S. Hoang; T. L. Nguyen; Rémy Baraille; Olivier Talagrand

Approximation approach for nonlinear filtering problem with time dependent noises. I. Conditionally optimal filter in the minimum mean square sense

Kybernetika, Vol. 33 (1997), No. 4, 409--425

Persistent URL: http://dml.cz/dmlcz/124888

KYBERNETIKA — VOLUME 33 (1997), NUMBER 4, PAGES 409 - 425

APPROXIMATION APPROACH FOR NONLINEAR FILTERING PROBLEM WITH TIME DEPENDENT NOISES

Part I: Conditionally Optimal Filter in the Minimum Mean Square Sense

S. HOANG, T. L. NGUYEN, R. BARAILLE AND O. TALAGRAND

An approximation approach is proposed to design a nonlinear recursive filter which is conditionally optimal in the minimum mean square (MMS) sense for a nonlinear filtering problem with dependent noises. Definition of an MMS estimator in a given class of estimators is introduced and its uniqueness (with probability 1) is established in Theorem 1. Efficiency of a new optimal filter is illustrated in Theorems 2, 3. Some numerical examples are presented.

1. INTRODUCTION

State estimation plays a very important role in the field of nonlinear multivariable stochastic control. Given noisy observations of the state of a nonlinear dynamical system, contaminated with random noise, the filtering problem consists in estimating as precisely as possible the system state. Most works on nonlinear filtering are based on the minimum mean square (MMS) or on the least squares (LS) approaches. The LS method, considering a state estimation as an output LS problem, is unrecursive and therefore a filtering problem then is solved numerically as such. In this context, many theoretical and practical questions remain open such as the existence and uniqueness of a global minimum, the possible existence of local minima (which are highly undesirable from a computational point of view), the stability of the global minimum with respect to data or to parameter uncertainties etc. We refer to [6], [28] for a detailed discussion of the LS approach.

The classical MMS approach proceeds via determination of the conditional distribution function (c.d.f.) for the system state given observations. A nonliner estimation problem then reduces to the computation of the time evolution of a c.d.f. Equations for the time evolution of a c.d.f. can be found in Bucy [5] and Jazwinski [11]. Unfortunately, the recursive computation of a c.d.f. is in fact unrealizable for a wide variety of nonlinear systems. Equations for c.d.f. are so complicated that there is little interest in evaluating them in practical structures. This explains the high interest in efforts to approach a nonliner filtering from an approximative point of view which aims at constructing simple realizable nonlinear filters for practical systems (cf. [2], [3], [5], [7], [11]). In several works, the optimal filtering equations are linearized around a current estimate and complicated expressions are discarded from computation. In others, an expansion in Fourier series is used and moments of orders higher than second are replaced by functions of conditional expectation and variance. Several authors try even to integrate the Bayes formulae (see [26]).

The following three approximate nonlinear filtering techniques appear very attractive: the linearized Kalman filter (LKF), the extended Kalman filter (EKF) and the iterative EKF. In the LKF, the gain coefficients are independent of the state and they are computed along a nominal solution. The estimate, however, may become very different from the nominal solution as time increases and as a consequence, nonlinearities may grow considerably. In the EKF the equations, linearized around the lastest estimate, are used to fit the estimator to the true state and the gain matrix has to be calculated in real time. This technique is very expensive for high dimensional systems [9]. Another difficulty is the possible divergence of the process [27]. As to iterative EKF, it attempts to decrease the nonlinear effects of measurements by employing in addition an iterative procedure. This procedure repeatedly recalculates an estimate until changes in one iteration become small enough. Several authors have modified the iterative EKF in order to decrease the system nonlinear effects by performing iterations between measurements (iterative EKF with smoothing). For a more complete survey of nonlinear filtering approaches, see [3], [15], [27].

The present paper is a continuation of linear and nonlinear filtering and smoothing problems with correlated noises studied in [17], [18]. In a series of papers [22], [23], [25] Pugachev has suggested and successfully developed a new approach to the design of nonlinear parameter and state estimation algorithms which are (conditionally) optimal in the MMS sense, for nonlinear systems. This approach allows to solve many complicated filtering problems from the point of view of "suboptimality". Many of the difficulties, discussed above, can be removed by introducing first a class of nonlinear filters of given structure; an optimal filter can then be found in that class of filters by solving some optimization problem. This approach enables to overcome traditional difficulties in nonliner filtering and to concentrate the attention on the choice of a structure for the filter and on optimization of the filter's performance. In [17], [18] a solution to filtering and smoothing problems with correlated noises is also proposed in the framework of an approximation method. In fact, when the noise sequence is correlated, the optimal (in MMS sense) estimator $\hat{x}(t)$, written in the recursive form, remains a function of all data measured up to and including t. In addition, in reality, time correlation between variables of the noise sequence decreases as a time lag increases [4], [12], [13]. These facts are in favour of an approximation method which seeks an optimal filter in some class of recursive filters whose current estimate $\hat{x}(t)$ depends only on the lastest observations [17], [20]. In fact, for large t, the computation of $\hat{x}(t|z_t^1)$, $z_t^1 := [z(1), \ldots, z(t)]$ is time-consuming in comparison with $\hat{x}(t|z_t^q), 1 \ll q \leq t$, while their performances, discussed under the above assumptions, should be comparable. Moreover, under some conditions like a Markov property of the noise sequence, the latter approach can yield an MMS optimal filter (see Theorem 3 in [20] for linear case). Finally, the following remark concerned with two approaches in [23] and [17] is worth mentioning: In the light of the works [22], [23], [25], for the case of dependent noises, an efficient recursive filter can be designed by using only the last $(t+1)$th observation $z(t+1)$ and involving some last filtered estimates $\hat{x}(r), \ldots, \hat{x}(t)$, $1 \ll r \leq t$. The estimator $\hat{x}(t)$ can be interpreted as an "observation" for $x(t)$ (cf. [17]) and thus the estimation of $x(t+1)$ can be regarded as an MMS filtering in a class the filters which are the functions of $\hat{x}(t), z(q), \ldots, z(t+1)$ or, following [23] – an MMS filtering in a class of filters depending on $\hat{x}(r), \ldots, \hat{x}(t), z(t+1)$. In fact, the "observations" $\hat{x}(r), \ldots, \hat{x}(t)$ contain a "rich" information on $x(t+1)$ since they are filtered estimates for the system states $x(r), \ldots, x(t)$ and one can considerably improve the quality of $\hat{x}(t+1)$ by using the set $\hat{x}(r), \ldots, \hat{x}(t)$. The two approximation approaches viewed above are of one common idea to choose a class of filters of a given structure and to optimize the filter's performance by solving an appropriate optimization problem. Their difference is the following: if the method in [17] is proposed to overcome the difficulties due to correlation in the noise sequences, Pugachev's method [23] intends to deal with difficulties due to nonlinearity. The purpose of the present paper is to show that a reasonable combination of these two methods can give a new efficient filtering algorithm for solving the complicated nonlinear filtering problems with dependent noises. Efficiency of this new approach will be illustrated on some simple numerical examples.

