ISSN 1064-2307, Journal of Computer and Systems Sciences International, 2016, Vol. 55, No. 1, pp. 36–58. © Pleiades Publishing, Ltd., 2016.
Original Russian Text © E.A. Rudenko, 2016, published in Izvestiya Akademii Nauk. Teoriya i Sistemy Upravleniya, 2016, No. 1, pp. 43–65.

DATA PROCESSING AND IDENTIFICATION

Finite-Dimensional Recurrent Algorithms for Optimal Nonlinear Logical–Dynamical Filtering

E. A. Rudenko

Moscow Aviation Institute (National Research University), Volokolamskoe shosse 4, Moscow, 125993 Russia
e-mail: rudenkoevg@yandex.ru

Received April 20, 2015; in final form, June 2, 2015

Abstract—The problem of the most accurate estimation of the current state of a multimode nonlinear dynamic observation system with discrete time based on indirect measurements of this state is considered. The general case when a mode indicator is available and the measurement errors depend on the plant disturbances is investigated. A comparative analysis of two known approaches is performed—the conventional absolutely optimal one based on the use of the posterior probability distribution, which requires the use of an unimplementable infinite-dimensional estimation algorithm, and a finite-dimensional optimal approach, which produces the best structure of the difference equation of a low-order filter. More practical equations for the Gaussian approximations of these two optimal filters are obtained and compared. In the case of the absolutely optimal case, such an approximation is finite-dimensional, but it differs from the approximation of the finite-dimensional optimal version in terms of its considerably larger dimension and the absence of parameters. The presence of parameters, which can be preliminarily calculated using the Monte-Carlo method, allows the Gaussian finite-dimensional optimal filter to produce more accurate estimates.

DOI: 10.1134/S1064230715060131

INTRODUCTION

The fundamental difficulty of the classical approach to the optimally accurate nonlinear estimation of the current state of stochastic logical–dynamical (LD) observation systems, which are also called Markov jump parameter systems, is well known. This difficulty is due to the infinite dimensionality of the Stratonovich absolutely optimal filter (AOF) [1–7]. The state vector of this filter consists of all posterior moments (sufficient statistics) of the state to be estimated. For this reason, the AOF cannot be implemented in real time, and implementable finite-dimensional approximations such as the fairly coarse bank of quasi-linear extended Kalman filters or more accurate normal approximation filters (NAFs) have to be used. Recently, the difficult to implement particle filter, which is an implementation of the AOF along one trajectory of the estimated stochastic process, has become popular. This filter uses the sequential Monte-Carlo method [5, 7] for calculating the integrals in the process of filtering.

The alternative Pugachev conditionally optimal filter [8, 9] makes it possible to construct nonlinear filters of the given structure with optimal parameters that are finite-dimensional from the very beginning. The nonparametric developments of this approach are *optimal structure filters* (OSFs) [10–14]. In contrast to AOF, the OSF's state vector is finite-dimensional by definition, and the Monte-Carlo method is used to construct the OSF before the estimation process is started.

The present paper is devoted to the generalization of the known LD versions of AOF [1, 4] and OSF [10] for the case of the statistical dependence of measurement errors on the plant disturbances, to the construction of Gaussian approximations of these filters, and to the analytical comparison of these approximations. Equations of the AOF and OSF and of their Gaussian approximations are presented. It is shown that these approximations use identical structure functions in the form of characteristics of the statistical linearization of the system nonlinearities, but they considerably differ in the amount of computer memory needed. Due to the lower order of the OSF, its Gaussian approximation has numerical parameters, which are found in advance using the Monte-Carlo method, thus adjusting the filter to the specific observation system better than the a priory NAF equations.

1. STATEMENT OF THE FILTERING PROBLEM

Let the logic of switching the observation plant structure (mode) and its operational dynamics in each mode be described in the discrete time $k = 0,1,2,\ldots$ by the Markov system of stochastic difference equations

$$I_{k+1} = \phi_k^{(I_k)}(X_k,V_k^{\triangle}), \quad X_{k+1} = a_k^{(I_k)}(X_k,V_k); \tag{1.1}$$

and let the indirect (incomplete and (or) inaccurate) structure indicator and state vector measurement of the plant be determined by explicit formulas

$$J_k = \mathrm{o}_k^{(I_k)}(X_k,W_k^{\triangle}), \quad Y_k = b_k^{(I_k)}(X_k,W_k). \tag{1.2}$$

Here, $I_k \in \{\overline{1,L}\}$ is the structure index; $X_k \in \mathbb{R}^n$ is the state vector; $J_k \in \{\overline{1,M}\}$ is the indicator variable; $Y_k \in \mathbb{R}^{n_y}$ is the vector of measurements; $V_k^{\triangle}$, V_k, $W_k^{\triangle}$, W_k are discrete white noises with a known joint distribution function; and $\phi_k^{(i)}(x,v^{\triangle})$, $\mathrm{o}_k^{(i)}(x,w^{\triangle})$ and $a_k^{(i)}(x,v)$, $b_k^{(i)}(x,w)$ are given pairs of Borel measurable deterministic functions with integer scalar and continuous vector values, respectively. The white noises are independent of the random initial state of the plant (I_0, X_0) determined by its joint distribution function

$$p_0(i,x) = P_0^{(i)} p_0^{(i)}(x),$$

where $P_0^{(i)} = \Pr\left[I_0 = i\right]$ is the probability of the discrete component I_0 and $p_0^{(i)}(x) = p_0(x|I_0 = i)$ is the conditional probability density of the absolutely continuous component X_0.

Note that Eqs. (1.1) and (1.2) are generalizations of the mathematical model of the ordinary (purely dynamic) observation system [3, 5, 7, 8], which does not include the discrete variables I_k and J_k. The discrete components require the extended vector state $\overline{X}_k = (I_k, X_k)$, measurement $\overline{Y}_k = (J_k, Y_k)$, plant disturbances $\overline{V}_k = (V_k^{\triangle}, V_k)$, and measurement disturbances $\overline{W}_k = (W_k^{\triangle}, W_k)$.

