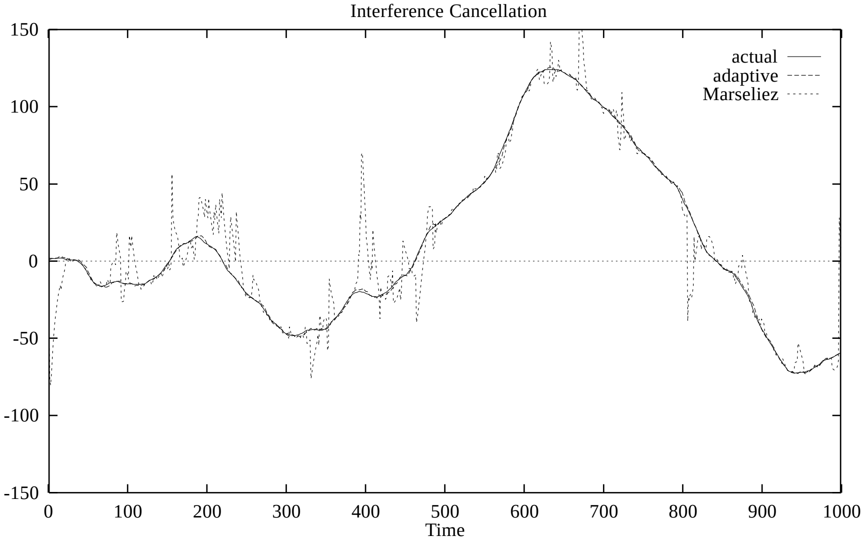

Figure 6. Nonlinear Filters: state estimation results ($\varepsilon = 0.2$, $\lambda = 10\,000$).

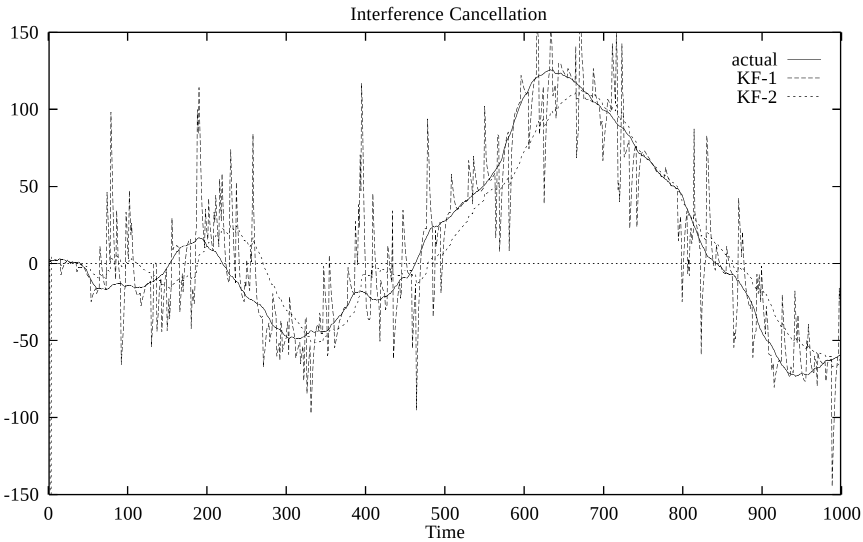

Figure 7. Linear Filters: state estimation results ($\varepsilon = 0.2$, $\lambda = 10\,000$).

In conclusion, comparing nonlinear filters it is obvious that for the case of additive non-Gaussian observation noise the proposed here adaptive filter should be used instead of the Masreliez filter since it delivers better results exhibiting at the same time significantly less computational complexity.

