2nd International Conference on Experiments/Process/System Modelling/Simulation & Optimization
2nd IC-EpsMsO
Athens, 4-7 July, 2007
© IC-EpsMsO

NON-STATIONARY NOISE ESTIMATION IN ADAPTIVE LINEAR AND EXTENDED KALMAN FILTERING

N. Moriya

siOnet Ltd. – Applied Modeling Research, Herzelia, 46445 Israel,

e-mail: netzer@si-O-net.com, web page: http://www.si-O-net.com

Keywords: adaptive, Kalman Filtering, noise, Stochastic process, Gaussian, non-stationary, variance, covariance.

Abstract. *When Optimal Linear Kalman Filtering is employed, the complete knowledge of all system parameters, including the forcing input functions and the noise statistics, is required. In Adaptive schemes, frequently employed in control, communications, and other applications where the statistical characteristics of the signals to be filtered are either totally unknown a priori or, as assumed in other cases, slowly varying in time, the noise component of the measurement equations may, in principle be estimated during the calculation, along with the track of the dynamic changes in both the measurement noise variance matrix (R) and the error covariance matrix (P). However, when the changes in the covariance matrices are severe, inconsistent with assumed stationary conditions, or, more commonly, when the model itself is highly uncertain, such dynamical changes may be accompanied by strong suboptimal conditions of the filter and may result unexpected behavior of the algorithm and even divergence of the calculations.*
In the present work, we employ a new, general method to estimate the noise characteristics of an unknown, arbitrary, smoothed pure signal contaminated with an additive measurement noise component, to the traditional adaptive linear and extended Kalman Filtering algorithm. In the following, we assume that the noise component is white, zero mean, and Gaussian in nature, with however, time-dependent $\sigma_v^2(t)$ *and show that the noise properties can be extracted in real-time, externally to the filter, and that the non-stationary variance of the measurement noise component can be used in the calculations of the filter. This is demonstrated by computer simulation to improve the estimation of the state parameters of the filter and contribute to the overall stabilization of the calculations as compared to the traditional algorithms of both linear and extended Kalman filters.*

1 INTRODUCTION

Kalman filter is a recursive method, usually used to estimate the states of an unknown physical system, based on the complete knowledge of the mathematical model, the noise statistics of all noise components, the input functions involved and the dynamics of the system. For optimal operation of the filter, stationarity is required as well as some other characteristics of the noise parameters, such as independency between the noise components (measurement and process) and the mandatory requirement of “white” nature of all sources of the noise components.

In reality, as is well documented in the literature [1], these conditions are rarely met in most practical problems, and many versions of the Kalman approach has been suggested in order to solve non-linear problems [2], non-Gaussian noise statistics [3] non-stationary conditions [4] and other cases where some or most of the parameters to estimate are not at hand *a-priori*.

Usually, a linear system is models by the following relations [5]:

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

where the index *k* refers to the current time step, $x_k \in \Re^n$ is the state vector, φ_k is the transition matrix (*n x n*), u_k is the vector of the input forcing functions (*d x 1*), B_k is a matrix of (*n x d*), w_k is the process white noise vector of (*n x 1*) and Δ_k is an (*n x n*) matrix. The process noise vector is assumes to be of zero mean and Gaussian.

The measurement equation is written as:

$$z_{k+1} = H_{k+1}x_{k+1} + D_{k+1}u_{k+1} + \Gamma_k v_{k+1} \tag{2}$$

where the dimensions of $z_k \in \Re^s$ are (*s x 1*), D_k is a (*n x d*) matrix, v_k is the measurement noise vector of (*s x 1*), again, assumed to be normal, white, zero mean in independent of the process noise and Γ_k is a corresponding (*s x d*) matrix.

Also, the covariance matrices of the noise components are given by:

$$E\left\langle w_i w_j^T \right\rangle = \begin{cases} Q_k & \text{for} \quad i = j = k \\ 0 & \phantom{\text{for}} \quad i \neq j \end{cases} \tag{3}$$

$$E\left\langle v_i v_j^T \right\rangle = \begin{cases} R_k & \text{for} \quad i = j = k \\ 0 & \phantom{\text{for}} \quad i \neq j \end{cases} \tag{4}$$

$$E\left\{ w_k v_i^T \right\} = 0 \qquad \text{for} \quad i \neq k \tag{5}$$

with $E\{\bullet\}$ as the expectation operator, Q_k and R_k are the respective process and measurements noise covariance matrices.

The optimal Linear Kalman estimator equations are given by [6, 1]:

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}\left(z_{k+1} - H_{k+1}\hat{x}_{k+1|k}\right) = \hat{x}_{k+1|k} + K_{k+1}N_{k+1|k+1} \tag{6}$$

where the notation $\hat{x}_{k+1|k}$ refers to the estimation of the (state) vector for the $k+1$ time step with the measurements only up to time step k, and $\hat{x}_{k+1|k+1}$ denotes the corresponding estimation of the state vector for time step $k+1$ and after the arrival of the $k+1$ measurement. Also,

$$\hat{x}_{k+1|k} = \phi_k \hat{x}_{k|k} + B_k u_k \tag{7}$$

$$N_{k+1|k+1} = \left(z_{k+1} - H_{k+1}\hat{x}_{k+1|k}\right) \tag{8}$$

are calculated to give

$$\hat{x}_{k+1|k+1} = \phi_k \hat{x}_{k|k} + B_k u_k + K_{k+1}\left(z_{k+1} - H_{k+1}\left(\phi_k \hat{x}_{k|k} + B_k u_k\right)\right) \tag{9}$$

as the accepted estimations for the states.

The error covariance matrix is given by

$$\hat{P}_{k+1|k} = \phi_k \hat{P}_{k|k} \phi_k^T + \Delta_k Q_k \Delta_k^T \tag{10}$$

and

$$\hat{P}_{k+1|k+1} = \left(I - K_{k+1}H_{k+1}\right)\hat{P}_{k+1|k} \tag{11}$$

where the filter gain, K_{k+1}, is given by

$$K_{k+1} = \hat{P}_{k+1|k} H_{k+1}^{\ T} \left[H_{k+1} \hat{P}_{k+1|k} H_{k+1}^{\ T} + \Gamma_{k+1} R_{k+1} \Gamma_{k+1}^{\ T} \right]^{-1} \tag{12}$$

In order to perform the regular operation of the filter, the initial conditions of the filter should be defined (i.e. $\hat{x}_0$, the state of the system at $k = 0$ and the respective assumption as for the error covariance matrix P_0).

As the indirect parameters of the filter, including the actual properties of the noise statistics are never completely known and in some, more complicated situations, are also time-depended, adaptive algorithms were developed in order to estimate the covariance matrices of the filter as well as some of its other hidden properties.