We want to obtain at each time $k \geq 0$ the optimal estimates of the mode index $\hat{I}_k$ and the plant's state vector $\hat{X}_k$ as functions of the whole set of pairs of measurements (of increasing size) $\overline{Y}_0^k = (\overline{Y}_0, \ldots, \overline{Y}_k)$ that is available by this time:

$$\hat{I}_k = e_k(\overline{Y}_0^k), \quad \hat{X}_k = g_k(\overline{Y}_0^k). \tag{1.3}$$

The criterion of the best estimation accuracy is the minimum of the Bayesian mean risk

$$\mathrm{M}\left[c_k^{(I_k,\hat{I}_k)}(X_k,\hat{X}_k)\right] \to \min, \tag{1.4}$$

where M is the expectation operator; and $c_k^{(i,\hat{i})}(x,\hat{x})$ is the element of the matrix function of joint losses due to the incorrect identification of the structure index I_k and inaccurate estimation of the state vector X_k; this function ensures that the structure estimate $\hat{I}_k$ is an integer and the state vector estimate $\hat{X}_k$ is continuous [15]. For definiteness, we will use the two most popular loss functions below—*the additive quadratically simple* function

$$c_k^{(i,\hat{i})}(x,\hat{x}) = (x-\hat{x})^{\mathrm{T}} C_k (x-\hat{x}) + (1-\delta_{i,\hat{i}}) \tag{1.5}$$

and *the simple multiplicative function*

$$c_k^{(i,\hat{i})}(x,\hat{x}) = 1 - \delta_{i,\hat{i}}\, \delta(x-\hat{x}). \tag{1.6}$$

Here, $C_k = C_k^{\mathrm{T}} > 0$ is the weighting coefficient matrix, $\delta_{i,l}$ is the Kronecker delta, and $\delta(x)$ is the Dirac delta function.

Instead of the switching functions $\phi_k^{(i)}(\cdot)$ and indication functions $\mathrm{o}_k^{(i)}(\cdot)$, equivalent conditional probabilities of the Markov chain consisting of the structure indices I_k, its indicators J_k

$$\mathrm{A}_k^{(l|i)}(x) = \Pr\left[I_{k+1} = l | I_k = i, X_k = x\right], \quad \mathrm{B}_k^{(j|i)}(x) = \Pr\left[J_k = j | I_k = i, X_k = x\right],$$

and the switching probability improved by using the last measurement $\bar{Y}_k$

$$Z_k^{(l|ij)}(x,y) = \Pr\left[I_{k+1} = l | I_k = i, X_k = x, J_k = j, Y_k = y\right]$$

may be specified. The first two probabilities can be represented by the formulas

$$A_k^{(l|i)}(x) = \int \alpha_k(l,\chi|i,x)d\chi, \quad B_k^{(j|i)}(x) = \int \beta_k(j,y|i,x)dy \tag{1.7}$$

in terms of the known (see [13, 14]) conditional distributions of *switching–transition* of plant (1.1) from the current state $\bar{X}_k$ to the new state $\bar{X}_{k+1}$

$$\alpha_k(l,\chi|i,x) = \mathrm{M}\left[\delta_{l,\phi_k^{(i)}(x,V_k^{\triangle})}\delta\left(\chi - a_k^{(i)}(x,V_k)\right)\right], \quad l = i_{k+1}, \quad \chi = x_{k+1} \tag{1.8}$$

and the *indication–measurement* $\bar{Y}_k$ of its current state

$$\beta_k(j,y \mid i,x) = \mathrm{M}\left[\delta_{j,o_k^{(i)}(x,W_k^{\triangle})}\delta\left(y - b_k^{(i)}(x,W_k)\right)\right]. \tag{1.9}$$

In (1.7) and below, all the *integrals are definite*—they are taken over the entire Euclidean space of the corresponding dimension. Taking into account (1.8) and (1.9), probabilities (1.7) can be found directly as

$$A_k^{(l|i)}(x) = \mathrm{M}\left[\delta_{l,\phi_k^{(i)}(x,V_k^{\triangle})}\right], \quad B_k^{(j|i)}(x) = \mathrm{M}\left[\delta_{j,o_k^{(i)}(x,W_k^{\triangle})}\right]. \tag{1.10}$$

The third probability is determined similarly as $Z_k^{(l|ij)}(x,y) = \int \zeta_k(l,\chi|i,x,j,y)d\chi$ in terms of the improved (taking into account the last measurement $\bar{Y}_k$) conditional distribution of the next plant state $\bar{X}_{k+1}$

$$\zeta_k(l,\chi|i,x,j,y) = \gamma_k(l,\chi,j,y|i,x)/\beta_k(j,y \mid i,x), \tag{1.11}$$

where the conditional distribution of the new state and the current measurement $\gamma_k(\cdot)$ is

$$\gamma_k(l,\chi,j,y|i,x) = \mathrm{M}\left[\delta_{l,\phi_k^{(i)}(x,V_k^{\triangle})}\delta\left(\chi - a_k^{(i)}(x,V_k)\right)\delta_{j,\,o_k^{(i)}(x,W_k^{\triangle})}\delta\left(y - b_k^{(i)}(x,W_k)\right)\right]. \tag{1.12}$$

The distributions $\alpha_k(\cdot)$ and $\beta_k(\cdot)$ are partial (marginal) with respect to $\gamma_k(\cdot)$. Using (1.12), we obtain an explicit representation of the third probability in terms of functions of the observation system

$$Z_k^{(l|ij)}(x,y) = \mathrm{M}\left[\delta_{l,\phi_k^{(i)}(x,V_k^{\triangle})}\,\delta_{j,o_k^{(i)}(x,W_k^{\triangle})}\delta\left(y - b_k^{(i)}(x,W_k)\right)\right]\Big/\beta_k(j,y|i,x). \tag{1.13}$$

R e m a r k. Finding expectations such as (1.12) that include the Dirac delta function with nonlinear supports $a_k^{(i)}(\cdot)$ and $b_k^{(i)}(\cdot)$ is reduced to the parametric (for arbitrary i and x) application of the known solution of the continuous nonlinear transformation of the random variables V_k and W_k with the given probability distribution. The construction of this solution requires finding the transformation monotonicity domains, summing over these domains, finding the inverse of each monotone transformation, and calculating their Jacobians.

2. ABSOLUTELY OPTIMAL FILTER

Consider the classical procedure [1, 4] for finding the optimal form of nonrecurrent dependences (1.3) of estimates on measurements to compare it with the method for building alternative recurrent finite-dimensional dependences [10] described in Section 4 and with the two-step analog of the latter method [13].

2.1. Relationship of the Optimal Estimates with the Posterior Distribution

Plug the explicit dependences (1.3) of the estimates on measurements into the optimality criterion (1.4) to conclude that finding the estimate functions $e_k(\cdot)$, $g_k(\cdot)$ reduces to the minimization of the average risk functional on the set of all measurable functions:

$$\mathrm{M}\left[c_k^{\left(I_k, e_k(\bar{Y}_0^k)\right)}\left(X_k, g_k(\bar{Y}_0^k)\right)\right] \to \min_{e_k(\cdot),\, g_k(\cdot)}, \quad k \ge 0.$$

In order to find this minimum, it suffices to minimize the *posterior risk* function

$$\mathfrak{R}_k^{(e)}(g|\bar{y}_0^k) = \sum_i \int c_k^{(i,e)}(x,g)\rho_k(i,x|\bar{y}_0^k)dx \to \min_{e\in\overline{1,L},\ g\in\mathbb{R}^n} \tag{2.1}$$

with respect to its integer parameter e and continuous parameter g. Here and in what follows, the summation is over all the values of the index, e.g., over $i = \overline{1,L}$, and $\rho_k(\cdot)$ is the posterior distribution of the structure index I_k and the state vector X_k, which directly takes into account the possible values of all accumulated measurements $\bar{y}_0^k$ whose number increases with time.