However, the question to be answered at this point is if one should use a nonlinear filter instead of a linear (e.g. Kalman) filter based on a fixed *a priori*

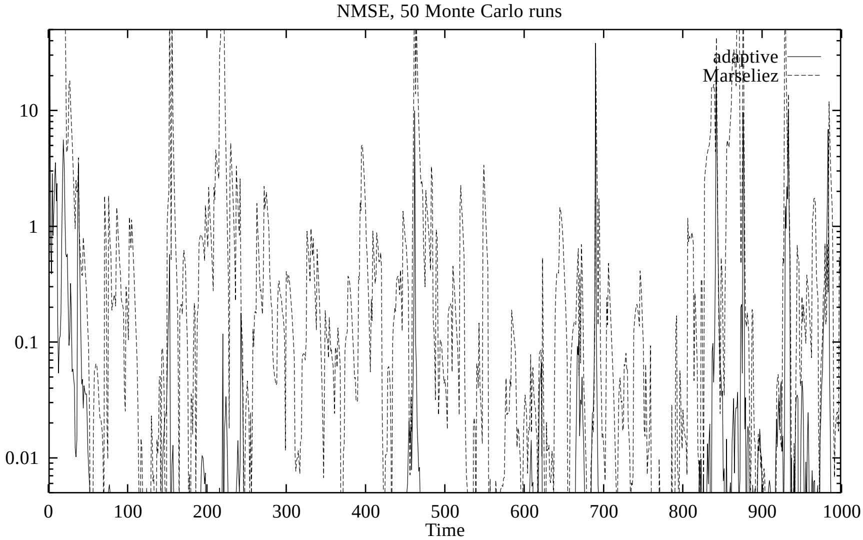

Figure 8. State estimation error analysis ($\varepsilon = 0.2$, $\lambda = 10\,000$).

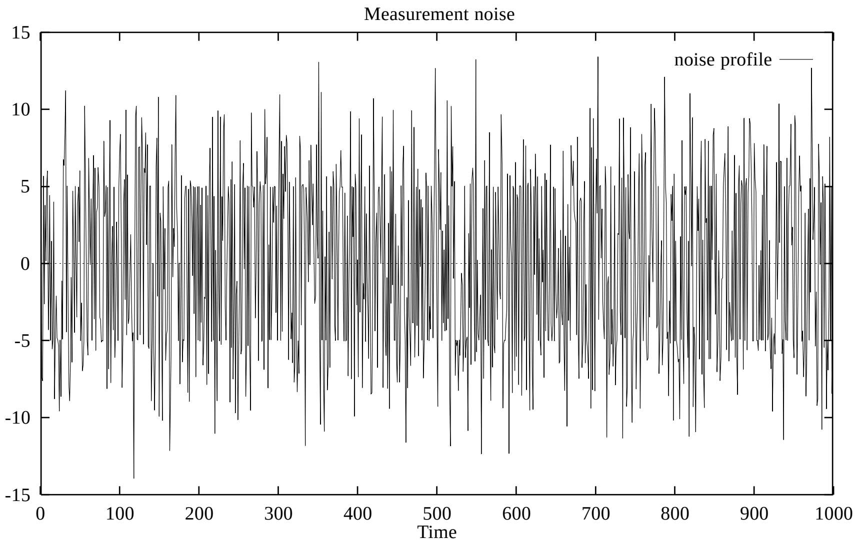

Figure 9. Measurement noise profile ($\varepsilon = 0.2$, $\lambda = 100$, $m = 5$).

Gaussian approximation to the non-Gaussian measurement noise. This question may be answered via Monte Carlo simulation of the systems that represent the particular design problems. As an example, Monte Carlo studies have been made on the linear Kalman filters, the Masreliez filter and the new filter when the

