Approximation Approach to Linear Filtering Problem with Correlated Noise

Hong Son Hoang, Rémy Baraille

SHOM/HOM, 42 Avenue Gaspard Coriolis 31100 Toulouse, France
Tel : +33 (0)5 67 76 68 23, Fax : +33 (0)5 61 40 52 58,
Email: hhoang@shom.fr

Abstract—The (sub)-optimal solution of linear filtering problem with correlated noises is considered. The special recursive form of the class of filters and criteria for selecting the best estimator are the essential elements of the design method. The properties of the proposed filter are studied. In particular, for Markovian observation noise, the approximate filter becomes an optimal Gevers-Kailath filter subject to a special choice of the parameter in the class of given linear recursive filters.

Keywords—Linear dynamical system, filtering, minimum mean square filter, correlated noise

I. INTRODUCTION

Consider a standard linear filtering problem

\[ x_{k+1} = \Phi_k x_k + G_k w_k, \]
\[ z_{k+1} = H_{k+1} x_{k+1} + v_{k+1}, \quad k = 0, 1, 2, \ldots \tag{1} \]

where \( x_k \) is the \( n \)-dimensional system state at \( k \) instant, \( \Phi_k \) is the \( (n \times n) \) fundamental matrix, \( z_k \) is the \( p \)-dimensional observation vector, \( H_k \) is the \( (p \times n) \) observation matrix, \( w_k, v_k \) are the model and observation noises. The statistical characteristics of the entering random variables are given as

\[ E[x_0] = \bar{x}_0, \quad E[x_0 x_0^T] = M_0, \]
\[ E[w_k] = 0, \quad E[w_k w_k^T] = Q_k, \]
\[ E[v_k] = 0, \quad E[v_k v_k^T] = R_k, \]
\[ E[(x_0 - \bar{x}_0) w_k^T] = K_{w_0}(k), \quad E[(x_0 - \bar{x}_0) v_k^T] = K_{v_0}(k). \tag{2} \]

Denote by \( \hat{x}_k \) a minimum mean square (MMS) estimator for the state \( x_k \). The Kalman filter (KF) yields the MMS solution to this filtering problem with white and uncorrelated process and observation noises [10]. The extension of the KF to the systems with colored noises that are Markovian is studied on the basis of innovation process [6]. The filtering problems with correlated noises are widely encountered in engineering applications (data assimilation in meteorology and oceanography [4], GPS position time series [2], halftoning systems with blue noise [5], speech signal processing [13], navigation [14], guidance [3] ... For many practical applications, the assumption on Markovian noises is nevertheless not necessary. The filtering problem in the form (1)(2)(3)-(6) has been considered in [12]. Generally speaking, due to assumptions (3)-(6) the estimator \( \hat{x}_{k+1} \), written in recursive form, depends on \( \hat{x}_k \) and all the observations \( \{z_1, \ldots, z_{k+1}\} \). This dependence makes implementation of the optimal filter extremely difficult for large \( k \).

The present paper aims to overcome the mentioned above difficulty. The approach follows that reported in [8], with emphasis on the linear filtering problem considered in [12]: Given the system dynamics and observations contaminated by correlated noises, the task is to construct an algorithm providing an (sub)-optimal filtered estimate of high quality. Concretely, according to [8], the class of filters \( \{\hat{x}_k(n_k)\} \) with \( n_k \leq k \) - some positive integer number, is introduced in a way such that \( \hat{x}_{k+1}(n_{k+1}) \) depends on \( \hat{x}_k(n_k) \) and \( n_{k+1} \) latest observations. One important requirement to the algorithm will be that the produced estimate \( \hat{x}_k(n_{k+1}) \) will be truly MMS if \( \hat{x}_k(n_k) = \hat{x}_k \) and \( n_{k+1} = k + 1 \). Such algorithm has a merit to be studied in more detail, noticing in practice the time correlation generally becomes weaker as the time difference increases. More importantly, for a particular case of the Markovian observation noise with memory \( m \), the sub-optimal filter becomes truly MMS in the class of filters being linear functions of the last estimate \( \hat{x}_k \) and \( m + 1 \) last observations. The case of Markovian noise sequence with memory \( m = 1 \) will be studied in detail in section 6.

The paper is organized as follows. In section 2, for the time-invariant system state, the main theoretical results on MMS filter optimal in a given class of linear filters are presented. These results will be extended to the general time-varying system state in section 3. The properties of the obtained filter are studied in section 4. Conditions for equivalence of two estimators obtained on the basis of the last estimator and two different numbers of latest observations are given in section 5. Application of the theoretical results to the design of the MMS filter subject to the Markovian noise sequence with memory \( m = 1 \) is considered in section 6. The conclusions are given in section 7.

II. PRELIMINARIES RESULTS : TIME-INVARIANT SYSTEM STATE

For simplifying the presentation, first consider the filtering problem (1),(3)-(5) under assumptions
\[ \Phi_k = I, \quad G_k = 0, \text{i.e., } x_{k+1} = x_k = x. \]  

\[ (7) \]

The system state is then time-invariant. Throughout this paper, \( I \) denotes a unit matrix of the appropriate dimension. Introduce the notations

\[ z_k^i = (z_i^1, ..., z_i^\nu)^T, \quad z_k = z_{k, k} \geq i, \]

\[ (8) \]

\[ H_k^i = (H_i^1, ..., H_i^{\nu_k})^T, \quad H_k = H_{k, k} \geq i, \]

\[ (9) \]

\[ v_k^i = (v_i^1, ..., v_i^{\nu_k})^T, \quad v_k = v_{k, k} \geq i, \]

\[ (10) \]

\[ V_k^i = E(v_i^1 v_i^{T}). \]

\[ (11) \]

Let \( \tilde{x}_k(n_k) \) be a sequence of estimators for \( x \) given by \( \{z_1, ..., z_k\} \) such that each next estimator \( \tilde{x}_{k+1}(n_{k+1}) \) is a linear function of \( \tilde{x}_k(n_k) \) and \( n_{k+1} \) last observations. According to notations in (8), we have

\[ \tilde{x}_{k+1}(n_{k+1}) = \delta_k [\tilde{x}_k(n_k)] + z_{k+2-n_k+1} \]

\[ \gamma_k = \delta_{k,0}[\xi_k \xi_{k,2}] + z_{k, k}, \]

\[ \xi_k = \tilde{x}_k(n_k). \]

For simplicity and without generality, assume that the sequence \( \{\tilde{x}_k(n_k)\} \) is unbiased. Then one can set \( \gamma_k = 0 \). In what follows we will use the following notation for the sequence \( \{\tilde{x}_k(n_k)\} \)

\[ \tilde{x}_{k+1}(n_{k+1}) = A_k \tilde{x}_k(n_k) + B_k \tilde{x}_{k+2-n_k+1}. \]

\[ (12) \]

where \( A_k, B_k \) are matrices of appropriate dimensions. Denote by \( X_k(n_k) \) the class of all unbiased estimators having the structure (12), where \( \tilde{x}_k(n_k) \) is unbiased estimator too.

