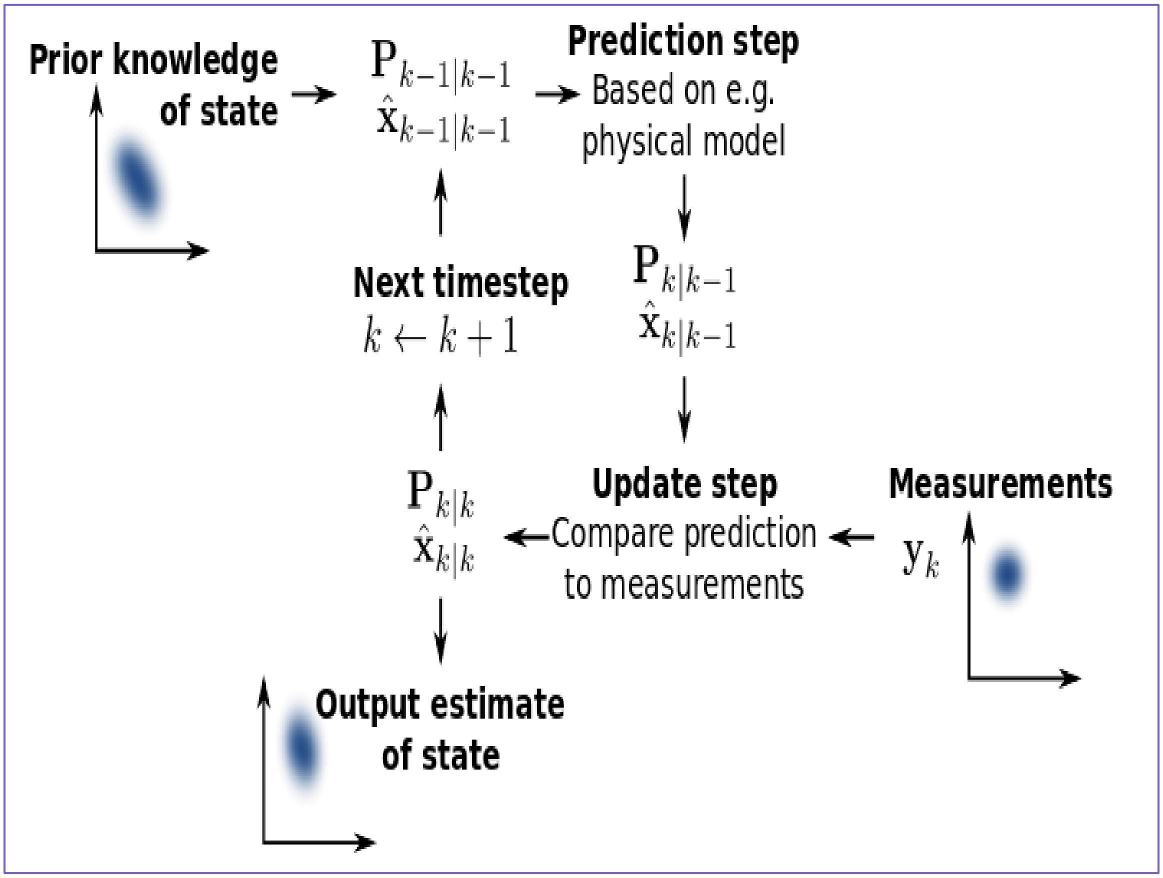


On the Identification of Variances and Adaptive Kalman Filtering

DET Paper Summary and Simulation Report

Fateme Noorzad | 810198271 | July 1, 2020

Abstract-A Kalman filter requires an exact knowledge of the process noise's covariance matrix represented by Q as well as that of the measurement's noise, represented by R. But in this paper it is considered that the true value of these 2 matrix is unknown. It is also assumed the system which we are dealing with is constant, and the inputs are random and stationary. First, a correlation test is given in order to check whether a particular Kalman filter is working optimally or not. In the case of suboptimal result a technique is given. Using this technique an approximated normal, unbiased, and consistent estimates of Q and R are found. This technique only works for special cases where Q's form is known and the number of its unknown elements is less than n*r. (n representing dimension of the state vector and r representing that of measurement vector.) For other cases, the optimal steady-state gain represented by K_{op} is calculated iterative by an iterative procedure without identifying Q. (It is shown that K_{op} depends only on n*r linear functional of Q.) The results are first derived for discrete systems, and then extended to continuous ones. At last a numerical example is given to make everything clearer.

I. Introduction

THE OPTIMUM filtering results of Kalman and Bucy [1], [2] for linear dynamic systems require an exact knowledge of the process noise's covariance matrix represented by Q as well as that of the measurement's noise, represented by R. However in a number of practical situations, Q and R are wither unknown or are known approximately. Heffes [3] and Nishinura [4] have shown the effects of errors in calculating Q and R on the performance of optimal filter. Several other investigators [5]-[9] have proposed on-line scheme to identify these 2 matrix. Most of these proposed schemes do well in identifying R but face problems in identifying Q. Besides, they don't represent a clear algorithm and scheme for continuous case.

In this paper a whole different approach has been taken. First these assumptions has been made:

- The system under consideration is time invariant.
- It is completely controllable and observable. [2]
- Both the system and filter (whether they are optimal or suboptimal) area assumed to have reached steady-state conditions.

The procedures of this paper is also listed as:

- A correlation test is performed on the filter to check whether it is optimal or not. This test is based on the innovation property of optimal filters.[10]
- If the filter is suboptimal, the innovation process' auto-correlation function is used to obtain the approximated, unbiased and consistent estimates of Q and R.

This method has a limitation. The number of unknown elements in Q must be less than n*r where n represents dimension of the state vector and r represents that of measurement vector. However, it is shown that in spite of this limitation, K_{op} is calculated using an iterative process. (These gains depends only on n*r linear relationship between elements of Q.)

- A numerical example is included to show the application of the results found in this paper.
- In the last part using all the relations shown for discrete forms, the continuous form is derived.

II. Statement of the Problem

SYSTEM:

Consider a multivariable linear discrete system:

$$x_{i+1} = \Phi x_i + \Gamma u_i \quad (1)$$

$$z_i = Hx_i + v_i \quad (2)$$

Where x_i is n*1 state vector, Φ is n*n nonsingular transition matrix, Γ is n*q constant input matrix, z_i is r*1 measurement vector, and H is r*n constant output matrix.

The sequences u_i with n*1 and v_i with r*1 as their size, are uncorrelated Gaussian white noise sequences with means and covariances as follows:

$$E\{u_i\} = o; E\{u_i u_j{}^T\} = Q\delta_{ij}$$

$$E\{v_i\} = o; E\{v_i v_j{}^T\} = R\delta_{ij}$$

$$E\{u_i v_i{}^T\} = o. for\ all\ i.j$$

Where $E\{.\}$ denotes the expected value of a random variable, and δ_{ij} denotes the Kronecker delta function.

Q and R are bounded positive definite matrices. (Meaning for all x vectors we have: $x^T Qx > o\ and\ x^T Rx > o$)

Initial state x_o is normally distributed with:

$$E\{x_o\} = o; Var\{x_o\} = E\left\{x_{o_i} x_{o_j}{}^T\right\} = P_o\delta_{ij}$$

As was mentioned the system is assumed to be completely observable and also controllable, meaning:

$$rank\{H^T.(H\Phi)^T.\dots\ .(H\Phi^{n-1})^T\} = n$$

$$rank\{\Gamma^T.\Phi\Gamma.\dots\ .\Phi^{n-1}\Gamma\} = n$$

Filter:

Let Q_o and R_o be the initial states of Q and R. Since Q and R are positive definite matrices, their initial states are also positive definite matrices. Using these estimate:

$$K_o = M_o H^T (H M_o H^T + R_o)^{-1} \tag{3}$$

$$M_O = \Phi[M_O - M_O H^T (H M_O H^T + R_O)^{-1} H M_O]\Phi^T + \Gamma Q_O \Gamma^T \tag{4}$$

Where the steady-state Kalman filter gain be K_o which is a n*r matrix[1], and M_o is the steady-state solution to the covariance equation of Kalman [1].

The filtering equations are:

$$\hat{x}_{i+1/i} = \Phi \hat{x}_{i/i} \tag{5}$$

$$\hat{x}_{i/i} = \hat{x}_{i/i-1} + K_O(z_i - H\hat{x}_{i/i-1}) \tag{6}$$

Where in the above equations $\hat{x}_{i+1/i}$ is the estimate of x_{i+1} based on all the measurements up to i, meaning all the measurements from set$\{z_O.\dots\ .z_i\}$.

Two cases can happen:

- Optimal Kalman Filter: when $Q = Q_0$ and $R = R_0$. In these cases M_0 is the covariance of the error in estimating the state.
- Suboptimal Kalman Filter: in this case, the covariance of the error called M_1 , is given by the following equation:

$$M_1 = E\{(x_i - \hat{x}_{i/i-1})(x_i - \hat{x}_{i/i-1})^T\}$$
$$= \Phi\left[M_1 - K_O H M_1 - M_1 H^T {K_O}^T + K_O(H M_1 H^T + R){K_O}^T\right]\Phi^T + \Gamma Q \Gamma^T \tag{7}$$

Problem:

The true values of Q and R are unknown so we need to check the below items:

[1] The conditions of complete controllability and observability together with the positive definiteness of Q_0 and R_0 ensure the asymptotic global stability of the Kalman filter. For better explanation check Deyst and Price [11].

- Check whether the Kalman filter constructed using some estimations of Q and R are close to optimal or not. → Hypothesis Testing
- Obtain unbiased and consistent estimates of Q and R. → Statistical Estimation
- Adapt the Kalman filter at regular intervals using all the previous information → Adaptive Filtering

To solve the stated problems, we make use of innovation property of an optimal filter [10].

III. The Innovation Property of an Optimal Filter[1]

Statement:

For an optimal filter, the sequence $\upsilon_i = z_i - H\hat{x}_{i/i-1}$ known as the innovation sequence, is a Gaussian white noise sequence.

Proof: A direct proof is obtained using the orthogonality principle of linear estimation [10][2].