Some methods for estimating the noise statistics are briefly reviewed in [7]. In this early paper, the Bayesian, maximum likelihood (ML), Correlation and the Covariance Matching are studied as alternatives for the adaptive algorithm. However the Bayesian estimation method is usually considered as impractically suitable to most applications due to its time consuming operation and its inherent assumption of time invariant conditions. The maximum likelihood (ML) criterion which is used to estimate the set of inputs that maximize the probability of obtaining the observed outputs given the initial set of inputs, is proved to diverge in many cases, and is mainly used for stationary processes. The Correlation and Covariance methods for estimating the noise statistics mentioned, where autocorrelations to known linear models of the outputs of the system are obtained, are less popular as such relations are not always easy to find with a satisfactory degree of confidence. In [8-10] several algorithms to estimate the Kalman gain value were proposed, based on stochastic approximation, as the new measurements are integrated with the past-step estimation of the gain. This approach, however, requires relatively long runtime before convergence, and close to stationary conditions are also desirable in order to yield reasonable results. Adaptive schemes to simultaneously estimate statistical properties of the noise components with the forcing functions were also suggested but are usually relay on other estimations of the filter states and may stimulate inconsistency conditions of the filter. In [11], a method to estimate the input forcing functions and the statistical moments of the noise parameters was tested. This algorithm which is based on LMS polynomial curve fitting to the measurements time sequence vectors, using a dynamical moving window, was shown to yield improved results and better robostic behavior of the filter to outliers. However, several problems with this method are still unsolved and may question the applicability of the method. The main of them are the unknown fitting parameters such as the polynomial order, the matching points of the fit during the moving of the dynamical window and the potential for ill posed conditions when the fitting order is high and the measurements series to fit contains fast changing components. All these conditions may cause unstable conditions for the filter to operate and thus may limit its use.

In the following, we suggest the application of a unique estimation algorithm for the measurement noise which is external to the filter, dynamical in time and extremely robostic to outliers. We use this estimation in order to obtain independent estimation for the measurement noise covariance matrix. The main advantage of the proposed method is that using independent calculation of the dynamic estimation of the measurement noise may allow non-stationary conditions (of the measurement noise statistics) to be considered with a total separation between other estimations of the filter parameters and the noise conditions. Assuming that the noise is normally distributed around the unknown deterministic part of the true measurement signal, the method proposed here for the external estimation of the measurement noise is based on high order stochastic differentiation of the measurement inputs and does not require any *a-priori* knowledge about the model of the measurement vector. It is shown that the estimation of the process covariance noise Q, which is related to R, can also be improved as well as the general behavior of the filter. Some other general properties of this approach, such as deviation from non-Gaussian conditions were not intensively investigated within the scope of this work, however, preliminary results show that the filter is less sensitive to such conditions and better recover from temporal appearance of such situations.

The remainder of the paper is organized as follows. In Section 2, we briefly summarize the conventional Adaptive Linear Kalman Filter Algorithm. Then, in Section 3, we describe in a more detailed fashion, the principles of the method for statistical noise estimation based on High Order Stochastic Differentiation (HOSD) and present the moving window procedure to incorporate the external measurement noise estimation into the Kalman algorithm. In section 4 we evaluate the performance of the filter based on numerical simulations for a simple linear example and a more complicated, non-linear case, solved by Adaptive Extended Kalman Filter (AEKF) as compared to the proposed HOSD method. In both cases, a non-stationary, Gaussian, measurement noise is simulated and is added to the deterministic exemplary signals. We conclude the paper in Section 5.

2 ADAPTIVE KALMAN FILTERING

Adaptive Kalman Filtering is a sub-optimal operation that uses the input sequential information to recursively estimate the state vector of a mathematically modeled system in order to simultaneously estimate the states and the unknown statistical parameters of the noise components of the system under consideration.

It is customize [12-13] to use the last N_s measurements of the system's errors $\hat{x}_{k+1|k} - \phi_k \hat{x}_{k|k}$ in order to average the contribution of the forcing functions $B_k u_k$ and use the result as a first order estimation for the unobserved input forcing constrain:

$$B_k \hat{u}_k = \frac{1}{N_s} \sum_{j=1}^{N_s} \gamma_{k+1-j} \tag{13}$$

where $\gamma_{k+1} = \hat{x}_{k+1|k} - \phi_k \hat{x}_{k|k}$. Note that $B_k \hat{u}_k$ is then used instead of $B_k u_k$.

Now, the state estimation can be written as

$$\hat{x}_{k+1|k} = \phi_k \hat{x}_{k|k} + B_k \hat{u}_k + K_{k+1}\left[z_{k+1} - H_{k+1}\left(\phi_k \hat{x}_{k|k} + B_k \hat{u}_k\right) - \hat{r}_{k+1}\right] \tag{14}$$

Here we have replaced $z_{k+1} = H_{k+1} x_{k+1} + D_{k+1} u_{k+1} + \Gamma_k v_{k+1}$ by $v_{k+1} = z_{k+1} - H_{k+1} x_{k+1}$, assuming that the measurement equation is independent on the forcing functions and that $\Gamma_k = I$.

In the above, $\hat{r}_{k+1}$ is the (unbiased [10]) estimation of the adaptive measurement based on the last N_s measurements and $x_{k+1} \to \hat{x}_{k+1|k}$ with $\hat{x}_{k+1|k} = \phi_k \hat{x}_{k|k} + B_k u_k$:

$$\hat{r}_{k+1} = \frac{1}{N_s} \sum_{j=1}^{N_s} z_{k-j+2} - H_{k-j+2} \hat{x}_{k-j+2|k} = \frac{1}{N_s} \sum_{j=1}^{N_s} z_{k-j+2} - H_{k-j+2}\left(\phi_{k-j+!} \hat{x}_{k-j+1|k} + B_{k-j+1} \hat{u}_{k-j+1}\right) \tag{15}$$

Substituting $v_{k+1} = z_{k+1} - H_{k+1} \hat{x}_{k+1|k} = z_{k+1} - H_{k+1}\left(\phi_k \hat{x}_{k|k} + B_k \hat{u}_k\right)$ and averaging all elements from $N_s + 1$ up to k.

For the case of linear H_k, it can be shown [14] that

$$\hat{Q}_{k+1} = \frac{1}{N_s - 1} \sum_{j=1}^{N_s} \left[\Xi_{k-j+2} \Xi^T_{k-j+2} - \left(\frac{N_s - 1}{N_s}\right)\left(\phi_{k-j+1} P_{k-j+1} \phi^T_{k-j+1} - P_{k-j+2}\right)\right] \tag{16}$$

and

$$\hat{R}_{k+1} = \frac{1}{N_s - 1} \sum_{j=1}^{N_s} \left[\Pi_{k-j+2} \Pi^T_{k-j+2} - \left(\frac{N_s - 1}{N_s}\right) H_{k-j+2}\left(\phi_{k-j+1} P_{k-j+1} \phi^T_{k-j+1} + \hat{Q}_{k-j+1}\right) H^T_{k-j+2}\right] \tag{17}$$

where $\Xi_{k-j+2} = \left(f'_{k-j+2} - B_{k+1} \hat{u}'_{k+1}\right)$ and $\Pi_{k-j+2} = z_{k-j+2} - H_{k-j+2}\left(\phi_{k-j} \hat{x}_{k-j+1|k} + B_{k-j+1} \hat{u}_{k-j+1}\right)$ are also unbiased (i.e. $E\left\langle \hat{Q}_k - Q_k \right\rangle = 0$ and $E\left\langle \hat{R}_k - R_k \right\rangle = 0$) [14].