**Definition 1.** We shall call \( \tilde{x}_{k+1} \) an optimal MMS estimator in the class \( X_k(n_k) \) if it satisfies

\[ (i) \quad E[\tilde{x}_{k+1}] = E(x); \]

\[ (ii) \quad \tilde{x}_{k+1} = \arg \min_{x' \in X_k(n_k)} E(x') \]

\[ J(x') = E(x' - x^2) \]

\[ \] for which \( X_k(n_k) = \{x' \in X_k(n_k) : E(x') = E(x)\} \), \( \text{tr} (.) \) denotes the trace operator.

In the present paper, for simplicity, we assume the existence of all figured inverse matrices.

**Lemma 1.** Let \( \tilde{x}_k(n_k) = \tilde{x}, \) \( P_k \) be its error covariance matrix (ECM). Then \( \tilde{x}_{k+1} \) is defined by

\[ \tilde{x}_{k+1}(n_{k+1}) = \tilde{x} + K_k^i [z_{k+2-n_k+1} - H_k^i x], \]

\[ (13) \]

\[ z_{k+1}(n_{k+1}) = z_{k+1} - \Sigma_{21} \Sigma_{11}^{-1} K_k^i x, \]

\[ (14) \]

\[ H_{k+1}^i = H_{k+1} - \Sigma_{21} \Sigma_{12}^{-1} K_k^i x, \]

\[ (15) \]

\[ K_{k+1} = (P_k H_k^i)^{-1} + N_{k+1} \Sigma_{22}^{-1}, \]

\[ (16) \]

\[ \Sigma_{22} = \Sigma_{22} - \Sigma_{21} \Sigma_{12}^{-1} \Sigma_{12}, \]

\[ (17) \]

\[ \Sigma_{11} = V_k^{k+2-n_k+1} - H_{k+2-n_k+1} E_k^{k+2-n_k+1}, \]

\[ (18) \]

\[ \Sigma_{12} = \Sigma_{12}^{T}, \]

\[ (19) \]

\[ \Sigma_{12} = E_k^{k+2-n_k+1} N_{k+1}, \]

\[ (20) \]

\[ N_{k+1} = x_{k+1} - x_{k+1}^T, \]

\[ (21) \]

\[ E_k^{k+2-n_k+1} = E[(\hat{x}_k - x_{k+1}) v_{k+1}^T], \]

\[ (22) \]

\[ K_k^{k+2-n_k+1} = E[v_k^{k+2-n_k+1} v_{k+1}^T], \]

\[ (23) \]

\[ P_k = E[(\hat{x}_k - x_{k+1}) (\hat{x}_k - x_{k+1})^T], \]

\[ (24) \]

\[ P_{k+1}(n_{k+1}) = E[(\hat{x}_k - x_{k+1}) (\hat{x}_k - x_{k+1})^T] = P_k - K_k^{k+1} P_k H_{k+1}^i (n_{k+1} - N_{k+1})^T. \]

\[ (25) \]

**Proof:** The proof is similar to that presented in [8]: From the requirement (i) on unbiasedness of \( \tilde{x}_{k+1}(n_k) \) we have \( A_k = I - B_k H_{k+1}^i \). Substituting \( A_k \) into (12) and taking the gradient of \( J(.) \) with respect to \( B_k \) leads to the equation for finding \( B_k \). Thus,

\[ A_k = I - B_k H_{k+1}^{k+2-n_k+1}, \]

\[ (26) \]

where \( \Sigma_1, \Sigma_3 \) are defined in (29). We have the ECM \( P_{k+1} \)

\[ P_{k+1}(n_{k+1}) = B_k \Sigma_4 B_k^T + B_k \Sigma_2 + \Sigma_3 B_k^T + \Sigma_4, \]

\[ (27) \]

\[ \Sigma_1 = H_{k+2-n_k+1} P_k (H_{k+2-n_k+1})^T + \]

\[ V_{k+1}^{k+2-n_k+1} + \Delta \Sigma_4, \]

\[ \Delta \Sigma_4 = -H_{k+2-n_k+1} B_{k+2-n_k+1} - \]

\[ (H_{k+2-n_k+1} B_{k+2-n_k+1})^T, \]

\[ (28) \]

\[ \Sigma_2 = -H_{k+1}^{k+2-n_k+1} P_k + (E_{k+1}^{k+2-n_k+1})^T, \]

\[ \Sigma_3 = \Sigma_2^T, \Sigma_4 = P_k, \]

\[ (29) \]

\[ E_{k+1}^{k+2-n_k+1} \]

is defined by (23). Using \( A_k \) from (27) the estimator \( \tilde{x}_{k+1} \) can be rewritten as

\[ \tilde{x}_{k+1} = \tilde{x} + B_k [z_{k+2-n_k+1} - H_{k+2-n_k+1} \tilde{x}], \]

\[ (30) \]

Compute the matrix \( B_k \). Since \( z_k^i = H_k^i x + v_k^i \), the MMS estimate \( \tilde{x}_k = T_k z_k^i, T_k = P_k H_k^i V_k^{k+1}, P_k = (H_k^i V_k^{k+1} H_k^T)^{-1}, V_k = E(v_k^i v_k^{T}) \) is an unbiased estimate, \( T_k H_k = I \). Hence

\[ E_{k+1}^{k+2-n_k+1} = E[(T_k (H_k^i x + v_k^i) - x) (v_k^{k+2-n_k+1})^T] = \]

\[ T_k E(v_k^i (v_k^{k+2-n_k+1})^T). \]

\[ (31) \]