To this end, in the case of an arbitrary loss function $c_k^{(\cdot)}(\cdot)$, the conditionally optimal estimates of the state vector under the assumption that each ith mode is true are found:

$$\hat{X}_k^{(i)} = g_k^{(i)}(\bar{Y}_0^k) \in \operatorname{Arg}\min_{g\in\mathbb{R}^n} \mathfrak{R}_k^{(i)}(g|\bar{Y}_0^k), \quad i = \overline{1,L}. \tag{2.2}$$

Knowing the conditionally optimal estimates, the estimate of the mode index and the unconditional estimate of the state vector are found as

$$\hat{I}_k = e_k(\bar{Y}_0^k) \in \operatorname{Arg}\min_{i=\overline{1,L}} \mathfrak{R}_k^{(i)}(\hat{X}_k^{(i)}|\bar{Y}_0^k), \quad \hat{X}_k = \hat{X}_k^{(\hat{I}_k)}. \tag{2.3}$$

In the particular case of the quadratically simple function (1.5), formulas (2.2) and (2.3) have the form (see [13, 15])

$$\hat{X}_k^{(i)} = \int x\rho_k^{(i)}(x|\bar{Y}_0^k)dx, \quad \hat{I}_k \in \operatorname{Arg}\max_{i=\overline{1,L}} \mathrm{P}_k^{(i)}(\bar{Y}_0^k), \quad \hat{X}_k = \sum_i \mathrm{P}_k^{(i)}(\bar{Y}_0^k)\,\hat{X}_k^{(i)}; \tag{2.4}$$

while in the case of the simple multiplicative function (1.6) more complicated expressions (see [13, 15])

$$\hat{X}_k^{(i)} \in \operatorname{Arg}\max_{x\in\mathbb{R}^n} \rho_k^{(i)}(x|\bar{Y}_0^k), \quad \hat{I}_k \in \operatorname{Arg}\max_{i=\overline{1,L}}\left[\mathrm{P}_k^{(i)}(\bar{Y}_0^k)\,\rho_k^{(i)}(\hat{X}_k^{(i)}|\bar{Y}_0^k)\right], \quad \hat{X}_k = \hat{X}_k^{(\hat{I}_k)} \tag{2.5}$$

hold. In these relations, the posterior probability of mode $\mathrm{P}_k^{(i)}(\cdot)$ and the conditionally posterior probability density $\rho_k^{(i)}(\cdot)$ of the plant state under the assumption the ith mode is true, which represent the posterior distribution $\rho_k(\cdot)$ as the product

$$\rho_k(i,x|\bar{y}_0^k) = \mathrm{P}_k^{(i)}(\bar{y}_0^k)\rho_k^{(i)}(x|\bar{y}_0^k), \tag{2.6}$$

are given in terms of this distribution by the formulas

$$\mathrm{P}_k^{(i)}(\bar{y}_0^k) = \int \rho_k(i,x|\bar{y}_0^k)dx,\ \rho_k^{(i)}(x|\bar{y}_0^k) = \rho_k(i,x|\bar{y}_0^k)\big/\mathrm{P}_k^{(i)}(\bar{y}_0^k). \tag{2.7}$$

2.2. Equation for the Posterior Distribution

In this section, we derive the well-known Stratonovich–Bukhalev [4] recurrent formula for the general case of dependent plant and measurer disturbances. To this end, we represent the posterior distribution $\rho_k(\cdot)$ using the Bayes formula

$$\rho_k(i,x|j,y,\bar{y}_0^{k-1}) = \omega_k(i,x,j,y|\bar{y}_0^{k-1})\Big/\sum_i \int numerator\ dx, \quad k \ge 1. \tag{2.8}$$

Here, the distribution $\omega_k(\cdot)$ predicts the joint behavior of the current states $\bar{X}_k$ and the observation $\bar{Y}_k$ based on the preceding measurements $\bar{Y}_0^{k-1}$ and the symbol *numerator* in the denominators of fractions everywhere denotes their numerators. Using the probability multiplication theorem, the numerator in (2.8) can be written as the product $\omega_k(i,x,j,y|\bar{y}_0^{k-1}) = \widehat{\beta}_k(j,y|i,x,\bar{y}_0^{k-1})\pi_k(i,x|\bar{y}_0^{k-1})$, which predicts the distributions of the state $\pi_k(\cdot)$ and the corresponding conditional distribution $\widehat{\beta}_k(\cdot)$. Since measurer (1.2) is conditionally Markovian, the latter distribution is independent of old measurements; therefore, it coincides with the known prior distribution (1.9). Hence, the numerator in ratio (2.8) is

$$\omega_k(i,x,j,y|\bar{y}_0^{k-1}) = \beta_k(j,y|i,x)\pi_k(i,x|\bar{y}_0^{k-1}), \tag{2.9}$$

and the distribution $\pi_k(\cdot)$ with respect to the ratio $\omega_k(\cdot)$ is marginal:

$$\pi_k(i,x|\bar{y}_0^{k-1}) = \sum_j \int \omega_k(i,x,j,y|\bar{y}_0^{k-1})dy. \tag{2.10}$$

Now we represent the current predictive distribution $\pi_k(\cdot)$ in terms of the posterior distribution of the preceding step $\rho_{k-1}(\cdot)$ (for convenience, we make a one step forward shift). Using the consistency property of the distributions and multiplication of the distributions, we represent the former (shifted by one step) as

$$\pi_{k+1}(l,\chi|\bar{y}_0^k) = \sum_i \int \hat{\zeta}_k(l,\chi|i,x,\bar{y}_0^k)\rho_k(i,x|\bar{y}_0^k)dx.$$

Here, due to the Markov property of the system of random variables $\bar{X}_{k+1}$, $\bar{Y}_k$ determined by distribution (1.12), the conditional distribution $\hat{\zeta}_k(\cdot)$ is independent of the old part $\bar{y}_0^{k-1}$ of all its conditions $\bar{y}_0^k = (j,y,\bar{y}_0^{k-1})$, and it coincides with distribution (1.11): $\hat{\zeta}_k(l,\chi|i,x,\bar{y}_0^k) = \zeta_k(l,\chi|i,x,j,y)$. Therefore, the next predicting distribution is represented in terms of the current posterior distribution as

$$\pi_{k+1}(l,\chi\,|\,j,y,\bar{y}_0^{k-1}) = \sum_i \int \zeta_k(l,\chi\,|\,i,x,j,y)\rho_k(i,x\,|\,j,y,\bar{y}_0^{k-1})dx. \tag{2.11}$$

As a result, plugging (2.9) into (2.8) and taking (2.11) into account, we obtain the following theorem.