If error in estimating a state is represented by e_i:

$$e_i = x_i - \hat{x}_{i/i-1}$$

$$\upsilon_i = He_i + v_i$$

$$E\{\upsilon_i \upsilon_j{}^T\} = E\{(He_i + v_i)(He_i + v_i)^T\} \tag{8}$$

For $i > j$, $v_i$is independent of e_j and v_j so:

$$E\{\upsilon_i \upsilon_j{}^T\} = E\left\{He_i\left(He_j + v_j\right)^T\right\} + E\{v_i\}E\left\{\left(He_j + v_j\right)^T\right\} = E\{He_i\left(z_j - H\hat{x}_{j/j-1}\right)^T\}$$

The orthogonality principle states that e_i is orthogonal to $\{z_k \, for \, k < i\}$ (based on orthogonality principle error is orthogonal to observations). Since $\hat{x}_{j/j-1}$ depends only on $\{z_k \, for \, k < i\}$, it is concluded that:

$$E\{\upsilon_i \upsilon_j{}^T\} = o. for \; i > j$$

Using the same approach the above result can be obtained, for $i < j$. For $i = j$:

$$E\{\upsilon_i \upsilon_j{}^T\} = HMH^T + R$$

---

[1] For detailed discussion, see Kailath [10].
[2] An alternative proof will be given in Section IV.

Since υ_i is a linear sum of Gaussian random variables, it is also Gaussian, or in better words Gaussian white noise sequence. Heuristically, the innovation υ_i represents the new information brought by z_i. There are also two papers showing other results as well:

- Kailath [10] shows that υ_i and z_i contain the same statistical information and are equivalent as far as linear operations are concerned.
- Schweppe [12] shows that υ_i can be obtained from z_i by a Gram-Schmidt orthogonalization (or a whitening) procedure.

In this paper, innovation sequence is used to check the optimality of Kalman filter and also to estimate Q and R. Using all the discussed information, investigation of suboptimal filter is discussed in the next part.

IV. Innovation Sequence for a Suboptimal Filter

Let K denote the steady-state filter gain. Using the above information it is shown that υ_i is also a stationary Gaussian sequence, under steady-state, like it is for the optimal filter case.

$$\upsilon_i = z_i - H\hat{x}_{i/i-1}$$

And also:

$$\upsilon_i = He_i + v_i$$

So for showing the dependence between υ_i and another sample of it:

$$E\{\upsilon_i \upsilon_{i-k}{}^T\} = E\{(He_i + v_i)(He_{i-k} + v_{i-k})^T\} = HE\{e_i e_{i-k}{}^T\}H^T + HE\{e_i v_{i-k}{}^T\}$$

The above equation holds for all $k > 0$. For a simpler equation for the above result we follow 2 steps:

- First Step: Finding the result of the first term: We need to find a recursive relationship for calculating e_i:
  As we know error is defined:

  $$e_i = x_i - \hat{x}_{i/i-1}$$

  For each state we have:

  $$x_i = \Phi x_{i-1} + \Gamma u_{i-1}$$

  Also the estimated value of state is defined as:

$$\hat{x}_{i/i-1} = \Phi\hat{x}_{i-1/i-1}$$
$$\hat{x}_{i-1/i-1} = \hat{x}_{i-1/i-2} + K(z_{i-1} - H\hat{x}_{i-1/i-2})$$

For the observations the blow equations holds as well:

$$v_{i-1} = z_{i-1} - H\hat{x}_{i-1/i-2}$$

So using the above equations we get:

$$e_i = \Phi x_{i-1} + \Gamma u_{i-1} - \Phi\hat{x}_{i-1/i-2} - \Phi K H x_{i-1} - \Phi K v_{i-1} + \Phi K H \hat{x}_{i-1/i-2}$$

The colored ones are added together, resulting in the below equation:

$$e_i = \Phi(I - KH)\, e_{i-1} + \Gamma u_{i-1} - \Phi K v_{i-1} \qquad (9)$$

If the above equation is stepped back k steps, the resulting recursive equation will be:

$$e_i = [\Phi(I - KH)]^k\, e_{i-k} + \sum_{j=1}^{k}[\Phi(I - KH)]^{j-1}\Gamma u_{i-j} - \sum_{j=1}^{k}[\Phi(I - KH)]^{j-1}\Phi K v_{i-j}$$
(10)

Now if the above equation is multiplied to ${e_{i-k}}^T$ and its expected value is taken:

$$E\{e_i {e_{i-k}}^T\} = E\{[\Phi(I - KH)]^k\, e_{i-k}{e_{i-k}}^T\} + E\left\{\left(\sum_{j=1}^{k}[\Phi(I - KH)]^{j-1}\Gamma u_{i-j}\right){e_{i-k}}^T\right\}$$

$$- E\left\{\left(\sum_{j=1}^{k}[\Phi(I - KH)]^{j-1}\Phi K v_{i-j}\right){e_{i-k}}^T\right\}$$

Based on each term's characteristics:

I. $E\{[\Phi(I - KH)]^k\, e_{i-k}{e_{i-k}}^T\}$
$= [\Phi(I - KH)]^k E\{e_{i-k}{e_{i-k}}^T\}$
$= [\Phi(I - KH)]^k\, M$

II. $E\{\left(\sum_{j=1}^{k}[\Phi(I - KH)]^{j-1}\Gamma u_{i-j}\right){e_{i-k}}^T\}$
$= \sum_{j=1}^{k}[\Phi(I - KH)]^{j-1}\Gamma\, E\{u_{i-j}{e_{i-k}}^T\}$
$= \sum_{j=1}^{k}[\Phi(I - KH)]^{j-1}\Gamma\, E\{u_{i-j}\}E\{{e_{i-k}}^T\}$
$= o$ → Because u and error are independent from each other and inputs are zero mean random variables the result is zero.

III. $E\{\left(\sum_{j=1}^{k}[\Phi(I - KH)]^{j-1}\Phi K v_{i-j}\right){e_{i-k}}^T\}$

$= \sum_{j=1}^{k}[\Phi(I-KH)]^{j-1}\Gamma\, E\{v_{i-j}e_{i-k}{}^{T}\}$
$= \sum_{j=1}^{k}[\Phi(I-KH)]^{j-1}\Gamma\, E\{v_{i-j}\}E\{e_{i-k}{}^{T}\}$
$= o \rightarrow$ Because v and error are independent from each other and inputs are zero mean random variables the result is zero.

$\rightarrow$ $\boxed{E\{e_i e_{i-k}{}^{T}\} = [\Phi(I-KH)]^{k}\, M}$

Where M is the steady-state error covariance matrix. An expression for M as below:

$$M = \Phi(I-KH)M(I-KH)^{T}\Phi^{T} + \Phi KRK^{T}\Phi^{T} + \Gamma Q\Gamma^{T} \qquad (11)$$

Which is obvious based on the equation mentioned for M_0.

- Second Step: If the recursive formula found for e_i is multiplied to $v_{i-k}{}^{T}$ and then its expected value is taken:

$$E\{e_i v_{i-k}{}^{T}\} = E\{\,[\Phi(I-KH)]^{k}e_{i-k}v_{i-k}{}^{T}\,\} + E\left\{\left(\sum_{j=1}^{k}[\Phi(I-KH)]^{j-1}\Gamma u_{i-j}\right)v_{i-k}{}^{T}\right\}$$

$$- E\left\{\left(\sum_{j=1}^{k}[\Phi(I-KH)]^{j-1}\Phi K v_{i-j}\right)v_{i-k}{}^{T}\right\}$$

Based on each term's characteristics:

I. $E\{[\Phi(I-KH)]^{k}\, e_{i-k}v_{i-k}{}^{T}\}$
$= [\Phi(I-KH)]^{k}E\{e_{i-k}v_{i-k}{}^{T}\}$
$= [\Phi(I-KH)]^{k}\, E\{e_{i-j}\}E\{v_{i-k}{}^{T}\}$
$= o \rightarrow$ Because v and error are independent from each other and inputs are zero mean random variables the result is zero.

II. $E\{\,(\sum_{j=1}^{k}[\Phi(I-KH)]^{j-1}\Gamma u_{i-j})v_{i-k}{}^{T}\,\}$
$= \sum_{j=1}^{k}[\Phi(I-KH)]^{j-1}\Gamma\, E\{u_{i-j}e_{i-k}{}^{T}\}$
$= \sum_{j=1}^{k}[\Phi(I-KH)]^{j-1}\Gamma\, E\{u_{i-j}\}E\{e_{i-k}{}^{T}\}$
$= o \rightarrow$ Because u and v are uncorrelated Gaussian white noise zero mean sequences.

III. $E\{(\sum_{j=1}^{k}[\Phi(I-KH)]^{j-1}\Phi K v_{i-j})e_{i-k}{}^{T}\}$
$= \sum_{j=1}^{k}[\Phi(I-KH)]^{j-1}\Phi \mathrm{K}\, E\{v_{i-j}v_{i-k}{}^{T}\}$
$= \sum_{j=1}^{k}[\Phi(I-KH)]^{j-1}\Phi \mathrm{K}\, R\delta_{jk}$

$= [\Phi(I - KH)]^{k-1}\Phi KR$ → This part is calculated using the given covariance of *v*.

$$\Rightarrow \boxed{E\{e_i {v_{i-k}}^T\} = -[\Phi(I - KH)]^{k-1}\Phi KR}$$

Summing up these 2 steps, we get:

$$E\{v_i {v_{i-k}}^T\} = H\big([\Phi(I - KH)]^k\, M \big)H^T + H\big(-[\Phi(I - KH)]^{k-1}\Phi KR \big)$$

$$= H[\Phi(I - KH)]^{k-1}\Phi[MH^T - K(HMH^T + R)].\ K > O$$

And for k = 0: $E\{v_i {v_{i-k}}^T\} = HMH^T + R$

Therefore, we can see that autocorrelation doesn't depend on *i*, so we can say that v_i is stationary. Also because it is a linear product of Gaussian random variables, it is in fact Gaussian.

Now we define a new variable:

$$C_k \equiv E\{v_i {v_{i-k}}^T\}$$

So:

$$C_k = \begin{cases} HMH^T + R\,. & k = o \\ H[\Phi(I - KH)]^{k-1}\Phi[MH^T - KC_o]. & k > o \end{cases} \qquad (12)\,,(13)$$

As can be seen: $C_{-k} = C_k^T$

(Note that based on innovation property, the optimal choice of $K = MH^T C_o^{-1}$ makes C_k vanish for all values of k, but *k=0*.)

