DOI 10.1007/s10958-020-05005-7
Journal of Mathematical Sciences, Vol. 250, No. 1, October, 2020

OPTIMAL STRUCTURE OF RECURRENT NONLINEAR FILTERS OF LARGE ORDER FOR DIFFUSION SIGNALS

E. A. Rudenko
Moscow Aviation Institute (National Research University)
4, Volokokamskoe Shosse, Moscow 125993, Russia
rudenkoevg@yandex.ru

UDC 519.246.2: 681.518.22

We consider the optimal mean-square estimation problem for the state variables of a continuous nonlinear stochastic object by using results of time-discrete measurements. To obtain clock and inter-clock estimates on a computer of limited power in real time, we propose a procedure for the synthesis of a nonlinear structure of a discrete finite-dimensional filter, the state vector of which is formed from the desired number of already obtained preceding clock estimates. We describe the synthesis algorithm for the filter and its suboptimal approximations. The advantage of the latter is shown in comparison with the corresponding generalizations of the Kalman filter. Bibliography: 8 *titles.*

To obtain estimates for the Markov state variables of a stochastic object of observation by an absolutely optimal filter [1]–[3], it is required to promptly find the posterior probability density of the estimated random process, which makes such a filter a distributed parameter system. Therefore, the state vector of such a filter is of infinite order, and it is difficult to implement an absolutely optimal filter in real time. Therefor, in practice, one has to use approximate finite-dimensional filtering algorithms such as various generalizations of the Kalman filter, with a loss of accuracy, or create poly-Gaussian banks of such filters, which complicates the computer. The implementation of the particle filter [3] requires a very powerful computer due to the use of the cumbersome Monte Carlo method at each trajectory.

A conditionally optimal filter is finite-dimensional and thereby can be easily realized [4, 5], but it is only parametric and its order is bounded by the order of the object of observation. Finite-dimensional filters of optimal structure of different orders, free from these restrictions, are synthesized in [6]–[8]. In the sense of potential accuracy, the optimal structure filters occupy an intermediate position between the absolutely and conditionally optimal filters $I_t^{\mathrm{AOF}} \leqslant I_t^{\mathrm{OSF}} \leqslant I_t^{\mathrm{COF}}$, where I_t is the mean-square estimation error at the time t. However, the accuracy of an optimal structure filter of small order [6, 7] is bounded exactly by its order, whereas an optimal finite memory filter [8] the order of which is a multiple of the dimension of the measurement vector, forgets the preceding measurements which could be more exact than the current ones.

Translated from *Problemy Matematicheskogo Analiza* **104**, 2020, pp. 121-128.

In this paper, we construct a rather simple recurrent optimal structure filter of a large order multiple to the dimension of the estimate vector. The memory of this filter is infinite, and a contradiction between accuracy and complexity is regulated by the choice of multiplicity. The estimate vector is still sought as the best function of the last measurement and the filter state vector. But the latter is proposed to form from the vectors of several preceding estimates, thereby increasing the operative filter memory, which allows us to obtain the best accuracy by extending the admissible set of estimates. Furthermore, the preceding measurements are not forgotten because information about them is accumulated in the stored estimates.

It is shown that the synthesis of a new filter is reduced to finding in advance the corresponding conditional probability densities from the recurrent chain of prediction-correction type transformations. This is based on the Fokker–Planck–Kolmogorov differential equation and the integral Bayesian formula. The same can be done numerically by the Monte Carlo method, but by cumbersome construction of the conditional expectation function histogram at each clock point; moreover, the sought function has a large number of arguments and histogram should be further smoothed. Therefore, we also discuss construction of analytic-numerical covariance approximations of the proposed filter. The computational advantage of such approximations over similar approximations of an absolutely optimal filter is shown.