The main problem of this algorithm, as also suggested in [11] is that all the estimated quantities are strongly affected by the inherent estimations of the states during the evolution progress of the calculations. If, due to unpredicted change of the noise conditions, one or more estimation may strongly deviate from the expected value, the whole behavior of the filter would possibly experience unstable conditions. This was demonstrated and discussed in [15] and references therein.

In order to minimize such effects, we therefore propose the use of a new method, originally developed to recover statistical properties of probability density functions (*pdf*) of high order differentiations of stochastic processes, for the purpose of independent estimation of the measurement noise component, to be obtained externally to the filter scheme, and in order to incorporate this estimation back into the filter so to improve its stability reaction to non-stationary characterizations of the noise conditions. Figure 1 illustrates the schematic operation of the proposed Kalman loop.

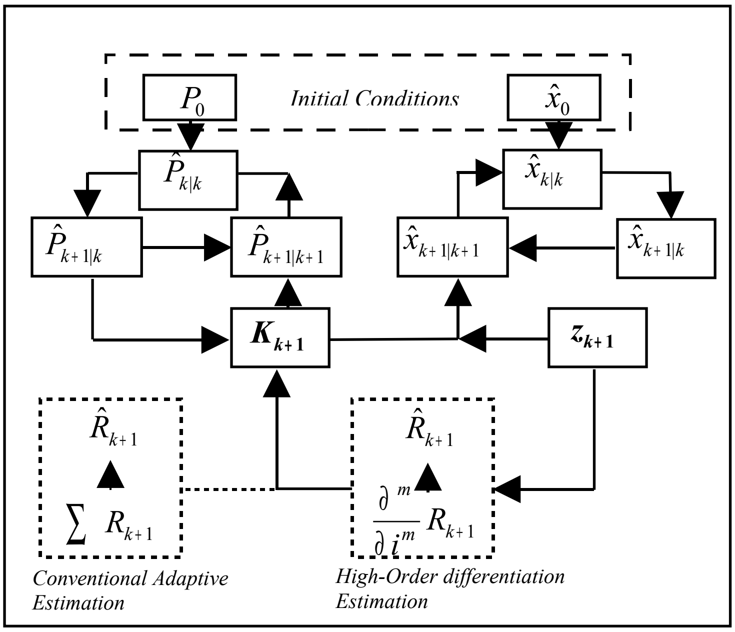

Figure 1. A schematic diagram of the Kalman loop with the HOSD estimation

3 THE METHOD FOR EXTERNAL ESTIMATION OF THE MEASUREMENT NOISE STATISTICS

3.1 The Stochastic High-Order Numerical Differentiation [16, 17]

In a recent publication, we have presented a general method to estimate the noise level of an arbitrary, smoothed pure signal, contaminated with an additive noise component. If the noise characteristics of the signal are known, for instance the type of the corresponding probability density function (*pdf*), the noise properties can be extracted. In such cases both the noise level, as may arbitrarily be defined, and a simulated noise component can be generated, such that the simulated noise component is statistically indistinguishable from the true noise component in the original signal. Further discussion is given for cases where the noise *pdf* is initially unknown.

In this section, we briefly present the main assumptions and skeleton of the method. The proposed algorithm is then applied, in the next section within the Kalman filtering process.

We start with considering a stochastic process $\xi\left(n_i\right)$ with n_i, the collection of stochastic events (e.g. experimental data measurements), in a bounded, measurable space (state-space) so that the variance of the stochastic variables considered are finite. A differentiating operator, operating on a signal state-vector, may then be defined with respect to the index of the signal data points in their sequenced order (or equivalently, treating the signal as a time series vector with a unit time step). By this, one may realize that a differentiation procedure, of the first order, is equivalent to subtracting the element n_i from the element n_{i+1} in the stochastic signal. Since in such random set of points each point is totally independent of all other points in the set and controlled only by the mutual statistics belongs to the sample space, denoted here by Ω (i.e. all points (i,j) are uncorrelated where $i \neq j$), the equivalence to subtracting the element n_i from the element n_{i+1} in the noise signal would be the subtraction of two independent Random Variables with identical statistical distribution (*IID*).

In contrast with the case of the first numerical differentiation, where one could assume that all individual data points were uncorrelated, higher order numerical differentiation involves correlated expressions that may lead, in the general case, to non-trivial expressions for the resultant probability functions.

Using the above definitions and referring to some arbitrary random variable function $z(n_i, \xi)$, considered here as the original data signal with ξ the stochastic random variable, one can now derive the second order numerical differentiation index series $z^{(2)}\left(n_i^{(2)}, \xi^{(2)}\right)$, with $\xi^{(2)}$ refers to the unknown stochastic random variable corresponding to the second-order numerical differentiation vector by realizing that $n_i^{(1)} = n_i - n_{i-1}$ and $n_{i+1}^{(1)} = n_{i+1} - n_i$ so that $n_i^{(2)} = n_{i+1}^{(1)} - n_i^{(1)} = n_{i+1} - 2n_i + n_{i-1}$ (we use the parenthesized superscript $n^{(m)}$ to denote differentiation of order m). These expressions imply that the probability density function of the second order numerical differentiation is the equivalent *pdf* of the sum of three independent, however non-identical, random variables (*InID*), all with similar, yet, not identical, probability density functions.

Generally, given two independent random variables ξ_1 and ξ_2 on the space R^k, with μ and ν corresponding to their respective distribution functions and f and g denote their respective density functions, the distribution of the sum $\xi_1 + \xi_2$ is the convolution $\mu * \nu$ operation between them and the analogue density function of the sum equals to the convolution integral denoted by $f * g$ in the index/time domain.

Using the notation $f_{\xi_1,\xi_2,\xi_3} = f_{\xi_1} * f_{\xi_2} * f_{\xi_3}$ this implies for the second order numerical differentiation may be written as:

$$f_{n_i^{(2)}} = \left(f_{n_i} * f_{(-2n_i)}\right) * f_{n_i} \tag{18}$$

Following the above arguments, for higher numerical differentiation, it can now easily be deduced that the m 'th order numerical differentiation of a random variable derived from an arbitrary statistically defined variable, can be obtained by noting that the correlation elements that dictate the numerical differentiation expressions are given by the non-zero elements (omitted) of following matrix denoted here as the Stochastic Derivative matrix S_k^m :

(19)

$$\begin{array}{ccccccccccccccccc}
 & & & & & & & & 1 & & & & & & & & \\
 & & & & & & & 1 & & -1 & & & & & & & \\
 & & & & & & 1 & & -2 & & 1 & & & & & & \\
 & & & & & 1 & & -3 & & 3 & & -1 & & & & & \\
 & & & & 1 & & -4 & & 6 & & -4 & & 1 & & & & \\
 & & & 1 & & -5 & & 10 & & -10 & & 5 & & -1 & & & \\
 & & 1 & & -6 & & 15 & & -20 & & 15 & & -6 & & 1 & & \\
 & 1 & & -7 & & 21 & & -35 & & 35 & & -21 & & 7 & & -1 & \\
1 & & -8 & & 28 & & -56 & & 70 & & -56 & & 28 & & -8 & & 1
\end{array}$$

. . . .

where each of the elements are given by:

$$S_j^{(m)} = (-1)^{j+1} \binom{m}{j} \tag{20}$$

and $\binom{m}{j}$ represents the elements of the binomial coefficients [17].