V. A Test of Optimality for a Kalman Filter

Based on what has been discussed, for a Kalman filter to be optimal, a necessary and sufficient condition is for v_i to be a white sequence. This condition can be tested with different methods stated in [13], [16]-[19]. But now we just take a look at what has been discussed in [13] by Jenkins and Watts.

In this method, provided by this 2 authors, an estimation of C_k denoted as $\hat{C}_k$ is calculated by using the ergodic property of a stationary random sequence:

$$\hat{C}_k = \frac{1}{N}\sum_{i=k}^{N} \nu_i \nu_{i-k}{}^T \tag{14}$$

Where N is the number of sample points.

These $\hat{C}_k$ are biased for finite sample sizes:

$$E\{\hat{C}_k\} = (1 - k/N)\, C_k \tag{15}$$

For the case where we need an unbiased estimate, we divide by (N-k) instead of N. However, it is shown in [13] since the estimated stated above is preferable, since it gives a smaller mean-square error than the unbiased case.

For a discussion on covariance of $\hat{C}_k$, an approximation of it from [14] is used. This approximation holds for large *N* which is a useful one for our application.

$$cov\left(\left[\hat{C}_k \right]_{ij} . \left[\hat{C}_l \right]_{pq} \right) \approx \frac{1}{N} \sum_{t=-\infty}^{\infty} \left([C_t]_{ip}[C_{t+l-k}]_{iq} + [C_{t+l}]_{iq}[C_{t-k}]_{ip} \right) \tag{16}$$

Where in this equation $\left[\hat{C}_k \right]_{ij}$ denotes the element in the *i*th row and the *j*th column of the matrix $\hat{C}_k$.

As the above equations suggest, it can be seen that $C_k \to \infty$ for large values of *k*. Using the same way a Kalman filter is proven to be stable [11], it can be seen that although the above sum is infinite, the result is finite and proportional to *1/N*. Thus $\hat{C}_k$ are asymptotically unbiased and consistent. Moreover, as [14] shows since all the eigenvalues of $\Phi(I - KH)$ lie inside the unit circle, ν_i belongs to the class of linear processes. Based on [15] $\hat{C}_k$ are asymptotically normal. So for the white noise case:

$$cov\left(\left[\hat{C}_k \right]_{ij} . \left[\hat{C}_l \right]_{pq} \right) = \begin{cases} o & k \neq l \\ \frac{1}{N}\ [C_o]_{ip}\ [C_o]_{iq} & k = l > o \\ \frac{1}{N}\ [C_o]_{ip}\ [C_o]_{iq} + [C_o]_{iq}\ [C_o]_{ip} & k = l = o \end{cases} \tag{17}$$

If we normalize the estimates of autocorrelation coefficients we get:

$$[\hat{\rho}_k]_{ij} = \frac{[\,\hat{C}_k\,]_{ij}}{\sqrt{[\,\hat{C}_o\,]_{ii}\,[\,\hat{C}_o\,]_{jj}}} \tag{18}$$

Since our interest is on white noise, for this case and using $cov\left(\left[\hat{C}_k \right]_{ij} . \left[\hat{C}_l \right]_{pq} \right)$ value found above we get:

$$var\{\, [\hat{\rho}_k]_{ii} \,\} = \frac{1}{N} + O(\frac{1}{N^2}) \tag{19}$$

As $[\ \hat{C}_k\ ]_{ii}$ were normally distributed, $[\hat{\rho}_k]_{ii}$ are asymptotically normal [15] and as a result the 95% confidence intervals for it when $k > 0$ are $\pm(1.96/\sqrt{N})$. Equivalently the 95% confidence intervals of $[\ \hat{C}_k\ ]_{ii}$ are $\pm(1.96/\sqrt{N})\ [\ \hat{C}_o\ ]_{ii}$.

Test:

In cases where N is large, for checking whether ν_i is a white sequence or not, we need to look at the set of values for $[\hat{\rho}_k]_{ii}$ while $k > 0$. If the number of times its values lie outside of the $\pm(1.96/\sqrt{N})$ bound is less than 5% of the total, then we can conclude that ν_i is white and as a result the Kalman filter is optimal. Otherwise ν_i is not a white sequence and as a result we are dealing with a suboptimal Kalman filter.

For cases when N is small other tests mentioned in [17] by Anderson and in [16] by Hannan may be used. Another frequency domain test is given in [13] by Jenkins and Watts which is useful in the case of slow periodic components in the time series.

VI. Estimation of Q and R

If the test of last section indicates that ν_i is not a white sequence and as a result we are dealing with a suboptimal Kalman filter, then the next step is to obtain better estimates of Q and R. The proposed way of doing so, is to compute C_k earlier. The method can be formulated as three steps listed below:

I. Obtain an estimate of MH^T:

$$C_1 = H\Phi M H^T - H\Phi K C_O$$
$$C_2 = H\Phi^2 M H^T - H\Phi K C_1 - H\Phi^2 K C_O$$
$$\vdots$$
$$C_n = H\Phi^n M H^T - H\Phi K C_{n-1} - H\Phi^2 K C_{n-2} - \cdots - H\Phi^n K C_o$$

Therefore:

$$MH^T = B^{\#} \begin{bmatrix} C_1 + H\Phi K C_O \\ C_2 + H\Phi K C_1 + H\Phi^2 K C_O \\ \vdots \\ C_n + H\Phi K C_{n-1} + H\Phi^2 K C_{n-2} + \cdots + H\Phi^n K C_o \end{bmatrix} \tag{20}$$

Where $B^{\#}$ is the pseudo-inverse of the matrix B defined as:

$$B = \begin{bmatrix} H \\ H\Phi \\ \vdots \\ H\Phi^{n-1} \end{bmatrix} \Phi$$

There is just one point to note. Since B is the product of the observability matrix and also the nonsingular transition matrix Φ, its rank is n. As a result for its pseudo-inverse we have:

$$B^{\#} = (B^T B)^{-1} B^T$$

So now if we want to find an estimation of MH^T, by using all the above equations we have:

$$\widehat{MH^T} = B^{\#} \begin{bmatrix} \hat{C}_1 + H\Phi K \widehat{C_O} \\ \hat{C}_2 + H\Phi K \hat{C}_1 + H\Phi^2 K \hat{C}_O \\ \vdots \\ \hat{C}_n + H\Phi K \hat{C}_{n-1} + H\Phi^2 K \hat{C}_{n-2} + \cdots + H\Phi^n K \hat{C}_o \end{bmatrix} \tag{21}$$

Another way of estimating MH^T is by using the following equations:

$$\widehat{MH^T} = K\hat{C}_O + A^{\#} \begin{bmatrix} \hat{C}_1 \\ \vdots \\ \hat{C}_n \end{bmatrix} . A = \begin{bmatrix} H\Phi \\ H\Phi(I - KH)\Phi \\ \vdots \\ H[\Phi(I - KH)]^{n-1}\Phi \end{bmatrix} \tag{22}$$

In numerical calculations, the second way of computing $\widehat{MH^T}$ is preferable, because A is better conditioned than B.

II. Obtain an estimation for *R*:

For doing so, as we saw earlier, $C_o = HMH^T + R$, so an estimation for R can be obtained:

$$\hat{R} = \widehat{C_o} - H\left(\widehat{MH^T}\right) \tag{23}$$

III. Obtain an estimation of *Q*:

Since we want to estimate Q using the following equation and only an estimation of MH^T is available rather than M, this step is complicated.

$$M = \Phi(I - KH)M(I - KH)^T\Phi^T + \Phi K R K^T \Phi^T + \Gamma Q \Gamma^T$$

In better words only n*r linear relationships between the unknown elements of Q are available. So in order to proceed, only n*r elements or less must be unavailable and calculated in this step. Otherwise we can't find a unique solution for Q's elements. However, as can be seen in the next step, even in these conditions, a unique solution for K_{op} can be found.

Assuming the case where only n*r or less elements pf Q are unknown, we get:

$$M = \Phi M \Phi^T + \Omega + \Gamma Q \Gamma^T$$

Where:

$$\Omega = \Phi[-KHM - MH^T K^T + KC_o K^T]\Phi^T$$

If M is replaced by itself in the right hand side of the above equation we get:

$$M = \Phi^2 M (\Phi^2)^T + \Phi\Omega\Phi^T + \Omega + \Phi\Gamma Q \Gamma^T \Phi^T + \Gamma Q \Gamma^T \quad (25)$$

Repeating the same procedure as above for n times, and separating the terms involving Q on the left-hand side, we get:

$$\sum_{j=o}^{k-1} \Phi^j \Gamma Q \Gamma^T \left(\Phi^j\right)^T = M - \Phi^k M (\Phi^k)^T - \sum_{j=o}^{k-1} \Phi^j \Omega \left(\Phi^j\right)^T . for\ k = 1. \dots .n$$

(26)

Multiplying both sides by H and $(\Phi^{-k})^T \mathrm{H}^T$, we get:

$$\sum_{j=o}^{k-1} \mathrm{H}\Phi^j \Gamma Q \Gamma^T \left(\Phi^{j-k}\right)^T H^T = HM(\Phi^{-k})^T \mathrm{H}^T - \mathrm{H}\Phi^k M H^T$$
$$- \sum_{j=o}^{k-1} \mathrm{H}\Phi^j \Omega \left(\Phi^{j-k}\right)^T \mathrm{H}^T \ . for\ k = 1. \dots .n$$

(27)

Now it can be seen that both sides of the above equations depend on estimation of MH^T instead of M, so the problem is solved. All we have to do now is to replace the exact values by the approximated ones and the equations for finding an estimation of Q is found.

$$\sum_{j=o}^{k-1} \mathrm{H}\Phi^j \Gamma \hat{Q} \Gamma^T \left(\Phi^{j-k}\right)^T H^T = \widehat{HM}(\Phi^{-k})^T \mathrm{H}^T - \mathrm{H}\Phi^k \widehat{MH^T}$$

$$- \sum_{j=o}^{k-1} \mathrm{H}\Phi^{j}\widehat{\Omega}\left(\Phi^{j-k}\right)^{T}\mathrm{H}^{T} \;.for\; k = 1.\dots\; .n$$

(28)

Where:

$$\widehat{\Omega} = \Phi[-K\widehat{HM} - \widehat{MH^{T}} + K\hat{C}_{O}K^{T}]\Phi^{T}$$

(29)

One point worth mentioning in this part is that all the k equations found above are not linearly independent. So when using them to find the estimated value of Q, one has to choose the ones which are independent and actually find a subset of these equations. This procedure is explained in better words in section IX with an example to make this point clearer.