1 Statement of the Continuous–Discrete Estimation Problem

Let a Markov diffusion type process $X_t \in \mathbb{R}^n$ defining the state of an object of observation on a finite time-interval $t \in [0, T]$ be described by the Itô stochastic equation

$$dX_t = a(t, X_t)dt + B(t, X_t)dW_t, \quad X_0 \sim p_0(x_0), \tag{1.1}$$

and let its measurements $Y_k \in \mathbb{R}^m$ be obtained only at known clock points $t_0 = 0 < t_1 < \ldots < t_K < T$ by the formula

$$Y_k = c_k(X_{t_k}, V_k), \quad V_k \sim q_k(v_k), \quad k \in 0, \ldots, K. \tag{1.2}$$

Here, $a(t, x)$ and $B(t, x)$ are the vector-valued drift function and the matrix-valued diffusion function of the object of observation respectively, W_t is the standard Wiener process vector, $c_k(x, v)$ is the vector-valued measurer function, the measurement number k runs from 0 to K, and the symbol $\sim$ means the correspondence between a random variable and its probability density, so that the initial value X_0 of the process is determined by the probability density $p_0(x_0)$, whereas the independent discrete white noise vector V_k has the probability density $q_k(v_k)$. We assume that the Itô stochastic equation (1.1) has a strong solution in the form of implementation of the process X_t and this solution is unique.

On each interval between neighboring clock points, it is required to promptly obtain an estimate $\widehat{X}'_t \in \mathbb{R}^{n'}$ for the part X'_t of the first $n' \leqslant n$ components of the estimated vector X_t, which is the most interesting for the user, as some explicit or implicit Borel measurable function of all accumulated measurements at that time

$$\widehat{X}'_t = \psi_{t,k}(Y_0^k), \quad Y_0^k = \{Y_0, Y_1, \ldots, Y_k\}, \quad t \in (t_k, t_{k+1}], \quad k \in 0, \ldots, K,$$

optimal in the sense that the mean-square estimation error is minimal:

$$I_t = \mathrm{M}\left[(X'_t - \widehat{X}'_t)^T C_t (X'_t - \widehat{X}'_t)\right] \to \min_{\psi_{t,k}(\cdot)}, \quad C_t = C_t^T > 0. \tag{1.3}$$

Here, M is the expectation operator, C_t is the weighted coefficient matrix, and the existence of I_t is assumed. The sufficient existence condition is formulated in Theorem 4.1.

2 Known Continuous–Discrete Filters of Optimal Structure

Definition 2.1 (cf. [6]). A *small order filter* is a recurrent algorithm for obtaining an estimate vector that is the state vector.

The order of a small order filter is equal to the dimension n' of the estimate vector, but does not exceed the dimension n of the entire vector X_t. We distinguish three types of such filters.

1. A *small order filter with constant prediction* in the case $\widehat{X}'_t = Z_k$, $t \in [t_k, t_{k+1})$, finds only the clock estimates $Z_k = \widehat{X}'_{t_k}$ by the recurrent formula for a small order discrete filter [6]–[7]

$$Z_k = f_k(Y_k, Z_{k-1}), \quad k \geqslant 1, \quad Z_0 = f_0(Y_0). \tag{2.1}$$

   Here, each function $f_k(\cdot)$ is found from the optimality conditions (1.3) for $t = t_k$.

2. A *small order filter with continuous prediction* is described in the interval between the measurement times by the autonomous differential equation

$$d\widehat{X}'_t = g(t, \widehat{X}'_t)dt, \quad t \in (t_k, t_{k+1}]$$

   with the initial condition $\widehat{X}'_{t_0^+} = f_0(Y_0)$, and, at each clock point, its solution is corrected by the new measurement Y_k according to the formula

$$\widehat{X}'_{t_k^+} = f_k(Y_k, \widehat{X}'_{t_k}).$$

   However, in this case, from the optimality conditions (1.3) for $t = t_k$ we can obtain only the correction function $f_k(\cdot)$, whereas the continuous prediction function $g(\cdot)$ is found only from the unbiased estimate condition $\mathrm{M}\,[X'_t - \widehat{X}'_t] = 0$.