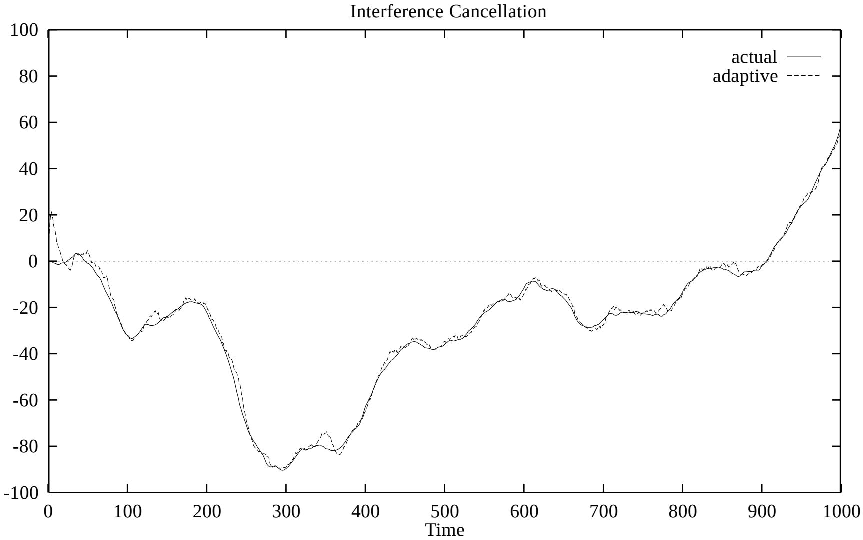

Figure 10. Nonlinear Filter: state estimation results ($\varepsilon = 0.2$, $\lambda = 100$, $m = 5$).

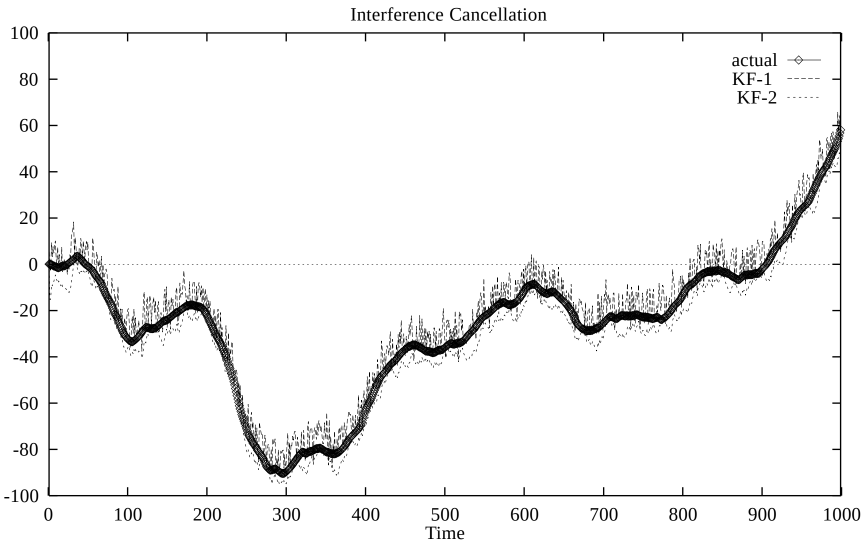

Figure 11. Linear Filters: state estimation results ($\varepsilon = 0.2$, $\lambda = 100$, $m = 5$).

observation noise is modeled by (26). For this case if the maximum allowable mean square deviation from the best achieved estimate is δ,

$$(\|\widehat{x}_b(k \mid k) - \widehat{x}_s(k \mid k)\|)^2 \leqslant \delta$$

one may switch from the nonlinear filter to a single Kalman filter without sacrificing performance. Performance look-up tables, such as the Tables I–VI included here can be used to guide the designer.