Theorem 1. The Stratonovich–Bukhalev integral recurrent formula for the posterior distribution generalized for dependent disturbances of the LD observation system (1.1), (1.2) is

$$\rho_{k+1}(l,\chi|\,j_0^{k+1},y_0^{k+1}) = \frac{\beta_{k+1}(j_{k+1},y_{k+1}|l,\chi)\sum_i \int \zeta_k(l,\chi|i,x,j,y)\rho_k(i,x|\,j_0^k,y_0^k)dx}{\sum_l \int numerator\ d\chi}. \tag{2.12}$$

The initial condition for this formula is the similarly derived distribution

$$\rho_0(i,x|\,j,y) = \beta_0(j,y|i,x)p_0(i,x)\Big/\sum_i \int numerator\ dx, \tag{2.13}$$

where $p_0(i,x)$ is the known distribution of the initial state of the plant.

In the particular case of statistical independence of the plant disturbance $\bar{V}_k$ on the noise measurer $\bar{W}_k$, the prior distribution $\zeta_k(\cdot)$ degenerates, according to (1.11) and (1.12), into the switching–transition distribution (1.8). Then, (2.11) takes the simpler form

$$\pi_{k+1}(l,\chi|\bar{y}_0^k) = \sum_i \int \alpha_k(l,\chi|i,x)\rho_k(i,x|\bar{y}_0^k)dx;$$

the numerator in (2.12) changes similarly. Such an independent version of the LD-extension of the recurrent Stratonovich formula [3] was given by Bukhalev in [4] for the case of the inertial measurer, which is more general than (1.2); in the absence of the structure indicator and inertia-free measurements, it was obtained by Klekis in [1].

2.3. Practical Implementation of the Filter

We consider three implementation methods.

1. Finding filtering functions. The difference equation (2.12) and the relations for the optimal estimates (2.2) and (2.3) completely solve the problem of AOF synthesis. Indeed, knowing the prior distributions (1.9) and (1.11), as well as the initial posterior distribution (2.13), one can theoretically find the posterior distribution $\rho_k(\cdot)$ for any $k \geq 1$ as a function of all its arguments using Eq. (2.12). By solving two finite-dimensional optimization problems (2.2) and (2.3), we determine the filtering functions $e_k(\cdot)$, $g_k(\cdot)$ of the growing number of arguments. These computations must be performed in the sequence

$$\omega_0(\cdot) \to \ldots \to \omega_k(\cdot) \to \rho_k(\cdot) \to \pi_{k+1}(\cdot) \to \omega_{k+1}(\cdot) \to \ldots \tag{2.14}$$
$$\searrow e_k(\cdot), g_k(\cdot)$$

and they begin with the known function

$$\omega_0(i,x,j,y) = \beta_0(j,y|i,x)p_0(i,x). \tag{2.15}$$

These computations should be performed before starting to process measurements in the AOF synthesis.

The actual filtering process, which must be performed in real time as the new measurements $\overline{Y}_k$ arrive, is reduced to their accumulation in a memory array and to the straightforward computation at each step k of the estimates $\hat{I}_k$, $\hat{X}_k$ using (1.3) by plugging the entire measurement sample $\overline{Y}_0^k$ into the known functions $e_k(\cdot)$, $g_k(\cdot)$ of $(k+1)(m+1)$ arguments. The expressions for these functions should also be kept in the filter memory.

However, such a hypothetical filter is almost impossible to implement in real time if the number of steps k is large due to the rapid increase in the size of the measurement sample $\overline{Y}_0^k$ and the complexity of memorizing nonlinear functions of the growing number of variables. These difficulties are due to the fact that formulas (1.3) for absolutely optimal estimates are not recurrent.

2. Trajectory-by-trajectory filtering. Another method for the practical implementation of the AOF with automatic accumulation of measurements can be obtained by using the Stratonovich equation (2.12) directly in the process of measurement processing. To this end, we plug into it a specific random sample $\overline{Y}_0^k$ and use the notation $\tilde{\rho}_k(i,x) = \rho_k(i,x|J_0^k,Y_0^k)$ for the corresponding realization of the posterior distribution; this realization is a function of the fixed number of arguments $n+1$. Then, Eq. (2.12) and its initial condition (2.13) for each realization have the form

$$\tilde{\rho}_{k+1}(l,\chi) = \beta_{k+1}(J_{k+1},Y_{k+1}|l,\chi)\sum_i \int \zeta_k(l,\chi|i,x,j,y)\tilde{\rho}_k(i,x)dx \Big/ \sum_l \int numerator\ d\chi,$$

$$\tilde{\rho}_0(i,x) = \beta_0(J_0,Y_0|i,x)p_0(i,x) \Big/ \sum_i \int numerator\ dx.$$

As a result, we obtain random realizations of probability and density (2.7)

$$\tilde{P}_k^{(i)} = \int \tilde{\rho}_k(i,x)\,dx, \quad \tilde{\rho}_k^{(i)}(x) = \tilde{\rho}_k(i,x)/\tilde{P}_k^{(i)},$$

which allows us to find the quadratically simple estimates using (2.4)

$$\hat{X}_k^{(i)} = \int x\,\tilde{\rho}_k^{(i)}(x)\,dx, \quad \hat{I}_k \in \operatorname{Arg}\max_{i=\overline{1,L}} \tilde{P}_k^{(i)}, \quad \hat{X}_k = \sum_i \tilde{P}_k^{(i)}\hat{X}_k^{(i)}$$

or

the simple multiplicative estimate using (2.5)

$$\hat{X}_k^{(i)} \in \operatorname{Arg}\max_{x\in\mathbb{R}^n} \tilde{\rho}_k^{(i)}(x), \quad \hat{I}_k \in \operatorname{Arg}\max_{i=\overline{1,L}} \left[\tilde{P}_k^{(i)}\tilde{\rho}_k^{(i)}(\hat{X}_k^{(i)})\right], \quad \hat{X}_k = \hat{X}_k^{(\hat{I}_k)}.$$

However, computations by these formulas, among which evaluation of $3L$ n-dimensional integrals is the most costly part, in real time when receiving new measurements requires a powerful computer for the accurate implementation of the AOF. Evaluation of these integrals by the Monet Carlo method yields a particle filter [4, 7].