3. A *small order filter with discrete predictions* allows us, in addition to the clock estimates Z_k, to obtain the optimal predictions $Z_k^i = \widehat{X}'_{\tau_k^i}$, $i = 1, \ldots, L$, at L additional points $t = \tau_k^i$ in the interval (t_k, t_{k+1}) between the last Y_k and next Y_{k+1} measurements

$$t_k = \tau_k^0 < \tau_k^1 < \ldots < \tau_k^L < \tau_k^{L+1} = t_{k+1}.$$

   As a result, the prediction is constant only on a small interval between the prediction points

$$\widehat{X}'_t = Z_k^i, \quad t \in [\tau_k^i, \tau_k^{i+1}), \quad i = 1, \ldots, L.$$

   These estimates are obtained by using the recurrent formulas

$$Z_k^0 = Z_k, \quad Z_k^i = g_k^i(Z_k^{i-1}), \quad i = 1, \ldots, L, \quad k \geqslant 0, \tag{2.2}$$

$$Z_0 = f_0(Y_0), \quad Z_k = f_k(Y_k, Z_{k-1}^L), \quad k \geqslant 1. \tag{2.3}$$

   Here, all the correction functions $f_k(\cdot)$ as well as the prediction functions $g_k^i(\cdot)$ can be obtained from the optimality conditions (1.3) at the time t_k and τ_k^i respectively.

Remark 2.1. The small order filter with constant prediction is a particular case of the small order filter with discrete predictions in the case $L = 0$.

However, the fixed form of equations for any small order filter does not allow us to improve in any way the potential accuracy of the filter. The only changeable parameter for these filters is their order $n' \leqslant n$ bounded from above by the dimension of the estimated vector X_t. It is obvious that the filter of the same order n as the object is the most accurate among all small order filters because of a wider class of correction functions and the greatest possible number of arguments [7].

Definition 2.2 (cf. [8]). A *finite memory filter* is a discrete estimation algorithm, where the clock estimates and, possibly, inter-clock predictions are functions of only last few measurements, the number of which is given.

Thus, the clock estimates Z_k and inter-clock predictions Z_k^i estimates for a finite memory filter with discrete predictions [8] are sought as Borel functions of the last no more than $(l+1)$ measurements

$$\begin{aligned} Z_k &= f_k(Y_{\max(0,k-l)}^k), \quad k \geqslant 0, \\ Z_k^i &= g_k^i(Y_{\max(0,k-l+1)}^k), \quad i = 1, \ldots, L, \quad k \geqslant 0. \end{aligned} \tag{2.4}$$

Here, the integer parameter $l \geqslant 0$ can be changed to change the accuracy of the finite memory filter and the processing time for the last measurement by this filter. For example, for $l = 0$ we have the trivial case of dependence of all clock estimates on one measurement $Z_k = f_k(Y_k)$, $k \geqslant 0$, but for $l = 1$ only the initial estimate $Z_0 = f_0(Y_0)$ depends on one measurement, whereas the remaining ones depend on two measurements $Z_k = f_k(Y_k, Y_{k-1})$, $k \geqslant 1$. Moreover, the explicit expressions in (2.4) can be written in the following difference form.

Lemma 2.1 (cf. [8]). *The finite memory filter* (2.4) *is a recurrent filter of order lm with the fixed difference equation $U_k = s_k(Y_k, U_{k-1})$ for its state vector $U_k = Y_{\max(0,k-l+1)}^k$, and the following formulas for the output, optimal in the sense of* (1.3), *hold*:

$$\begin{aligned} &Z_0 = f_0(Y_0), \quad Z_k = f_k(Y_k, U_{k-1}), \quad k \geqslant 1, \\ &Z_k^i = g_k^i(U_k), \quad i = 1, \ldots, L, \quad k \geqslant 0. \end{aligned}$$

3 Equations for New Continuous–Discrete Large Order Filter

Definition 3.1. A *large order filter* is a recurrent algorithm for obtaining clock estimates $Z_k = \widehat{X}'_{t_k}$ and, possibly, inter-clock predictions $Z_k^i = \widehat{X}'_{\tau_k^i}$, regarded as functions of the last measurement Y_k and at most $l \in \mathbb{N}$ preceding clock estimates

$$\begin{aligned} &Z_0 = f_0(Y_0), \quad Z_k = f_k(Y_k, Z_{\max(0,k-l)}^{k-1}), \quad k \geqslant 1, \\ &Z_k^i = g_k^i(Z_{\max(0,k-l+1)}^k), \quad i = 1, \ldots, L, \quad k \geqslant 0. \end{aligned} \tag{3.1}$$