Let

\[ V_k^i = [v_k^T(1), ..., v_k^T(k)]^T, \]

\[ V_i = E[v_k^i v_k^{T}], \]

\[ (32) \]

Then we have \( V_i (V_j) = \delta_{ij} \) where \( \delta_{ij} \) is the Kronecker symbol. But \( E(v_k^i v_k^{T}) \) hence

\[ ... \]

\[ \]
\[ P_k^{k+2-n_k+1} = T_k E[(V_{k+2-n_k+1})^T] = (V_{k+2-n_k+1}, ..., V_{k})^T [V_k(1), ..., V_k(k)] \]
\[ H_k^T P_k T = \{ [H_k^{T+1-k}, H_k^{T+2-k}] \} P_k \]
\[ = P_k H_k^{T+2-k+1} T \]

Taking into account (29)(31)(32) one can write
\[ B_k = -\Sigma_1 \Psi_1^{-1} \]
\[ = (E_k^{k+2-n_k+1} - P_k H_k^{T+2-k+1} T) \Psi_1^{-1} \]

or
\[ B_k = (P_k H_k^{T+2-k+1} - N_{k+1}) \Psi_1^{-1} \]

here \( N_{k+1} \) is defined as in (22) and for
\[ \Psi_1 = \left\{ \begin{array}{c} \Sigma_1 \\ \Sigma_2 \\ \Sigma_3 \\ \Sigma_4 \end{array} \right\} \]
\[ \Psi_1^{-1} = \left\{ \begin{array}{c} \Sigma_1 \\ \Sigma_2 \\ \Sigma_3 \\ \Sigma_4 \end{array} \right\} \]

the lemma on Inversion of block matrix [10] yields
\[ \Psi_2 = \left\{ \begin{array}{c} \Sigma_2 \Psi_1^{-1} \Sigma_1 \Psi_1^{-1} \end{array} \right\} \]

which shows (17). Substituting (33) into (30) yields (13) with \( K_{k+1} \) defined in (16). The formula (26) is obtained by using (27),(28) and (33).

To show (18)-(20), noticing from (29) that \( \Psi_1 \) can be written as the ECM of the following random vector
\[ \Psi_1 = E[\xi^T \xi] \]
\[ \xi = H_k^{T+2-k+1} \Psi_{k+1} \]

Represent \( \xi = (\xi_1, \xi_2)^T \), from (34) one sees that
\[ \Sigma_{11} = V_{k+2-k+1} + H_k^{T+2-k+1} P_k (H_k^{T+2-k+1})^T - H_k^{T+2-k+1} E_k^{T+2-k+1} E_k^{T+2-k+1} \]
\[ \Sigma_{22} = R_k^{*+1} + H_k^{T+2-k+1} P_k H_k^{T+2-k+1} - H_k^{T+2-k+1} E_k^{T+2-k+1} \]
\[ \Sigma_{12} = [H_k^{T+2-k+1} P_k H_k^{T+2-k+1} - H_k^{T+2-k+1} E_k^{T+2-k+1}] \]

\( \Sigma_{12} \) is defined by (24). These formulas imply (18)-(20) noticing from (32) that \( \Psi_1, \Sigma_1 \) can be simplified. ■

**Comment 1.** As shown by [8], when \( \Sigma_1 \) in (27) is singular, the matrix \( B_k \) is defined by \( B_k = -\Sigma_1 \Psi_1^{-1} \). The solution \( \tilde{x}_{k+1} \) then exists and is unique (almost surely). The uniqueness of \( \tilde{x}_{k+1} \) for non-singular \( \Sigma_1 \) follows automatically.

**Comment 2.** The estimate \( \tilde{x}_{k+1} \) can be obtained in the following way [9]: Interpreting \( z^*: = \tilde{x}_k \) as the “observation” available after arriving \( z_{k+1} \),

\[ z^* = x + \varepsilon_k, E[\varepsilon_k] = 0, E[\varepsilon_k \varepsilon_k^T] = P_k, z^* = \tilde{x}_k \]

and introducing \( \tilde{z} = (z^T, z_{k+2-n_k+1}^T, ..., z_{k+1}^T)^T \), one has the following system of observations

\[ \tilde{z} = \tilde{H} \tilde{x} + \tilde{v}, \]

Then with probability 1 the estimator \( \tilde{x}_{k+1} \) in Lemma 1 is equal to

\[ \tilde{x}_{k+1} = (H^T \tilde{V}^{-1} H)^{-1} H^T \tilde{V}^{-1} \tilde{z}. \]

Really, the estimator (37) is a linear function of \( \tilde{x}_k \) and \( z_{k+1} \). From Theorem 6.1.11 of [1] it is the BLUE (unbiased and of minimum variance). Thus (37) must be also a MMS estimator by Definition 1.

**Theorem 1.** Let \( \{x_k(n_k)\} \) be a sequence of unbiased estimators for the unknown vector \( x \) such that each estimator is obtained on the basis of the previous one and the \( n_k \) latest observations. Let these estimators be MMS according to Definition 1 subject to \( n_{k+1} \leq n_k + 1 \). Then

\[ \tilde{x}_{k+1}(1) = \tilde{x}_k(1) + K_{k+1}^* (z_{k+1}^* - H_{k+1}^* \tilde{x}_k(1)) \]

where \( z_{k+1}^* = H_{k+1}^* x_k + v_k(n_k), \tilde{V}_k(n_k) = E[v_k(n_k) v_k(n_k)^T] \),

\[ z_k(n_k) = [\tilde{x}_k(n_k-1)]^T, H_k(n_k) = [H_k^T(n_{k-1})]^T \]

and from \( n_{k+1} \leq n + 1 \) the matrix \( E_k^{T+2-k+1} \) is simplified to the form (32). Thus the condition \( k_{k+1} \leq n + 1 \) is introduced only for having the compact formulas (14)-(26).