VII. Direct Estimation of the Optimal Gain

The method stated in the last section only works for when the number of unknown parameters in Q is less than n*r. So in the cases where that of Q are more than n*r or Q is completely unknown the above algorithm cannot be used. However, the optimal Kalman filter gain can still be calculated. This calculation is based on an iterative procedure that we want to discuss in this section.

Based on what has been discussed in section II, this following equation holds:

$$M_1 = \Phi\left[M_1 - K_O H M_1 - M_1 H^T {K_O}^T + K_O(HM_1H^T + R){K_O}^T\right]\Phi^T + \Gamma Q \Gamma^T \quad (30)$$

Where K_0 is the initial gain of the Kalman filter and M_1 is the error covariance matrix corresponding to K_0. Now we define a new parameter as below:

$$K_1 \equiv M_1 H^T (HM_1H^T + R)^{-1}$$

Now if we call M_2 the error covariance matrix corresponding to the new defined gain K_1, we will have:

$$M_2 = \Phi\left[M_2 - K_1 H M_2 - M_2 H^T {K_1}^T + K_1(HM_2H^T + R){K_1}^T\right]\Phi^T + \Gamma Q \Gamma^T \quad (32)$$

If we subtract (30) from (31), we get:

$$M_2 - M_1 = \Phi(I - K_1H)(M_2 - M_1)(I - K_1H)^T\Phi^T - \Phi(K_1 - K_O)$$

(32)

$$(HM_1H^T + R)(K_1 - K_O)^T\Phi^T$$

Based on how Kalman shows the positive definiteness of M and since we assumed our system is observable and controllable:

$$M_2 - M_1 < o \quad or \quad M_2 < M_1$$

Similarly by defining

$$K_2 \equiv M_2H^T(HM_2H^T + R)^{-1}$$

And M_3 as its corresponding error covariance matrix and by attributes stated above, we conclude that:

$$M_3 < M_2 < M_1$$

In addition to above inequalities, we know that M is a positive definite matrix, meaning this sequence of decreasing matrices are bounded from below. As a result the sequence of $K_o. K_1. \dots$ must converge to K_{op}. So we can come up with a 4 step algorithm to estimate K_{op}:

I. Obtain an estimate K_1:

This estimated is denoted as $\widehat{K_1}$:

$$\widehat{MH^T} = K\hat{C}_O + A^{\#}\begin{bmatrix}\hat{C}_1\\ \vdots\\ \hat{C}_n\end{bmatrix} \rightarrow \widehat{K_1} = K_o + A^{\#}\begin{bmatrix}\hat{C}_1\\ \hat{C}_2\\ \vdots\\ \hat{C}_n\end{bmatrix}\hat{C}_o^{-1} \tag{33}$$

Using the same equations estimates of M_lH^T and R can be found too.

II. Obtain $\widehat{\delta M}$ an estimate of $M_2 - M_1$:

Based on what has been calculated for $M_2 - M_1$, its estimated value's equations can be written as below:

$$\widehat{\delta M_1} = \Phi\left(I - \widehat{K_1}H\right)\widehat{\delta M_1}\left(I - \widehat{K_1}H\right)^T\Phi^T - \Phi\left(\widehat{K_1} - K_O\right)$$
$$\hat{C}_O(\widehat{K_1} - K_O)^T\Phi^T \tag{34}$$

$\widehat{\delta M_1}$ can be calculated recursively in the same way as M_o is calculated for Kalman filter. Also note that for convergence considerations, $\Phi\left(I - \widehat{K_1}H\right)$ must have eigenvalues in the unit circle in order to be stable.

III. Obtain $\widehat{M_2H^T}$ and $\widehat{K_2}$:

$$\widehat{M_2 H^T} = \widehat{M_1 H^T} + \delta\widehat{M}_1 H^T \tag{35}$$

$$\widehat{K_2} = \widehat{M_2 H^T}\left(H\widehat{M_2 H^T} + \hat{R}\right)^{-1} \tag{36}$$

IV. Repeat steps II and III until $||\delta\widehat{M_\iota}||$ or $||\widehat{K}_\iota - \widehat{K_{\iota-1}}||$ became small compared to $||\widehat{M_\iota}||$ or $||K_i||$ where ||.|| denotes a suitable matrix norm. There is also an alternative way. K_2 can be calculated by filtering z using K_1 and then performing step I.

This procedure for finding optimal Kalman filter gain reveals an interesting relationship between this gain and Q. As can be seen in the equation leading to the value of $M_2 - M_1$ there is not Q and Q is only needed to calculate$M_1 H^T$. By considering this fact:

Corollary: It is sufficient to know n*r linear functions of Q in order to obtain the optimal gain of a Kalman filter.

Proof:

$$\begin{aligned} M_1 &= \Phi\left[M_1 - K_O H M_1 - M_1 H^T K_O{}^T + K_O(HM_1H^T + R)K_O{}^T\right]\Phi^T + \Gamma Q \Gamma^T \\ &= \Phi\,(I - K_O H) M_1 (I - K_O H)^T \Phi^T + \Phi K_O R K_O^T \Phi^T + \Gamma Q \Gamma^T \end{aligned} \tag{37}$$

Multiplying it by H^T and writing it as a recursive infinite series we get:

$$M_1 H^T = \sum_{j=o}^{\infty} [\Phi\,(I - K_O H)]^j (\Phi K_O R K_O^T \Phi^T + \Gamma Q \Gamma^T)[\,(I - K_O H)^T \Phi^T]^j H^T$$

(38)

As can be seen, if these $\sum_{j=o}^{\infty}[\Phi\,(I - K_O H)]^j (\Gamma Q \Gamma^T)[\,(I - K_O H)^T \Phi^T]^j H^T$ linear functions are given, we don't need Q to obtain the value of $M_1 H^T$. Also since the equations for optimal gain doesn't need a complete knowledge of Q to be calculated, *n*r* values of Q is enough.

On point needs to be cleared. A complete knowledge of Q is needed to calculate *M*, but if our goal is to find the optimal gain of Kalman filter, as has been stated a complete knowledge of Q is not needed. But we need to have it in mind that since our iterative scheme tries to identify Kalman filter gain by whitening ν_i, it fails to identify the complete Q matrix if unknowns in Q is more than n*r.

VIII. Statistical Properties of the Estimates

It was shown in section V that the estimates $\hat{C}_k$ are asymptotically normal, unbiased, and consistent. Since the estimators of MH^T, R, and Q are linearly related to $\hat{C}_k$, it is easy to show that they are also asymptotically normal, unbiased, and consistent. So in this following section, by finding their mean and variance, their statistical properties are shown.

I. Statistical Properties of MH^T 's estimator:

$$\widehat{MH^T} = K\hat{C}_O + A^{\#}\begin{bmatrix}\hat{C}_1\\ \vdots \\ \hat{C}_n\end{bmatrix} . E\{\hat{C}_k\} = (1-\frac{k}{N})C_k$$

Using the stated equations:

$$E\{\widehat{MH^T}\} = E\left\{K\hat{C}_o + A^{\#}\begin{bmatrix}\hat{C}_1\\ \vdots \\ \hat{C}_n\end{bmatrix}\right\} = KE\{\hat{C}_o\} + A^{\#}\begin{bmatrix}E\{\hat{C}_1\}\\ \vdots \\ E\{\hat{C}_n\}\end{bmatrix} = KC_o + A^{\#}\begin{bmatrix}\left(1-\frac{1}{N}\right)C_1\\ \vdots \\ \left(1-\frac{n}{N}\right)C_n\end{bmatrix}$$

$$= KC_o + A^{\#}\begin{bmatrix}C_1\\ \vdots \\ C_n\end{bmatrix} - \frac{A^{\#}}{N}\begin{bmatrix}C_1\\ \vdots \\ nC_n\end{bmatrix} = MH^T - \frac{A^{\#}}{N}\begin{bmatrix}C_1\\ \vdots \\ nC_n\end{bmatrix}$$

(39)

As the above equation states, for $n << N$, the bias on the estimator of MH^T is negligible. So we can conclude that for these values, it is an unbiased estimator.

Now we find the covariance of this estimator as the proceeding steps: (This variance is calculated for large values of N)

$$cov\{\widehat{MH^T}\} \approx K\,var(\hat{C}_o)K^T + A^{\#}cov\left(\begin{bmatrix}\hat{C}_1\\ \vdots \\ \hat{C}_n\end{bmatrix}\right){A^{\#}}^T + Kcov\left(\hat{C}_o . \begin{bmatrix}\hat{C}_1\\ \vdots \\ \hat{C}_n\end{bmatrix}\right){A^{\#}}^T$$

$$+A^{\#}cov\left(\begin{bmatrix}\hat{C}_1\\ \vdots \\ \hat{C}_n\end{bmatrix} . \hat{C}_o\right)K^T \qquad (40)$$

The covariances stated in the above equation can be obtained by the equations found in part V. If these values are used, it can be see that as N grows, $cov\{\widehat{MH^T}\}$ decreases.

II. Statistical Properties of R's estimator:

$$\hat{R} = \widehat{C_o} - H\left(\widehat{MH^T}\right)$$

Based on above equations and what we found for $\widehat{MH^T}$:

$$E\{\hat{R}\} = E\{\widehat{C_o} - H(\widehat{MH^T})\} = E\{\widehat{C_o}\} - HE\{\widehat{MH^T}\} = R + \frac{1}{N}HA^{\#}\begin{bmatrix} C_1 \\ 2C_2 \\ \vdots \\ nC_n \end{bmatrix} \quad (41)$$

$$Var\{\hat{R}\} = var\{\hat{C}_o\} + H\,cov\{\widehat{MH^T}\}H^T - cov\{\hat{C}_o.\widehat{MH^T}\}H^T - Hcov\{\widehat{MH^T}.\hat{C}_o\} \quad (42)$$

III. Statistical Properties of Q's estimator:

Property of this estimator is also the same as the other ones.