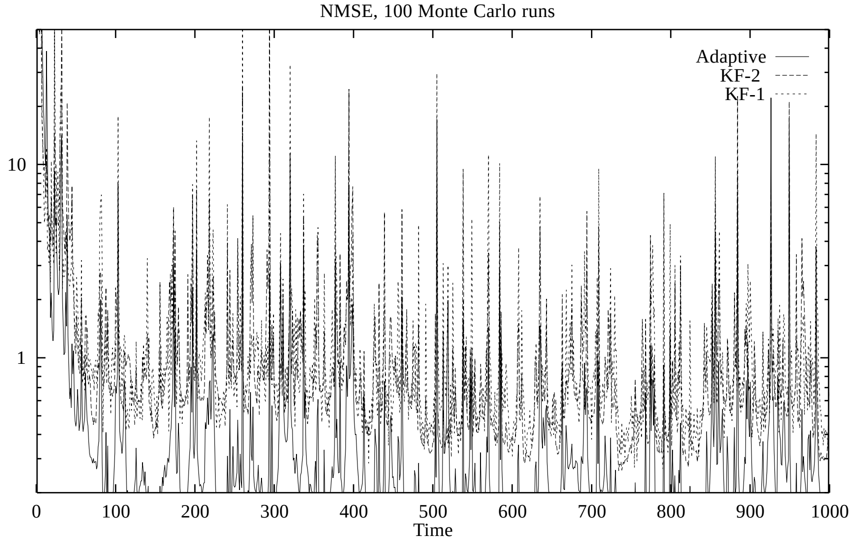

Figure 12. State estimation error analysis ($\varepsilon = 0.2$, $\lambda = 100$, $m = 5$).

Table VII. 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 VIII. Filter comparison

Filter	a	b	c	d
Kalman	0	0	0	2
Masreliez	2	1	–	0
Bayesian adaptive	2	2	2	0

In general, 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 (Table VIII).

5. Conclusions

The paper has addressed the important problem of state estimation in non-Gaussian observation channels. A new, intelligent, robust and computationally efficient filter was introduced. The proposed design constitutes a form of intelligent signal processing system. Characteristics unique to intelligent designs can be identified in our estimator. Namely,

- it is able to cope with uncertainties which may occur due to large variations in the parameter values, environmental conditions and signal (noise) inputs;
- can adapt to changing conditions through its Bayesian, adaptive (self learning) module of (16)–(18);
- is robust in the sense that its performance is insensitive to changes in the environment (noise outliers);
- it can resist failures due to its natural decoupled structure.

The problem of narrowband interfere suppression in impulsive noise channels has been also discussed. Extensive experimentation has been introduced to demonstrate the effectiveness of the new filter in this problem. Results indicate that the new filter performs better than presently available linear or nonlinear filters. In addition, the new adaptive filter is computationally attractive, has a natural decoupled structure suitable for parallel processing and does not require any problem-dependent nonlinearities in its design. As such the new filter constitutes an excellent general purpose real-time filtering system which can deliver acceptable results in a variety of noise filtering applications.

Appendix

In this appendix we outline the derivation of the new filter. We start with some results used in the derivation of the optimal estimator and the linear Kalman filter.

The probabilistic description of the optimal filter is as follows:

$$\begin{aligned} f(x(k), z(k) \mid Z^{k-1}) &= f(x(k) \mid z(k), Z^{k-1}) f(z(k) \mid Z^{k-1}) \\ &= f(z(k) \mid x(k), Z^{k-1}) f(x(k) \mid Z^{k-1}), \end{aligned} \tag{31}$$

or

$$\begin{aligned} f(x(k) \mid z(k), Z^{k-1}) &= \frac{f(z(k) \mid x(k), Z^{k-1}) f(x(k) \mid Z^{k-1})}{f(z(k) \mid Z^{k-1})} \\ &= \frac{f(z(k) \mid x(k), Z^{k-1}) f(x(k) \mid Z^{k-1})}{\int f(z(k) \mid x(k), Z^{k-1}) f(x(k) \mid Z^{k-1}) \, \mathrm{d}x(k)}. \end{aligned} \tag{32}$$