As will be seen later, the case \( k_{k+1} = 2 \) is of special interest and it is formulated in the form of the following Corollary

**Corollary 1.** Let in Theorem 1, \( n_{k+1} = 2 \). Then the following relations hold for the estimator \( \tilde{x}_{k+1}(2) \) satisfying Definition 1,

\[ \tilde{x}_{k+1}(2) = \tilde{x}_k(n_k) + K_{k+1}(2) (z_{k+1}^* - H_{k+1}^* \tilde{x}_k(n_k)) \]

\[ z_{k+1} = z_{k+1}^* - \Sigma_1 \Sigma_1^{-1} \tilde{z}_k \]

\[ H_{k+1} = H_{k+1}^* - \Sigma_2 \Sigma_1^{-1} \tilde{z}_k \]

\[ K_{k+1}(2) = [P_k(n_k) H_k^T(n_k) \Sigma_2] \]

\[ \Sigma_2 = [\Sigma_2 - \Sigma_1 \Sigma_1^{-1} \Sigma_1]^{-1} \]

\[ z_k(n_k) = [\tilde{x}_k(n_k-1)]^T, H_k(n_k) = [H_k^T(n_{k-1})]^T \]

\[ z_{k+1} = z_{k+1}^* - \Sigma_1 \Sigma_1^{-1} \tilde{z}_k \]

\[ H_{k+1} = H_{k+1}^* - \Sigma_2 \Sigma_1^{-1} \tilde{z}_k \]

\[ K_{k+1}(2) = [P_k(n_k) H_k^T(n_k) \Sigma_2] \]

\[ \Sigma_2 = [\Sigma_2 - \Sigma_1 \Sigma_1^{-1} \Sigma_1]^{-1} \]

\[ z_k(n_k) = [\tilde{x}_k(n_k-1)]^T, H_k(n_k) = [H_k^T(n_{k-1})]^T \]

where \( n_{k+1}, \Sigma_2 \) are defined in Theorem 1.

Mention that the structure of the filter (38)-(43) is similar to that of the Gevers-Kailath filter [6].

**Corollary 2.** Let \( \tilde{x}_k(n_k) = \tilde{x}_k \). For \( n_{k+1} = k + 1 \), Theorem 1 yields \( \tilde{x}_{k+1} = \tilde{x}_{k+1} \) and the following equality holds
\[ \Delta \Sigma_{22} := -\Sigma_{12}^{T} \Sigma_{12}^{-1} \Sigma_{12} = L_{1} - L_{2}. \]
\[ L_{1} := \tilde{K}_{k}^{T} V_{k}^{-1} \cdot H_{k}^{T} P_{k} \cdot H_{k}^{T} V_{k}^{-1} \cdot \tilde{K}_{k}, \quad L_{2} := \tilde{K}_{k}^{T} V_{k}^{-1} \cdot \tilde{K}_{k}. \]

(44)

Here \( \tilde{K}_{k} = \tilde{K}_{k}^{1} \) and \( \tilde{K}_{k}^{1} \) is defined by (24). The equality (44) will be used in the further.

### III. TIME-VARYING SYSTEM STATE

Consider the filtering problem in its general formulation (1)(3)-(5). A natural way to generalize (12) in this case is to introduce the class of recursive filters

\[ \tilde{x}_{k+1}(n_{k+1}) = A_{k} \tilde{x}_{k+1/k}(n_{k+1}) + B_{k} z_{k+1-n_{k+1}}, \]
\[ \tilde{x}_{k+1/k}(n_{k+1}) := \Phi_{k} \tilde{x}_{k}(n_{k}) \]  

(45)

where \( \{ \tilde{x}_{k}(n_{k}) \} \) is a sequence of filtered estimates for the system state \( x_{k}, k = 1, 2, \ldots \). The results to be presented below can be established by the same technique as done in section 2.1.

Introduce the notations: Let \( \Phi(i, j) \) be the transition matrix for the system (1). Then [10]

\[ x_{j} = \Phi(j, i) x_{i} - \sum_{i=0}^{j-1} \Phi(i, j+1) w_{i}, \quad i < j, \]
\[ H_{k+1}^{1} = (H^{T}(1, k+1), H^{T}(2, k+1), \ldots), \]
\[ H^{T}(k+1, k+1))^{T}, \quad H_{j}(x, j) = H_{j}(\Phi(i, j), i), \]
\[ w(i, k+1) = v_{i} - \sum_{i=0}^{k} H_{i}(\Phi(i, k+1) v_{i}, \sum_{i=0}^{k} v_{i} = 0, \]
\[ w_{k+1}^{1} = [w^{T}(1, k+1), \ldots, w^{T}(k+1, k+1)]^{T} = \eta_{k}^{1} \eta_{k+1}^{1}, \]
\[ E[w_{k+1}^{1} w_{k+1}^{T}] = W_{k+1}, \]
\[ E[\eta_{k}^{1} \eta_{k+1}^{T}] = 0, \quad \eta_{k}^{1} = 0, \quad E[\eta_{k}^{1} \eta_{k+1}^{T}] = \Lambda_{k}. \]

(49)  

Using the notations above we have

\[ z_{j} = H_{j}(j, k+1) x_{k+1} + w(j, k+1), \quad j = 1, 2, \ldots, k+1. \]
\[ z_{k+2-n_{k+1}} = \tilde{H}_{k+1} \tilde{x}_{k+1} + w_{k+1-n_{k+1}}, \]
\[ z_{k+1}^{1} = \tilde{H}_{k+1} \tilde{x}_{k+1} + w_{k+1}^{1}, \]
\[ \tilde{H}_{k+1}^{1} = [H_{k}^{1} H_{k}^{1} T], \quad \tilde{H}_{k+1}^{1} = H_{k+1}^{T} \Phi(k, k+1), \]
\[ \eta_{k}^{1} = \tilde{u}_{k+1}^{1} - H_{k}^{T} \Gamma_{k} u_{k}. \]

(52)  

In the further, according to [11] we will refer to the model (53) as of high initial uncertainty if the information on \( x_{k+1} \) is contained only in the observation vector \( \tilde{x}_{k+1}^{1} \) (equivalently to assuming \( M_{0} = \infty \) - there is no priori information on \( x_{k+1} \)). For the case (3)-(5)(7) are given, from (1) \( x_{k+1} = \Phi(k+1, 0) x_{0} + \sum_{i=0}^{k} \Phi(k+1, j+1) w_{j} \)

\[ \tilde{x}_{k+1/0} = \Phi(k+1, 0) x_{0} + \sum_{i=0}^{k} \Phi(k+1, j+1) w_{j} \]

(53)

This information can be represented by the additional “observation” \( z^{*} := \tilde{x}_{k+1/0} \) (Comment 2) and instead of (53) we have the following model

\[ \tilde{z}_{k+1/0}^{1} = \tilde{H}_{k+1}^{1} \tilde{x}_{k+1/0} + v_{k+1/0}, \]
\[ \tilde{z}_{k+1/0} = \begin{pmatrix} \ldots \tilde{z}_{k+1} \ldots \end{pmatrix}, \quad \tilde{H}_{k+1}^{1} = \begin{pmatrix} I \ldots \end{pmatrix}, \quad \tilde{v}_{k+1/0} = \begin{pmatrix} \ldots \tilde{v}_{k+1} \ldots \end{pmatrix}. \]

(56)

As the model (56) includes the information (3)-(5) in the form of \( z^{*} \), it can be considered as that of high initial uncertainty. Later on, for simplicity we shall derive the filtering algorithms for the vector \( x_{k+1} \) in the model (53) remembering that the similar results can be deduced for \( x_{k+1} \) in the model (56).