As the above equations suggest, all the calculated values depend on the actual and accurate values of Q and R. In the case where these matrices are known and we have the knowledge of their range, plotting variance of their estimator, helps us give a better view of that estimator. But in the case of not having the knowledge of these matrices, we need another solution. To remove the dependence on Q and R:

$$cov\{\hat{K}\} = cov\left\{\widehat{MH^T}\hat{C}_o^{-1}\right\} \approx A^{\#}cov\left(\begin{bmatrix} \hat{\rho}_1 \\ \vdots \\ \hat{\rho}_n \end{bmatrix}\right)A^{\#T}$$

(43)

Now all we need to know, is the statistical properties of $\hat{\rho}_k$. It is shown in [14] and [15] that $\hat{\rho}_k$ are asymptotically normal with mean ρ_k and:

$$cov(\hat{\rho}_k.\hat{\rho}_l) \approx \frac{1}{N}\sum_{j=-\infty}^{\infty} \rho_j\rho_{j+l-k}$$

A good estimation for the above covariance is:

$$cov(\hat{\rho}_k.\hat{\rho}_l) \approx \frac{1}{N}\sum_{j=-(N-1)}^{N-1} \rho_j\rho_{j+l-k}$$

Using this estimation for $cov\{\hat{K}\}$ results in the below expression:

$$cov\{\hat{K}\} \approx \frac{1}{N}A^{\#}A^{\#T} = \frac{1}{N}(A^TA)^{-1}$$

(44)

This equation is a good approximation for the minimum value of $cov\{\hat{K}\}$. It can help us decide the minimum value of N, in order to achieve the goal we are looking for in the gain of the Kalman filter.

V. Statistical Properties of M_2-M_1's estimator

If we consider the asymptotic convergence in which N is considered to be large enough, in section VII, $E\{\widehat{\delta M_1}\}$ was calculated and based on its equation, second and higher order moments of $\widehat{K}_1$, which in case of it being a normal process, it is finite and tend to zero asymptotically. Therefore:

$$\lim_{n\to\infty} E\{\widehat{\delta M_1}\} = \delta M_1$$

Meaning its estimator is unbiased for large values of n. Similarly covariance of $\widehat{\delta M}_1$ tends to be zero asymptotically. Thus, $\widehat{\delta M}_1$ tends to be δM_1 with probability one.

Extending the same argument, $\widehat{K}_2 \rightarrow K_2 . \widehat{K}_3 \rightarrow K_3 \ldots . \widehat{K}_{op} \rightarrow K_{op}$ with probability one.

IX. A Numerical Example from Inertial Navigation

In this section, the results of section V and VI are applied to a damped Schuler loop forced by an exponentially correlated stationary random input. Two measurements are made on the system, both of which are corrupted by exponentially correlated as well as white noise type errors. (Although here the distribution is considered to be exponential, in the beginning of the paper input is considered to have normal distribution and this normal distribution is used in many other parts of the paper. So for simulation, input's distribution is considered to be normal.)

The state of the system is augmented to include all the correlated random inputs so that the augmented state vector x is 5*1 and the random input vector u is 3*1 and the measurement noise vector is 2*1. The system is then discretized using a time step of 0.1 and the resultant system matrix can be seen as below:

$$\Phi = \begin{bmatrix} 0.75 & -1.74 & -0.3 & 0 & -0.15 \\ 0.09 & 0.91 & -0.0015 & 0 & -0.008 \\ 0 & 0 & 0.95 & 0 & 0 \\ 0 & 0 & 0 & 0.55 & 0 \\ 0 & 0 & 0 & 0 & 0.905 \end{bmatrix} \rightarrow \text{Nonsingular Transition Matrix}$$

$$\Gamma = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 24.64 & 0 & 0 \\ 0 & 0.835 & 0 \\ 0 & 0 & 1.83 \end{bmatrix} \rightarrow \text{Constant Input Matrix}$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix} \rightarrow \text{Constant Output Matrix}$$

$$Q = \begin{bmatrix} q_1 & 0 & 0 \\ 0 & q_2 & 0 \\ 0 & 0 & q_3 \end{bmatrix} \rightarrow \text{Input Covariance Matrix}$$

$$R = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} \rightarrow \text{Noise Covariance Matrix}$$

The actual values of $q_1 . q_2 . q_3 . r_1 .$ and r_2 are unity, but they are assumed unknown. It is required to identify these values using measurements $\{ z_i . i = 1. \ldots \ . N\}$.

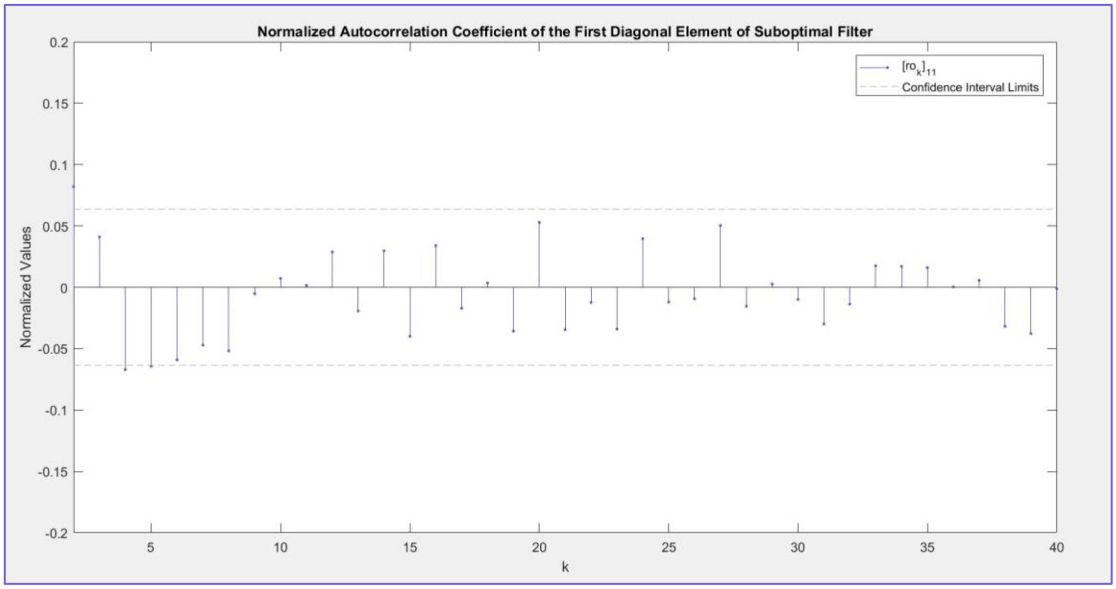
Based on what has been discussed, first the test explained in section V is executed on the given data. So in MATLAB first the given data including the number of samples ($N = 950$) is added. The given data also includes the initial values stated below:

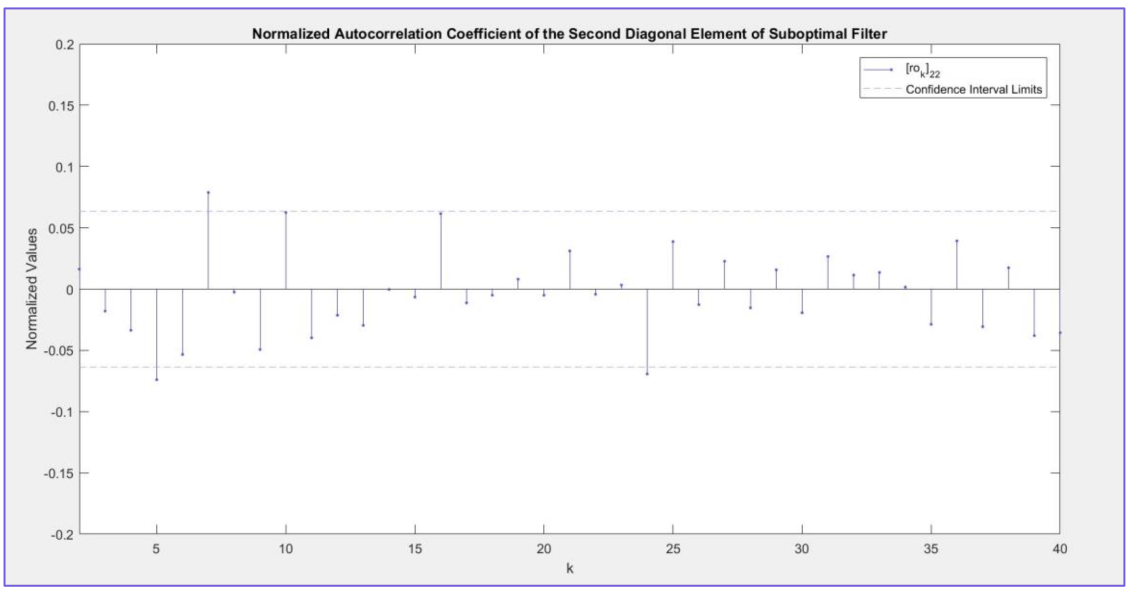
$$M_O = 100\, I = 100 \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} . \ Q_o = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.75 \end{bmatrix} . R_o = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.6 \end{bmatrix}$$

Then based on system equations given in the beginning of the report, the system is simulated. Now based on what initial value is selected for Q and R, we will have an optimal or suboptimal filter.

Optimality Test:

If the above values are given as initial values, then we are dealing with a suboptimal filter and as a result, we expect to see more than 5 percent of $[\hat{\rho}_k]_{ii}$ outside of 95% confidence interval. In order to find $[\hat{\rho}_k]_{ii}$, sequence of v_i is simulated, and then the estimator of C_k meaning $\hat{C}_k$ is calculated with $k \in \{0.1. \ldots \ .40\}$. After that, $[\hat{\rho}_k]_{ii}$ is computed with $i \in \{1.2\}$. Below plots of these estimators are plotted:






```
The number of samples outside 95 percent confidence interval for the first diagonal element of suboptimal filter is 4
The number of samples outside 95 percent confidence interval for the second diagonal element of suboptimal filter is 4
```

The above statement shows that for both estimators, more than 5% of data is outside confidence interval, meaning the filter is not optimal and is suboptimal as suspected.

