# ADAPTIVE KALMAN FILTER FOR NOISE IDENTIFICATION

**M. Oussalah and J. De Schutter**

K. U. Leuven, PMA, Celestijnenlaan 300B, 3001 Heverlee, Belgium
E-mail: oussalah.mourad@mech.kuleuven.ac.be / joris.deschutter@mech.kuleuven.ac.be

**Abstract** This paper deals with an on line identification of the values of these entities by a suitable adjusting of the filter gain and incorporating information about the quality of innovation sequence. The result yields a new adaptive Kalman filter. The proposal is mainly based on statistical properties of the innovation sequence of the filter and improves some previous results pointed out by Bellanger and Mehra. Particularly, the autocorrelation function of the innovation sequence is used. Furthermore, information about the quality of the autocorrelation parameter is incorporated through a weighted least squares methodology. The determination of the weights is based on a distance criterion, which involves the ideal probability distribution and the current probability referring to the first and second order statistics of autocorrelation functions. The estimation of the a priori noise statistics $Q$ and $R$ is obtained straightforwardly from the preceding and the optimal gain and the innovation covariance of the filter.

**Keywords**: Kalman filter, Adaptive Kalman filter, weighted least squares, noise statistics.

## 1. Introduction

Analysis of divergence of optimal estimators like Kalman filter (KF) [1, 5] was appeared since the first papers given rise to the theoretical foundations of KF [4]. Indeed, it is recognized from navigation experiments, among others, that the performance of KF under actual operating conditions can be seriously degraded from the theoretical performance indicated by the state covariance matrix inducing a divergence of KF. Typically, divergence is said to occur when the covariance matrix produced by KF no longer provides a reasonable characterization of the error in estimate. Effects that may cause divergence are various, for example: i) ill known parameters of the state or measurement models, i.e., the model itself is wrong or not adequate; ii) noisy errors are ill known; iii) initial guesses; vi) strong non linearity.

Assuming that the models are linear and roughly acceptable up to some Gaussian additive noise with known variance, then it was proven that the filter converges to the steady state regardless of the initial conditions. This justifies why particular interest has been focused on the estimation of the noise variance, or, equivalently, variance-covariance matrices $Q$ and $R$ pertaining respectively to the state and the measurement models. The resulting algorithm is known in the literature as Adaptive Kalman Filter (AKF), which has been investigated by several researchers [2,3,6,7,8]. Intuitively, in AKF, the filter adjusts its knowledge about the $Q$ and $R$ values according to the gap between the predicted estimates and the current measurements. There are several ways to this end. Mehra [7] classified these methods into four categories, Bayesian, maximum likelihood, correlation (autocorrelation) and covariance matching. The approach described in this paper belongs to third above class where the autocorrelation functions of innovation sequence are constructed. It improves Mehra's [6] and Carrew's and Bellanger's [3] approaches by incorporating information about the quality of the innovation estimates leading to a weighted least squares methodology. The latter permits to generate a convergent sequence to the steady state filter, which after some manipulations allows to determine the values of a priori noise statistics $Q$ and $R$. Strictly speaking, the convergence of the sequence towards the steady state depends also on the quality of the innovation values. Consequently, even if in practice it may happen that Bellanger's as well Mehra's approaches do not converge, our proposal can not be worse than the former. This is, in some ways, similar to robust statistical tools used in robust statistics [9]. In the next section, we introduce the problem statement

and general notations. The third section outlines Carrew's and Bellanger's approaches based on the autocorrelation functions of the innovation sequence. The fourth section describes an improvement of the previous approach based on a weighted least squares methodology. Particularly, an approach for determining the weight factors taking into account both the first and second order statistics of autocorrelation functions is put forward. The determination of the values of Q and R noise matrices in the case of weighted square approach is pointed out in the fifth section. In section 6, some simulations examples similar to that used by Mehra [7] are carried out and the significant improvement of the proposal are illustrated.