### Application of Theorem 1 to the model (53) leads to the following

**Theorem 2.** Consider the class of recursive filters (45) and let \( \{ \tilde{x}_{k+1/n_{k+1}} \} \) be a sequence of unbiased estimators for \( x_{k+1} \) such that each estimator \( \tilde{x}_{k+1} = \tilde{x}_{k+1/n_{k+1}} \) is a function of the previous \( \tilde{x}_{k} \) and \( n_{k+1} \) last observations. Assume that these estimators are optimal in the sense of Definition 1. Then we have

\[ \tilde{x}_{k+1/n_{k+1}} = \Phi_{k} \tilde{x}_{k+1} + B_{k} z_{k+1-n_{k+1}}, \]

\( \tilde{x}_{k+1} \) can be determined as done in the proof of Lemma 1.

**Corollary 3.** Let \( \tilde{x}_{k+1/k} = \tilde{x}_{k+1/1} = \Phi_{k} \tilde{x}_{k+1} \). Then

\[ \tilde{x}_{k+1/n_{k+1}} = \tilde{x}_{k+1/k} + K_{k+1} \tilde{z}_{k+1}^{*} - \tilde{H}_{k+1}^{1} \tilde{x}_{k+1/k}. \]

\[ \tilde{x}_{k+1/k} = \begin{pmatrix} \tilde{x}_{k+1/1} \ldots \tilde{x}_{k+1} \ldots \end{pmatrix}, \quad \tilde{H}_{k+1}^{1} = \begin{pmatrix} \tilde{H}_{k+1}^{1} \tilde{H}_{k+1}^{1} \ldots \tilde{H}_{k+1}^{1} \end{pmatrix}, \]

\[ z_{k+1}^{*} = z_{k+1}^{1} + z_{k+1-n_{k+1}} - \tilde{H}_{k+1}^{1} \tilde{x}_{k+1/n_{k+1}}, \]

\[ \tilde{K}_{k+1} = (M_{k+1} \tilde{H}_{k+1}^{1} - N_{k+1}) \tilde{\Sigma}_{k+1}^{22}, \]

\[ \tilde{\Sigma}_{k+1}^{22} = \tilde{\Sigma}_{k+1}^{22} - \tilde{E}_{k+1} \tilde{E}_{k+1}^{T}, \]

\[ R_{k+1} + H_{k+1} M_{k+1} H_{k+1}^{T} - H_{k+1} N_{k+1} + \tilde{H}_{k+1}^{1} \tilde{x}_{k+1/k}, \]

\[ \tilde{H}_{k+1}^{1} \tilde{x}_{k+1/k} = E[c_{k+1} \tilde{c}_{k+1}^{T}], \quad E[c_{k+1}^{2}] = \tilde{x}_{k+1/k} - \tilde{x}_{k+1}, \]

\[ \tilde{K}_{k+1} = (M_{k+1} \tilde{H}_{k+1}^{1} - N_{k+1}) \tilde{\Sigma}_{k+1}^{22}, \]

\[ \tilde{\Sigma}_{k+1}^{22} = \tilde{\Sigma}_{k+1}^{22} - \tilde{E}_{k+1} \tilde{E}_{k+1}^{T}, \]

(55)

**Corollary 4** (Case \( n_{k+1} = 2 \)). Under the conditions of Corollary 3, for \( n_{k+1} = 2 \)

\[ \tilde{x}_{k+1/2} = \tilde{x}_{k+1/1} + K_{k+1} \tilde{z}_{k+1}^{*} - \tilde{H}_{k+1}^{1} \tilde{x}_{k+1/1}, \]

\[ \tilde{z}_{k+1}^{*} = z_{k+1} + \tilde{E}[\tilde{z}_{k+1/0}] \tilde{E}[\tilde{z}_{k+1/0}]^{T}, \]

\[ \tilde{H}_{k+1}^{1} = \tilde{H}_{k+1}^{1} - \tilde{E}[\tilde{z}_{k+1/0}] \tilde{E}[\tilde{z}_{k+1/0}]^{T}, \]

\[ K_{k+1} = (M_{k+1} \tilde{H}_{k+1}^{1} - N_{k+1}) \tilde{\Sigma}_{k+1}^{22}. \]
in the recursive filter with the estimates depending on \( m_{k+1} \) latest observations. Let

\[
\Sigma(m_{k+1}) := \Sigma^{-1}(m_{k+1}).
\]

**Lemma 3.** For all \( n_{k+1} \geq m_{k+1} \),

\[
\Sigma(n_{k+1}) - \Sigma(m_{k+1}) \leq 0. \quad (60)
\]

**Proof:**

It is not hard to see that one can represent

\[
\Sigma_{21}(n_{k+1}) \Sigma_{11}^{-1}(m_{k+1}) \Sigma_{12}(m_{k+1}) = \\
\Sigma_{21}(n_{k+1}) \Sigma_{11}^{-1}(m_{k+1}) \Sigma_{22}(n_{k+1}). \quad (61)
\]

\[
\Sigma_{11}^{n_{k+1}}(m_{k+1}) = \left\{ \begin{array}{c}
0 \\
0 \\
\Sigma_{11}^{-1}(m_{k+1})
\end{array} \right\}. \quad (62)
\]