2. PROBLEM STATEMENT

Consider the following standard nonlinear filtering problem. We are given a dynamical system described by

$$x(t+1) = \phi_t(x(t), w(t)), \quad t = 0, 1, \ldots \tag{1}$$

where $\phi_t(\cdot)$ is a known deterministic function, and observations

$$z(t+1) = h_{t+1}(x(t+1), v(t+1)), \quad t = 0, 1, 2, \ldots \tag{2}$$

where $h_t(\cdot)$ is also a known function. In (1), (2) $x(t)$ denotes the nth dimensional system state, $z(t)$ is an observed vector of dimension p, $w(t)$, $v(t)$ are random vectors with respective dimensions n_w and n_v. We will assume that the initial condition $x(0)$ and the vector processes $\{w(t)\}$ and $\{v(t)\}$ are independent with known distribution functions (d.f.). The sequence $\{w(t)\}$ is supposed to be white with zero-mean and covariance matrix $K_w(t)$. We will denote $F(v_t^1)$ and $g(\lambda, v_t^1)$ the d.f. and characteristic function (c.f.) of the vector v_t^1, $v_t^1 := [v(1), \ldots, v(t)]$. The problem to be considered in this paper is to estimate the state $x(t)$ using all the observations z_t^1.

According to the ideas developed in [17], [23], we introduce the class of filters

$$\hat{x}(t+1) = \delta_t \xi_t[\hat{x}(t-r), \ldots, \hat{x}(t), z(t-q), \ldots, z(t+1)] + \gamma_t, \quad t = 0, 1, \ldots \tag{3}$$

where r, q are integers $0 \leq r \leq t$, $-1 \leq q \leq t-1$. We want to find among the class of filters (3) a filter which is optimal in some sense. In (3) the n_ξ-dimensional vector-function ξ_t is assumed to be given (its choice is usually based on practical experience we have in solving concrete linear or nonlinear suboptimal filtering problems [22], [23]) and $\delta_t \in R^{n \times n_\xi}, \gamma_t \in R^n$ are functions to be determined as a solution of some optimization problem. In what follows equation (3) for the filter will be referred to as an (r, q)-model. It can be easily seen that the (r, q)-model is a mixture of the two models discussed in [17], [23] which are particular cases of (3) subject to $q = -1$ [17] and $r = 0$ [23] respectively.

3. MMS FILTER FOR AN (r, q)–MODEL

1. To avoid any possible confusion which may arise from the generality of the assumptions related to the noise sequence $\{v(t)\}$ we will give here a more direct proof of the theorem which establishes basic relationships for an MMS filter (denoted as MMSF) for an (r, q)-model. We mention that these relationships can be obtained also by the method described in [23]. For simplicity, the class of filters (3) will be denoted by $\hat{X}(\delta, \gamma)$,

$$\hat{x} \in \hat{X}(\delta, \gamma), \hat{x} := \delta\xi + \gamma. \tag{4}$$

The MMS estimator in the class $\hat{X}(\delta, \gamma)$ is denoted by

$$\hat{x}^0 = \delta^0\xi + \gamma^0. \tag{5}$$

Let $\mathrm{tr}(A)$ denote the "trace" operator for a symmetric matrix A.

Definition 1. The estimator $\hat{x}^0$ is said to be MMS optimal in the class $\hat{X}(\delta, \gamma)$ if it satisfies the two following conditions:

(C1) $E[\hat{x}^0] = E[x]$

(C2) $\hat{x}^0 = \arg\min_{\hat{x} \in \hat{X}_u} J[\hat{x}]$

$$J[\hat{x}] = \mathrm{tr}\, E[ee^T], e := \hat{x} - x, \quad \hat{X}^u := \left[x' \in \hat{X}(\delta, \gamma) : E[x'] = E[x]\right].$$

Thus, $\hat{X}^u$ includes all unbiased estimators from the class $\hat{X}$. We mention that the two conditions above are introduced also in [17] for the definition of an optimal estimator. In what follows for any two random vectors ξ, η, let $\bar{\xi} := E[\xi], K_\xi := E[\xi\xi^T], K_{\xi\eta} := E[\xi\eta^T]$.

Theorem 1. There exists an MMSF for the (r, q)-model for the filtering problem (1)–(3). This filter is unique w.p.1 (with probability 1) and it is defined by the relations

$$\hat{x}^0(t+1) = \delta_t^0\xi_t(\hat{x}^0(t-r), \ldots, \hat{x}^0(t), z(t-q), \ldots, z(t+1)) + \gamma_t^0 \tag{6}$$
$$\gamma_t^0 = \bar{x}(t+1) - \delta_t^0\bar{\xi}_t$$
$$\delta_t^0 = K_{x\eta}K_\eta^+, K_\eta^+ \text{ is the pseudoinverse of } K_\eta$$

$$\chi := x(t+1) - \bar{x}(t+1), \ \eta := \xi_t - \bar{\xi}_t$$
$$P(t+1) = E\left[(\hat{x}^0(t+1) - x(t+1))(\hat{x}^0(t+1) - x(t+1))^T\right]$$
$$= M(t+1) - \delta_t^0 K_{\chi\eta}^T$$
$$M(t+1) = E[\chi\chi^T] = K_\chi.$$

Proof. The requirement (C1) for unbiasedness of the estimator $E[\hat{x}] = E[\delta\xi + \gamma] = \delta\bar{\xi} + \gamma = \bar{x}$ leads to

$$\gamma^0 = \bar{x} - \delta\bar{\xi}. \tag{7}$$

Substituting (7) into (3) one sees that the optimal matrix δ^0 is found by solving the following optimization problem

$$\delta^0 = \arg\min J(\delta), \ J(\delta) = \operatorname{tr} E\left[(\delta(\xi - \bar{\xi}) + \bar{x} - x)(\delta(\xi - \bar{\xi}) + \bar{x} - x)^T\right].$$

Taking the derivative of $J(\delta)$ with respect to δ leads to the following matrix equation for an optimal δ^0 (cf. [17])

$$K_\eta X^T = K_{\eta\chi} \tag{8}$$

where $\delta^0 := X$ and χ, η are defined above.

Consider the general case when K_η may be singular. Note that equation (8) is always solvable. Indeed, for the vector $\nu := (\eta^T, \chi^T)^T$ it can be easily seen $E(\nu) = 0$ and

$$\Sigma := E(\nu\nu^T) = \begin{vmatrix} K_\eta & K_{\eta\chi} \\ K_{\chi\eta} & K_\chi \end{vmatrix}.$$

Since Σ is a covariance matrix, $\Sigma \geq 0$ (i. e. Σ is symmetric nonnegative definitive). On the other hand, a necessary and sufficient condition for $\Sigma \geq 0$ is (cf. [1]): (i) $K_\eta \geq 0$;, (ii) $K_{\eta\chi} = K_\eta K_\eta^+ K_{\eta\chi}$; (iii) $K_\chi - K_{\chi\eta} K_\eta^+ K_{\chi\eta}^T \geq 0$. The condition (ii) implies that $K_{\eta\chi} \in (K_\eta)$ where $R(K_\eta)$ is the linear space spanned by the columns of K_η. Hence the equation (8) has a solution.