Note that the above expression was derived in order to account for the weights of the individual contributions in the convolution integrals (i.e. the elements of the Stochastic Derivative Matrix) and represents the factors that are needed in order to calculate the probability density function of the differentiated set, and not the differentiated noise values themselves (that increase rapidly in value as will be shown in the following).

In terms of a summation of the individual elements needed to account for the probability density function of the *m*'th order numerical differentiation, one may write the summation as:

$$f_m = \sum_{j=1}^{m} S_j^{(m)} f(z) \tag{21}$$

with $f(z)$ representing the probability density function of the original random variable.

It can be shown [16-17] that for the general case:

$$F_Z(z) = \iint_{D_z} \prod_{j=1}^{m} f_{\xi_j} d\xi_1 d\xi_2 \ldots d\xi_N = \iint_{D_z} \left[\prod_{j=1}^{m} S_j^m f_j(\xi) \right] d\xi^{(m)} . \tag{22}$$

where $F_Z(z)$ is the probability distribution function of the new random variable z over its volume of existence D_z. Since the density function used here is the same for all individual elements of the multiplication term under the integral, only weighted properly by the suitable elements of the stochastic derivative matrix, this can symbolically be written as:

$$F_{n_i}^{(m)} = \iint_{D_z} \left\{ \prod_{j=1}^{m} S_j^m f(\xi) \right\} d\xi^{(m)} \tag{23}$$

where $F_{n_i}^{(m)}$ represents the probability distribution function of $\frac{\partial^m z(n_i, \xi)}{\partial i^m}$ that can easily be evaluated to derive the respective density function, recalling that the term $\prod_{j=1}^{m} S_j^m f(\xi)$ really represents a convolution of the original probability function weighted accordingly, based on the Stochastic-Derivative matrix.

We focus now our discussion to the case where the probability density function of the noise statistics in the original signal is Gaussian. For the Gaussian case, the analysis yields relatively straightforward expressions as the Gaussian *pdf* belongs to the few probability functions that convolve into similar functions. We therefore consider a Gaussian distribution, where ξ is referred to as the random variable, $N(0, \sigma_0^2)$, i.e. a Gaussian distribution with the first moment equals to zero and the variance is given by σ_0^2 as an illustrative probability density function (the numerical differentiation of the following with mean values other than zero is straightforward). For the above, the corresponding expression is of the form [16]:

$$\frac{d^m N(0, \sigma_0^2)}{di^m} = \alpha(m) N(0, \beta(m) \sigma_0^2) \tag{24}$$

with $\beta(m)$ given by the sum of the squares of the elements of the $m+1$'s row in the Stochastic-Derivative matrix and with $\alpha(m)$ given by the inverse of square-root of the sum of the squares of the elements of the *m+1*'s row of the Stochastic-Derivative matrix. Note that for a Gaussian distribution function, as used above, the condition $\alpha \propto 1/\sqrt{\beta}$ is required by the normalization condition.

Using the equations and the arguments above, one can now derive the probability density function of a zero mean Normal distribution for any differentiation of order m of a discrete random signal. For instance, the analytical expressions for the cases of the first, second and fifth numerical differentiations are simply:

$$\frac{dN(a, \sigma_0^2)}{di} = \frac{1}{\sqrt{1^2 + 1^2}} N\left(0, \left(\sqrt{2}\sigma_0\right)^2\right) = \frac{1}{\sqrt{2}} N\left(0, \left(\sqrt{2}\sigma_0\right)^2\right) \tag{25}$$

$$\frac{d^2N\left(0,\sigma_0^2\right)}{di^5} = \frac{1}{\sqrt{1^2+2^2+1^2}}N\left(0,\left(\sqrt{1^2+2^2+1^2}\sigma_0\right)^2\right) = \frac{1}{\sqrt{6}}N\left(0,\left(\sqrt{6}\sigma_0\right)^2\right)$$

$$\frac{d^5N\left(0,\sigma_0^2\right)}{di^5} = \frac{1}{\sqrt{1^2+5^2+10^2+10^2+5^2+1^2}}N\left(0,\left(\sqrt{252}\sigma_0\right)^2\right) = \frac{1}{\sqrt{252}}N\left(0,\left(\sqrt{252}\sigma_0\right)^2\right)$$

To demonstrate one of the proposed motivations for the use of high-order numerical differentiation of a stochastic signal, we may refer to the numerical differentiation of the noise-level of a noisy signal, where noise, either due to experimental set-up or to the process itself (or from both sources), is added to the signal and is screening the, usually required, pure data signal.

For simplicity we assume that the arbitrary noisy signal $P = S + N$, with N being the additive noise that is added to the pure signal *S*, can be represented by a smooth and continuous signal contaminated by noise. Let us further assume that on the interval of validity of *S*, one can approximate the stochastic signal *S* (for instance, in the Least Mean Square sense) by an m-degree polynomial function that may belong to a complete orthogonal polynomial basis. Note that no restrictions were set on the order of the numerical differentiation and that m can, in principle, be arbitrarily chosen. This can be proved to be possible for any bounded, smoothed and continuous function S, but may be of practical use only when the interval is short, as compared to the structure of the signal, and for a relatively low polynomial degree. Assuming the above, it turns that one can use the approximation: $\frac{d^{m+1}P}{dk^{m+1}} = \frac{d^{m+1}N}{dk^{m+1}}$, as the *m*'th numerical differentiation of *S*, under the above assumptions, is constant and thus vanishes for higher orders. It is thus clear that the purpose of the differentiation is to terminate the effect of the "pure" signal (which is, of course an unknown contribution in the stochastic system). However, the differentiation does leave the differentiated noise component untouched (as we have assumed additive noise) and when assuming that a prior knowledge about the type of the noise is at hand (e.g. Gaussian, Poisson or other type), the above analysis may suggest a general method to obtain the statistical parameters of the original noise components (i.e. before the differentiation). When applying this method in relatively limited time intervals, where the full signal is sliced into a moving "window", such a window should however be short enough so that a low degree polynomial would be sufficient to represent a reasonable fit to the pure component in the signal (or, more precisely, the deviation will lay safely within the statistical error covered by the noise component).

Now, if the characteristics of the statistical properties of the high-order numerical differentiation of the original noise $\frac{\partial^{m+1}}{\partial i^{m+1}}\left(N\right)$ are known, i.e. the probability density function that statistically describes the initial noise subject to high-order numerical differentiation, in terms of the parameters (assumed to be unknown) of the statistical nature of the noise (assumed to be known), one can obtain the specific parameters of the original noise and thus to deduce the noise-level in the original signal *P*. It should also be mentioned that the effect of outliers is drastically suspended as such events are usually propagates to the external regions of the axis in the differentiated domain and have only minor contribution to the *pdf* under study.