Let us assume that a linear, discrete time, time varying system, such as the one described in (1)–(2) is available. Let us for a moment also assume that the measurement noise $f(v(k))$ is a white Gaussian density,

$$N(\mu, R) = \frac{1}{(2\pi)^m} |R|^{-0.5} \exp(-0.5 \|v(k) - \mu\|^2_{R^{-1}}), \tag{33}$$

where $E[(v(k)-\mu)(v(j)-\mu)^\tau] = R\delta_{ij}$. For such a model the predicted measurement density $f(z(k) \mid x(k), Z^{k-1})$ required for the implementation of the estimator can be defined as:

$$\begin{aligned} f(z(k) \mid x(k), Z^{k-1}) &= f(z(k) \mid x(k)) = f_{v|x}(z(k) - x(k)) \\ &= f_v(z(k) - x(k)) \\ &= N(\widehat{z}(k \mid k-1), P_z(k \mid k-1)), \end{aligned} \tag{34}$$

where $f_{v|x}(\cdot)$ denotes the conditional pdf of v given x and $f_{v|x}(\cdot) = f_v(\cdot)$ when v and x are independent. The predicted measurement is defined as:

$$\begin{aligned} \widehat{z}(k \mid k-1) &= E(z(k) \mid Z^{k-1}) \\ &= E(H(k)x(k) + v(k) \mid Z^{k-1}) \\ &= H(k)\widehat{x}(k \mid k-1), \end{aligned} \tag{35}$$

yielding the measurement prediction error

$$\begin{aligned} \tilde{z}(k \mid k-1) &= z(k) - \widehat{z}(k \mid k-1) \\ &= H(k)(x(k) - \widehat{x}(k \mid k-1)) + v(k) \end{aligned} \tag{36}$$

with the corresponding measurement prediction covariance

$$P_z(k \mid k-1) = H(k)P(k \mid k-1)H^\tau(k) + R(k) \tag{37}$$

with all the quantities defined in the Theorem. These two moments sufficiently describe the density $f(z(k) \mid x(k), Z^{k-1})$ and are used in the recursive calculations involved in the derivation of the Kalman filter.

However, in our case $f(v(k))$ is not simply one Gaussian density, but a linear combination of such densities resulting in a density $f(z(k) \mid x(k), Z^{k-1})$ which can be similarly expressed. Let us denote $f(v(k))$ as:

$$f(v(k)) = \sum_{i=1}^{l} a_i N(\mu_i, R_i) \tag{38}$$

with μ_i, R_i known means and covariance matrices, and a_i weighting coefficients in the mixture defined in the closed interval $[0, 1]$.

For the same reasoning as in the Gaussian case, the density $f(z(k) \mid x(k), Z^{k-1})$ can now be expressed as a linear combination of l Gaussian terms each one of which corresponds to a density in the definition of $f(v(k))$. We now examine the question of defining the mixture regulating weights when a measurement $z(k)$ becomes available. From the Bayes' rule the posterior, after the measurement $z(k)$ is available, probability which describes the contribution of the ith elemental Gaussian term to the density $f(z(k) \mid x(k), Z^{k-1})$ is given as:

$$\alpha(k) = f(z(k) \mid x(k), Z^{k-1}, b_i) f(b_i \mid x(k), Z^{k-1}), \tag{39}$$

$$\alpha_\tau(k) = \sum_{i=1}^{l} f(z(k) \mid x(k), Z^{k-1}, b_i) f(b_i \mid x(k), Z^{k-1}), \tag{40}$$

$$f(b_i \mid z(k), x(k), Z^{k-1}) = \frac{\alpha(k)}{\alpha_\tau(k)}. \tag{41}$$

However, conditioned on b_i (the indicator which defines the ith member) the density $f(z(k) \mid x(k), Z^{k-1}, b_i)$ is a Gaussian density defined as:

$$f(z(k) \mid x(k), Z^{k-1}, b_i) = N(\widehat{z}_i(k \mid k-1), P_{zi}(k \mid k-1)), \tag{42}$$

where the elemental mean and covariance are calculated through an elemental Kalman filter as follows:

$$\widehat{z}_i(k \mid k-1) = H(k)\widehat{x}(k \mid k-1) + \mu_i, \tag{43}$$

$$P_{zi}(k \mid k-1) = H(k)P(k \mid k-1)H^\tau(k) + R_i. \tag{44}$$

The *a priori* density $f(b_i \mid x(k), Z^{k-1})$ reflects the prior knowledge we have about the weighting coefficient before the measurement $z(k)$ becomes available. Since we assume that at each time instant a mixture of white Gaussian densities with fixed regulators a_i used to model $f(v(k))$ the probability reduces to $f(b_i \mid x(k), Z^{k-1}) = a_i$. Thus, the Bayesian estimate of the density after the measurement $z(k)$ has become available is given as:

$$\widehat{f}_B(z(k) \mid x(k), Z^{k-1}) = \sum_{i=1}^{l} f(z(k) \mid x(k), Z^{k-1}, b_i) f(b_i \mid z(k), x(k), Z^{k-1}), \tag{45}$$

where $\widehat{f}_B(\cdot)$ stands for the *a posteriori* Bayesian estimate of the density $f(z(k) \mid x(k),\ Z^{k-1})$.

However, a Gaussian sum like this can not be easily accommodated in a recursive Kalman like filter. Thus, it has to be replaced by an equivalent Gaussian term.

The best mean-square approximation of the $f(z(k) \mid x(k), Z^{k-1})$ is a Gaussian density with mean:

$$\widehat{z}(k \mid k-1) = \sum_{i=1}^{l} \widehat{z}_i(k \mid k-1) f(b_i \mid z(k), x(k), Z^{k-1}), \tag{46}$$

covariance:

$$b(k) = (\widehat{z}(k \mid k-1) - \widehat{z}_i(k \mid k-1)), \tag{47}$$

$$b_a(k) = P_{zi}(k \mid k-1) + b(k)b(k)^\tau, \tag{48}$$

$$P_z(k \mid k-1) = \sum_{i=1}^{l} b_a(k) f(b_i \mid z(k), x(k), Z^{k-1}). \tag{49}$$

The above equations are obtained by applying the smoothing property of the conditional expectation operator on the predicted measurements resulting from each member of the Gaussian mixture. Specifically:

$$\begin{aligned} \widehat{z}(k \mid k-1) &= E(z(k) \mid Z^{k-1}) \\ &= E(E(z(k) \mid b_i) \mid Z^{k-1}) \\ &= \sum_{i=1}^{l} \widehat{z}_i(k \mid k-1) f(b_i \mid Z^{k-1}), \end{aligned} \tag{50}$$

assuming that the weighting coefficient is also used to indicate the ith member of the mixture. Similarly, the error covariance is calculated as:

$$\begin{aligned} P_z(k \mid k-1) &= E(\tilde{z}(k \mid k-1)\tilde{z}^{\tau}(k \mid k-1) \mid Z^{k-1}) \\ &= \sum_{i=1}^{l} P_{zbi}(k \mid k-1) f(b_i \mid Z^{k-1}) \end{aligned} \tag{51}$$

and

$$\begin{aligned} &P_{zbi}(k \mid k-1) \\ &\quad = E(\tilde{z}(k \mid k-1)\tilde{z}^{\tau}(k \mid k-1) \mid Z^{k-1}, b_i) \\ &\quad = E((z(k) - \widehat{z}(k \mid k-1))(z(k) - \widehat{z}(k \mid k-1))^{\tau} \mid Z^{k-1}, b_i) \\ &\quad = E(z(k)z^{\tau}(k) \mid Z^{k-1}, b_i) - E(\widehat{z}(k \mid k-1)z(k) \mid Z^{k-1}, b_i) - \\ &\qquad - E(z(k)\widehat{z}^{\tau}(k \mid k-1) \mid Z^{k-1}, b_i) + \\ &\qquad + E(\widehat{z}(k \mid k-1)\widehat{z}^{\tau}(k \mid k-1) \mid Z^{k-1}, b_i) \\ &\quad = E(z(k)z^{\tau}(k) \mid Z^{k-1}, b_i) + \widehat{z}(k \mid k-1)\widehat{z}^{\tau}(k \mid k-1) - \\ &\qquad - \widehat{z}(k \mid k-1)\widehat{z}_i^{\tau}(k \mid k-1) - \widehat{z}_i^{\tau}(k \mid k-1)\widehat{z}(k \mid k-1) \\ &\quad = E((z(k) - \widehat{z}_i(k \mid k-1))(z(k) - \widehat{z}_i(k \mid k-1))^{\tau} \mid Z^{k-1}, b_i) + \\ &\qquad + (\widehat{z}_i(k \mid k-1) - \widehat{z}(k \mid k-1))(\widehat{z}_i(k \mid k-1) - \widehat{z}(k \mid k-1))^{\tau} \\ &\quad = P_{zi}(k \mid k-1) + (\widehat{z}_i(k \mid k-1) - \widehat{z}(k \mid k-1)) \times \\ &\qquad \times (\widehat{z}_i(k \mid k-1) - \widehat{z}(k \mid k-1))^{\tau} \end{aligned} \tag{52}$$