## 2. Statement of the problem

Consider a discrete-time dynamic system described by the following vector difference equations

$$X_{k+1} = FX_k + Gw_k . \tag{1}$$

$$Y_{k+1} = HX_k + v_k . \tag{2}$$

Where

$X_i$ is n x 1 state vector, A is n x n state transition matrix, $w_i$ is q x 1 process noise, $Y_i$ is m x 1 measurement vector, $v_i$ is m x 1 measurement noise, H is m x n measurement matrix. Both $w_i$ and $v_i$ are assumed to be uncorrelated zero-mean Gaussian white noise sequences with time-invariant covariances respectively Q and R, i.e., $E[w_i]=E[v_i]=0$. $E[w_i w_j^T] = Q\delta_{ij}$ and $E[v_i v_j^T] = R\delta_{ij}$,

in which Q and R are nonnegative definite matrices whose true values are assumed to be *unknown*. Note that E [ ] denotes the expectation, and $\delta_{ij}$ corresponds to the Kronecker delta function, i.e., $\delta_{ij}=1$ if i=j and zero otherwise.

The adaptive filtering problem consists of obtaining on-line estimates $\hat{X}_{k|k}$ of $X_k$ based on observation set $Z^k = \{Y_1, Y_2, ..., Y_k\}$, and values of Q and R are identified.

Let $P_{k|k} = E[(X_k - \hat{X}_{k|k})(X_k - \hat{X}_{k|k})^T]$ be the variance-covariance matrix of $\hat{X}_{k|k}$.

In the case where Q and R are completely known, then the solution is provided by the steady state filter:

$$\hat{X}_{k|k-1} = F\hat{X}_{k-1|k-1} . \tag{3}$$

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k(Y_k - H\hat{X}_{k|k-1}) . \tag{4}$$

$$K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1} . \tag{5}$$

$$P_{k|k-1} = FP_{k-1|k-1}F^T + GQG^T . \tag{6}$$

$$P_{k|k} = (I - K_k H)P_{k|k-1} . \tag{7}$$

$K_k$ refers to the Kalman gain and $v_k = Y_k - H\hat{X}_{k|k-1}$ to the innovation or one-step- ahead prediction error. The initial state $\hat{X}_{0|0}$ is also assumed to be Gaussian with mean $X_0$ and variance matrix $P_{0|0}$. I refers to the identity matrix.

The system is assumed to be completely observable and controllable if

Rank $[H^T, (HF)^T, ..., (HF^{n-1})^T] = n$ and
Rank $[G, F^T, ..., F^{n-1}G] = n$

In the case of a steady state filter, the state prediction matrix $P_{k|k-1}$ fulfills the following time discrete Riccati matrix [1]

$$P_{k+1|k} = F[P_{k|k-1} - P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}HP_{k|k-1}]F^T + GQG^T . \tag{8}$$

Particularly, if the steady state is reached, then the solution of (8) is unique and $P_{k|k-1}=P_{k|k}$, which means that the state covariance coincides with its prediction leading to an optimal value for the gain $K_k$. The counterpart of (8) when using the state estimate instead of covariance matrix is

$$\hat{X}_{k+1|k} = F\hat{X}_{k|k-1} + FK_k(Y_k - H\hat{X}_{k|k-1}) . \tag{9}$$

However, when the a priori noise statistics Q and R are unknown or ill defined, the solution of (8) is not unique and leads to a suboptimal filter, or, equivalently, to a suboptimal gain filter. In what follows, we will describe how the innovation as well the autocorrelation functions can be used to provide values of Q and R leading to optimal steady state filter.

## 3. Use of filter innovation and its autocorrelation function

Let $X_{k|k}^*$ be denote the suboptimal estimate of $X_k$ (due to the unknown Q and R) leading to a suboptimal gain $K_D$. More formally,

$$\hat{X}_{k+1|k}^* = F\hat{X}_{k|k-1}^* + FK_D(Y_k - H\hat{X}_{k|k-1}^*). \quad (10)$$

It is then clear that the gap between $X_{k|k}^*$ and $\hat{X}_{k|k}$ provides a faithful information about the convergence of the suboptimal filter to the steady state filter. Let

$$P_{k|k-1}^* = E[(\hat{X}_{k|k-1} - X_{k|k-1}^*)(\hat{X}_{k|k-1} - X_{k|k-1}^*)^T] \quad (11)$$

be its covariance matrix. Then using $v_k$ and expressions (2), (9), (10) and (11), we have

$$\hat{X}_{k+1|k} - X_{k+1|k}^* = F(\hat{X}_{k|k-1} - X_{k|k-1}^*) + (K_k - K_D)v_k$$
$$- K_D H((\hat{X}_{k|k-1} - X_{k|k-1}^*). \quad (12)$$