The Borel estimation function $f_k(\cdot)$ and discrete prediction function $g_k^i(\cdot)$ of this filter are also found from the optimality conditions (1.3) at the time t_k and τ_k^i respectively. Moreover, the prediction function can be left to be constant only on the first subinterval $\widehat{X}'_t = Z_k$, $t \in [t_k, \tau_k^1)$, while it can be made more accurate on other subintervals, for example, it can be linear starting from the second subinterval

$$\widehat{X}'_t = Z_k^i + (Z_k^i - Z_k^{i-1})\frac{t-\tau_k^i}{\tau_k^i - \tau_k^{i-1}}, \quad t \in [\tau_k^i, \tau_k^{i+1}), \quad i = 1, \dots, L.$$

Another more economic variant of construction of the prediction function that does not require intermediate values of Z_k^i, $i = 1, \dots, L$, is its at least linear extrapolation by two only clock estimates Z_k and Z_{k-1}, starting with the second subinterval

$$\widehat{X}'_t = Z_k + (Z_k - Z_{k-1})\frac{t-t_k}{t_k - t_{k-1}}, \quad t \in [t_k, t_{k+1}), \quad k \geqslant 1.$$

Remark 3.1. The large order filters generalizes the notion of small order filters because not one, but several preceding estimates are used to derive a new estimate. The large order filters differ from the finite memory filters by replacing the preceding measurements $Y_{\max(0,k-l)}^{k-1}$ with the estimates $Z_{\max(0,k-l)}^{k-1}$.

Lemma 3.1. *The large order filter* (3.1) *is a recurrent filter of order ln' with the fixed structure $s_k(\cdot)$ of its linear equation of state*

$$U_0 = Z_0, \quad U_k = s_k(Z_k, U_{k-1}) \triangleq \begin{cases} \begin{bmatrix} Z_k \\ U_{k-1} \end{bmatrix}, & k = 1, \dots, l-1, \\ \begin{bmatrix} Z_k \\ CU_{k-1} \end{bmatrix}, & k \geqslant l, \end{cases} \tag{3.2}$$

where $C = [E_{(l-1)n'} \ 0_{(l-1)n' \times n'}]$, E is the identity matrix, and the output formulas, optimal in the sense of (1.3)*, hold*:

$$\begin{aligned} Z_0 &= f_0(Y_0), \\ Z_k &= f_k(Y_k, U_{k-1}), \quad k \geqslant 1, \\ Z_k^i &= g_k^i(U_k), \quad i = 1, \dots, L, \quad k \geqslant 0. \end{aligned} \tag{3.3}$$

Proof. We combine the preceding estimates used in (3.1) for obtaining Z_k to the block column-vector $U_{k-1} = Z_{\max(0,k-l)}^{k-1}$ of the operative filter memory. Then these higher order recurrent dependences take the form of the final output formulas (3.3). Analyzing the filling process of this block column-vector at the next clock point

$$U_k = Z_{\max(0,k-l+1)}^k = \begin{cases} Z_0^k, & k = 0, \dots, l-1 \text{ (accumulation stage)}, \ \dim Z_0^k = n'(k+1), \\ Z_{k-l+1}^k, & k \geqslant l \text{ (renewal stage)}, \ \dim Z_{k-l+1}^k = n'l, \end{cases}$$

we obtain the first order difference equation (3.2), where C is the matrix for removing the last and already obsolete block Z_{k-l} from the vector U_{k-1} As a result, the column U_k is the state vector of the large order filter (3.1) and its dimension, unchanged starting with the lth clock $k = l - 1$, is the filter order. ☐

4 Finding Output Functions of Continuous-Discrete Large Order Filters

Substituting (3.3) into the criterion (1.3) for the corresponding time moments t_k, τ_k^i and using theorem on the best mean-square regression, we easily obtain the following assertion.