3. The method of sufficient statistics. The posterior distributions of the discrete I_k and continuous X_k random variables are replaced with a finite set of posterior probabilities $P_k^{(i)}(\overline{y}_0^k)$ of the discrete variable and with an infinite set of conditional numerical characteristics of the continuous variable (its sufficient statistics in the form of conditional moments, semi-invariants, quasi-moments, etc.). This allows us to obtain an equivalent filtering algorithm in the form of an infinite system of difference equations for the random values of these variables such as the probabilities $\tilde{P}_k = P_k^{(i)}(\overline{Y}_0^k)$ and their relations to the optimal estimates. An advantage of the method of sufficient statistics is that the infinite sets of semi-invariants or quasi-moments can be truncated to obtain a finite-dimensional approximation of the exact but infinite-dimensional filter. A simple truncation of this kind with the minimally acceptable number of sufficient statistics is the Gaussian approximation discussed in Section 3. A more general and hence more accurate and sophisticated method on non-Gaussian two-moment parametric approximation of probability densities is described in [4].

3. THE BANK OF NORMAL APPROXIMATION FILTERS

In this section, we obtain the Gaussian approximation of the LD AOF determined by the truncation of the infinite number of sufficient statistics of the state vector X_k so that only the first two conditional semi-invariants that coincide with the conditional expectation (mean value) and covariance remain. This corresponds to the approximation of the posterior distribution $\rho_k(\cdot)$ by the product of the probability $P_k^{(i)}(\cdot)$ and the Gaussian approximation of the density $\rho_k^{(i)}(\cdot)$ similarly to (2.6):

$$\rho_k(i, x|\bar{y}_0^k) \approx P_k^{(i)}(\bar{y}_0^k) N\left(x \| \sigma_k^{(i)}(\bar{y}_0^k); \Upsilon_k^{(i)}(\bar{y}_0^k)\right);$$

here, $N(u\|m; D)$ is the density of the normal (Gaussian) distribution of the random vector U with the parameters $m = \mathrm{M}[U]$ and $D = \operatorname{cov}[U, U]$.

3.1. Equations for the Probabilities and Conditional Densities

We also replace the numerator $\omega_k(\cdot)$ in the Stratonovich equation and the predicting distribution $\pi_k(\cdot)$ with the products of the corresponding probabilities and conditional densities

$$\begin{gathered}\omega_k(i, x, j, y|v_k) = \Omega_k^{(ij)}(v_k)\omega_k^{(ij)}(x, y|v_k), \quad \Omega_k^{(ij)}(v_k) = \int\int \omega_k(i, x, j, y|v_k)dx\,dy, \\ \pi_k(i, x|v_k) = \Pi_k^{(i)}(v_k)\pi_k^{(i)}(x|v_k), \quad \Pi_k^{(i)}(v_k) = \int \pi_k(i, x|v_k)dx,\end{gathered} \tag{3.1}$$

where $v_k = \bar{y}_0^{k-1}$ is their common condition. Then, the consistency of these two distributions (2.10) implies the following relationships between their probabilities and densities:

$$\begin{gathered}\Pi_k^{(i)}(v_k) = \sum_j \Omega_k^{(ij)}(v_k), \\ \pi_k^{(i)}(x|v_k) = \sum_j \Omega_k^{(ij)}(v_k) \int \omega_k^{(ij)}(x, y|v_k)dy \Big/ \Pi_k^{(i)}(v_k).\end{gathered} \tag{3.2}$$

Let us find formulas for the posterior probability and density (2.7) implied by the Bayes formula in which we also move the discrete argument j to the superscript:

$$P_k^{(i|j)}(y, v_k) = P_k^{(i)}(j, y, v_k), \; \rho_k^{(i|j)}(x|y, v_k) = \rho_k^{(i)}(x|j, y, v_k).$$

For this purpose, we integrate (2.8) with respect to x and use the product in (3.1). As a result, we obtain

$$P_k^{(i|j)}(y, v_k) = \Omega_k^{(ij)}(v_k) \int \omega_k^{(ij)}(x, y|v_k)dx \Big/ \sum_i \mathit{numerator}, \tag{3.3}$$

$$\rho_k^{(i|j)}(x|y, v_k) = \omega_k^{(ij)}(x, y|v_k) \Big/ \int \mathit{numerator}\; dx. \tag{3.4}$$

Thus, the posterior probability density $\rho_k^{(i|j)}(\cdot)$ is conditional with respect to the jointly predicting density $\omega_k^{(ij)}(\cdot)$.

3.2. Gaussian Approximations of the Optimal Estimates

We approximate the conditional probability density in (3.1) by the Gaussian density

$$\omega_k^{(ij)}(x, y|v_k) \approx N\left(x, y \| \lambda_k^{(ij)}(v_k), \mu_k^{(ij)}(v_k); \Psi_k^{(ij)}(v_k), \Phi_k^{(ij)}(v_k), \Delta_k^{(ij)}(v_k)\right). \tag{3.5}$$

As the proximity measure of these densities, we use the equality of their conditional expectations

$$\lambda_k^{(ij)}(v_k) = \mathrm{M}[X_k|i, j, v_k] = \iint x\omega_k^{(ij)}(x, y|v_k)dx\,dy, \tag{3.6}$$

$$\mu_k^{(ij)}(v_k) = \mathrm{M}[Y_k|i, j, v_k] = \iint y\omega_k^{(ij)}(x, y|v_k)dx\,dy \tag{3.7}$$

and conditional covariances

$$\Psi_k^{(ij)}(v_k) = \mathrm{cov}[X_k, X_k|i, j, v_k] = \iint xx^{\mathrm{T}}\omega_k^{(ij)}(x, y|v_k)dx\,dy - \lambda_k^{(ij)}(v_k)\lambda_k^{(ij)\mathrm{T}}(v_k), \tag{3.8}$$

$$\begin{gathered}\Delta_k^{(ij)}(v_k) = \mathrm{cov}[X_k, Y_k|i, j, v_k] = \iint xy^{\mathrm{T}}\omega_k^{(ij)}(x, y|v_k)dx\,dy - \lambda_k^{(ij)}(v_k)\mu_k^{(ij)\mathrm{T}}(v_k),\\ \Phi_k^{(ij)}(v_k) = \mathrm{cov}[Y_k, Y_k|i, j, v_k] = \iint yy^{\mathrm{T}}\omega_k^{(ij)}(x, y|v_k)dx\,dy - \mu_k^{(ij)}(v_k)\mu_k^{(ij)\mathrm{T}}(v_k).\end{gathered} \tag{3.9}$$

As a result, the Gaussian approximation of the distribution $\omega_k(\cdot)$ is

$$\omega_k(i, x, j, y|v_k) \approx \Omega_k^{(ij)}(v_k)N\left(x, y\|\lambda_k^{(ij)}(v_k), \mu_k^{(ij)}(v_k); \Psi_k^{(ij)}(v_k), \Phi_k^{(ij)}(v_k), \Delta_k^{(ij)}(v_k)\right); \tag{3.10}$$

i.e., it is completely determined by the probability $\Omega(\cdot)$ and two moments (3.6)–(3.9).