Using the othogonality principle, which entails $E(\hat{X}_{i|i-1}v_j^T) = E(v_j\hat{X}_{i|i-1}^T) = 0$ (similarly for $X_{i|i-1}^*$, with $v_j = Y_j - H\hat{X}_{j|j-1}$) for all i, j. While, the covariance of the innovation is

$$E(v_jv_j^T) = HP_{j|j-1}H^T + R. \quad (13)$$

Putting this in (11), leads after some manipulations to

$$P_{k+1|k}^* = (F - K_D H)P_{k|k-1}^*(F - K_D H)^T +$$
$$(K_D - K_k)(HP_{k|k-1}H^T + R)(K_D - K_k)^T. \quad (14)$$

Now following the autocorrelation methodology initiated by Mehra [6, 7], we shall find

$$C_j = E[v_k^*v_{k-j}^{*~T}] = E[(Y_k - HX_{k|k-1}^*)(Y_{k-j} - HX_{k-j|k-j-1}^*)^T] \quad (15)$$

which, after some manipulations becomes

$$C_j = \begin{cases} H(P_{k|k-1}^* + P_{k|k-1})H^T + R, & \text{if } j = 0 \\ H(F - K_D H)^{j-1}(F - K_D H)[P_{k|k-1}^*H^T \\ \quad + (K_k - K_D)(HP_{k|k-1}H^T + R)] \text{ for } j > 1. \end{cases} \quad (16)$$

In the optimal case, of course, $K_D = K_k$ and $P_{k|k-1}^* = P_{k|k}^* = 0$, thereby, $C_j = 0 \quad \forall j$. Using $S_k = HP_{k|k-1}H^T + R$, then (15) can be rewritten for j>1 as

$$C_j = H(F - K_D H)^{j-1}[FP_{k|k-1}^*H^T + K_kS_k - K_DC_0]. \quad (17)$$

$C_j$ are m x m matrices. Noticing that the coefficients $C_j$ can be assessed directly from experiment data and the given model (let $\overline{C}_j$ be such measured coefficient), then the quantity under square brackets in (17) may be estimated. The use of least squares method leads to

$$FP_{k|k-1}^*H^T + K_kS_k = K_D\overline{C}_0 + A^{tt}\begin{bmatrix} \overline{C}_1 \\ . \\ \overline{C}_n \end{bmatrix}. \quad (18)$$

With

$$A = \begin{bmatrix} H \\ H(F - K_D H) \\ . \\ H(F - K_D H)^{n-1} \end{bmatrix}. \quad (19)$$

$A^{tt} = (A^TA)^{-1}A^T$ is a pseudo inverse of a matrix A.

The existence of $A^{tt}$ follows from the controllability and observability constraint.

Strictly speaking, the quality of such estimation is mainly related to the goodness of the autocorrelation estimates $C_j$. Indeed, the latter are also subjected to the effect of the suboptimal filter since $v_i^*$ are dependent of $X_{i|i-1}^*$. This makes the convergence of such algorithm difficult even with some repetitive procedures in the spirit of those proposed by Mehra [6, 7]. This motivates the use of the weighted least squares introduced in the next paragraph.

## 4. Use of weighted least squares

One idea to overcome the above difficulty, or at least to yield better estimates, is to introduce information about the quality of the measured autocorrelation $C_j$ as a weighting factor, leading to a weighted least squares scheme. While the determination of the weights obeys the general statistical behavioral of the autocorrelation

functions as it will be pointed out. Indeed, using general statistics properties, assuming N samples, we have

$$\overline{C}_j = \frac{1}{N} \sum_{i=1,N} v_i^* v_{i-j}^{*\ T} .$$ 

(20)

While the associated variance can be determined using

$$Cov(C_j) = \frac{1}{N-1} \sum_{i=1}^{N} (v_i^* v_{i-j}^{*\ T} - \overline{C}_j)(v_i^* v_{i-j}^{*\ T} - \overline{C}_j)^T .$$

(21)

On the other hand, it has been proved (see for instance [1]) that the normalized autocorrelation

$$C'_j = \sum_{i=1}^{N} v_i^{*T} v_{i+j}^* \left[ (\sum_{i=1}^{N} v_i^{*T} v_i^*).(\sum_{i=1}^{N} v_{i+j}^{*\ T} v_{i+j}^*) \right]^{-1/2}$$

(22)

for large N, is, in view of central limit theorem, normally distributed with zero mean and variance $1/N$. Note that expressions $v_i^{*T} v_{i+j}^*$ and $v_i^* v_{i-j}^{*\ T}$ in (21) and (22) are mathematically equivalent due to the property of the autocorrelation function $C_j = (C_{-j})^T$. Consequently, a rational for obtaining the weight $w_j$, which will be incorporated into the above least squares method, should take account for the closeness of the estimate in terms of its first and second order statistics to the ideal distribution having zero and 1/N as respectively first and second order statistics.