Theorem 4.1. *If the estimated vector X_t' has the finite second moment $\mathrm{M}\,|X_t'|^2 < \infty$, then the Borel output functions (3.2), (3.3), optimal in the sense of (1.3) for t_k, τ_k^i, exist and, under the assumption that the corresponding random variables are absolutely continuous, can be found as the conditional expectations calculated in terms of the corresponding conditional probability densities $\rho_k(\cdot)$, $\pi_k(\cdot)$ by the formulas*

$$
\begin{aligned}
f_0(y_0) &= \mathrm{M}\,[X_0'|y_0] = \int x_0'\rho_0(x_0|y_0)dx_0, \\
f_k(y_k, u_{k-1}) &= \mathrm{M}\,[X_{t_k}'|y_k, u_{k-1}] = \int x_k'\rho_k(x_k|y_k, u_{k-1})dx_k, \quad k \geqslant 1, \\
g_k^i(u_k) &= \mathrm{M}\,[X_{\tau_k^i}'|u_k] = \int x'\pi_k(\tau_k^i, x|u_k)dx, \quad i = 1, \ldots, L, \quad k \geqslant 1.
\end{aligned}
\tag{4.1}
$$

Moreover, the estimates Z_k and Z_k^i are unbiased, $\mathrm{M}\,[Z_k - X_{t_k}'] = 0$, $\mathrm{M}\,[Z_k^i - X_{\tau_k^i}'] = 0$, and have the finite second moments $\mathrm{M}\,|Z_k|^2 < \infty$ and $\mathrm{M}\,|Z_k^i|^2 < \infty$.

Hereinafter, for the sake of simplicity we write integrals over the whole space in the notation of indefinite integrals, for example,

$$\int f(x)dx = \int_{\mathbb{R}^n} f(x)dx.$$

We note that the assumption and last assertion of Theorem 4.1 guarantee the existence of the mean-square estimation error (1.3) at the required times.

In the first relation in (4.1), the conditional probability density is known:

$$\rho_0(x|y) = \frac{\beta_0(y|x)p_0(x)}{\displaystyle\int \text{numerator}\, dx}.$$

Hereinafter, $\beta_k(\cdot)$ denotes the likelihood function obtained from (1.2) and $p_0(x)$ is the density of the initial object state (1.1). The remaining probability densities in (4.1) are found by the Bayesian formula as follows.

Theorem 4.2. *If there exists a joint probability density $r_k(t, x, u_k)$ of the states of the object X_t and the filter U_k, then the prediction $\pi_k(\cdot)$ and correction $\rho_k(\cdot)$ probability densities in (4.1) are represented in terms of r_k by the formulas*

$$
\begin{aligned}
\pi_k(t, x|u_k) &= \frac{r_k(t, x, u_k)}{\displaystyle\int \text{numerator}\, dx} \quad t \in (t_k, t_{k+1}], \quad k \geqslant 0, \\
\rho_k(x|y_k, u_{k-1}) &= \frac{\beta_k(y_k|x)\pi_{k-1}(t_k, x|u_{k-1})}{\displaystyle\int \text{numerator}\, dx}, \quad k \geqslant 1.
\end{aligned}
\tag{4.2}
$$

Finally, let us find a sequence of probability densities $r_k(t, x, u_k)$, $k \geqslant 0$. For this purpose we use the Fokker–Planck–Kolmogorov equation for the transition probability density of the random diffusion process X_t defined by the Itô equation (1.1). Furthermore, to describe the replacement of the old state vector U_{k-1} with the new one U_k at the measurement time t_k by formulas (3.2) and (3.3), we again apply known properties of probability densities. In particular, we use the formula, convenient because of its shortness, for the probability density $q(y)$ of the result of the nonlinear transformation $Y = h(X)$ of the random variable X with probability density $p(x)$

$$q(y) = \int \delta[y - h(x)]\, p(x)dx,$$

where $\delta(\cdot)$ is the Dirac function. Then, as in [8], we obtain the following assertion.