Then, we conclude from (3.5) that both marginal densities with respect to $\omega_k^{(ij)}(\cdot)$ are also approximated by the Gaussian densities as

$$\int \omega_k^{(ij)}(x, y|v_k)dy \approx N\left(x\|\lambda_k^{(ij)}(v_k); \Psi_k^{(ij)}(v_k)\right), \tag{3.11}$$

$$\int \omega_k^{(ij)}(x, y|v_k)dx \approx N\left(y\|\mu_k^{(ij)}(v_k); \Phi_k^{(ij)}(v_k)\right). \tag{3.12}$$

The conditional (with respect to (3.5)) posterior probability (3.4) is also approximated by the Gaussian one as

$$\rho_k^{(i|j)}(x \mid y, v_k) \approx N\left(x\|\sigma_k^{(i|j)}(y, v_k), \Upsilon_k^{(i|j)}(v_k)\right); \tag{3.13}$$

moreover, the parameters of the Gaussian density are found using the normal correlation theorem (see [16])

$$\begin{gathered}\sigma_k^{(i|j)}(y, v_k) = \lambda_k^{(ij)}(v_k) + \Delta_k^{(ij)}(v_k)\Phi_k^{(ij)\oplus}(v_k)\{y - \mu_k^{(ij)}(v_k)\},\\ \Upsilon_k^{(i|j)}(v_k) = \Psi_k^{(ij)}(v_k) - \Delta_k^{(ij)}(v_k)\Phi_k^{(ij)\oplus}(v_k)\Delta_k^{(ij)\mathrm{T}}(v_k),\end{gathered} \tag{3.14}$$

where $\oplus$ is the Moore–Penrose matrix pseudoinversion operation. According to (3.3) and (3.12), the posterior probability has the fractional Gaussian approximation

$$\mathrm{P}_k^{(i|j)}(y, v_k) \approx \Omega_k^{(ij)}(v_k)N\left(y\|\mu_k^{(ij)}(v_k); \Phi_k^{(ij)}(v_k)\right)\Big/\sum_i \mathit{numerator}. \tag{3.15}$$

As a result, the filter state is determined by probabilities (3.15) and moments (3.14).

Since the Gaussian density is symmetric and unimodal, the conditional quadratically simple and simple multiplicative estimates of the state vector in (2.4) and (2.5) have the same approximation

$$\hat{X}_k^{(i)} \approx \sigma_k^{(i|J_k)}(Y_k, \overline{Y}_0^{k-1})$$

as a random value of the conditionally posterior mean value in (3.14). This function corrects the corresponding value of the prediction function $\lambda_k^{(ij)}(\cdot)$ using the last measurement and its prediction functions $\mu_k^{(ij)}(\cdot)$, $\Delta_k^{(ij)}(\cdot)$, $\Phi_k^{(ij)}(\cdot)$. The Gaussian quadratically simple estimates are found, according to (2.4), by the formulas

$$\hat{I}_k \in \operatorname{Arg}\max_{i=\overline{1,L}} \mathrm{P}_k^{(i)}, \quad \hat{X}_k \approx \sum_i \mathrm{P}_k^{(i)}\hat{X}_k^{(i)}; \tag{3.16}$$

the Gaussian simple multiplicative estimates are found, according to (2.5), by the formulas

$$\hat{I}_k \in \operatorname{Arg}\max_{i=\overline{1,L}}\left[\mathrm{P}_k^{(i)}\Big/\left(2\pi\det\Upsilon_k^{(i)}\right)^{n/2}\right], \quad \hat{X}_k \approx \hat{X}_k^{(\hat{I}_k)}. \tag{3.17}$$

Here, for brevity we use the notation $\mathrm{P}_k^{(i)} = \mathrm{P}_k^{(i|J_k)}(Y_k, \overline{Y}_0^{k-1})$, $\Upsilon_k^{(i)} = \Upsilon_k^{(i|J_k)}(\overline{Y}_0^{k-1})$.

Thus, we have proved the following result, which is similar to Proposition 2 in [14].

L e m m a 1. Let finite conditional moments (3.6)–(3.9) of the random variables X_k and Y_k exist at any time $k \in \mathbb{N}$, and let their joint conditional probability density $\omega_k^{(ij)}(\cdot)$ have the Gaussian approximation (3.5). Then, the optimal *estimation functions* (2.4) and (2.5) are represented in terms of these moments and the conditional probability $\Omega_k^{(ij)}(\cdot)$ in (3.1) by the approximate formulas (3.16) and (3.17), respectively, with functions (3.14) and (3.15).

3.3. Relations between the Measurement and State Predictions

Due to approximation (3.5), the prediction probability according to (3.2), (3.11) has the poly-Gaussian approximation

$$\pi_k^{(i)}(x|\nu_k) \approx \sum_\gamma \Omega_k^{(i\gamma)}(\nu_k) N\left(x \| \lambda_k^{(i\gamma)}(\nu_k); \Psi_k^{(i\gamma)}(\nu_k)\right) \Big/ \Pi_k^{(i)}(\nu_k). \tag{3.18}$$

Here and in what follows, the summation is over $\gamma = \overline{1, M}$. This relation and dependence (2.9) of distribution $\omega_k(\cdot)$ on $\pi_k(\cdot)$ allows us to represent the conditional prediction functions $\mu_k^{(ij)}(\cdot)$, $\Delta_k^{(ij)}(\cdot)$, and $\Phi_k^{(ij)}(\cdot)$ of the measurement vector in terms of the conditional prediction functions $\lambda_k^{(ij)}(\cdot)$ and $\Psi_k^{(ij)}(\cdot)$ of the state vector.

Indeed, we can plug the fraction $\omega_k^{(ij)}(x, y|\nu_k) = \omega_k(i, x, j, y|\nu_k) \big/ \Omega_k^{(ij)}(\nu_k)$ into (3.7) and (3.9) and use products (2.9) and (3.1) to obtain

$$\begin{aligned}
\mu_k^{(ij)}(\nu_k) &= \Pi_k^{(i)}(\nu_k) \int \nu_k^{(j|i)}(x) \pi_k^{(i)}(x|\nu_k) dx \Big/ \Omega_k^{(ij)}(\nu_k), \\
\Delta_k^{(ij)}(\nu_k) &= \Pi_k^{(i)}(\nu_k) \int x \nu_k^{(j|i)}(x) \pi_k^{(i)}(x|\nu_k) dx \Big/ \Omega_k^{(ij)}(\nu_k) - \lambda_k^{(ij)}(\nu_k) \mu_k^{(ij)\mathrm{T}}(\nu_k), \\
\Phi_k^{(ij)}(\nu_k) &= \Pi_k^{(i)}(\nu_k) \int \mathrm{H}_k^{(j|i)}(x) \pi_k^{(i)}(x|\nu_k) dx \Big/ \Omega_k^{(ij)}(\nu_k) - \mu_k^{(ij)}(\nu_k) \mu_k^{(ij)\mathrm{T}}(\nu_k),
\end{aligned} \tag{3.19}$$

where $\nu(\cdot)$ and $\mathrm{H}(\cdot)$ are the first and the second initial moments of the indication–measurement distribution

$$\nu_k^{(j|i)}(x) = \int y \beta_k(j, y|i, x) dy, \quad \mathrm{H}_k^{(j|i)}(x) = \int y y^{\mathrm{T}} \beta_k(j, y|i, x) dy.$$