When any prior knowledge, with respect to the original probability function of the noise is not at hand, one may, based on the above approach, conduct a simple test, to deduce the specific probability density function of the original stochastic process in the observed data signal by reproducing only the noise component from the noisy signal and further process it so that no unintentional influences, related to the pure signal rather than to the noise, are affecting the probability testing procedure.

It should also be noted that the theory and method described here are totally insensitive to the sampling rate in the original discrete signal. This is due to the normalization of the time axis (in a time series signal) into an indexed axis where each of the sampled points are given integer indexes (i.e. 1,2,…,k). The assumption that the separation between two adjacent points is one unit (i.e. that the sampled points are equally separated in the index axis), as mentioned above, allows the further numerical differentiation of the differentiation process with no need to consider any specific case (with respect to sampling rates) for the sake of the generality of the discussion.

3.2 The Kalman Filtering External Measurement Covariance Estimation

As an alternative to applying the conventional method to estimate the measurement noise covariance matrix (in equation 17 above) we propose an m'th order differentiation process (within the moving window) of the measurement vectors, a separate operation for each of the s measurement inputs. This will lead to (in principle) s probability density functions (each for every one of the s measurement input), ${}^{s}\left\{F_{n_n}^{(m)}\right\}_{k-N_s:k}$ of the

sequence $\left\{ {}^{s}z(n_i,\xi) \right\}_{k-N_s:k}$ after the differentiation $\dfrac{\partial^m \left\{ {}^{s}z(n_i,\xi) \right\}_{k-N_s:k}}{\partial i^m}$ where the operator $\left\{ \varepsilon \right\}_{k-N_s:k}$ represents the vector ε in a window from element $k-N_s$ until k.
This can be expressed as:

$$ {}^{s}\left\{ F_{n_i}^{(m)} \right\}_{k-N_s:k} = \iint_{D_z} \left\{ \prod_{j=1}^{m} S_j^m f(\xi) \right\} d\xi^{(m)} \tag{26} $$

Assuming that the initial probability density functions of the corresponding measurement input are all Gaussian and given by ${}^{s}f_{init}(z) = N\left(0, {}^{s}\sigma_0^2\right)$, and the differentiated probability density function is ${}^{s}f_m(z) = N\left(0, {}^{s}\sigma_m^2\right)$ the variance ${}^{s}\sigma_0^2$ to be recognized as the elements of the measurement covariance matrix may be written as ${}^{s}\sigma_0^2 = {}^{s}\sigma_m^2 \left[\sum_{j=1}^{m} \Phi_j^m \right]^{-1}$ and:

$$ \hat{R}_{k+1} = {}^{s}\sigma_0 = {}^{s}\sigma_m \left[\sum_{j=1}^{m} \Phi_j^m \right]^{-1/2} \tag{27} $$

were Φ is the stochastic derivative matrix.

A special consideration is needed for the choice of the differentiation order in the above derivation. In principle, it can be shown [16] that given a sufficient statistics, the covariance estimation, as derived above for the normal case, is almost insensitive to the order of differentiation ($m > 5$). For the examples in the following, we have set the order to be defined by $m = round(N_s/10)$ with the arbitrary requirement $50 < N_s < (N_{Total}/10)$, were N_s and N_{Total} are the number of data elements in the moving window and the total number of measurements respectively. It should also be noted that for non-Gaussian noise models the above derivation is still valid, only with the proper formulation of the expressions used.
Although the systematic analysis of the stability evaluation of the method proposed here is practically impossible in an analytic manner, due to its complexity, the main ideas of the algorithm is easy to follow and are summarized in the next few steps:

a. define a window for each of the s measurement sequences (not necessarily the same window),
b. define the differentiation order m for each of the s measurement sequences (not necessarily the same order),
c. estimate the *pdf* of the resulting distribution after the differentiation under the assumption that the original *pdf* is Gaussian. In particular estimate ${}^{s}\sigma_m^2$,
d. used the stochastic derivative matrix and the relations given above to extract the original variance ${}^{s}\sigma_0^2$ from ${}^{s}\sigma_m^2$,
e. using all ${}^{s}\sigma_0^2$ (for all s sequences) to construct $\hat{R}_{k+1}$.

4 SIMULATION RESULTS OF THE KALMAN FILTERING WITH EXTERNAL MEASUREMENT COVARIANCE ESTIMATION

We begin the presentation of the results with a very simple example of the estimation of a constant value. Although trivial, the injection of a non-stationary measurement noise to the system and the response of the filter with the conventional adaptive scheme to the change of the noise conditions, as compared to the operation of the proposed algorithm, can emphasize some of the advantages in the later without the involvement of further complications that may blur the effect. In a second example, we demonstrate the operation of the proposed method for a much complicated, non-linear case.

4.1 Estimation of a Constant

We use the model and measurement equations:

$$x_{k+1} = x_k + w_k \tag{28}$$
$$z_{k+1} = x_{k+1} + v_{k+1} \tag{29}$$

with $H_k = 1$ for all k.

The update state equation is

$$\hat{x}_{k+1|k} = \hat{x}_{k|k} \tag{30}$$

and the update error process covariance equation

$$\hat{P}_{k+1|k} = \phi_k \hat{P}_{k|k} \phi_k^T + \Delta_k Q_k \Delta_k^T \tag{31}$$

For this case, $Q_k = E\left[w^2\right] = \sigma_s^2 = q$ and equation (31) thus reduces to

$$\hat{P}_{k+1|k} = \hat{P}_{k|k} + q \tag{32}$$

The filter gain can be written as

$$K_{k+1} = \hat{P}_{k+1|k}\left[\hat{P}_{k+1|k} + R_{k+1}\right]^{-1} \tag{33}$$

with $R_k = E\left[v^2\right] = \sigma_m^2 = r$. The state estimator is

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1} N_{k+1|k+1} \tag{34}$$

and the covariance matrix corresponding to the state estimation error

$$\hat{P}_{k+1|k+1} = \left(I - K_{k+1}\right)\hat{P}_{k+1|k} \tag{35}$$

In the following simulation we have used a constant value of $100V$ (with an arbitrary unit system V). We have also run the simulation over 300 simulated data points using $q = 0.1$ and $\begin{cases} r = 20 & 100 < N < 201 \\ r = 2 & elsewhere \end{cases}$. The window length was set to $N_s = 50$ and the differentiation order used was $m = 5$. The non-stationary measurement noise sequence used is shown in figure 2. As initial values, we have used $\hat{P}_0 = q/100$ and $\hat{x}_0 = 0$.

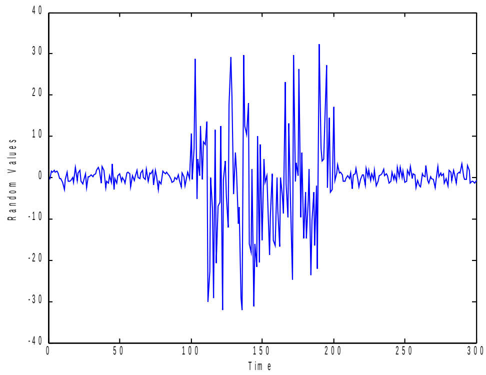

Figure 2. The non-stationary measurement noise used in the linear Kalman example

The noise covariance dynamics is shown in figure 3. It is clear that the conventional adaptive scheme (a) has worse response for the sudden change in the measurement noise conditions, as compared to the proposed algorithm, where the external estimation of the noise has been benefited from the High-Order Stochastic Differentiation calculation.