It is known (cf. [1]) that the class of all solutions of the equation (8) is given by

$$\delta^{Y,T} = X^{Y,T} = K_\eta^+ K_{\eta\chi} + (I - K_\eta^+ K_\eta) Y \tag{9}$$

where Y is an arbitrary matrix of appropriate dimensions. For $Y = 0$ one particular solution is

$$X^0 = K_{\chi\eta} K_\eta^+. \tag{10}$$

Taking the second derivative of $J(X)$ with respect to X gives the Hessian matrix which will be positive if $\det(K_\eta) \neq 0$. Under this condition, the class of solutions (9) consists of only one element $\delta^0 := \delta^{Y=0} = X^0 = K_{\chi\eta} K_\eta^{-1}$. In general, for $\det(K_\eta) = 0$, let $\hat{x}^Y := \delta^Y(\xi - \bar{\xi}) + \bar{x} = \delta^Y \eta + \bar{x}$. We show now that for any Y, all the estimators $\hat{x}^Y$ are identical in the sense that the set of sample points at which they can assume different values, has probability zero. Indeed, write $\hat{x}^Y = \bar{x} + \delta^Y \eta = \hat{x}^0 + Y^T(I - K_\eta^+ K_\eta)\eta$ where $\hat{x}^0$ is defined by (10), i. e. $\hat{x}^0 = \hat{x}^{Y=0}$. Let us compute the mean square deviation of $e := \hat{x}^Y - \hat{x}^0$. It is equal to

$$E[ee^T] = Y^T(I - K_\eta^+ K_\eta) K_\eta (I - K_\eta^+ K_\eta) Y = 0$$

where the last equality follows from the property of the pseudoinverse matrix. Thus $\hat{x}^Y = \hat{x}^0$ for all Y (w.p.1). The formula for the filtered error covariance matrix $P(t+1)$ can be obtained by direct calculation of the matrix $E[(\hat{x}^0 - x)(\hat{x}^0 - x)^T]$.□

Comment 1. In the proof of Theorem 1 we did not use the condition on independence of the sequences $\{w(t)\}, \{v(t)\}$ of the initial state $x(0)$ as well as the properties of whiteness of the process $\{w(t)\}$. These assumptions are made only for simplifying the computational procedure described below.

Comment 2. It is evident that the *a priori* information we are given on all the random variables, as assumed in the statement of the problem, is sufficient to compute the MMS estimator $\hat{x}(t)$ (for simplicity, the upper index "0" from now are dropped from the notation for the MMS estimator). However, due to dependence of the sequence $\{v(t)\}$, the algorithm (3) may become very expensive. The complexity of the filtering algorithm (3) depends on the degree of dependence between the variables of the sequence $\{v(t)\}$. For example, when $\{v(t)\}$ is the output of some dynamical system affected by a white noise, one can use the procedure described in [23] to calculate δ_t^0, γ_t^0. This type of possible simplification can be taken into account by a suitable approximation for the statistical characteristics of the sequence $\{v(t)\}$. The next section will be devoted to this important question.

2. The following theorem establishes the fact that by allowing for a "rich" structure for the vector-function ξ one can improve the performance of the MMS filter. This fact justifies the usefulness of the (r, q)-model for the nonlinear filtering problem.

Theorem 2. Let $\hat{x}$ and $\hat{x}^1$ be the MMS estimators for the class $\hat{X}$ (4) and the class $\hat{X}^1 = [x' : x' = \delta^1\xi^1 + \gamma^1]$ respectively, where $\xi = (\xi^{1,T}, \xi^{2,T})^T$ is an m-vector, $m = m_1 + m_2$, $\delta = (\delta^1, \delta^2)$, $\delta^1 \in R^{n\times m_1}$, $\delta^2 \in R^{n\times m_2}$. If $\hat{x} \neq \hat{x}^1$ in some set of positive probability (w. p. > 0) then $J(\hat{x}) < J(\hat{x}^1)$ where the cost function $J(\cdot)$ is defined in Definition 1.

Proof. Since $\hat{x}^1 = \delta^1\xi^1 + \gamma^1 = (\delta^1, 0)\xi + \gamma^1 = \delta^{10}\xi + \gamma^1, \delta^{10} := (\delta^1, 0)$ therefore $\hat{X}^1 \subset \hat{X}$. However, the unbiased estimator $\hat{x}^1$ belongs to $\hat{X}^{1,u} \subset \hat{X}^u$ (recall, that by definition, $\hat{X}^u$ is the set of all unbiased estimators in the set $\hat{X}$; see Definition 1) and $\hat{x} \neq \hat{x}^1$ (w. p. > 0) by assumption, the proof of the theorem follows from the uniqueness (w.p.1) of the MMS estimate $\hat{x}$ in the class $\hat{X}$. □

3. Theorem 2 shows that using ξ instead of ξ^1 naturally leads to improvement of the MMSF performance. More precisely, this theorem states that $\hat{x}$ is "better" than $\hat{x}^1$. We now derive a condition under which taking ξ from a subspace with higher dimension cannot give a more efficient MMS estimator. We mention that the analogous question for a linear filtering problem is also investigated in Theorems 2 – 3 [18] added and their corollaries.

Theorem 3. Let $\hat{x} = \delta\xi + \gamma$ and $\hat{x}^1 = \delta^{10}\xi + \gamma^1$ be determined as in Theorem 2. Then a necessary and sufficient condition for $\hat{x} = \hat{x}^1$ (w.p.1) is

$$\Delta\delta^T \in R(P),\ \Delta\delta = \delta^{10} - \delta,\ P := I - K_\eta^+ K_\eta. \tag{11}$$

Proof. If $\hat{x} = \hat{x}^1$ (w.p.1) then $J(\hat{x}) = J(\hat{x}^1)$ and from the proof of Theorem 1, there exists some matrix Y such that $\delta^{10,T} = \delta^T + (I - K_\eta^+ K_\eta)Y$, $\delta = K_{\chi\eta} K_\eta^+$. The last equation means that $\Delta\delta^T = (I - K_\eta^+ K_\eta)Y$ or we have what stated in (11).

Conversely, let $\Delta\delta^T \in R(P)$. It is evident that there exists some matrix Y such that $(I - K_\eta^+ K_\eta)Y = \Delta\delta^T$ or $\delta^{10,T} = \delta^T + (I - K_\eta^+ K_\eta)Y$. Thus δ^{10} may be written as $\delta^{10,T} = K_\eta^+ K_{\eta\chi} + (I - K_\eta^+ K_\eta)Y$ for some matrix Y, i.e. $\hat{x} = \hat{x}^1$ (w.p.1) (see the proof of Theorem 1). □

Corollary. Let $K_\eta > 0$. Then a necessary and sufficient condition for $\hat{x} = \hat{x}^1$ (w.p.1) is $\delta^{10} = \delta$.

The Corollary allows to conclude that if $K_\eta > 0$, a necessary and sufficient condition for $\hat{x} = \hat{x}^1$ (w.p.1) is $\delta^{2,0} = 0$ where $\delta^{2,0}$ is the optimal matrix for δ^2 (Theorem 2). It implies that if there exists at least one non-zero element of the matrix $\delta^{2,0}$, increasing the number of components of the vector-function ξ must lead to a more efficient MMS estimator. Detailed consideration of the condition (11) with the special structure of δ, δ^{10} is of importance and is left for the future study.