Calculating Estimators of Q and R:

We now proceed to the identification of Q and R. Since the number of unknowns in Q is 3 and $n \times r$ is 10 in this example, we can identify Q completely. The set of equations in (28) give us a large number of linear equations for $\hat{q}_1$, $\hat{q}_2$, and $\hat{q}_3$. However, the most important equations which are linearly independent happen along the diagonal for $k = 1$ and $k = 5$.

For $k = 1$ the left hand-side of (28) is:

$$\begin{bmatrix} 4.37\hat{q}_3 & -0.0326\hat{q}_3 \\ 0 & 1.27\hat{q}_2 \end{bmatrix}$$

and for $k = 5$ it will be:

$$\begin{bmatrix} -8.38\hat{q}_1 + 22.3\hat{q}_3 & 1.22\hat{q}_1 - 1.47\hat{q}_2 \\ -1.25\hat{q}_1 - 0.87\hat{q}_3 & 0.141\hat{q}_1 + 20\hat{q}_2 + 0.023\hat{q}_3 \end{bmatrix}$$

The diagonal elements of the first matrix are used to calculate $\hat{q}_3$and $\hat{q}_2$ and the first diagonal element of the second matrix is then used to calculate $\hat{q}_1$. Using these equations, the unknowns are determined.

It is also possible to use all linear equations form (28) and in that case we get the best answers with the minimum least-squares error. But as it is shown in Table I, using these 3 equations doesn't influence MSE significantly.

The results obtained in Table I are obtained by using the identification scheme repeatedly on the same batch of data with 950 samples. As can be seen further iterations do not increase the likelihood function[1] much, even though the changes in Q and R are significant. A check case using the true values of Q and R is also shown in Table I. As can be seen the value of likelihood function in the check case is very close to that in the first iteration. It indicates that the estimates obtained in different iterations are quite close to one another and close to the maximum likelihood estimates as well. Also changing the value of the starting point doesn't change the converging value.

Also as can be seen in this table an estimate of the actual MSE and calculated MSE is calculated to, with the following formulas respectively:

$$Estimate\ of\ MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_{i|i-1})^{\mathrm{T}}(x_i - \hat{x}_{i|i-1})$$

$$Calculated\ MSE = tr(M_0)$$

[1] The likelihood function has been given by Scheweppe [12] as below:

$$L(Q.R) = -\frac{1}{N}\sum_{i=1}^{N}\upsilon_i^T(HMH^T + R)^{-1}\upsilon_i - \ln|HMH^T + R|$$

Table I

Estimates of Q and R Based on a Set of 950 Points

```
The calculated value for r1 for 4 iterations on the same batch of data is:
    0.6305
    0.6231
    0.6218
    0.6218
```

```
The calculated value for r2 for 4 iterations on the same batch of data is:
    0.9645
    0.9062
    0.9043
    0.9043
```

```
The calculated value for q1 for 4 iterations on the same batch of data is:
    1.0147
    1.0159
    1.0161
    1.0161
```

```
The calculated value for q2 for 4 iterations on the same batch of data is:
    1.0153
    1.1588
    1.1620
    1.1622
```

```
The calculated value for q3 for 4 iterations on the same batch of data is:
    1.5685
    1.5677
    1.5675
    1.5675
```

```
The calculated value for the calculated MSE for 4 iterations on the same batch of data is:
  1.0e+03 *

    0.3556
    1.3403
    1.3433
    1.3436
```

```
The check set value for the calculated MSE is:
   1.2994e+03
```

```
The calculated value for the actual value of MSE for 4 iterations on the same batch of data is:
  1.0e+03 *

    1.4394
    1.4266
    1.4266
    1.4266
```

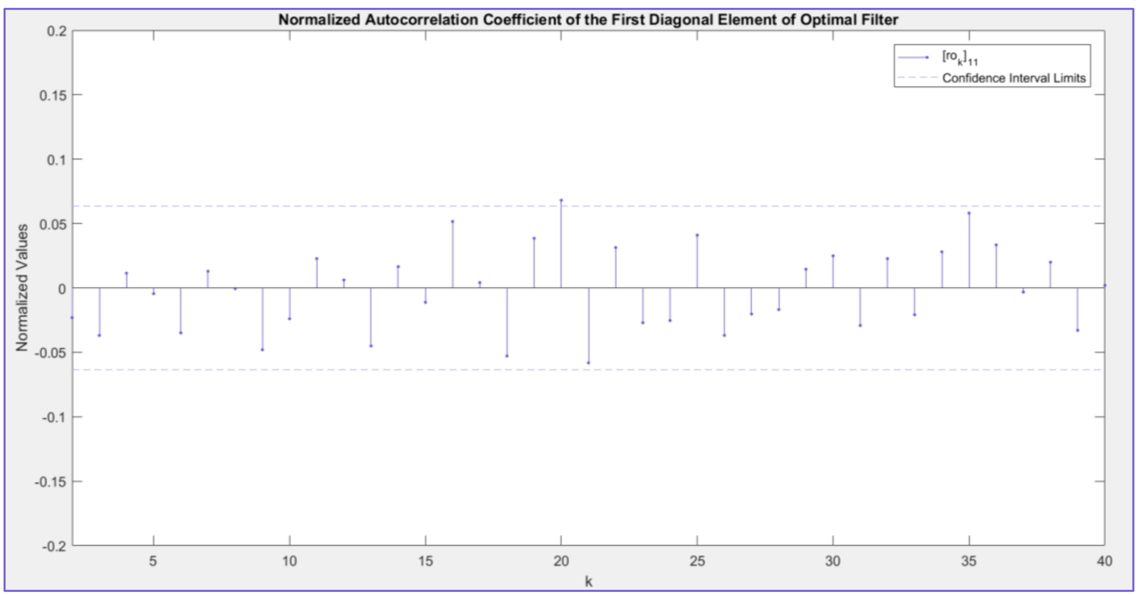
```
The check set value for the actual value of MSE is:
   1.4234e+03
```

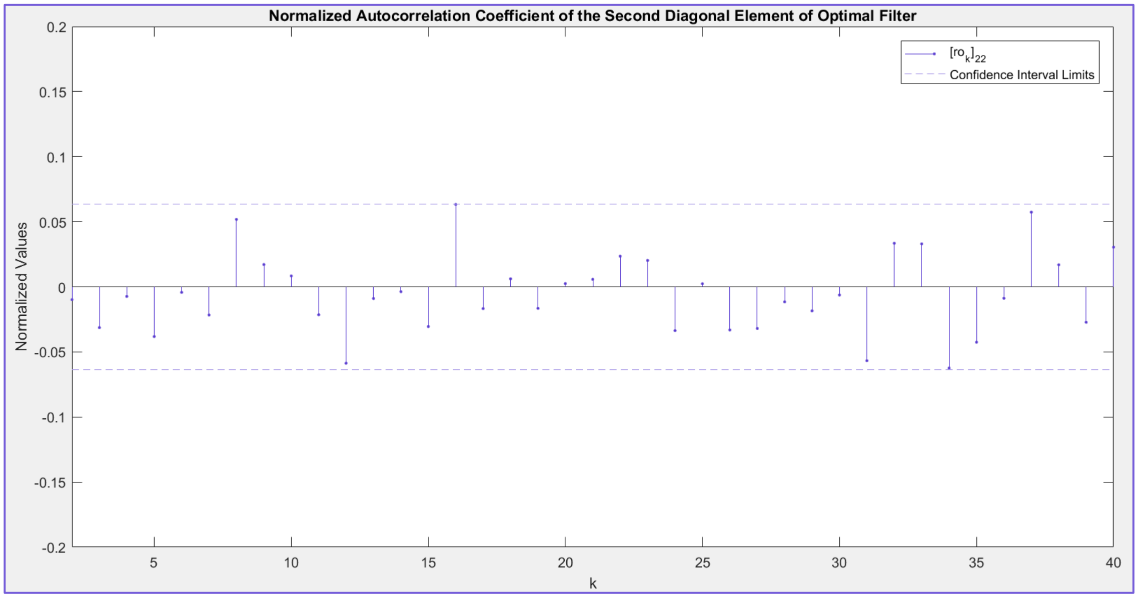
```
The calculated value for likelihood function for 4 iterations on the same batch of data is:
   -8.6519
   -7.1406
   -7.1416
   -7.1417
```

```
The check set value for likelihood function is:
   -7.1339
```

Optimal Case:

Using the calculated coefficients, now we expect to have an optimal filter. To check this statement, the optimality test is carried out using the new calculated coefficients. The result can be found below:






```
The number of samples outside 95 percent confidence interval for the first diagonal element of optimal filter is 1
The number of samples outside 95 percent confidence interval for the second diagonal element of optimal filter is 0
```

As can be seen from the above statement, 2.5% of data is out of confidence interval resulting in an optimal filter. This supports the hypothesis that υ_i is white.