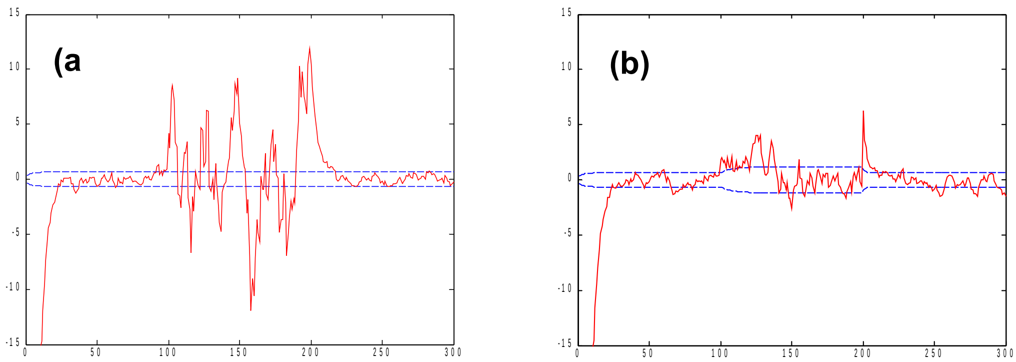

Figure 3. The noise covariance matrix in the case of (a) the conventional adaptive scheme and (b) with the High-Order Stochastic Differentiation algorithm

4.2 Estimation of a Sine Function with Unknown Amplitude, Phase and Angular Frequency

In this simulation we apply the Adaptive Extended Kalman Filtering with a non-linear system model. We simulate a 2D curve sine wave with several regions of parametric changes including the angular frequency, the amplitude, the phase and the bias. This system is solves by a non-linear model. In order to examine the proposed algorithm for external measurement noise estimation, we have added to the signal a time-dependent noise where the region of changes are abrupt and the level of change is sever (i.e. 10 times the background noise). Two simulations were performed similarly to the first example above. In the first simulation, the traditional Adaptive filter was applied. In the second simulation, the HOSD noise estimation was tested. Both simulations were performed with the same input data (measurements) in order to minimize any sources for differences other than the algorithms themselves.

The model equation is

$$y_1 = A\sin(\omega t + \varphi) + B + n(t) \tag{36}$$

and we assume that the angular frequency is a linear function in time i.e.:

$$\ddot{y}_2 = \frac{d^2}{dt^2}\left(\omega^2\right) = 0 \tag{37}$$

The state vector is defined to be:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} y_1 \\ \dot{y}_1 \\ y_2 \\ \dot{y}_2 \end{pmatrix} = \begin{pmatrix} A\sin(\omega t + \varphi) + B \\ \omega A\cos(\omega t + \varphi) \\ \omega^2 \\ B \end{pmatrix} \tag{38}$$

The corresponding state-vector derivative is:

$$\dot{x} = f(x,t) \Rightarrow \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{pmatrix} = \begin{pmatrix} x_2 \\ -A\omega^2\sin(\omega t + \varphi) \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} x_2 \\ -\omega^2(x_1 - B) \\ 0 \\ 0 \end{pmatrix} \tag{39}$$
$$= \begin{pmatrix} x_2 \\ -x_3(x_1 - x_4) \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} f_1 = x_2 \\ f_2 = -x_3(x_1 - x_4) \\ f_3 = 0 \\ f_4 = 0 \end{pmatrix}$$

The linearization process on the functions f_i 's yields the matrix $F = \frac{\partial}{\partial x_j} f_{i,j}$:

$$F = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -x_3 & 0 & (-x_1 + x_4) & x_3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{40}$$

We now use a second order Taylor expansion to get:

$$\phi(t) = I + Ft + \frac{1}{2}F^2t^2 = \begin{pmatrix} 1 - \frac{1}{2}t^2x_3 & t & \frac{1}{2}t^2(x_4 - x_1) & \frac{1}{2}t^2x_3 \\ -x_3t & 1 - \frac{1}{2}t^2x_3 & t(x_4 - x_1) & tx_3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{41}$$

or, in discrete form:

$$\phi_k(t = T_s) = \begin{pmatrix} 1 - \frac{1}{2}T_s^{\,2}x_3 & t & \frac{1}{2}T_s^{\,2}(x_4 - x_1) & \frac{1}{2}T_s^{\,2}x_3 \\ -x_3T_s & 1 - \frac{1}{2}T_s^{\,2}x_3 & T_s(x_4 - x_1) & T_sx_3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{42}$$

where T_s defines the time interval between two successive measurements, and

$$Q_k = \int_0^{T_s} \phi(t) Q \phi^T(t) dt \tag{43}$$

The measurement equation is assumed to be the actual position of the 2D curve points:

$$z_1(t) = A\sin(\omega t + \varphi) + B \tag{44}$$

This implies that $h_1(x) = x_1$. Linearization to obtain H will give trivially

$$H = \frac{\partial}{\partial x_j} h_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \tag{45}$$

We now define initial conditions for the simulation to have

$$\hat{x}_{k|k}(0) = (x\_corr(0)) \quad \text{and} \quad \hat{P}_{k|k}(0) = (P\_corr(0)) \tag{46}$$

The Kalman algorithm will start with the calculation of F following by the calculation of the matrix ϕ, each and every time, with the most updated state variables. The process of updating the covariance matrix will include time integration of all state variables (i.e. the state equations and the matrix F), and with the definition of the estimation for the covariance matrix corresponding to the state estimation error

$$\hat{P}_{k+1|k} = \phi_k \hat{P}_{k|k} \phi_k^T + \Delta_k Q_k \Delta_k^T \tag{47}$$

which will be simulated as $P\_pred = Phi * P\_corr * Phi' + Q\_discrete$

The filter gain is defined as

$$\begin{aligned} K_{k+1} &= \hat{P}_{k+1|k} H_{k+1}^T \left[H_{k+1} \hat{P}_{k+1|k} H_{k+1}^T + \Gamma_{k+1} R_{k+1} \Gamma_{k+1}^T \right]^{-1} = \\ &= \hat{P}_{k+1|k} H_{k+1}^T \left[H_{k+1} \hat{P}_{k+1|k} H_{k+1}^T + R_{k+1} \right]^{-1} \end{aligned} \tag{48}$$

and is simulated as

$$K = (P\_pred * H') * (H * P\_pred * H' + R_{HOSD})^{-1} \tag{49}$$

with $\hat{R}_{HOSD} = {}^s\sigma_m \left[\sum_{j=1}^{m} \Phi_j^m \right]^{-1/2}$ given according to equation (27) above.

The innovation is now written as $N = z - H * x\_pred$ and the correction to the state vector:

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1} N_{k+1|k+1} \tag{50}$$

which is simulated as $x\_corr = x\_pred + K * N$.