with the weighting coefficients defined above.

Acknowledgment

The authors wish to acknowledge all the comments and suggestions made by the reviewers which clearly improved the clarity in presentation.

References

1. Candy, J. V.: *Signal Processing: The Model-Based Approach*, McGraw-Hill, New York, 1986.
2. Pitas, I. and Venetsanopoulos, A. N.: *Nonlinear Digital Filters: Principles and Applications*, Kluwer Academic Publishers, Norwell, MA, 1990.
3. Huber, P. J.: Robust estimation of a location parameter, *Ann. Math. Statist.* **35** (1964), 73–101.
4. Middleton, D.: Man-made noise in urban environments and transportation systems, *IEEE Trans. on Communications* **COM-21** (1973), 1232–1241.
5. Middleton, D.: Canonical non-Gaussian noise models: Their implication for measurement and for prediction of receiver performance, *IEEE Trans. Electromagn. Compat.* **EMC-21** (1979), 209–219.
6. Achieser, N. I.: *Theory of Approximation*, translated by C. T. Hyman, Frederic Ungar, New York, 1956.
7. Namera, T. and Stubberud, A. L.: Gaussian sum approximation for nonlinear fixed-point prediction, *Internat. J. Control* **38**(5) (1983), 1047–1053.
8. Vastola, K. S.: Threshold detection in narrowband non-Gaussian noise, *IEEE Trans. on Communications* **COM-32** (1984), 134–139.
9. Garth, L. M. and Poor, H. V.: Narrowband interference suppression in impulsive channels, *IEEE Trans. on Aerospace and Electronic Systems* **AES-28** (1992), 15–35.
10. Sorenson, H. and Alspach, D. L.: Recursive Bayesian estimation using Gaussian sum, *Automatica* **7** (1971), 465–479.
11. Alspach, D. L. and Sorenson, H.: Nonlinear Bayesian estimation using Gaussian sum approximations, *IEEE Trans. on Automatic Control* **AC-17**(4) (1972), 439–448.
12. Masreliez, C. J.: Approximate non-Gaussian filtering with linear state and observation relations, *IEEE Trans. on Automatic Control* **AC-20** (1975), 107–110.
13. Masreliez, C. J. and Martin, R. D.: Robust Bayesian estimation for the linear model and robustifying the Kalman filter, *IEEE Trans. on Automatic Control* **AC-22** (1977), 361–371.
14. Wu, Weng-Rong and Yu, F. F.: New nonlinear algorithms for estimating and suppressing narrowband interference in DS spread spectrum systems, *IEEE Trans. on Communications* **44**(4) (1996), 508–515.
15. Fukunaga, K.: *Introduction to Statistical Pattern Recognition*, Second Edition, Academic Press, London, 1990.
16. Vijayan, R. and Poor, H. V.: Nonlinear techniques for interference suppression in spread-spectrum systems, *IEEE Trans. on Communications* **COM-38** (1990), 1060–1065.
17. Plataniotis, K. N.: Distributed parallel processing state estimation algorithms, PhD Dissertation, Florida Institute of Technology, Melbourne, Florida, 1994.