4. PRACTICAL COMPUTATION OF AN MMS ESTIMATOR

1. It is not hard to show that in general the knowledge of the d.f. for the vector y,

$$\begin{aligned} y &:= [x(t-q), \hat{x}_t^{t-r}, v_{t+1}^{t-q}] \\ \hat{x}_t^{t-r} &:= [\hat{x}(t-r), \dots, \hat{x}(t)], v_{t+1}^{t-q} := [v(t-q), \dots, v(t+1)] \end{aligned} \tag{12}$$

is sufficient for the computation of $\hat{x}(t+1)$. Evidently, the computation can be very expensive because of the dependence of the sequence $\{v(t)\}$.

2. Suppose that $\{v(t)\}$ can be represented or approximated by the equation

$$v(t+1) = f_v[v(t), \eta(t)] \tag{13}$$

where $\{\eta(t)\}$ is a white noise sequence and f_v – a deterministic function. Under this condition, it is sufficient to know the d.f. for the following vector y,

$$y := [x(t-q), \hat{x}_t^{t-r}, v(t-q)] \tag{14}$$

whose evaluation is considerably less expensive in comparison with that of (12). In particular, for $r = 0$,

$$y = [x(t-q), \hat{x}(t), v(t-q)] \tag{15}$$

and, if in addition, $q = 0$ we have

$$y = [x(t), \hat{x}(t), v(t)]. \tag{16}$$

Let us look in detail at the recursive procedure for computation of the d.f. for the vector y defined by (16). Note that analogous procedures can be constructed for the cases (14) and (15).

Let $g_t(\lambda, \mu, \nu) = E\exp\{i\lambda^T x(t) + i\mu^T \hat{x}(t) + i\nu^T v(t)\}$. Inserting (1), (2), (13) into the right hand side of $g_t(\lambda, \mu, \nu)$ gives $g_t(\lambda, \mu, \nu) = E\exp\{i\lambda^T \phi_{t-1}(x(t-1), w(t-1)) + i\mu^T[\delta_{t-1}\xi_{t-1}(\hat{x}(t-1), h_t[x(t), v(t)]) + \gamma_{t-1}] + i\nu^T f_v(v(t-1), \eta(t-1))\}$. The last expression represents the mathematical expectation of a known function of $(x(t-1), \hat{x}(t-1), v(t-1))$ (after substitution of expressions for $x(t), v(t)$ into $h_t[x(t), v(t)]$) and $\{w(t-1), \eta(t-1)\}$. Moreover, $(x(t-1), \hat{x}(t-1), v(t-1))$ does not depend on $(w(t-1), \eta(t-1))$. Hence $g_t(\lambda, \mu, \nu)$ can be evaluated from $g_{t-1}(\lambda, \mu, \nu)$ and the d.f. of $\{w(t-1), \eta(t-1)\}$ which is known *apriori* by assumption.

3. One particular important method for approximation of the d.f. is known as a Gaussian approximation [14], [25] which consists in approximation of the initial d.f. by a Gaussian d.f. The Gaussian approximation method is efficient when there are given only two first moments of random variables.

Consider the situation when we know only the correlation function $K_v(t, \tau)$ of the $n_v = p$ dimensional vector-process $\{v(t)\}$. Suppose that $K_v(t, \tau)$ can be represented in the following form

$$K_v(t, \tau) = \sum_{i=1}^{N} Ki^{(N)}(t)\, Pi^{(N)}(\tau). \tag{17}$$

When $N = 2$ the formula (17) reduces to

$$K_v(t, \tau) = \sum_{i=1}^{2} Ki^{(2)}(t)\, Pi^{(2)}(\tau). \tag{18}$$

Let in (18) $K_1^{(2)}$, $P_1^{(2)}$ be $p \times p$ matrices, $K_1^{(2)}$ being nonsingular; $K_2^{(2)}$, $P_2^{(2)}$ are the $p \times p_1$ and $p_1 \times p$ matrix respectively. Let

$$v^{(1)}(t+1) = C^{(1)}(t)\, v^{(1)}(t) + \eta^{(1)}(t), t = 0, 1, \ldots \tag{19}$$

where $\{\eta^{(1)}(t)\}$ is a white noise sequence uncorrelated with $v^{(1)}(0)$. Then $K_{v^{(1)}}(t, \tau)$ has the representation (cf. [19])

$$K_{v^{(1)}}(t, \tau) = K_1^{(1)}(t)\, P_1^{(1)}(\tau) \tag{20}$$

here $K_1^{(1)}(t)$ is a $p_1 \times p_1$ nonsingular matrix. Let $K_1^{(1)}(t)$, $P_1^{(1)}(\tau)$ satisfy the equation

$$K_2^{(2)}(t)\, P_1^{(1)}(t) = P_2^{(2),T}(t)\, K_1^{(1),T}(t). \tag{21}$$

Under conditions (20), (21), the process $\{v(t)\}$ is solution of the following recursive equation

$$v(t+1) = C(t)\, v(t) + D(t)\, v^{(1)}(t) + \eta(t) \tag{22}$$

where $\{\eta(t)\}$ is another white noise sequence such that

$$E[\eta(t)\, \eta^{(1),T}(\tau)] = \Theta_t \delta_{t\tau}, \delta_{t\tau} \text{ is a Kronecker function.} \tag{23}$$

Detailed procedures for the computation of $C(t), C^{(1)}(t), D(t)$ and of the covariance matrices of the random sequences $\{\eta(t)\}, \{\eta^{(1)}(t)\}$ can be found in [19].

We note that the class of random processes (17) is large enough to include stationary as well as nonstationary processes. Moreover, random processes, possessing a canonical form, also belong to the class (17) (see [25]). It means therefore that large computational savings in the implementation of MMS nonlinear filters can be achieved by using the representation of the type (17) for correlated noise sequences.

We now return to the problem of the practical implementations of an MMSF for the (r,q)-model when the sequence $\{v(t)\}$ is verifying (22), (19). The algorithm for the filter is given in Theorem 1. For simplicity, consider the case when $r = 0,\ q = 0$, i.e. a first order nonlinear difference equation is employed with two last observations. Then the filter requires, at each time instant, the determination of the d.f. of the following vector

$$y := \{x(t), \hat{x}(t), v(t), v^{(1)}(t)\}. \tag{24}$$

This can be proved by a method analogous to the method used in Subsection 4.2, noticing that instead of (13) we have now (22), (19) and that the two white noise sequences $\{\eta(t)\}, \{\eta^{(1)}(t)\}$ are correlated only at $t = \tau$ (see (23)). Hence $\{\eta^{(1)}(t-1)\}$ depends only on $\{\eta(t-1)\}$ and $\{\eta^{(1)}(t-1)\}$ is independent of $v(t-1)$. A recursive procedure for the determination of the c.f. of the vector y defined by (16) can be obtained in a similar manner.