Using (1.9) and the rule for the Dirac function integration, we can explicitly represent these moments in terms of the measurement functions (1.2) as

$$\begin{aligned}
\nu_k^{(j|i)}(x) &= \mathrm{M}\left[\delta_{j, o_k^{(i)}(x, W_k^{\triangle})} b_k^{(i)}(x, W_k)\right], \\
\mathrm{H}_k^{(j|i)}(x) &= \mathrm{M}\left[\delta_{j, o_k^{(i)}(x, W_k^{\triangle})} b_k^{(i)}(x, W_k) b_k^{(i)\mathrm{T}}(x, W_k)\right].
\end{aligned} \tag{3.20}$$

We will call them *conditional means of the indicator–measurer.*

Now, take into account the poly-Gaussian approximation (3.18) to finally obtain

$$\begin{aligned}
\mu_k^{(ij)}(\nu_k) &\approx \sum_\gamma \Omega_k^{(i\gamma)}(\nu_k) h_k^{(j|i)}\left(\lambda_k^{(i\gamma)}(\nu_k); \Psi_k^{(i\gamma)}(\nu_k)\right) \Big/ \Omega_k^{(ij)}(\nu_k), \\
\Delta_k^{(ij)}(\nu_k) &\approx \sum_\gamma \Omega_k^{(i\gamma)}(\nu_k) G_k^{(j|i)}\left(\lambda_k^{(i\gamma)}(\nu_k); \Psi_k^{(i\gamma)}(\nu_k)\right) \Big/ \Omega_k^{(ij)}(\nu_k) - \lambda_k^{(ij)}(\nu_k) \mu_k^{(ij)\mathrm{T}}(\nu_k), \\
\Phi_k^{(ij)}(\nu_k) &\approx \sum_\gamma \Omega_k^{(i\gamma)}(\nu_k) F_k^{(j|i)}\left(\lambda_k^{(i\gamma)}(\nu_k); \Psi_k^{(i\gamma)}(\nu_k)\right) \Big/ \Omega_k^{(ij)}(\nu_k) - \mu_k^{(ij)}(\nu_k) \mu_k^{(ij)\mathrm{T}}(\nu_k),
\end{aligned} \tag{3.21}$$

where $h(\cdot)$, $G(\cdot)$, and $F(\cdot)$ are the *structure correction functions.* Two of them are determined as the result of Gaussian averaging of the conditional means (3.20)

$$\begin{aligned}
h_k^{(j|i)}(m, D) &= \int \nu_k^{(j|i)}(x) N(x \| m, D) dx, \\
F_k^{(j|i)}(m, D) &= \int \mathrm{H}_k^{(j|i)}(x) N(x \| m, D) dx,
\end{aligned} \tag{3.22}$$

the third function $G_k^{(j|i)}(m, D) = \int x \nu_k^{(j|i)\mathrm{T}}(x) N(x \| m, D) dx$ can be represented in terms of $h(\cdot)$ and its partial derivative with respect to the first arguments using the property of the Gaussian density $DN'_m(x \| m, D) = (x - m) N(x \| m, D)$ as the sum

$$G_k^{(j|i)}(m, D) = m h_k^{(j|i)\mathrm{T}}(m, D) + D\left(h_k^{(j|i)}{}'_m(m, D)\right)^{\mathrm{T}}. \tag{3.23}$$

We formulate this result as the following lemma.

L e m m a 2. Under the assumptions of Lemma 1, the functional parameters of the measurement prediction (3.7) and (3.9) are analytically represented in terms of the parameters of the state prediction (3.6) and (3.8) and the conditional probability $\Omega_k^{(ij)}(v_k)$ by formulas (3.21) with the functions of Gaussian means (3.22) and (3.23).

3.4. Dependence of the State Prediction Functions on Estimates

Since the prediction distribution $\pi_{k+1}(\cdot)$ is determined by formula (2.11) in terms of the preceding posterior $\rho_k(\cdot)$, the next state prediction functions $\Omega_{k+1}^{(\cdot)}(\cdot)$, $\lambda_{k+1}^{(\cdot)}(\cdot)$, and $\Psi_{k+1}^{(\cdot)}(\cdot)$ can be represented in terms of the current estimation functions $\mathrm{P}_k^{(\cdot)}(\cdot)$, $\sigma_k^{(\cdot)}(\cdot)$, and $\Upsilon_k^{(\cdot)}(\cdot)$.

To this end, we represent the desired functions (3.1), (3.6), and (3.8) in terms of the distribution $\omega_k(\cdot)$ as

$$\Omega_k^{(ij)}(v_k) = \int dx \int \omega_k(i,x,j,y|v_k)dy,$$

$$\lambda_k^{(ij)}(v_k) = \int x dx \int \omega_k(i,x,j,y|v_k)dy \Big/ \Omega_k^{(ij)}(v_k),$$

$$\Psi_k^{(ij)}(v_k) = \int xx^{\mathrm{T}} dx \int \omega_k(i,x,j,y|v_k)dy \Big/ \Omega_k^{(ij)}(v_k) - \lambda_k^{(ij)}(v_k)\lambda_k^{(ij)\mathrm{T}}(v_k).$$

With regard to (2.9) and (1.7), the inner integral here can be represented in terms of the indication probability (1.10) and the prediction distribution as $\int \omega_k(i,x,j,y|v_k)dy = \mathrm{B}_k^{(j|i)}(x)\pi_k(i,x|v_k)$. Therefore, for time $k+1$, we have an explicit dependence of these functions on the distribution $\pi_{k+1}(\cdot)$:

$$\begin{aligned}
\Omega_{k+1}^{(lf)}(v_{k+1}) &= \int \mathrm{B}_{k+1}^{(f|l)}(\chi)\pi_{k+1}(l,\chi|v_{k+1})d\chi, \quad f = j_{k+1}, \\
\lambda_{k+1}^{(lf)}(v_{k+1}) &= \int \chi \mathrm{B}_{k+1}^{(f|l)}(\chi)\pi_{k+1}(l,\chi|v_{k+1})d\chi \Big/ \Omega_{k+1}^{(lf)}(v_{k+1}), \\
\Psi_{k+1}^{(lf)}(v_{k+1}) &= \int \chi\chi^{\mathrm{T}} \mathrm{B}_{k+1}^{(f|l)}(\chi)\pi_{k+1}(l,\chi|v_{k+1})d\chi \Big/ \Omega_{k+1}^{(lf)}(v_{k+1}) - \lambda_{k+1}^{(lf)}(v_{k+1})\lambda_{k+1}^{(lf)\mathrm{T}}(v_{k+1}).
\end{aligned} \tag{3.24}$$