The obtained expressions (61)(62) allow us to present the difference \( \Sigma(m_{k+1}) - \Sigma(n_{k+1}) \) in the form

\[
\Delta \Sigma_{k+1}(n; m) = \Sigma(n_{k+1}) - \Sigma(m_{k+1}) = \\
\Sigma_{12}^{m_{k+1}}(n_{k+1}) A \Sigma_{12}(n_{k+1}), \quad (63)
\]

\[
A := \Sigma_{11}^{-1}(n_{k+1}) - \Sigma_{11}^{n_{k+1}}(m_{k+1}). \quad (64)
\]

Denote by \( \tilde{\Sigma}_{11}(n_{k+1}) \) the inverse matrix for \( \Sigma_{11}(n_{k+1}) \),

\[
\Sigma_{11}^{-1}(n_{k+1}) = \Sigma_{11}(n_{k+1}) = \left\{ \begin{array}{cccc}
\Sigma_1 \\
\Sigma_2 \\
\Sigma_3 \\
\Sigma_4
\end{array} \right\}. \quad (65)
\]

The formulas (65)(62) imply

\[
A = \left\{ \begin{array}{c}
\Sigma_1 \\
\Sigma_2 \\
\Sigma_3 \\
\Sigma_4 - \Sigma_{11}^{-1}(m_{k+1})
\end{array} \right\}. \quad (66)
\]

The Lemma will be proven if we can show that \( A \geq 0 \). According to Lemma 2, we need to establish

(a) \( \Sigma_1 \geq 0 \),

(b) \( \Sigma_2 \Sigma_4 + \Sigma_2 = \Sigma_2 \),

(c) \( [\Sigma_4 - \Sigma_{11}^{-1}(m_{k+1})] - \Sigma_{11}^{T} \Sigma_{11} \Sigma_{22} \geq 0 \).

First mention that from the existence of \( \Sigma_{11}^{-1}(m_{k+1}) \) we have \( \Sigma_{11}(n_{k+1}) > 0 \). The structure (34) implies then \( \Sigma_{11}(n_{k+1}) > 0 \) hence \( \Sigma_{11}^{-1}(n_{k+1}) > 0 \). This fact and (65) prove \( \Sigma > 0 \) and we have (a), (b). It remains to show (c). In fact the left-hand side of (c) is equal to 0. First mention that by construction,

\[
\Sigma_{11}(n_{k+1}) = (S_{ij})^2_{i,j=1}
\]

where \( S_{ij} \) are block matrices of appropriate dimension, with \( S_{22} = \Sigma_{11}(m_{k+1}) \). Now, as \( \Sigma_{11}^{-1}(m_{k+1}) = \Sigma_{11}(n_{k+1}) \), according to Lemma on inversion of block matrix [10] we have

\[
S_{22} = \Sigma_{11}(m_{k+1}) = [\Sigma_4 - \Sigma_{11}^{T} \Sigma_{11} \Sigma_{22}]^{-1}, \quad \text{or} \quad \Sigma_{11}^{-1}(m_{k+1}) = \\
\Sigma_4 - \Sigma_{11}^{T} \Sigma_{11} \Sigma_2 \Sigma_2^T \Sigma_{11}^{-1} \Sigma_4 - \Sigma_{11}^{T} \Sigma_{11} \Sigma_2 = 0. \quad \text{Thus all three conditions (a)-(c) hold.}
\]

The property 2 follows immediately from the formulas similar to (26)(16)(17) and Lemma 3.

Analogously one can establish the properties 1-2 for the filter in section 3.
V. ON EQUIVALENCE OF TWO ESTIMATORS BASED ON DIFFERENT NUMBERS OF LAST OBSERVATIONS

Consider two estimators \( \hat{x}_{k+1}(n_k+1) \) and \( \hat{x}_{k+1}(m_k+1) \) derived according to Theorem 1 on the basis of \( n_k+1 \) and \( m_k+1 \) last observations respectively, with \( n_k+1 > m_k+1 \).

**Definition 2.** Two estimators \( \hat{x}_{k+1}(n_k+1) \) and \( \hat{x}_{k+1}(m_k+1) \) are called equivalent if the following equality holds:

\[
\text{tr} P_{k+1}(n_k+1) = \text{tr} P_{k+1}(m_k+1). \tag{66}
\]

The Definition 2 is correct since the solution of the problem in Theorem 1 (Definition 1) is unique (see Comment 1).

In what follows a new definition, equivalent to Definition 2, will be introduced. This new definition allows us to easier examine the equivalence of two estimators. First we need some preliminary results.

**Lemma 4.** Let the matrix \( C = A - B \geq 0 \) be symmetric. Then

\[
\text{tr} C = 0 \iff A = B.
\]

**Proof:** It is well known if \( S \) is an orthonormal matrix then

\[
\text{tr}(S^T CS) = \text{tr}(C).
\]

Let \( S \) be orthonormal matrix diagonalizing \( C \). Then \( \text{tr}(C) = \text{tr}(S^T CS) = \text{tr}(\Lambda) = \sum_{i=1}^{n} \lambda_i(C) \) where \( \lambda_i(C) \) are the eigenvalues of \( C \), \( \lambda_i(C) \geq 0 \), \( n \) is the dimension of \( C \). Suppose \( \text{tr}(C) = 0 \). Then \( \sum_{i=1}^{n} \lambda_i(C) = 0 \) or \( \lambda_i(C) = 0 \), \( \forall i = 1, ..., n \). The last means that \( C = 0 \) since the number of non-zero eigenvalues of \( C \) is equal to the rank of \( C \). Hence \( A - B = 0 \) or \( A = B \).

The inverse implication is trivial.

According to Property 2, for \( n_k+1 > m_k+1 \) we have

\[
\text{tr}[P_{k+1}(n_k+1)] \leq \text{tr}[P_{k+1}(m_k+1)] \quad \text{or} \quad P_{k+1}(m_k+1) - P_{k+1}(n_k+1) \geq 0.
\]

Taking into account Lemma 4 and the properties of \( P_{k+1}(n_k+1) \), Definition 2 can be replaced by the following

**Definition 3.** Two estimators \( \hat{x}_{k+1}(n_k+1) \) and \( \hat{x}_{k+1}(m_k+1) \) are called equivalent if the following equality holds:

\[
P_{k+1}(n_k+1) = P_{k+1}(m_k+1). \tag{67}
\]

**Lemma 5.** Let \( A - B \geq 0 \) be symmetric matrix. Then

\[
C^T (A - B) C = 0 \iff (A - B) C = 0.
\]

**Proof:** According to the definition of a symmetric non-negative definitive matrix \( D \) there exists a matrix \( D \) such that \( A - B = D^T D \). But then \( C^T D^T D C = 0 \) and this takes place iff \( DC = 0 \). We have then \( D^T DC = 0 \) or \( (A - B) C = 0 \). The inverse implication is trivial since from \( D^T DC = 0 \) it follows \( C^T D^T DC = 0 \).