5. EXAMPLES

1. Example 1 (Chapter 10, Example 1 [24]). Let

$$\begin{aligned} x(t+1) &= [x(t) + w(t)]^2 \\ z(t) &= x(t) + v(t). \end{aligned} \tag{25}$$

To estimate the state of the system (25), let us use the $(0,0)$-model. The filter is then of the structure

$$\begin{aligned} &\hat{x}(t+1) = \delta_t \xi_t[\hat{x}(t), z(t), z(t+1)] + \gamma(t) \\ &\delta_t = [\delta_{t,1}, \delta_{t,2}, \ldots, \delta_{t,6}], \quad \xi_t = [\xi_{t,1}, \xi_{t,2}, \ldots, \xi_{t,6}]^T \\ &\xi_{t,1} = \hat{x}(t)^2,\ \xi_{t,2} = \hat{x}(t)[z(t) - \hat{x}(t)],\ \xi_{t,3} = [z(t) - \hat{x}(t)]^2,\ \xi_{t,4} = \hat{x}(t)[z(t+1) - \hat{x}^2(t)] \\ &\xi_{t,5} = [z(t+1) - \hat{x}^2(t)][z(t) - \hat{x}(t)],\ \xi_{t,6} = [z(t+1) - \hat{x}^2(t)]^2. \end{aligned} \tag{26}$$

As in [24] the vector-function ξ_t is chosen to be a quadratic function of the vector $[\hat{x}(t), z(t) - \hat{x}(t), z(t+1) - \hat{x}^2(t)]$. The component $z(t+1) - \hat{x}^2(t)$ is taken as a term approximating $z(t+1) - \hat{x}(t+1|t)$ where $\hat{x}(t+1|t)$ is a one-step ahead forecast for $x(t+1)$. Indeed, an MMS forecast is given by $\hat{x}(t+1|t) = E[x(t+1)|z_1^t] = E\{[x(t)+w(t)]^2|z_1^t\} = E[x(t)^2 + w(t)^2 + 2x(t)\, w(t)|z_1^t]$. Since $w(t)$ does not depend on $[x(t), z_1^t]$ hence $\hat{x}(t+1|t) = E[x(t)^2|z_1^t] + E[w(t)^2]$ if $E[w(t)] = 0$. Thus the approximation $\hat{x}(t+1|t) \simeq \hat{x}^2(t)$ is valid if $\hat{x}(t)$ is a good estimate for $x(t)$ (and $w(t)$ is small) since then $E([x(t) - \hat{x}(t)]^2|z_1^t)$ is small from which one can conclude $\hat{x}(t+1|t) \simeq \hat{x}^2(t)$. In general, however, we are free to assume the structure described above for ξ_t independently on how close is $\hat{x}(t)$ to $x(t)$.

Optimal parameters (δ_t^0, γ_t^0) are then given by Theorem 1,

$$
\begin{aligned}
&\gamma_t^0 = \bar{x}(t+1) - \delta_t^0 \bar{\xi}_t \\
&\delta_t^0 = K_{\chi\eta} K_\eta^+ \\
&K_{\chi\eta} = [k_1(t), \dots, k_6(t)], \\
&k_1(t) := E[x(t+1)\,(\xi_{t,1} - \bar{\xi}_{t,1})], \dots, k_6(t) := E[x(t+1)\,(\xi_{t,6} - \bar{\xi}_{t,6})] \\
&K_\eta = E[(\xi_t - \bar{\xi}_t)\,(\xi_t - \bar{\xi}_t)^T] \\
&\quad = \begin{vmatrix} E[(\xi_{t,1} - \bar{\xi}_{t,1})\,\xi_{t,1}] & \dots & E[(\xi_{t,1} - \bar{\xi}_{t,1})\,\xi_{t,6}] \\ \dots & \dots & \dots \\ E[(\xi_{t,6} - \bar{\xi}_{t,6})\,\xi_{t,1}] & \dots & E[(\xi_{t,6} - \bar{\xi}_{t,6})\,\xi_{t,6}] \end{vmatrix}.
\end{aligned}
\tag{27}
$$

The computation of $K_{\chi\eta}, K_\eta$ is similar to that described in [24]. For instance, we write here several formulae

$$
\begin{aligned}
k_1(t) &= E[x^2(t)\,\hat{x}^2(t)] - E[x^2(t)]\,E[\hat{x}^2(t)] \\
k_2(t) &= E[x^3(t)\,\hat{x}(t)] - E[x^2(t)]\,E[\hat{x}(t)\,x(t)] - k_1(t) \\
&\quad + E[x^2(t)\,\hat{x}(t) v(t)] - E[x^2(t)]\,E[\hat{x}(t)\,v(t)] \\
k_3(t) &= E[x^4(t)] - \{E[x^2(t)]\}^2 - k_1(t) - 2k_2(t) \\
k_4(t) &= E[x^2(t)\,\hat{x}^3(t)] - E[x^2(t)]\,E[\hat{x}^3(t)] + E[x^5(t)] - E[x^2(t)]E[x^2(t)\,\hat{x}(t)] \\
&\quad + E[x^2(t)\,v(t+1)\,\hat{x}(t)] - E[x^2(t)]\,E[v(t+1)\,\hat{x}(t)] \\
k_5(t) &= E[x^5(t)] - E[x^2(t)]\,E[x^3(t)] - \{E[x^3(t)\,\hat{x}^2(t)] - E[x^2(t)]\,E[x(t)\,\hat{x}^2(t)] \\
&\quad + E[x^2(t)\hat{x}^2(t)v(t)] - E[x^2(t)]\,E[\hat{x}^2(t)v(t)] - k_4(t)
\end{aligned}
$$

and so on.

Comment 1. As follows from Theorem 1 and Example 1, Theorem 1 in fact gives the filtered estimate for the system state while the procedure described in [24] computes a one-step ahead forecast.

Comment 2. For the filtering problem (25) with dependent noises one can employ a simpler class of $(0, -1)$-models as described in [24],

$$
\begin{aligned}
\hat{x}^{(1)}(t+1) &= \delta_t^{(1)} \xi_t^{(1)}[\hat{x}^{(1)}(t), z(t+1)] + \gamma_t^{(1)} \\
\xi_t^{(1)} &= \left\{\hat{x}^{(1)2}(t), \hat{x}^{(1)}(t)[z(t+1) - \hat{x}^{(1)2}(t)], [z(t+1) - \hat{x}^{(1)2}(t)]^2\right\}^T.
\end{aligned}
\tag{28}
$$

From Theorem 2 we have the inequality

$$E\left\{[\hat{x}(t+1)-x(t+1)]^2\right\} \leq E\left\{[\hat{x}^{(1)}(t+1)-x(t+1)]^2\right\} \tag{29}$$

with strict inequality if $\hat{x}(t+1) \neq \hat{x}^{(1)}(t+1)$ (w. p. > 0) where $\hat{x}(t+1)$ is defined by (26) (see Theorem 3). It is seen that one can improve the accuracy of any filter by involving a wider class of (r,q)-models. In particular, a $(0,0)$-model is certainly preferable over a $(0,-1)$-model. The stronger is the dependence between the variables of the noise sequence $\{v(t)\}$, the better is the performance of $\hat{x}(t+1)$ in comparison with that of $\hat{x}^{(1)}(t+1)$.