The corresponding correction to the covariance matrix:

$$\hat{P}_{k+1|k+1} = (I - K_{k+1} H_{k+1}) \hat{P}_{k+1|k} \tag{51}$$

simulated as $P\_corr = (I - K*H)*P\_pred$, is now calculated and the loop is closed where the updated values are used to calculate the estimations for the next step.

The simulation was designed with 7000 data points and assumes

$$
\begin{aligned}
y_k &= A_1 \sin(2\pi f_1 t_k + \varphi_1) + B_1 && \text{for} \quad 0 \le k \le K_1 \\
y_k &= A_2 \sin(2\pi f_2 t_k + \varphi_2) + B_2 && \text{for} \quad K_1 + T_s \le k \le K_2 \\
y_k &= A_3 \sin(2\pi f_3 t_k + \varphi_3) + B_3 && \text{for} \quad K_2 + T_s \le k \le K_3
\end{aligned}
$$

and

$$
\begin{aligned}
&f_1 = 200 \quad f_2 = 100 \quad f_3 = 150 \quad [\text{Hz}] \\
&A_1 = 2 \quad A_2 = 3 \quad A_3 = 1 \\
&B_1 = 1 \quad B_2 = 3 \quad B_3 = 2 \\
&\varphi_1 = 0.5 \quad \varphi_2 = 1.5 \quad \varphi_3 = 2.5
\end{aligned}
$$

The variance of the background measurement noise, as simulates here, was set to $r = 0.5$ and the conventional adaptive procedure was applied and compared to the HOSD scheme.

The window length in this simulation was set to $N_s = 100$ and the differentiation order used was $m = 10$. The non-stationary measurement noise is shown in figure 4a according to:

$$
\begin{cases} r = 5 & \text{with} \quad 3000 < N < 4001 \\ r = 0.5 & \quad\quad\quad elsewhere \end{cases}
$$

where the simulation signal is also shown (b) with the three regions of different sine parameters according to the simulated signal as set above.

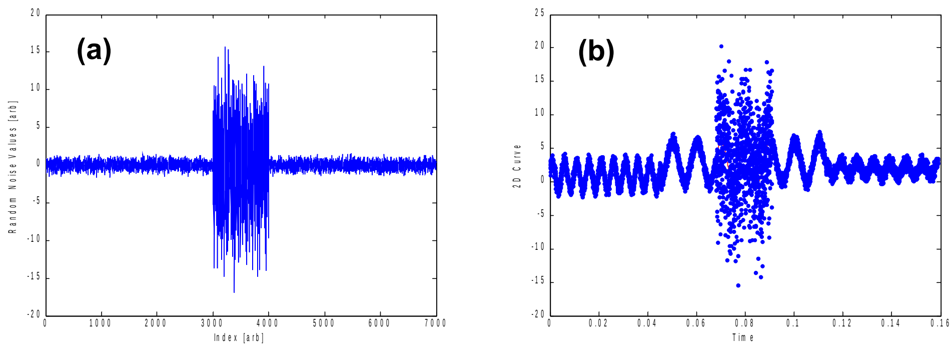

Figure 4. The non-stationary measurement noise used in the EKF sine-wave example (a). Sown in (b) is the noisy signal used

The covariance matrix was assumed to be

$$
Q = \begin{pmatrix} 100 & 0 & 0 & 0 \\ 0 & 300 & 0 & 0 \\ 0 & 0 & (5e6)^2 & 0 \\ 0 & 0 & 0 & 10 \end{pmatrix}
$$

The results, as obtained for the first state variable (the simulated signal) are shown in figure 5. The filtered signal after the HOSD scheme (b) shows some slight improvement over the conventional adaptive algorithm (a) with both filters respond nicely to the non-linearity of the model and to the changes of in the noise conditions.

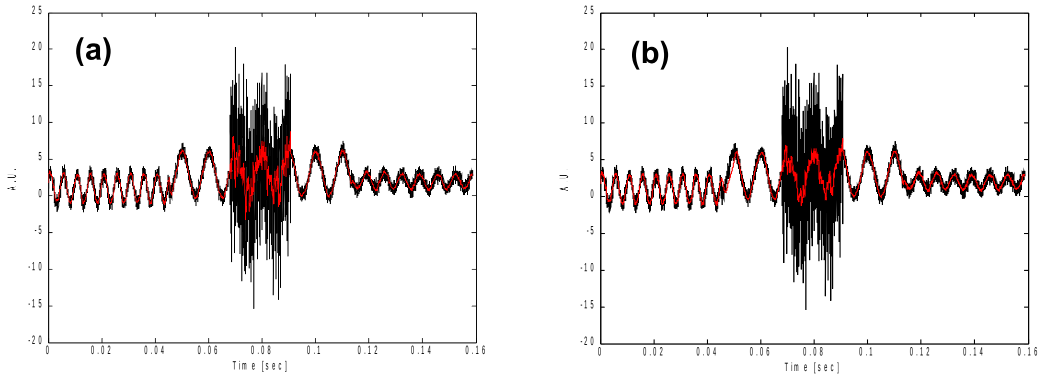

Figure 5. The estimation of the first state using the conventional filter (a) and the proposed HOSD algorithm (b)

It is more interesting however, to note that the angular frequency state variable did show a significant improvement with the HOSD method as compared to the conventional adaptive filtering. This is shown in figure 6 where the angular frequency after the adaptive filtering (a) is shown with the HOD result (b) along with the bias estimation using the conventional method (c) as compared to the proposed algorithm (d).

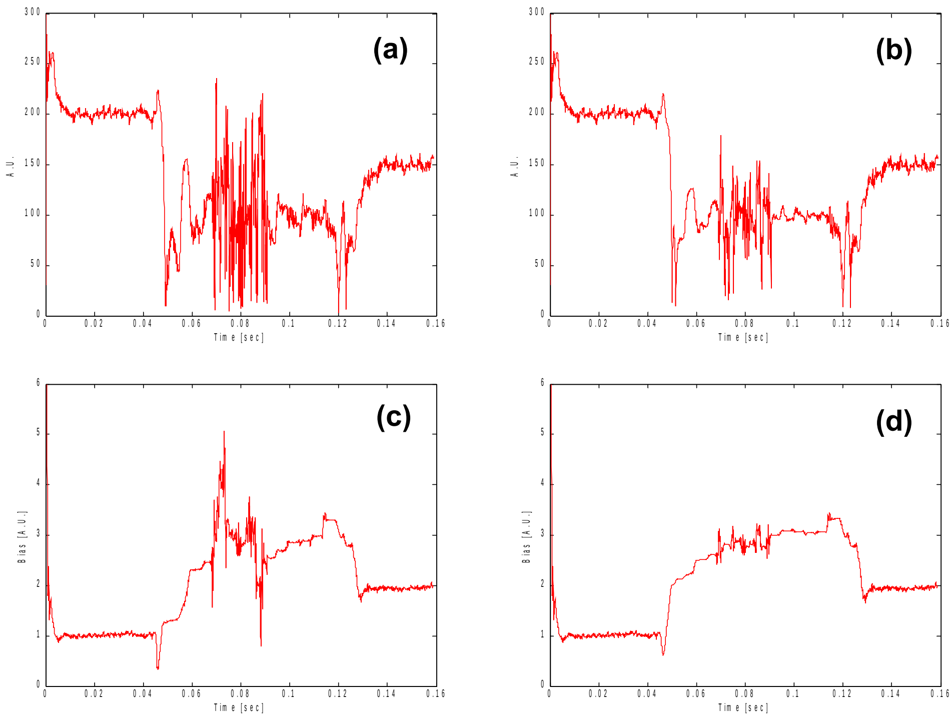

Figure 6. The Angular Frequency state (a) and the Bias (c) after the conventional filtering and the corresponding Angular Frequency (b) and the Bias (d) using the HOSD scheme

In figure 7 we also show the corresponding results of $Q(1,1)$ for the case of the conventional adaptive scheme (a) and the high-order differentiation method, as proposed here (b).