**Proposition 1.** Two estimators \( \hat{x}_{k+1}(n_k+1) \) and \( \hat{x}_{k+1}(m_k+1) \), obtained from Corollary 2, are equivalent iff

\[
\Delta \Sigma_{k+1}(n; m) \Sigma^{-1}(m_k+1) [P_{k+1}(n_k) H_{k+1}^T - N_{k+1}]^T = 0 \tag{68}
\]

where \( \Delta \Sigma_{k+1}(n; m) \) is defined by (63)-(64).

**Proof:** Inserting \( P_{k+1}(n_k+1), P_{k+1}(m_k+1) \) into (67), from (26) it follows

\[
\begin{align*}
A[\Sigma^{-1}(n_k+1) - \Sigma^{-1}(m_k+1)]A^T &= 0, \\
A &:= P_{k+1}(n_k) H_{k+1}^T - N_{k+1}
\end{align*}
\]

According to Lemma 3, \( \Sigma(n_k+1) \leq \Sigma(m_k+1) \) hence \( \Sigma^{-1}(n_k+1) - \Sigma^{-1}(m_k+1) \geq 0 \). Using Lemma 5 one can see that

\[
[\Sigma^{-1}(n_k+1) - \Sigma^{-1}(m_k+1)]A^T = 0. \tag{69}
\]

As

\[
\Sigma^{-1}(n_k+1) - \Sigma^{-1}(m_k+1) = -\Sigma^{-1}(m_k+1) \Delta \Sigma_{k+1}(n; m) \Sigma^{-1}(m_k+1),
\]

the condition (68) follows from substituting the last equality into (69).

**Proposition 2.** Two estimators \( \hat{x}_{k+1}(n_k+1) \) and \( \hat{x}_{k+1}(m_k+1) \), obtained from Corollary 3, are equivalent iff

\[
\Delta \Sigma_{k+1}(n; m) \Sigma^{-1}(m_k+1) [M_{k+1} H_{k+1}^T - N_{k+1}]^T = 0 \tag{70}
\]

where \( \Delta \Sigma_{k+1}(n; m) \) are defined as in Proposition 1, \( M_{k+1}, N_{k+1} \) are computed from Proposition 1 too.

VI. APPLICATIONS

A. Optimal in MMS filtering with the Markovian observation noise

Consider the filtering problem (1)(2) with the following conditions

\[
Q_{ij} = Q_i \delta_{ij}, K_{xv}(i) = 0, K_{xv}(i) = 0, \forall i, j, \tag{71}
\]

where \( \nu_0 \) is uncorrelated with \( \{w_i\}, \{x_i\}, x_0 \) is a random vector, \( \{\xi_i\} \) is a white random sequence with

\[
E(\nu_0) = 0, E(\nu_0^T \nu_0) = R_0, \tag{73}
\]

\[
E(\xi_0) = 0, E(\xi_0^T \xi_0) = \Xi. \tag{74}
\]

The filtering problem (1)(2)(71)-(74) is studied in [6]. On the basis of the results in section 3 we will derive here the solution to the filtering problem (1)(2)(71)-(74).

1) The filter for time-invariant system state: Denote by \( \Psi (i, j) \) the transition matrix for the system (72). We have then

\[
v_i = \Psi (i, j) v_j + \sum_{l=j}^{i-1} \Psi (l, l+1) \xi_l, i \geq j + 1. \tag{75}
\]

**Lemma 6.** The following relations hold

\[
R_{ij} = E(v_i v_j^T) = \Psi (i, j) R_{ij} + \Sigma (j), \tag{76}
\]

\[
R_{j+1} = E(v_{j+1} v_{j+1}^T) = \Psi_j R_{j+1} + \Xi_j. \tag{77}
\]

The Lemma 6 is proven by direct calculation of \( R_{ij}, R_{i+1} \) using the formulas (71)-(74), (75).

We proceed now to demonstrate that the condition (68) holds for \( \hat{x}(n_k) = \hat{x}_k, n_k+1 = k + 2, m_k+1 = 2 \). Let us first compute \( \Sigma (k + 2), \Sigma (2) \).

Make use of (44) for computing \( \Sigma (k + 2) \). From Lemma 6 and Eq. (24) one has
\[ K_k^T = [\Psi(k + 1, 0)R_0, \ldots, \Psi(k + 1, k)R_k] = \Psi_k[K_{k+1}^T, R_k] \]

For \( V_k := V_k^T = (V_{k+1}^T, V_{k+2}^T)^T \), \( V_k \) is defined in (11), \( V_{k,2} \) has \( p \) rows, it is not hard to see that

\[ K_k^T = \Psi_k V_{k,2}. \tag{78} \]

Let \( V_k^{-1} := (V_{k+1}^T, V_{k+2}^T)^T \). Then

\[ V_{k,2} V_{k,2} = 0, \quad V_{k,2} V_{k,2} = I \tag{79} \]

hence

\[ K_k^T V_k^{-1} = (0, \Psi_k). \tag{80} \]

Substituting (80) into (44) yields

\[ \Sigma(k + 1) = \Sigma_{22} + \Psi_k H_k P_k (\Psi_k H_k)^T - \Psi_k R_k \Psi_k^T. \tag{81} \]

Consider \( \Sigma(2) \) defined by (42)-(43). By Lemma 6 and (22) it follows

\[ N_{k+1} = P_k H_k \Psi_k^T \tag{82} \]

\[ \Sigma_{12}(2) = (R_k - H_k P_k H_k^T) \Psi_k^T \tag{83} \]

Now the equality \( \Sigma(k + 1) = \Sigma(2) \) holds by inserting (83)(43) into (42), taking into account (81). Thus \( \Delta \Sigma_{k+1}(k + 1) = 0 \) is valid in this case. It means, in virtue of Proposition 1, that \( \bar{x}_{k+1}(2) = \bar{x}_{k+1}(k + 1) = \bar{x}_{k+1}, \) i.e. this Corollary yields in this case the optimal in MMS filter for \( n_i = 2, \forall i \). We have hence