2. Example 2. Let in (1), (2)

$$\begin{aligned} x(t+1) &= \phi_t[x(t)] = x(t) = x \\ z(t+1) &= h_{t+1}[x(t+1), v(t+1)]. \end{aligned} \tag{30}$$

The filtering problem for the system (30) reduces to the estimation of the unknown parameter x. It is interesting to demonstrate here that by choosing an appropriate structure for the vector-function ξ_t one can construct a simple convergent algorithm for the estimation of x. Let us choose

$$\xi_t := [\hat{x}(t), z(t+1) - h_{t+1}(\hat{x}(t))]^T. \tag{31}$$

Introduce

$$\tilde{x}(t+1) = \tilde{\delta}\hat{x}(t) + \tilde{\gamma}. \tag{32}$$

Statement. Consider a filtering problem for the system (30) and suppose that $\hat{x}(t+1)$ is an MMS estimator produced by the $(0,-1)$-model, ξ_t being defined by (31). In addition, let $\tilde{x}(t+1)$ be an MMS estimator in the class (32) and $\hat{x}(t+1) \neq \tilde{x}(t+1)$ (w.p. > 0). Then

$$P(t+1) < P(t) \tag{33}$$

where $P(t) := E[\hat{x}(t) - x(t)]\,[\hat{x}(t) - x(t)]^T$ is the error covariance matrix for $\hat{x}(t)$.

Proof. By construction, from Theorem 2 $P(t+1) < \tilde{P}(t+1)$ where $\tilde{P}(t+1)$ is the error covariance matrix for $\tilde{x}(t+1)$. The strict inequality (33) holds if we can show that $\tilde{P}(t+1) \leq P(t)$. From Theorem 1, $\tilde{\gamma} = (1-\tilde{\delta})\,\bar{x}$. Hence

$$\begin{aligned} \tilde{x}(t+1) &= \tilde{\delta}\hat{x}(t) + (1-\tilde{\delta})\,\bar{x} = \tilde{\delta}[\hat{x}(t) - \bar{x}] + \bar{x} \\ \tilde{x}(t+1) - x &= \tilde{\delta}[\hat{x}(t) - \bar{x} + x - x] + \bar{x} - x = \tilde{\delta}[\hat{x}(t) - x] + (1-\tilde{\delta})\,(\bar{x} - x) \end{aligned}$$

or

$$\tilde{x}(t+1) - x = d^T\eta \tag{34}$$

here for simplicity we use the notation $d := (\delta, 1-\delta)^T$, $\delta := \tilde{\delta}, \hat{x} := \hat{x}(t), \eta := (\hat{x} - x, \bar{x} - x)^T$. Then

$$\tilde{P}(t+1) = d^T\Sigma d \tag{35}$$

$$\Sigma := E(\eta\eta^T) \tag{36}$$

where

$$\Sigma := E(\eta\eta^T) = \begin{vmatrix} P & n \\ n & M \end{vmatrix}, \; P := P(t), \; n := E[\hat{x} - x]\,[\bar{x} - x], \; M := E[\bar{x} - x]^2.$$

Consider the difference $\Delta P := \tilde{P}(t+1) - P(t)$. From (35)

$$\Delta P = a\delta^2 + 2b\delta + c \qquad (37)$$
$$a := P - 2n + M; \; b := n - M; \; c := M - P.$$

On the other hand, Theorem 1 gives

$$\delta := \tilde{\delta} = K_{\chi\eta}K_\eta^{-1} = (M - n)/(P + M - 2n) \qquad (38)$$

since $K_{\chi\eta} = E(x - \bar{x})(\hat{x} - x) = M - n, K_\eta = E(\hat{x} - \bar{x})^2 = E(\hat{x} - x + x - \bar{x})^2 = P + M - 2n$.

Inserting (38) into (37) leads to $\Delta P \leq 0$ which proves the Statement. □

Comment. In general, instead of (33) we have

$$P(t+1) \leq P(t). \qquad (39)$$

Considering $\nu(t) = \operatorname{tr} P(t)$ as a stochastic Lyapunov function [8], the relation (33) shows that $\Delta\nu(t) := \nu(t+1) - \nu(t) < 0$ and therefore the sequence $\nu(t)$ must converge since $\nu(0)$ is bounded from below by 0. It means that the estimation error for the sequence $\{\hat{x}(t)\}$ is bounded and that the filter is stable. A detailed study of the possibility to design a stable nonlinear filter for the filtering problem (1) (2) will be given in Part II.

3. Example 3. To illustrate the theoretical results presented in Theorem 2, let us return to the filtering problem of Example 2 subject to $z(t) = x^2(t) + v(t), t = 1, 2$. We limit ourselves to the case of two observations $z(t), t = 1, 2$ since the purpose of the present example is only to give a numerical illustration of Theorem 2. The question of the more appropriate structure for ξ_t and of the asymptotic behaviour of the estimate $\hat{x}(t)$ are not considered here. They are worthy of special investigation in the future.

Suppose that the vector $\tilde{v} := [x, v(1), v(2)]$ has the Gaussian d.f. with zero mean and covariance matrix

$$\tilde{V} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

Let us compute by Theorem 1 MMS estimators for the following three classes of filters

(i) $\hat{x}_1 := \delta_1\bar{x} + \delta_2[z(1) - \bar{x}^2] + \gamma$

(ii) $\hat{x}_P := \delta_1\hat{x}_1 + \delta_2[z(2) - \hat{x}_1^2] + \gamma$

(iii) $\hat{x} := \delta_1\hat{x}_1 + \delta_2[z(2) - \hat{x}_1^2] + \delta_3[z(1) - \hat{x}_1^2] + \gamma$.

Thus, $\hat{x}_1$ belongs to a class of estimates at $t = 1$, $\hat{x}_P$ belongs to a class of $(0, -1)$-models at $t = 2$, and $\hat{x}$ belongs to a class of $(0,0)$-models (at $t = 2$ too). The class (ii) is of the Pugachev form while (iii) is of the form (3).

3.1. Since $\bar{x} = 0$ by assumption, $\gamma = -2\delta_2$ and $\hat{x}_1 = \delta_2[z(1) - 2]$. For $\xi = z(1)$, Lemma A3 (see the Appendix) implies $E(x\xi) = 1, E(\xi - 2)^2 = 10$ hence by Theorem 1, $\delta_2^0 = 0.1$. The mean square error (MSE) of the estimator $\hat{x}_1$ is computed by Theorem 1 (see equation for $P(t)$ in Theorem 1) and is equal to $P_1 = E(\hat{x}_1 - x)^2 = 1.9$.

3.2. For the class (ii), since $E(\hat{x}_1) = 0$ we have $\gamma = -\delta_2 E[z(2) - \hat{x}_1^2] = -1.9\delta_2$ (see Lemma A3), hence $\hat{x}_P = \delta_1\hat{x}_1 + \delta_2[z(2) - \hat{x}_1^2 - 1.9]$.