Using 10 Batch of Data:

In this part using 10 batch of data, the convergence of Q and R is shown. These estimates of Q and R are updated after every batch of 950 points using the equations below:

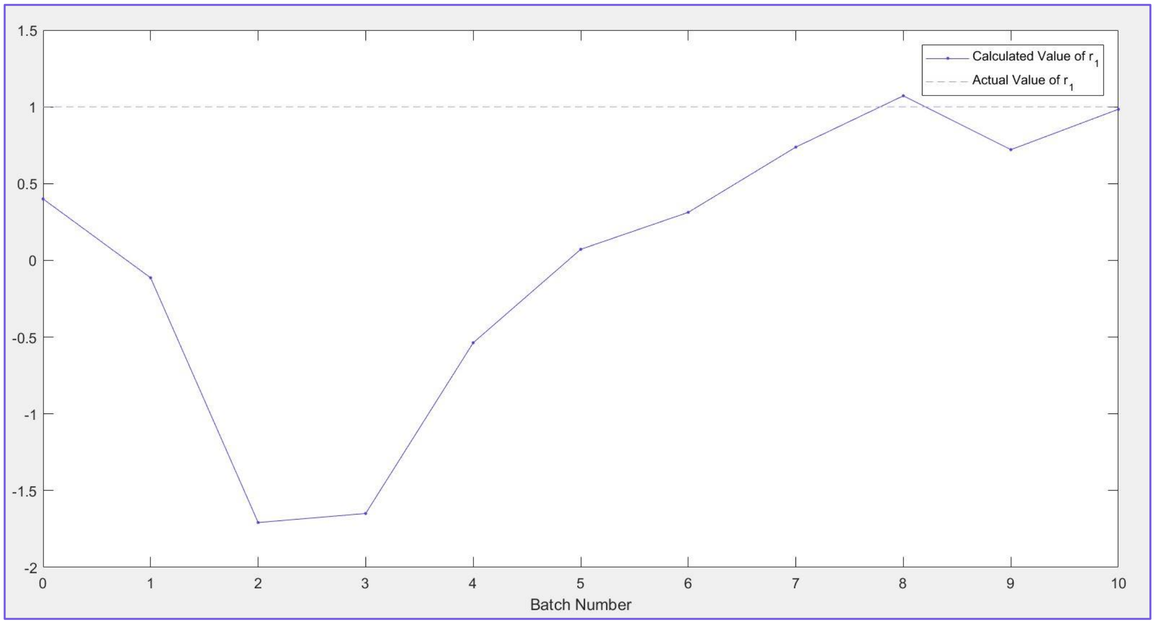
$$\hat{Q}_{k+1} = \hat{Q}_k + (\hat{Q}_{k+1.k} - \hat{Q}_k)/(k+1) \qquad (45)$$

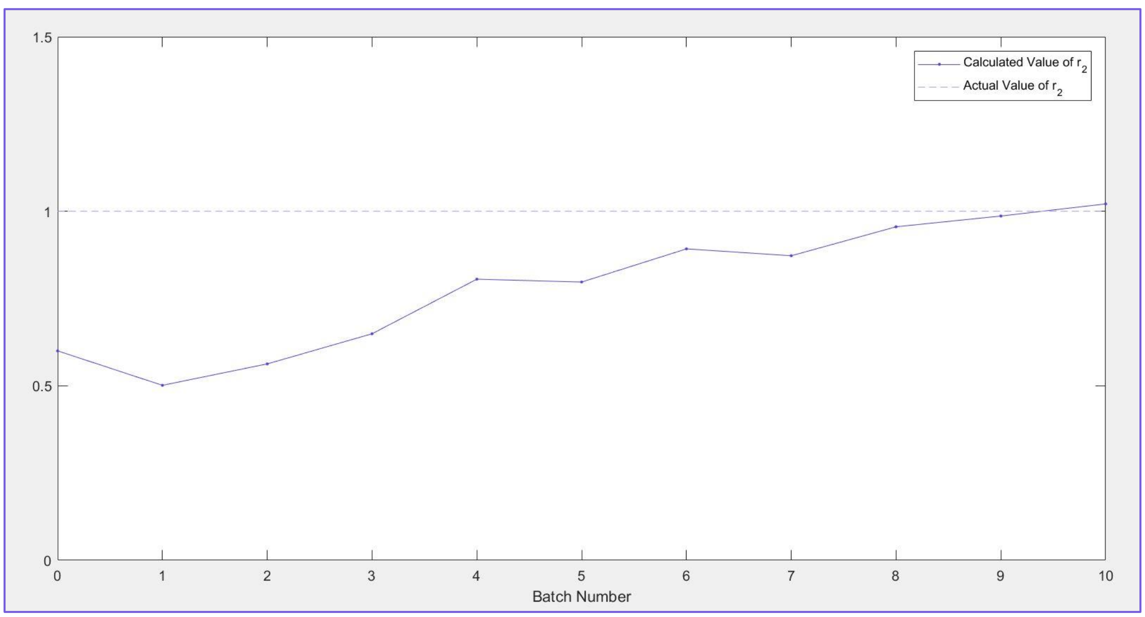
Where:

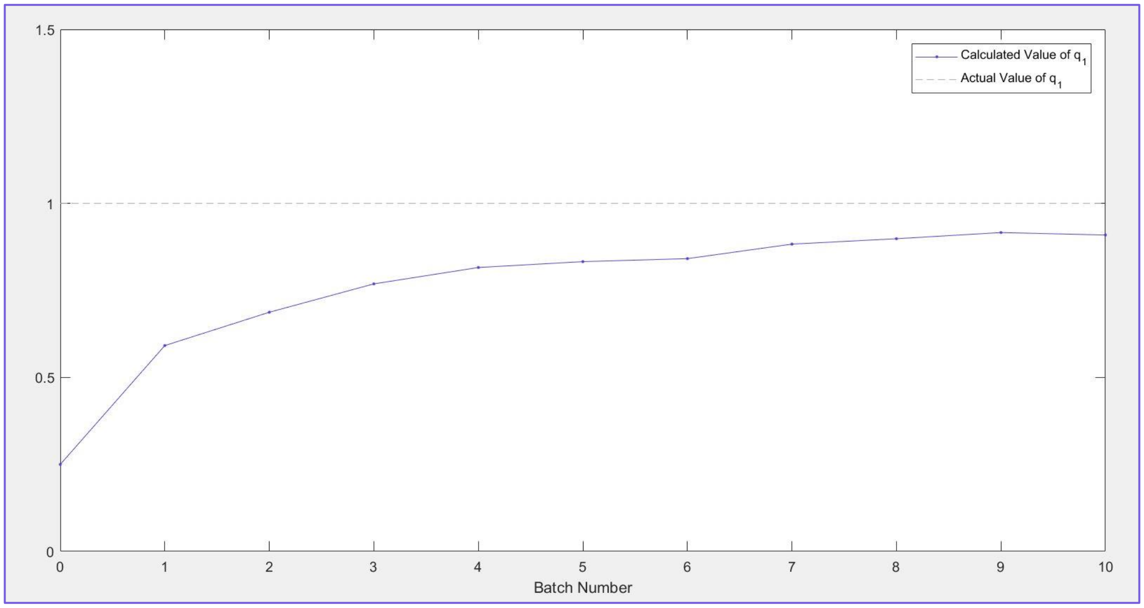
- $k \rightarrow$ the batch number
- $\hat{Q}_k \rightarrow$ the estimate of Q after k batches
- $\hat{Q}_{k+1.k} \rightarrow$ the estimate of Q based on the $k+1$ batch
- $\hat{Q}_{k+1} \rightarrow$ the estimate of Q after $k+1$ batches

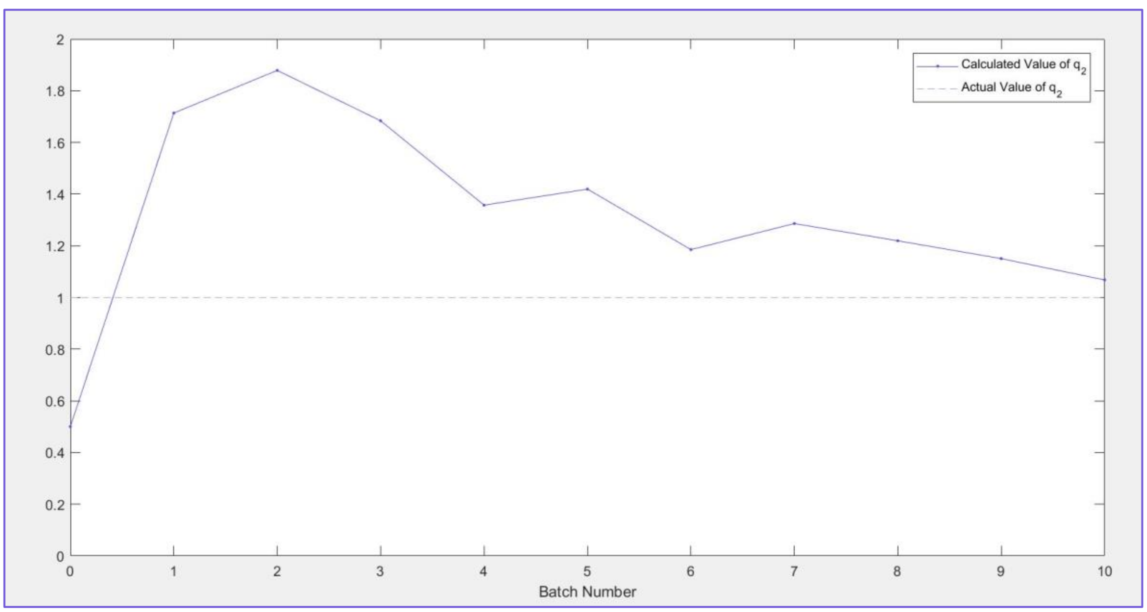
The same relation holds for R.