For this purpose, we first compute Bhattacharyya distance [9] between these two distributions. The latter argues that given two normal distributions provided by their means (may be in vector form) and covariance (may be matricial), say, if $G_1 = (M_1, P_1)$ and $G_2 = (M_2, P_2)$, then

$$d(G_1, G_2) = (M_1 - M_2)^T \left[ \frac{P_1 + P_2}{2} \right]^{-1} (M_1 - M_2)^T$$
$$+ \frac{1}{2} \ln \left[ \frac{|0.5(P_1 + P_2)|}{\sqrt{|P_1||P_2|}} \right].$$

(23)

where the notation |. | stands here for the determinant of the matrix.

As seen in (23), the distance employs two terms. The first one is cancelled if both distributions have the same mean or variance providing the class of separability due the mean difference. While the

second gives the separability class due to the covariance difference.

In our case the counterparts of $G_1$ and $G_2$ are respectively $(\overline{C}_j, Cov(C_j))$ and (zeros, $(1/N).I$), where zeros stands for a null matrix, i.e., zeros elements. That is, the first component in each pair is in matricial form instead of vectorial form as it is the case in (23). Therefore, in order to accommodate to the constraint imposed by (23), we will restrict to the diagonal elements of $\overline{C}_j$ Clearly, this restriction will not change drastically the performance of the result for many reasons. i) The main information supported by the autocorrelation functions $C_j$ are concentrated within the diagonal elements. ii) The use of scalar distances in (23) will significantly simplify the forthcoming procedure for the determination of the weight values. Particularly, it permits to avoid some inconsistencies due to singularity or lack of positive definiteness. iii) The check for positiveness will be significantly simplified, since it boils down to a comparison of values of diagonal elements, while otherwise, it entails the search for the eigen values of the matrix. iv) The values of weights themselves are not so significant as the order of magnitude is. That is, the relevant information is rather the extent to which one element is more significant than another one. This means that the weights are rather defined in relative scale instead of absolute scale. Let

$$d_j = d\left( \begin{bmatrix} diag(\overline{C}_j) \\ Cov(C_j) \end{bmatrix}, \begin{bmatrix} zeros(m,l) \\ (1/N).I \end{bmatrix} \right),$$

(24)

where zeros(m,1) stands for the zeros vector of m elements and I for the identity with the same size as Cov($C_j$), i.e., m x m matrix.

Substituting (24) in (23) leads to expression

$$d_j = [diag(\overline{C}_j)]^T \left[ \frac{1}{2}.(Cov(C_j) + Id.\frac{1}{N}) \right]^{-1} [diag(\overline{C}_j)]$$
$$+ \frac{1}{2}.\ln \left[ \frac{|0.5(Cov(C_j) + Id.[1/N]|}{\sqrt{\frac{|Cov(C_j)|}{N^m}}} \right].$$

(25)

The latter arises from noticing that $|Id.[1/N]| = \dfrac{1}{N^m}$, where the identity matrix Id has the same size as the covariance matrix $Cov(C_j)$, i.e., m x m matrix.

Now in order to emphasize the meaning of the weights with respect to the distances, one should have

$$d_i \le d_j \Rightarrow w_i \ge w_j . \qquad (26)$$

In other words, the weight sequence is non-increasing with respect to the distance sequence. This is, of course, very natural, since small distance means high agreement between the ideal distribution and the current estimated distribution, thereby, high degree of weight should be attached to.

Consequently, any decreasing function f such that $w_i = f(d_i)$ should be a potential candidate. We assume the following representation

$$w_j = \max_i [d_i] - d_j \qquad (j=1 \text{ to } n). \qquad (27)$$

The expression (27) permits to avoid negative values for weights $w_j$. While, the use of maximum operation allows a maximum compensation, in the sense that the worst alternative is fully forbidden since it will be attached zero weight. Of course, the choice in (27) can be questioned and the use of other alternative still is possible.