Theorem 4.3. *If the drift function $a(t, x)$ and diffusion function $B(t, x)$ of the object of observation* (1.1) *are once and twice continuously differentiable with respect to x respectively, then the probability density $r_k(\cdot)$ satisfies the equation*

$$\frac{\partial r_k(t, x, u_k)}{\partial t} = -\nabla_x^T\{a(t, x) r_k(t, x, u_k)\} + 0.5\, \mathrm{tr}\, [\nabla_x \nabla_x^T \{B(t, x) B^T(t, x) r_k(t, x, u_k)\}] \qquad (4.3)$$

for $t \in (t_k, t_{k+1}]$, where ∇_x is the gradient operator with respect to $x \in \mathbb{R}^n$ and tr *denotes the matrix trace, and the initial conditions with known initial data in the first interval $t \in (t_0, t_1]$:*

$$r_0(t_0^+, x, u_0) = p_0(x) \int \delta[z_0 - f_0(y_0)] \beta_0(y_0|x) dy_0.$$

For each next following interval the corresponding initial condition $r_k(t_k^+, x, u_k)$ is calculated from the finite section $r_{k-1}(t_k, x, u_{k-1})$ of the solution to the same equation on the preceding interval and for already known from (4.1) *and* (4.2) *optimal clock output function $f_k(y_k, u_{k-1})$ by one of the following formulas:*

$$r_k(t_k^+, x_k, z_0^k) = \int \delta[z_k - f_k(y_k, z_0^{k-1})]\ \beta_k(y_k|x_k) r_{k-1}(t_k, x, z_0^{k-1}) dy_k, \quad k = 1, \ldots, (l-1),$$

$$r_k(t_k^+, x_k, z_{k-l+1}^k) = \iint \delta[z_k - f_k(y_k, z_{k-l}^{k-1})]\ \beta_k(y_k|x_k) r_{k-1}(t_k, x, z_{k-l}^{k-1}) dy_k dz_{k-l}, \quad k \geqslant l.$$

In the last expression, the integration with respect to z_{k-l} corresponds to removing the obsolete block Z_{k-l} from the vector $U_{k-1} = Z_{k-l}^{k-1}$ in (3.2).

Remark 4.1. The strong smoothness conditions on the drift and diffusion functions in Theorem 4.3 can be omitted if a solution to Equation (4.3) is understood as a generalized solution in the sense of Galerkin.

The relations of this section constitute a synthesis algorithm for optimal functions of a large order filter. This algorithm does not use the measurement results and, consequently, can be realized before the estimation process. This fact favorably distinguishes this and other filters of optimal structure from the classical absolutely optimal filter. This is similar to the advantage of the linear Kalman filter over its nonlinear generalizations since its measurement gain matrix can be found in advance by solving the Riccati matrix equation for covariances on a powerful computer long enough.

5 Covariance Approximations of Continuous–Discrete Large Order Filters

By the above-mentioned connection between large order filters and finite memory filters, the Gaussian approximations of the numerators of both fractions in (4.2) leads, as in [8], to the following assertion.

Proposition 5.1. *If the measurements Y_k have the finite second moment $\mathrm{M}\,|Y_k|^2 < \infty$ and the conditional means of the sensor* (1.2)

$$\nu_k(x) = \mathrm{M}\,[c_k(x, V_k)] \triangleq \int c_k(x, v) q_k(v) dv, \quad \Pi_k(x) = \mathrm{M}\,[c_k(x, V_k) c_k^{\mathrm{T}}(x, V_k)]$$

have the Gaussian moments

$$\begin{aligned} h_k(m, D) &= \mathrm{M}\,{}_N^{m,D}[\nu_k(X)] \triangleq \int \nu_k(x) N(x||m, D) dx, \\ F_k(m, D) &= \mathrm{M}\,{}_N^{m,D}[\Pi_k(X)] - h_k(m, D) h_k^{\mathrm{T}}(m, D), \\ G_k(m, D) &= \frac{\partial h_k(m, D)}{\partial m}, \end{aligned} \tag{5.1}$$

where $N(x||m, D)$ is the Gaussian distribution density of the random variable X with parameters m, D, then the following equations of a suboptimal large order filter hold:

$$\begin{aligned} &Z_0 = H_0 Y_0 + e_0, \quad U_0 = Z_0, \quad U_k = s_k(Z_k, U_{k-1}), \quad \Lambda_k = \Gamma_k U_{k-1} + \varkappa_k, \\ &Z_k = \Lambda_k' + T_k' G_k^T(\Lambda_k; T_k) F_k^{\oplus}(\Lambda_k; T_k)[Y_k - h_k(\Lambda_k; T_k)], \quad Z_k^i = \Gamma_k^i U_k + \varkappa_k^i, \quad i = 1, \dots, L, \end{aligned} \tag{5.2}$$

where $k \geqslant 1$, Λ_k' and T_k' are the first n' rows of the matrices Λ_k and T_k respectively, $\oplus$ denotes the Moore–Penrose pseudoinversion symbol for matrices, and the function $s_k(\cdot)$ is defined in (3.2). *Moreover, the initial H_0, e_0 and current parameters, the estimates Γ_k, $\varkappa_k$, T_k, and predictions Γ_k^i, $\varkappa_k^i$, are expressed only in terms of the first two moments m, D of known random variables by the formula*

$$\begin{aligned} &H_0 = D_{0\ 0}^{x'\ y}(D_0^y)^{\oplus}, \quad e_0 = m_0^{x'} - H_0 m_0^y, \\ &\Gamma_k = D_{t_k, k-1}^{x,u}(D_{k-1}^u)^{\oplus}, \quad \varkappa_k = m_{t_k}^x - \Gamma_k m_{k-1}^u, \quad T_k = D_{t_k}^x - \Gamma_k (D_{t_k, k-1}^{x,u})^T, \\ &\Gamma_k^i = D_{\tau_k^i, k}^{x', u}(D_k^u)^{\oplus}, \quad \varkappa_k^i = m_{\tau_k^i}^{x'} - \Gamma_k^i m_k^u, \quad i = 1, \dots, L. \end{aligned} \tag{5.3}$$

Since the calculation of sampled values of the parameters m and D in (5.3) does not cause difficulties, these parameters of the covariance large order filter (5.2) can be easily found in advance by the Monte Carlo method. For this purpose one can perform the trivial processing of the results of statistical modeling of the object (1.1), sensor (1.2), and filter (5.2).

As in the case of the finite memory filter, the following assertion holds [8].

Proposition 5.2. *If the sensor* (1.2) *can be linearized at the point of clock prediction Λ_k and the mean value of its noise m_k^v, i.e.,*

$$c_k(X_{t_k}, V_k) \approx c_k(\Lambda_k, m_k^v) + C_k^x(\Lambda_k)(X_{t_k} - \Lambda_k) + C_k^v(\Lambda_k)(V_k - m_k^v),$$

where $C_k^x(x)$ *and* $C_k^v(x)$ *denote the sections of the Jacobi matrix of partial derivatives of the sensor function*

$$C_k^x(x) = \left.\frac{\partial c_k(x,v)}{\partial x}\right|_{v=m_k^v}, \quad C_k^v(x) = \left.\frac{\partial c_k(x,v)}{\partial v}\right|_{v=m_k^v},$$

then the Gaussian correction functions (5.1) *are approximated by*

$$\begin{aligned} &h_k(m,D) \approx c_k(m,m_k^v), \quad G_k(m,D) \approx C_k^x(m), \\ &F_k(m,D) \approx C_k^x(m) D C_k^{x\mathrm{T}}(m) + C_k^v(m) R_k C_k^{v\mathrm{T}}(m), \end{aligned} \tag{5.4}$$

where $R_k = \operatorname{cov}[V_k]$ *is the covariance matrix for the measurement noise.*

Thus, both covariance approximations of the large order filter have the same equation (5.2) and the same formula (5.3) for parameters, but the correction functions are defined by (5.1) for the Gaussian filter and by (5.4) for the linearized one. Moreover, the adjustment process of the approximate large order filter on the solution to a particular estimation problem consists in not only obtaining analytically the structure functions (5.1) or (5.4) from the models of object (1.1) and sensor (1.2), but also it is required to find the parameters of the filter (5.3). The latter can be easily implemented by numerical successive statistical simulation of the values of the signal X_t at clock points and, possibly, inter-clock times, as well as the values of the measurement Y_k and the filter state vector U_{k-1} consisting of the preceding estimates $Z_{\max(0,k-l)}^{k-1}$ with further processing the results of this simulation by simple formulas of mathematical statistics.