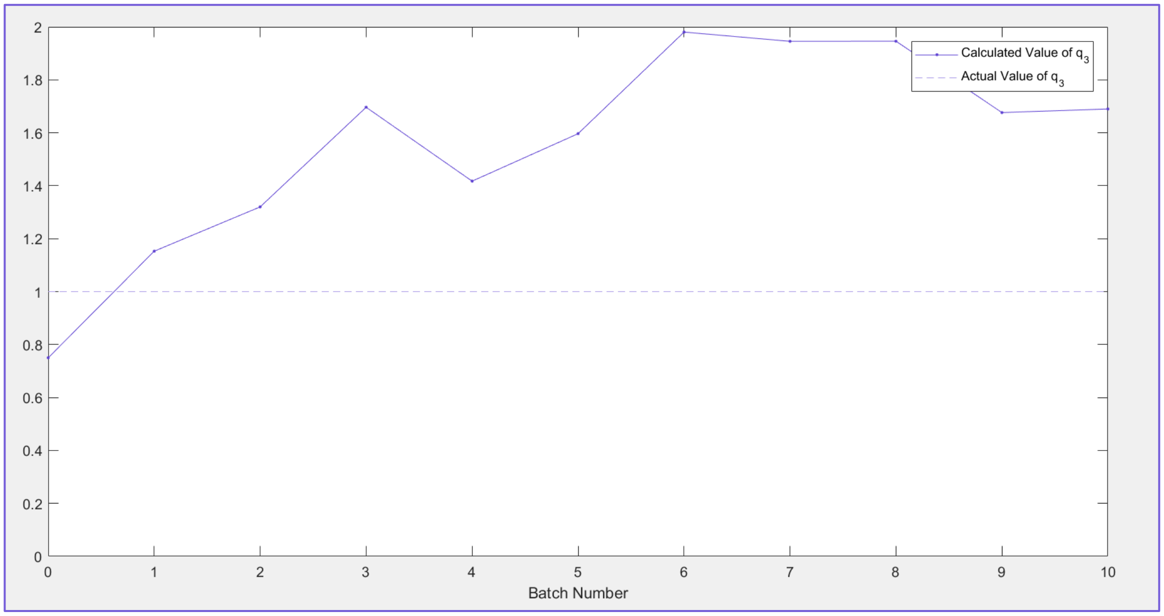
So after using the above iterations, the results are plotted as below:






1.5
1
0.5
0
0
1
2
3
4
5
6
7
8
9
10
Batch Number
Calculated Value of q_1
Actual Value of q_1


2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
1
2
3
4
5
6
7
8
9
10
Batch Number
Calculated Value of q_2
Actual Value of q_2



One point should be mentioned. Since sequences u_i and v_i are created randomly, the values and plots stated in this report, can vary. But what is important is that in mostly all cases convergence to the goal value occurs and also MSE and LF of the estimations are near that of the actual and goal values.

X. Continuous System

The results found and calculated in previous sections can be easily extended to the continuous case. We simply state the results as following[1]:

System:

$$\dot{x} = Fx + Gu$$

$$z = Hx + v$$

Filter:

$$\dot{\hat{x}} = F\hat{x} + K_0(z - H\hat{x})$$

$$K_0 = P_0 H^T R_0^{-1}$$

And

[1] The results stated in this section have not been applied to a practical example so far.

$$FP_0 + P_0F^T + GQ_0G^T - P_0H^TR_0^{-1}HP_0 = 0$$

Also the error covariance represented by P_1 is given as below:

$$(F - K_0H)P_1 + P_1(F - K_0H)^T + GQG^T + K_0RK_0^T = 0$$

Innovation Process:

$$\upsilon = z - H\hat{x} = He + v$$

where

$$e = x - \hat{x}$$

In the optimal filter case, υ is a white with the same covariance as v in [10]. However, in the suboptimal case:

$$\dot{e} = (F - K_0H)e + Gu - K_0v$$

and the autocorrelation function of υ is given as:

$$C(\tau) = E\{\upsilon(t)\upsilon^T(t-\tau)\}$$

$$= E\left\{\big(He(t) + v(t)\big)\big(He(t-\tau) + v(t-\tau)\big)^T\right\}$$

$$= E\{He(t)e^T(t-\tau)H^T\} + E\{He(t)v^T(t-\tau)\} + E\{v(t)e^T(t-\tau)H^T\} + E\{v(t)v^T(t-\tau)\}$$

$$= He^{\tau F'}[P_1H^T - K_0R] + R\delta(\tau)\,. \quad F' = F - K_0H \qquad \tau > 0$$

If we take the Fourier transform of $C(\tau)$, the spectrum of υ is given as below:

$$S(\omega) = H(i\omega - F')^{-1}(P_1H^T - K_0R) + \left(HP_1 - RK_0{}^T\right)\left(-i\omega - F'^T\right)^{-1}H^T + R$$

Test of Optimality and the Estimation of Q and R:

In order to check the optimality of Kalman filter we can use $C(\tau)$ or $S(\omega)$ and then identify Q and R. The following estimates use the procedure methods given in [13].

P_1H^T and R can be calculated using $C(\tau)$ or $S(\omega)$ stated above and its method is analogous to that of discrete case. If the number of unknowns in Q is $n \times r$ or less, Q can be obtained using the equation (52).

$$\sum_{j=0}^{k-1}(-1)^j HF^jGQG^TF^{k-j}H^T \quad for\ k = 0.1.\ldots$$

which are the set of equations similar as (28).

If the number of unknowns in Q be more than $n \times r$, then K_{op} is obtained directly without the estimation of Q. The procedure is as follow:

$$K_1 = P_1 H^T R^{-1} \tag{47}$$

Defining P_2 as the error covariance corresponding to K_1, we get:

$$(F - K_1 H)(P_2 - P_1) + (P_2 - P_1)(F - K_1 H)^T - (K_1 - K_0) R (K_1 - K_0)^T = 0 \tag{48}$$

Same as what we concluded in the discrete format, we get:

$$P_2 < P_1$$

Similarly defining $K_2 = P_2 H^T R^{-1}$and P_3 as the error covariance corresponding to K_2 we get:

$$P_3 < P_2 < P_1$$

In this way, P is decreasing as we go along and as a result K converges to its optimal value, i.e. the sequence $K_0 . K_1 . K_2 . \ldots$ reaches to K_{op}.

Based on the discussed equations and their convergence, now an estimation of K_{op} represented by $\widehat{K}_{op}$is calculated as follow:

Defining:

$$\widehat{K}_1 = \widehat{P}_1 \widehat{H}^T \widehat{R}^{-1}$$

and replacing it to equation (48) results:

$$\left(F - \widehat{K}_1 H\right)\left(\hat{\delta} \widehat{P}_1\right) + \left(\hat{\delta} \widehat{P}_1\right)(F - \widehat{K}_1 H)^T - \left(\widehat{K}_1 - K_0\right) \widehat{R} \left(\widehat{K}_1 - K_0\right)^T = 0 \tag{49}$$

with: $\hat{\delta} \widehat{P}_1 = \widehat{P}_2 - \widehat{P}_1$

Then we get:

$$\widehat{P}_2 \widehat{H}^T = \widehat{P}_1 \widehat{H}^T + \hat{\delta} \widehat{P}_1 H^T \tag{50}$$

$$\widehat{K}_2 = \widehat{P}_2 \widehat{H}^T \widehat{R}^{-1} \tag{51}$$

and so on until the relative changes in $\widehat{K}$ become small.

The asymptotic convergence of these estimates since they are essentially similar to the discrete case. All the estimates obtained are asymptotically unbiased and consistent.

XI. Summary and Conclution

The problem of optimal filtering for a linear time invariant system with unknown Q (process noise covariance matrix) and R (measurement noise covariance matrix) is considered. Using the innovation property of an optimal filter, a statistical test is given. This test checks whether a particular filter is working optimally or not. In the case of non-optimal or in better words suboptimal filter, an identification scheme is given to obtain asymptotically unbiased and consistent estimates of Q and R. This scheme works in cases where the form of Q is known or the number of unknown parameters in Q is less than or equal to $n \times r$ (where n is the dimension of state vector and r is that of the measurement vector.). In other cases an alternative scheme is given, in which an estimation of the optimal filter gain is calculated directly and without knowing Q. A numerical example is given to illustrate the proposed algorithm and to show how useful this scheme is. In that example the values of this algorithm is shown both on the same batch of data and also 10 batches of it. In the end using the results stated for discrete system are extended in order to demonstrate the scheme in continuous systems as well.

References

[1] R. E. Kalman, "New methods and results in linear prediction and filtering theory," *Proc. Symp. on Engineering Applications of Random Function Theory and Probability*. New York: Wiley, 1961.

[2] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Trans.* ASME, J. Basic Engrg., ser.D, vol. 83, pp. 95-108, March 1961.

[3] H. Heffes, "The effects of erroneous models on the Kalman filter response," *IEEE Trans. Automatic Control* (Short Papers), vol.AC_11, pp.541-543, July 1966

[4] T. Nishimura, "Error bounds of continuous Kalman filters and the application to orbit determination problems," *IEEE Trans Automatic Control,* vol.AC_12, pp. 268_275, June 1967

[5] R. C. K. Lee, "Optimal estimation identification and control," MIT, Cambridge, Res. Mono 28,1964.

[6] D. T. McGill, "Optimal adaptive estimation of sampled stochastic process," Stanford Electronics Labs, Stanford, California, Tech.Rep.SEL_63143. (TR.6302_3), December 1963.

[7] J. S. Shellenberger, **"A** multivariance learning technique for improved dynamic system performance," *1967 Pro. NEC,* vol. 23, pp.146-151

[8] G. L. Smith, "Sequential estimation of observation error variances in a trajectory estimation problem," *AIAA J,* vol.5, pp. 1964-1970, November 1907

[9] R. L. Kashyap, “Maximum likelihood identification of stochastic linear systems,” Purdue University: Lafayette, Ind.,Tech. Rept. TR-EE 68-28, August 1968

[10] T. Kailath, “An innovations approach to least-squares estimation, pt. I: linear filtering in additive white noise,” *IEEE Trans. Automatic Control*, vol. AC-13, pp. 646-655, December 1968.

[11] J. J. Deyst., Jr., and C. F. Price, “Conditions for asymptotic stability of the discrete minimum variance linear estimator,” *IEEE Trans. Automatic Control* (Short Papers), vol. AC13, pp. 702-705, December 1968.

[12] F. C. Schweppe, “Evaluation of likelihood functions for Gaussian signals,” *IEEE Trans. Information Theory*, vol. IT-11,

[13] G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications*. San Francisco: Holdon Day Publ., 1968

[14] M. S. Bartlett., *An Introdustion to Stochastic Processes*. London: Cambridge University Press, 1962.

[15] E. Parxen, ”An approach to time series analysis,” *Ann. Math. Statist.*, vol. 32, no.4 December 1961.

[16] E. J. Hannan, *Time Series Analysis*. New York: Wiley, 1960.

[17] R. L. Anderson, “Distribution of the serial correlation coefficient,” *Ann. Math. Statist.*, vol. 13, 1942.

[18] G. S. Watson and J. Durbin, “Exact tests of serial correlation using noncircular statistics,’’ *Ann. Math. Statist.*, vol. 22, pp. 446-451, 1951.

[19] C. W. J. Granger. ‘‘A quick test for serial correlation suitable for use with non-stationary time series,” *Am. Statist. Assoc. J.*, pp.728-736, September 1963.