Finally, in order to accommodate with the size of $C_j$, the weight matrix W is given as a diagonal matrix in the following way

$$W = \begin{bmatrix} \begin{bmatrix} \overbrace{\begin{matrix} w_1 & 0 & ...0 \\ 0 & w_1 & 0 & ...0 \\ . & & & \\ 0 & ... & & w_1 \end{matrix}}^{m} \end{bmatrix} \Big\} m. & 0. & ... & 0 \\ & . & & \\ 0 & & 0 & \begin{bmatrix} \overbrace{\begin{matrix} w_n & 0 & ...0 \\ 0 & w_n & 0 & ...0 \\ . & & & \\ 0 & ... & & w_n \end{matrix}}^{m} \end{bmatrix} \Big\} m \end{bmatrix} . \qquad (28)$$

Note that in the matricial weighted least squares, it is required that W possesses a symmetric positive definite square root $W^{1/2}$ satisfying $W = W^{1/2} . W^{1/2}$ (W is m.n x m.n matrix).

Clearly, this condition is satisfied using expression (28) since all diagonal elements in (28) are positively valued. Consequently, the square root $W^{1/2}$ corresponds to the matrix obtained when replacing all the diagonal elements by their square roots $\sqrt{w_j}$, which leads obviously to a definite positive matrix.

Another remark stipulates that, in view of (28), the entailment in (26) involves rather matrix expressions of $w_i$ and $w_j$. Consequently, the inequality $w_i \ge w_j$ is equivalent to argue that the matrix $(w_i - w_j)$ is positive definite, which, due the diagonalization, boils down to a matrix with diagonal elements of positive values. This provides another justification issue for our above simplification about the nature of the weight factors $w_i$.

Now, taking account for the weight matrix W, the counterpart of (18) is provided by the weighted least squares. The latter can easily be proved that it corresponds to

$$FP_{k|k-1}^* H^T + K_k S_k = K_D \overline{C}_0 + A_w^{tt} W \begin{bmatrix} \overline{C}_1 \\ . \\ \overline{C}_n \end{bmatrix} . \qquad (29)$$

With
$$A_w^{tt} = (A^T W A)^{-1} A^T . \qquad (30)$$

## 5. Determination of Q and R values

The determination of the a priori noise statistics Q and R follows from the above analysis and the properties of the steady state filter. Let S, K and $P^*$ be the optimal values of respectively $S_k$, $K_k$ and $P_{k|k-1}^*$. Then, from (16), (28) and (14)

$$S = C_0 - HP^* H^T . \qquad (31)$$

$$K = (K_D \overline{C}_0 + A_w{}^{tt}W\begin{bmatrix} \overline{C}_1 \\ . \\ \overline{C}_n \end{bmatrix} - FP^*H^T).S^{-1}. \qquad (32)$$

$$P^* = (F - K_DH)P^*(F - K_DH)^T + (K_D - K)S(K_D - K)^T. \qquad (33)$$

The above sequence (31-33) should be repeated several times to ensure the convergence to the optimal steady state filter. The proof for the convergence is outside the scope of this paper, but the reader can be inspired mainly from Carrew's and Bellanger's [3] methodology.

Assuming, that the filter reaches roughly its optimal state, then R can be estimated using expressions of S and gain K, i.e., $S=HPH^T+R$ and $K=PH^TS^{-1}$ (P stands for optimal $P_{k|k-1}$). Indeed, from the last expression, we get $H.PH^T = HK.H$, which reported in the first expression leads to

$$R = S - H.K.S = (I - H.K)S. \qquad (34)$$

The estimation of the matrix Q is more difficult since it is not involved explicitly in any expressions (31-33). Besides the error covariance $P^*$ is different from the error covariance P where Q is involved (see expression (6)). Indeed, P* refers to the covariance of the error between the optimal steady state estimate and the suboptimal estimate (current estimate), while P refers to the error between the current estimate and the true value of the state. To circumvent, this difficult, one may determine an estimate of $PH^T$ from the optimal gain K, i.e.,

$$PH^T = K.S. \qquad (35)$$

Then, we adopt the same procedure, as the one followed by Mehra in the case where the number of unknown in Q are less than n.m, obtained after carrying n steps back and substituting in Ricatti equation (8). It leads to

$$\sum_{j=0}^{k-1} HF^jGQG^T(F^{j-k})^TH^T = HP(F^{-k})^TH^T - HF^kPH^T$$
$$- \sum_{j=0}^{k-1} HF^j\Omega(F^{j-k})^TH^T, \quad k = 1, n \qquad (36)$$

With
$$\Omega = F[K\overline{C}_0K^T - KHP - PH^TK^T]. \qquad (37)$$

Note that (36-37) is fully determined from $\overline{C}_0$ and $PH^T$ (remark that HP is the transpose of $PH^T$ since P is symmetric matrix).