Minimizing $P_P = E(\hat{x}_P - x)^2$ leads to the following equations for the determination of $\delta := (\delta_1, \delta_2)$,

$$\Sigma\delta^T = b, \quad \Sigma := (\Sigma_{ij})_{i,j=1}^2, \quad b := (b_1, b_2)^T$$
$$\Sigma_{11} = E(\hat{x}_1^2), \quad \Sigma_{12} = \Sigma_{21} = E[(z(2) - \hat{x}_1^2 - 1.9)\hat{x}_1]$$
$$\Sigma_{22} = E[z(2) - \hat{x}_1^2 - 1.9]^2, \quad b_1 = E(x\hat{x}_1), \quad b_2 = E\{x[z(2) - \hat{x}_1^2 - 1.9]\}.$$

From Lemma A4 (see the Appendix) one finds $\Sigma_{11} = 0.1, \Sigma_{12} = 0.83$, $\Sigma_{22} = 8.7064$, $b = (0.1, 0.92)^T$. Solving the corresponding system of equations for δ gives $\delta^0 = (0.5852501,\ 0.0498793)$. Thus the MMS estimator in the class (ii) is of the form $\hat{x}_P = 0.5852501\hat{x}_1 + 0.0498793[z(2) - \hat{x}_1^2 - 1.9]$. The MSE of $\hat{x}_P$ is equal to $P_P = E(\hat{x}_P - x)^2 = P_1 - \delta^0 K_{\chi\eta}^T = P_1 - \delta^0 b = 1.7955861$.

3.3. For the class (iii), Theorem 1 and Lemma A3 yield $\gamma = -1.9(\delta_2 + \delta_3)$ hence

$$\hat{x} = \delta_1\hat{x}_1 + \delta_2[z(2) - \hat{x}_1^2 - 1.9] + \delta_3[z(1) - \hat{x}_1^2 - 1.9].$$

Analogously, minimizing the MSE of $\hat{x}$ leads to the following system of equations for finding δ: $\Sigma\delta^T = b$, $\Sigma = \Sigma := (\Sigma_{ij})_{i,j=1}^3$, $b := (b_1, b_2, b_3)^T$. It is easily to see that Σ_{ij}, $i, j = 1, 2; b_i, i = 1, 2$ remain unchanged. Lemma A4 gives $\Sigma_{13} = E[z(1) - \hat{x}_1^2 - 1.9]\hat{x}_1 = 0.93, \Sigma_{23} = E[z(2) - \hat{x}_1^2 - 1.9]\,[z(1) - \hat{x}_1^2 - 1.9] = 7.7064, \Sigma_{33} = E[z(1) - \hat{x}_1^2 - 1.9]^2 = 8.7064, b_3 = E[z(1) - \hat{x}_1^2 - 1.9]\,x = 0.92$. This leads to

$$\delta^0 = (\delta_1^0, \delta_2^0, \delta_3^0) = (1.888863,\ 0.04921551,\ -0.1396555).$$

Then the optimal MMS estimator in the class (iii) is

$$\hat{x} = 1.888863\hat{x}_1 + 0.04921551[z(2) - \hat{x}_1^2 - 1.9] - 0.1396555[z(1) - \hat{x}_1^2 - 1.9]$$

which has MSE equal to

$$P = E(\hat{x} - x)^2 = P_1 - \delta^0 b = 1.7943185 < P_P = 1.7955861.$$

This simple numerical example supports Theorem 2 and shows that using in addition the observation $z(1)$ yields an estimator which has a smaller MSE than $\hat{x}_P$.

6. CONCLUSION

We have considered nonlinear filtering problems with time-dependent noise sequences for the model and observation errors. A new approximate MMS filtering approach is proposed which seeks an MMSF in the class of (r, q)-models by solving an appropriate minimization problem. It is proved theoretically that an MMSF exists and is unique (w.p.1). A detailed algorithm for its computation is given. We prove theoretically and illustrate by a numerical example that the proposed method can produce a more accurate estimator than previously developed methods [17], [23]. This is obtained through the choice of a more appropriate structure for the filter. This fact is important since due to the nonlinearity and the time-dependence of the basic random processes, there is no possibility to obtain in practice a strictly optimal (in a statistical sense) structure for the nonlinear filter.

As seen from Section 4, practical implementation of the MMSF for an (r, q)-model is considerably simplified if dependent noise sequence is represented or approximated as the output of a dynamical system driven by a white noise sequence. This representation can be used for a wide class of random processes, either stationary or non-stationary, and either Markovian or non-Markovian. However, it should be emphasized that dynamical representation of a given random process produces another process which is *equivalent* to the original one only to the degree that their first- and second- order moments are identical. Since in practice statistics are given only approximately, dynamical representation approach can yield satisfactory results only if a Gaussian approximation is valid for the filtering problem at hand. More precisely, there exist other classes of random processes (for example, heavy-tailed non-Gaussian processes [10], [21]) for which a Gaussian Bayesian estimation is inefficient for small deviations from Gaussian models. In such situations, a possible way to avoid instability (with respect to deviations from a Gaussian model) is to operate on the innovation vector with a nonlinear optimal transformation [21] (or *influence function*, in terms of [16]).

Another important question is related to the design of a stable MMS nonlinear filter, i.e. a nonlinear filter which produces an MMS estimator with a bounded MSE. The importance of this question arises from the well known fact that even in a linear case with white noises, optimality of a Kalman filter does not imply its stability [11]. We will show in Part II that by imposing an additional constraint on the optimality condition, it is possible to design a nonlinear filter whose performance is comparable in accuracy to that of an MMSF with no parameter uncertainty. In contrast, when uncertainty is present, a stable MMSF will certainly behave better than the corresponding MMSF.

ACKNOWLEDGEMENT

The authors wish to thank the reviewers for their valuable comments and suggestions for improving the paper.

APPENDIX

In order to employ Theorem 1 to compute MMS estimators in Example 3 first we need Lemma A1. For its proof we refer an interested reader to [24].

Lemma A1. Let $x := (x_1, \ldots, x_n)^T$ be a random Gaussian vector with zero mean and covariance matrix $K_x = [k_{ij}]$. Let $m_{h_1,\ldots,h_n} := E(x_1^{h_1}.x_2^{h_2}\ldots x_n^{h_n})$ be a moment of order $h, h = \sum_1^n h_i$. Then for all $s = 1, 2, \ldots,$ we have $m_{2s-1} = 0$,

$$m_{2s} = m_{h_1,\ldots,h_n} = \frac{h_1!h_2!\ldots h_n!}{2^s s!} \sum k_{p,q}, \quad k_{p,q} = k_{p_1,q_1} \cdots k_{p_s,q_s}$$

where the sum $\sum$ includes all the different permutations of $2s$ indices $p_1, q_1, \ldots, p_s, q_s$ from which h_1 indices are equal to $1, \ldots, h_n$ indices are equal to n. In particular, for scalar x we have $m_{2s-1} = 0$,

$$m_{2s} = \frac{(2s)!}{2^s s!} D_x^s, \; D_x := E(x^2).$$

Lemma A2. Let $\tilde{v} := (x, v_1, v_2)$ be a Gaussian vector defined in Example 3. Then

$$m_{200} = 2, \quad m_{400} = 12, \quad m_{600} = 120, \quad m_{800} = 1680,$$
$$m_{220} = 6, \quad m_{310} = 6, \quad m_{211} = 4, \quad m_{420} = 48.$$

Lemma A2 is an application of Lemma A1 to computation of the moments of the vector $\tilde{v}$. Using Lemma A2 one can obtain