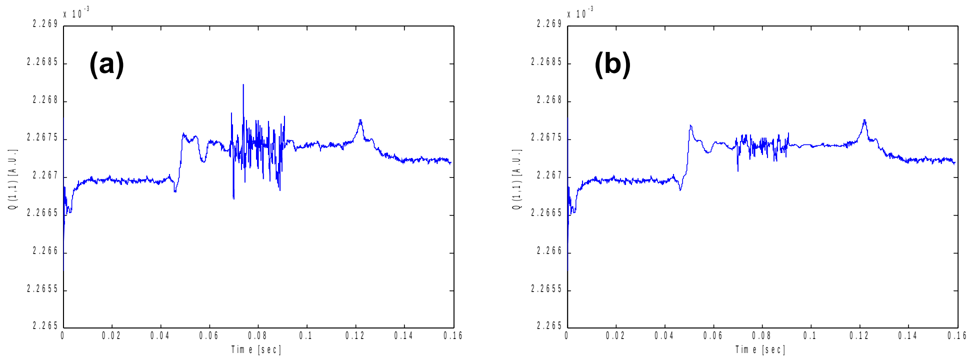

Figure 7. The covariance $Q(1,1)$ in the conventional run (a) as compared to the HOSD scheme

In general, although both methods were able to recover from the relatively sever measurement noise changes during the filter operation, it is clear that the results are better for the proposed algorithm. The effect is more pronounced for the hidden states and in particular for the angular frequency dynamics as reflected from the results in figure 6. This is also true for the error covariance matrix which is shown to have a better stability performance with the external estimation of the measurement noise as compared to the HOSD estimation. We have performed a set of over 100 simulations of this same example, with different noise conditions, to verify that the overall improvement is indeed achieved.

Also, although not specifically shown here, the HOSD algorithm can be shown to be extremely robust to outliers [16]. This is due, of course to the fact that the differentiation process suspends the weight of the outliers as their effect is shifted to the tail of the differentiated statistics of the respective *pdf*. This was indeed verified in [17].

5 CONCLUSIONS

The method presented here to estimate the measurement noise within the algorithm of the Kalman scheme is shown to improve the overall performance of the filter. We have simulated the results in two examples: a simple, linear, constant estimation test and a more complicated, non-linear problem of sine-function tracking with all parameters functions of time. In both simulations, a non-stationary, measurement noise was arbitrarily added to the deterministic signal and the Kalman method was applied in two modes: using the conventional adaptive version, and, using the proposed algorithm which uses an external estimation of the measurements noise by High-Order Stochastic Differentiation of the input signals to extract the noise statistics. This procedure assumed Gaussian probability density function for the noisy measurement signal, however with unknown time-varying variance.

The results clearly show that the later method can yield better results and improve the covariance matrix and the state estimations, as well as the overall performance of the filter.

The proposed method is based on the incorporation of the measurement noise estimations externally into the filter. This avoids the potentially closed positive feedback influence between the states estimation and the adaptation of the filter to the changes of the noise conditions with the appearance of possible instabilities that may arise by estimating the state variables based on adaptively estimated noise covariance quantities and than use these values to estimate the states. The algorithm is well suited to real-time applications and to situations where the noise statistics is a strong time-dependent.

REFERENCES

[1] Maybeck, P. S., (1979), *Stochastic Models, Estimation and Control*, Vol. 1. and Vol. 2. Academic Press,

[2] Anderson B.D. and Moore J.B. (1979), *Optimal Filtering*. Prentice-Hall, Englewood Cliffs. New-Jersey,

[3] Chen Y. and Singpurwalla N.D. (1994), *A Non-Gaussian Kalman Filter Model for Tracking Software Reliability*, Statistica Sinica Vol. 4, pp. 535-548,

[4] Krishnan Venkatarama, Non Linear Filtering and Smoothing: An Introduction to Martingales, Stochastic Integrals and Estimation, John Wiley & Sons, 1984,

[5] see for instance, Gelb A. Ed. (1996), *Applied Optimal Estimation*, MIT Press,

[6] Kalman R.E. (1960), *A New Approach to Linear Filtering and Prediction Problems,* ASME Journal of Basic Engineering, Series D, Vol. 82, pp. 34 ; Kalman R.E. and Bucy (1960), *New Results in Linear Prediction and Filtering Theory*, ASME Journal of Basic Enginiring, Series D, Vol. 83, pp. 95,

[7] Mehra R.K. (1972), *Approaches to adaptive filtering*, IEEE Trans. Autom. Control, Vol. 17(5), pp. 693–698,

[8] Sinha N.K. and Tom A. (1977), *Adaptive State Estimation for systems with unknown noise covariances,* Ibt. J. Syst. Sci., Vol. 8(4), pp. 377,

[9] Carew B. and Bellanger P.R. (1973), *Identification of Optimum Filter steady-State Gain for Systems with unknown noise covariances*, IEEE Trans. Automat. Control, Vol. AC-18,

[10] Myers K.A. and Tapley B.D. (1976), *Adaptive Sequential Estimation with Unknown noise Statistics*, IEEE Trans. Automat. Control, Vol. 21, pp. 520,

[11] Moghaddamjoo A. and Kirlin R.L. (1989), *Robust* A*daptive Kalman Filtering with Unknown Inputs,* IEEE Trans. on Acoust., Speech and Sig. Proc., Vol. 37, pp. 1166,

[12] Jia P.Z. and Zhu Z.T. (1984), *Optimal estimate and its application*, Science publishing company, China,

[13] Gustafsson F. (2000), *Adaptive filtering and change detection*, John Wiley & Sons,

[14] Yan Liang, De Xi An, Dong Hua Zhou and Quan Pan (2004), *A finite-horizon adaptive Kalman filter for linear systems with unknown disturbances*, Signal Processing Vol. 84(11), pp. 2175,

[15] Chen G. (1993), *Approximate Kalman Filtering*, University of Houston, World Scientific.

[16] Moriya N. (2006), *On high-order discrete derivatives of stochastic variables*, Applied Mathematical Modelling, Vol. 30(9), pp. 816,

[17] Moriya N. (2005), *Noise Level Estimation of Stochastic Processes*, Presented at the SIAM conference on Computational Science & Engineering, February 12-15 Orlando, FL., U.S.A.,