Consequently, the determination of the values of the parameters pertaining to Q may be given using like least square method, or one may restrict to values of k that induce more important weight in the diagonal elements of the matrix ascribed to the left hand side of (36) as it was suggested by Mehra [6].

Note that for the implementation purpose, the sequence (31-33) may be stopped when the difference of the gain K between two successive steps becomes significantly small, which means that the system has roughly reached the steady state filter. This can be adjusted by the user according to the desired accuracy and the time constraint. Besides, the above reasoning may either be carried out at each step, meaning that Q and/or R are time invariant, or be done only at the end of the process. But of course, in practice, it is better to perform it before the end in order to avoid strong divergence of the filter.

Finally, the following algorithm summarizes the different steps leading to the identification of the noise statistics Q and R.

*Step 1.* Use the model of the state estimation (3-7) with suboptimal R and Q and the true value of the observations to assess the innovations $v_{i-k}^*$ (i=1,N (number of observations) k=1,n (dimension of vector X)

*Step 2.* Determine the autocorrelation functions $\overline{C}_k$ (k=1, n) using (20).

*Step 3.* Determine the weight matrix W using (25), (27) and (28).

*Step 4.* Check consistency of W with respect to mean values $\overline{C}_k$.

*Step 5.* Determine $FP_{k|k-1}^*H^T + K_kS_k$ using (28-29).

*Step 6.* Repeat iteratively (31-33) to determine optimal values of S, K and $P^*$.

*Step 7.* Estimate R via (34) with previous estimate of S and K, then determine Q by solving (36-37).

Clearly, in Step 4, the consistency is understood in the sense that the mean values are considered to be favored over the simultaneous use of mean and covariance estimates. This is mainly justified by some observed incoherence due to the effect of non-diagonal elements in $\text{cov}(\overline{C}_k)$ as well the calculus of inverse matrix. This makes, in some ways, a bridge between the proposal and the ordered weighted averaging operator introduced by Yager [10]. With the difference that the normalization condition in our case is different from the Yager's proposal where $\sum_i w_i = 1$. While it required in our case that the matrix W possesses a symmetric positive definite square root $W^{1/2}$ such that $W = W^{1/2} \, W^{1/2}$.

Note that in general case where the autocorrelation innovation functions are close to diagonal matrices, the consistency still is hold in the sense that the ordering property with respect to mean values is fulfilled.

## 6. Simulation example

We will consider the same example tackled by Mehra [6] to prove the feasibility of his approach of adaptive filtering. The same example had also been handled by Carrew and Belanger [3]. The system arises from inertial navigation. The state model was augmented to include all the correlated parameters leading to the following matrices

$$F = \begin{bmatrix} 0.75 & -1.74 & -0.3 & 0 & -0.15 \\ 0.09 & 0.91 & -0.0015 & 0 & -0.008 \\ 0 & 0 & 0.95 & 0 & 0 \\ 0 & 0 & 0 & 0.55 & 0 \\ 0 & 0 & 0 & 0 & 0.905 \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 24.64 & 0 & 0 \\ 0 & 0.835 & 0 \\ 0 & 0 & 1.83 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix},$$

$Q = \text{diag}(1, 1, 1)$
$R = \text{diag}(1, 1)$

The above values of Q and R are the known values. It is required to identify these values using the algorithm of adaptive filtering. While the starting values of Q and R used in the system are

$$Q_0 = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.75 \end{bmatrix} \quad \text{and} \quad R_0 = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.6 \end{bmatrix}$$

Using these values, the innovation sequence $v_i$, and next the autocorrelation functions $\overline{C}_0, \overline{C}_1 ... , \overline{C}_n$ are estimated, for a number N=950 time increments (points).