Lemma A3. Let $z(1), z(2)$ be the observations given in Example 3. Under the conditions of Lemma A2,

$$Ez(1) = Ez(2) = 2$$
$$E[xz(1)] = E[xz(2)] = 1, \; E[z(1)z(2)] = 13, \; E[z(1)^2] = E[z(2)^2] = 14$$
$$E[\hat{x}^2(1)] = 0.1, \; E[x\hat{x}(1)] = 0.1, \; E[\hat{x}(1)z(1)] = 1, \; E[\hat{x}(1)z(2)] = 0.9$$
$$E[\hat{x}^3(1)] = 0.07, \; E[x\hat{x}^2(1)] = 2125, \; E[z(1)^3] = 138$$
$$E[\hat{x}^2(1)z(1)] = E[\hat{x}^2(1)z(2)] = 0.9, \; E[z(2)z^2(1)] = 134$$
$$E[\hat{x}^4(1)] = 0.1164, \; [z^4(1)] = 1980.$$

Finally from Lemma A3 it follows

Lemma A4. Under conditions of Lemma A3, elements Σ_{ij} of the matrix Σ and b_i of vector b, defined in the Example 3, are equal to

$$\Sigma_{11} := E(\hat{x}_1^2) = 0.1, \quad \Sigma_{12} := E\hat{x}(1)[z(2) - \hat{x}_1^2 - 1.9] = 0.83,$$
$$\Sigma_{13} := E[z(1) - \hat{x}_1^2 - 1.9]\hat{x}_1 = 0.93$$
$$\Sigma_{22} := E[z(2) - \hat{x}_1^2 - 1.9]^2 = 8.706,$$

$$\Sigma_{23} := E[z(2) - \hat{x}_1^2]\,[z(1) - \hat{x}_1^2 - 1.9] = 7.7064,$$
$$\Sigma_{33} := E[z(1) - \hat{x}_1^2 - 1.9]^2 = 8.7064$$
$$b_1 := E(x\hat{x}_1) = 0.1, \quad b_2 := E\{x[z(2) - \hat{x}_1^2 - 1.9]\} = 0.92,$$
$$b_3 := E[z(1) - \hat{x}_1^2 - 1.9]\,x = 0.92.$$

(Received July 31, 1995.)

REFERENCES

[1] A. Albert: Regression and the Moore–Penrose Pseudoinverse. Academic Press, New York 1972.
[2] D. L. Alspach and H. W. Sorenson: Nonlinear Bayesian estimation using Gaussian sum approximations. IEEE Trans. Automat. Control *AC-17* (1972), 4, 439–448.
[3] B. D. O. Anderson and J. B. Moore: Optimal Filtering. Prentice–Hall, Englewood Cliffs, NJ 1979.
[4] G. P. Box and G. M. Jenkins: Time Series Analysis. Holden–Day, San Francisco 1970.
[5] R. S. Bucy et al: An Engineer's Guide to Building Nonlinear Filters. F. J. Seiler Research Lab., Project 7904-00-37, May 1972.
[6] G. Chavent: On the uniqueness of local minima for general abstract nonlinear least squares problems. Inverse Problems *4* (1988), 417–433.
[7] G. C. Goodwin and K. S. Sin: Adaptive Filtering and Control. Prentice–Hall, Englewood Cliffs, NJ 1984.
[8] S. Hoang, R. Baraille, O. Talagrand, L. Nguyen and P. DeMey: Approximation approach for nonlinear filtering problem with time dependent noises. Part II: Stable nonlinear filters. Kybernetika *33* (1997), No. 5 (to appear).
[9] S. Hoang, P. DeMey, O. Talagrand and R. Baraille: A new reduced-order adaptive filter for state estimation in high dimensional systems. In: Proc. IFAC Internat. Symp. Adaptive Systems in Control and Signal Processings (Cs. Banyasz, ed.), Budapest 1995, pp. 153–158, to appear in Automatica *33* (1997), No. 8.
[10] P. J. Huber: Robust Statistics. Wiley, New York 1981.
[11] A. H. Jazwinski: Stochastic Processes and Filtering Theory. Academic Press, New York 1970.
[12] V. V. Kondrachev and A. A. Kocheshkov: Filtering and analysis of linear sample–data control systems with nonparametric specification of correlated noises (in Russian). Avtomat. i Telemekh. (1985), No. 6, 67–76.
[13] A. Kowalski and D. Szynal: Filtering in discrete–time systems with state and observation noises correlated on a finite time interval. IEEE Trans. Automat. Control *AC-31* (1986), 4, 381–384.
[14] I. E. Kazakov and X. V. Malchikov: Approximate design of Pugachev's filters of specified complexity. Automat. Remote Control *42* (1981), 1309.
[15] F. L. Lewis: Optimal Estimation. Wiley, New York 1986.
[16] C. J. Masreliez and D. Martin: Robust Bayesian estimation for the linear model and robustifying the Kalman filter. IEEE Trans. Automat. Control *AC-22* (1977), 3, 361–371.
[17] T. L. Nguyen and S. Hoang: On optimal filtering with correlated noises and singular correlation matrices. Automat. Remote Control *43* (1982), 5 (2), 660–669.
[18] T. L. Nguyen and S. Hoang: Optimal adaptive smoothing with correlated noises. Automat. Remote Control *43* (1982), 7 (1), 887–895.
[19] T. L. Nguyen and S. Hoang: On one model of non–Markovian random process and its application to optimal estimation theory. Automat. Remote Control *43* (1982), 8 (1), 1021–1032.

[20] T. L. Nguyen and S. Hoang: On solution of ill-posed optimal linear filtering with correlated noises. Automat. Remote Control *44* (1983), 4 (1), 453–466.
[21] B. T. Polyak and Ja. Z. Tsypkin: Robust identification. Automatica *16* (1980), 1, 53–63.
[22] V. S. Pugachev: Estimation of variables and parameters in discrete-time nonlinear systems. Automat. Remote Control *40* (1979), 512–521.
[23] V. S. Pugachev: Conditionally optimal estimation in stochastic differential systems. Automatica *18* (1982), 685–696.
[24] V. S. Pugachev: Probability Theory and Mathematical Statistics for Engineers. Pergamon Press, New York 1984.
[25] V. S. Pugachev, I. N. Sinitsyn and V. I. Shin: A software method of normal approximation in analysis of nonlinear stochastic systems (in Russian). Avtomat. i Telemekh. (1987), No. 2, 62–67.
[26] G. T. Schmidt: Linear and nonlinear filtering techniques. In: Control and Dynamical Systems (C. T. Leondes, ed.), Academic Press, New York 1976.
[27] H. W. Sorenson: Survey on filtering techniques and stochastic control in dynamical systems. In: Control and Dynamical Systems (C. T. Leondes, ed.), Academic Press, New York 1976.
[28] J. Sternby: On consistency for the method of least squares using martingale theory. IEEE Trans. Automat. Control *AC-22* (1977), 3, 346–352.

Dr. Hong Son Hoang, SHOM/GRGS/CNRS/CNES, 18 Avenue Edouard Belin, 31401 Toulouse Cédex 4. France.

Dr. Thuc Loan Nguyen, Institute of Physics, National Center for Natural Sciences and Technology, Hanoi. Vietnam.

Dr. Rémy Baraille, SHOM/CMO/GRGS/CNRS, 18 Avenue Edouard Belin, 31401 Toulouse Cédex 4. France.

Dr. Olivier Talagrand, LMD/ENS, 24 rue Lhomond, 75231 Paris Cédex. France.