Remark 5.1. Covariance approximations similar to (5.2) (the generalized Kalman filter and the filter of normal approximation) of an absolutely optimal filter can have the suboptimal estimate $\widehat{X}_t$ only for the whole object state vector (1.1) and only together with the positive definite approximation P_t of the posterior covariance of the estimation error [1, 2, 8]. For this purpose it is necessary to alternate integration of the autonomous prediction system of ordinary differential equations on each interval between the measurement times

$$\begin{aligned} &\frac{d}{dt}\widehat{X}_t = \tau(t,\widehat{X}_t,P_t), \quad t \in (t_k, t_{k+1}], \\ &\frac{d}{dt}P_t = A(t,\widehat{X}_t,P_t)P_t + P_t A^T(t,\widehat{X}_t,P_t) + \Theta(t,\widehat{X}_t,P_t), \quad t \in (t_k, t_{k+1}], \end{aligned} \tag{5.5}$$

with recalculation of the final values of their solutions on the preceding interval $\widehat{X}_{t_k}, P_{t_k}$ into the initial data for the next one by the correction formulas

$$\widehat{X}_{t_k^+} = \widehat{X}_{t_k} + H_k[Y_k - h_k(\widehat{X}_{t_k}, P_{t_k})], \quad P_{t_k^+} = P_{t_k} - H_k G_k(\widehat{X}_{t_k}, P_{t_k}) P_{t_k},$$

where the current measurement Y_k is taken into account, $H_k = P_{t_k} G_k^T(\widehat{X}_{t_k},\ P_{t_k}) F_k^{\oplus}(\widehat{X}_{t_k}, P_{t_k})$, and the initial values of variables $\widehat{X}_{t_0} = \mathrm{M}[X_0]$ and $P_{t_0} = \operatorname{cov}[X_0]$ are deterministic. Here, the correction functions $h_k(\cdot)$, $G_k(\cdot)$, $F_k(\cdot)$ are also defined by (5.1) or (5.4), and the additional prediction functions $\tau(\cdot)$, $\Theta(\cdot)$, $A(\cdot)$ are found in a similar way, but by using the object functions $a(\cdot)$, $B(\cdot)$. However, we have to solve the differential equations (5.5) with small step and the estimate vector, together with different entries of the symmetric covariance matrix, constitutes a rather large state vector of the filter, so that the orders of the generalized Kalman filter and the filter of normal approximation are equal to $n(n+3)/2$.

Acknowledgments

The work is supported by the Russian Foundation for Basic Research (project No. 17-08-00530-a).

References

1. A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York etc. (1970).

2. M. S. Yarlykov and M. A. Mironov, *Markov Theory of Estimation of Stochastic Processes* [in Russian], Radio Svyaz', Moscow (1993).

3. A. Budhiraja, L. Chen, and C. Lee, "A survey of numerical methods for nonlinear filtering problems," *Physica D* **230**, No. 1–2, 27–36 (2007).

4. V. S. Pugachev, I. N. Sinitsyn, and V. I. Shin, "Conditionally optimal discrete filtering of processes in continuous-discrete stochastic systems," *Sov. Math., Dokl.* **34**, 88–92 (1987).

5. I. N. Sinitsyn, *Kalman and Pugachev Filters* [in Russian], Logos, Moscow (2007).

6. E. A. Rudenko, "Low-order discrete linear filters: Their optimal structure," *Autom. Remote Control* **60**, No. 9, 1261–1272 (1999).

7. E. A. Rudenko, "Optimal discrete nonlinear filters of the object's order and their Gaussian approximations," *Autom. Remote Control* **71**, No. 2, 320–338 (2010).

8. E. A. Rudenko, "Optimal continuous-discrete nonlinear finite memory filter with discrete predictions," J. Comput. Syst. Sci. Int. **55**, No. 6, 878–893 (2016).

Submitted on November 7, 2019