Figure 1 and Figure 2 provide the estimation of the noise statistics Q and R based on weighted least squares approach (curve 2) and without weight vectors (curve 1). The estimation process is fired at each batch of N=950 measurements, which permit to determine the autocorrelation functions. It is clear from these two figures that in the estimation of either Q or R parameters, the weighted least squares approach provides better results in terms of the closeness of the output to the true value ('1') of Q and R parameters.
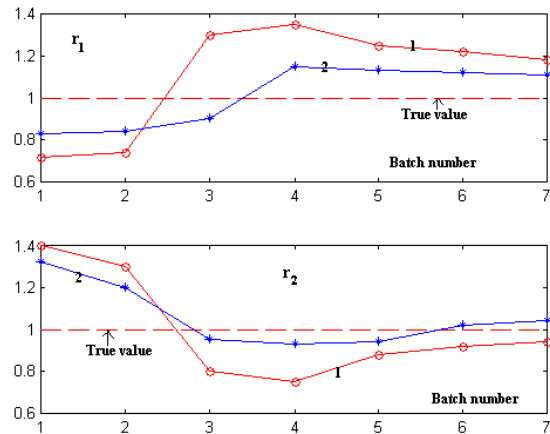


Figure 1. Estimation of R values ($r_1$ and $r_2$) based on a weighted least squares approach (curve 2) and without weight factors (curve 1).

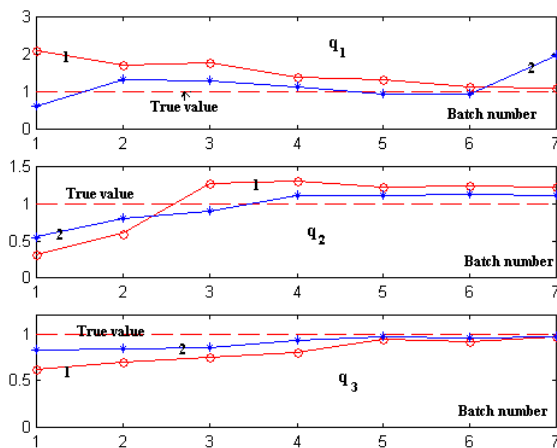Figure 2. Estimation of Q values ($q_1$, $q_2$ and $q_3$) based on a weighted least squares approach (curve 2) and without weight factors (curve 1).

## 7. Conclusion

In the present paper, we have investigated a Kalman filter approach when the a priori noise statistics Q and R are unknown. This refers to adaptive Kalman filter. We have proposed a new methodology based on the use of weighted least squares methodology, instead of least square approach, that generalizes Carrew's and Bellanger's method. The rational behind the proposal is to take account for the quality of the autocorrelation function of the innovation sequence. The weights are determined using a distance criterion between the ideal probability and the distribution referring to the current first and second order statistics of autocorrelation functions. Simulation examples are performed to prove the feasibility of the proposal.

## Acknowledgements

## References

1. Y. Bar-Shalom and X. R. Lie, *Estimation and Tracking Principles, Techniques and Software*, Artech House, Boston, London, 1993.
2. P. R. Bellanger, *Estimation of noise covariance matrices for a linear time varying stochastic process*, Automatica, Vol. 10, pp. 267-275 (1974).
3. B. Carrew and P. Bellanger, *Identification of optimum filter steady state gain for systems with unknown noise parameters*, IEEE Trans. Automatic Control, Vol. 18, N.6, pp. 582-587 (1973).
4. G. Chen, *Approximate Kalman Filtering*, Word Scientific, 1993
5. A. Gelb, *Applied Optimal Estimation*, MIT Press, 1974
6. R. K. Mehra, *On the identification of variance and adaptive Kalman filtering*, IEEE Trans. on Automatic Control, Vol. 15, N.2, pp. 175-184 (1970).
7. R. K. Mehra, *Approaches to adaptive filtering,* IEEE Trans. on Automatic Control, Vol. 17, pp. 693-698 (1972)
8. K. A. Myers and B. D. Taply, *Adaptive sequential estimation with unknown noise statistics*, IEEE Trans. on Automatic Control Vol. 21, pp. 520-523 (1976).
9. R. G. Staudte and S. J. Sheather, *Robust Estimation and Testing*, John Wiley & Sons, Inc, 1984
10. R. Yager, *Ordered weighted averaging operators*, IEEE Trans. Man Cybernetics and Systems, Vol. 18, pp. 183-190 (1988).