**Theorem 3.** The optimal in MMS filter for the filtering problem (1)(2)(7)(71)-(74) is given in the form

\[ \bar{x}_{k+1} = \bar{x}_k + K_{k+1}(z_{k+1} - H_{k+1} \bar{x}_k), \]

\[ K_{k+1} = P_k H_{k+1}^T \Xi_k = P_k H_{k+1}^T [H_{k+1} P_k H_{k+1}^T + \Xi_k]^{-1}, \]

\[ P_{k+1} = [P_k - H_{k+1}^T \Sigma_k H_{k+1}]^{-1}, \]

\[ H_{k+1} = H_k + \Psi_k H_k, \]

\[ z_{k+1} = z_k - \Psi_k \bar{x}_k. \tag{84} \]

2) The filter for time-varying system state: With the notations (46)(46)(48)-(50) and the assumption (71)-(74), from Corollary 3 we have (for simplicity, \( W_k := W_k^T, \Lambda_k := \Lambda_k^T \)) and the matrix \( W_k^{-1} \) is equal to

\[ W_k^{-1} = V_k^{-1}[I - \Delta V_k(I + V_k^{-1} \Delta V_k)^{-1} V_k^{-1}]. \tag{90} \]

As (79)(86) imply \( \bar{K}_k V_k^{-1} \Delta V_k = 0 \), taking into account (90)(80) we have

\[ \bar{K}_k W_k^{-1} = (0, \Psi_k), \tag{91} \]

\[ \bar{K}_k^T \Lambda_k^{-1} = (0, \Psi_k)B_1. \tag{92} \]

Let \( \bar{H} := \bar{H}_k^T \). From (88)(84) it implies

\[ B_1 \bar{H} = B_2 \Delta W_k W_k^{-1} \bar{H} = \bar{H} - B_2 B_4 = \bar{H} - B_3, \tag{93} \]

\[ B_4 := \Delta W_k W_k^{-1} \bar{H} = \bar{H} \Gamma Q_k^T(\Phi \Psi \Phi)^{-1}. \tag{94} \]

where

\[ B_3 := B_2 B_4 = \bar{H} \Gamma Q_k^T(\Phi \Psi \Phi)^{-1} - B_2 B_4 = \bar{H} I - \Phi \Psi \Phi T M^{-1} \Gamma Q_k^T(\Phi \Psi \Phi)^{-1} = \bar{H} I - \Gamma Q_k^T(\Phi \Psi \Phi)^{-1}. \tag{95} \]

In derivation of (95), for simplicity, the sub-index \( k \) is omitted for the matrices \( \Phi, P, \Gamma, \ldots \) and the following formulas have been used

\[ M = \Lambda_{k+1} = \Phi \Psi \Phi T + \Gamma Q_k^T, \tag{96} \]

\[ B_2 := I - B_3 = I - \bar{H} \Phi \Psi \Phi T M^{-1} \Gamma Q_k^T \bar{H} \Gamma W_k^{-1}, \]

\[ \Phi \Psi \Phi T M^{-1} = [I - \bar{H} \Gamma Q_k^T(\Phi \Psi \Phi)^{-1}]. \tag{97} \]

with \( M_{k+1} \) defined in Corollary 3. The proof of (96) will be given later. Taking into account (95), the formula (93) is equivalent to

\[ B_1 \bar{H} = \bar{H} \Phi \Psi \Phi T M^{-1}. \tag{98} \]

Return to (57). Taking into account (95)(97) one can transform \( L_1 \) in (57) into

\[ L_1 = (0, \Psi) B_1 \bar{H} \Gamma M \bar{H} \Gamma B_1^T(0, \Psi) = (0, \Psi) \bar{H} \Gamma Q_k^T M^{-1} M^{-1} \Phi \Psi \Phi T \bar{H} \Gamma W_k^{-1} (0, \Psi) \]

or

\[ L_1 = \Psi H_k R \Phi T P \bar{M}^{-1} \Phi \Psi \Phi T P \bar{H}_k T\Psi. \tag{99} \]

Let us calculate \( B_2 \Delta W_k \) using (94)(95),

\[ B_2 \Delta W_k = B_2 \bar{H} \Gamma Q_k^T \bar{H} \Gamma W_k^{-1} = B_2 \bar{H} \Gamma Q_k^T(\Phi \Psi \Phi T P \bar{M}^{-1} \Phi \Psi \Phi T P \bar{H}_k T\Psi) \]

or

\[ B_2 \Delta W_k = \bar{H} [P - \Phi \Psi \Phi T M^{-1} \Phi \Psi \Phi T P \bar{H}_k T\Psi] \tag{99} \]

The formula (99) can be used for simplifying \( L_2 \) as follows

\[ L_2 = \bar{K}_k^T \Lambda_k^{-1} \bar{K}_k = \bar{K}_k^T W_k^{-1} [W_k - B_2 \Delta W_k] W_k^{-1} K_k = (0, \Psi)[W_k - H^T (P - \Phi \Psi \Phi T M^{-1} \Phi \Psi \Phi T P \bar{H}_k T\Psi)] (0, \Psi) \]
\[ E[\{x_1, x_2, \ldots, x_m\}] = \Psi H_k P \Phi^T M^{-1} \]

such that the observation Markovian noise sequence \( \{x_m\} \) and \( \{z_m\} \) are the same. For this class of filters, it was clear that the approximate filter along with the truly optimal one, in order to show the main benefits of the proposed approximate approach.

\section{Conclusions}

An approximation approach to the solution of a linear filtering problem with correlated noises was presented. A new type of class of linear recursive filters is proposed together with definition of an optimal MMS estimator among the members of this class of filters. It was clear that the approximate filters have interesting and different properties to their truly optimal MMS filter. Thank to simplified recursive structure, a substantial reduction in computational burden and storage requirements is achieved compared to truly optimal MMS filter. This is important when there are non-Markovian noise processes. For the Markovian memory noise sequence, the proposed sub-optimal filter will yield the truly optimal MMS estimates if the filter is chosen as a function of the last estimate \( \hat{x}_m \) and \( m + 1 \) last observations.

\section{References}