Plug (2.11) into these expressions and interchange the order of integration to obtain

$$\begin{aligned}
\Omega_{k+1}^{(lf)}(v_{k+1}) &= \sum_i \int \varsigma_k^{(lf|ij)}(x,y)\rho_k(i,x|j,y,v_k)dx, \quad v_{k+1} = \bar{y}_0^k = (j,y,v_k), \\
\lambda_{k+1}^{(lf)}(v_{k+1}) &= \sum_i \int \varphi_k^{(lf|ij)}(x,y)\rho_k(i,x|j,y,v_k)dx \Big/ \Omega_{k+1}^{(lf)}(v_{k+1}), \\
\Psi_{k+1}^{(lf)}(v_{k+1}) &= \sum_i \int \Sigma_k^{(lf|ij)}(x,y)\rho_k(i,x|j,y,v_k)dx \Big/ \Omega_{k+1}^{(lf)}(v_{k+1}) - \lambda_{k+1}^{(lf)}(v_{k+1})\lambda_{k+1}^{(lf)\mathrm{T}}(v_{k+1}),
\end{aligned} \tag{3.25}$$

where the other conditional means are

$$\varsigma_k^{(lf|ij)}(x,y) = \int \mathrm{B}_{k+1}^{(f|l)}(\chi)\zeta_k(l,\chi|i,x,j,y)d\chi,$$

$$\varphi_k^{(lf|ij)}(x,y) = \int \chi \mathrm{B}_{k+1}^{(f|l)}(\chi)\zeta_k(l,\chi|i,x,j,y)d\chi,$$

$$\Sigma_k^{(lf|ij)}(x,y) = \int \chi\chi^{\mathrm{T}} \mathrm{B}_{k+1}^{(f|l)}(\chi)\zeta_k(l,\chi|i,x,j,y)d\chi.$$

Using (1.11), (1.2), and the rule for the integration of the Dirac function, these conditional means can be explicitly represented in terms of the known functions of the entire observation system

$$\begin{aligned}
\varsigma_k^{(lf|ij)}(x,y) &= \mathrm{M}\left[e_k^{(lf|ij)}(x,y,\bar{V}_k,\bar{W}_k)\right] \Big/ \beta_k(j,y|i,x), \\
\varphi_k^{(lf|ij)}(x,y) &= \mathrm{M}\left[a_k^{(i)}(x,V_k)e_k^{(lf|ij)}(x,y,\bar{V}_k,\bar{W}_k)\right] \Big/ \beta_k(j,y|i,x), \\
\Sigma_k^{(lf|ij)}(x,y) &= \mathrm{M}\left[a_k^{(i)}(x,V_k)a_k^{(i)\mathrm{T}}(x,V_k)e_k^{(lf|ij)}(x,y,\bar{V}_k,\bar{W}_k)\right] \Big/ \beta_k(j,y|i,x),
\end{aligned} \tag{3.26}$$

where the common factor $e(\cdot)$ is

$$e_k^{(lf|ij)}(x, y, \bar{V}_k, \bar{W}_k) = \mathrm{B}_{k+1}^{(f|l)}\left(a_k^{(i)}(x, V_k)\right)\delta_{l,\phi_k^{(i)}(x, V_k^{\triangle})}\delta_{j,o_k^{(i)}(x, W_k^{\triangle})}\delta\left(y - b_k^{(i)}(x, W_k)\right).$$

We formulate this result as the following lemma.

L e m m a 3. If the finite conditional moments (3.6) and (3.8) exist, then the state prediction functions are represented in terms of the posterior distribution $\rho_k(\cdot)$ by formulas (3.25) with the conditional means (3.26).

Now, let us take into account the Gaussian approximation of the posterior distribution

$$\rho_k(i, x | j, y, v_k) \approx \mathrm{P}_k^{(i|j)}(y, v_k) N\left(x \| \sigma_k^{(i|j)}(y, v_k), \Upsilon_k^{(i|j)}(v_k)\right), \tag{3.27}$$

which follows from (2.6) and (3.13).

Plug it into (3.25) to finally obtain

$$\begin{aligned}
\Omega_{k+1}^{(lf)}(v_{k+1}) &\approx \sum_i \mathrm{P}_k^{(i|j)}(y, v_k)\eta_k^{(lf|ij)}\left(\varepsilon_k^{(i|j)}(y, v_k)\right), \\
\lambda_{k+1}^{(lf)}(v_{k+1}) &\approx \sum_i \mathrm{P}_k^{(i|j)}(y, v_k)\tau_k^{(lf|ij)}\left(\varepsilon_k^{(i|j)}(y, v_k)\right) \Big/ \Omega_{k+1}^{(lf)}(v_{k+1}), \\
\Psi_{k+1}^{(lf)}(v_{k+1}) &\approx \sum_i \mathrm{P}_k^{(i|j)}(y, v_k)\mathrm{X}_k^{(lf|ij)}\left(\varepsilon_k^{(i|j)}(y, v_k)\right) \Big/ \Omega_{k+1}^{(lf)}(v_{k+1}) - \lambda_{k+1}^{(lf)}(v_{k+1})\lambda_{k+1}^{(lf)\mathrm{T}}(v_{k+1}),
\end{aligned} \tag{3.28}$$

where the common argument $\varepsilon_k^{(i|j)}(y, v_k) = \left(y, \sigma_k^{(i|j)}(y, v_k), \Upsilon_k^{(i|j)}(v_k)\right)$ and the *structure prediction functions* $\eta(\cdot)$, $\tau(\cdot)$, and $\Theta(\cdot)$ are the results of the Gaussian averaging of the conditional means (3.26)

$$\begin{aligned}
\eta_k^{(lf|ij)}(y, m, D) &= \int \varsigma_k^{(lf|ij)}(x, y) N(x \| m, D) dx, \\
\tau_k^{(lf|ij)}(y, m, D) &= \int \varphi_k^{(lf|ij)}(x, y) N(x \| m, D) dx, \\
\mathrm{X}_k^{(lf|ij)}(y, m, D) &= \int \Sigma_k^{(lf|ij)}(x, y) N(x \| m, D) dx.
\end{aligned} \tag{3.29}$$

L e m m a 4. If the Gaussian approximation (3.27) holds for the posterior distribution $\rho_k(\cdot)$, then the state prediction functions (3.1), (3.6), and (3.8) are analytically represented in terms of its functional parameters (3.14) and (3.15) by formulas (3.28) with the Gaussian means (3.29).

As a result, instead of the chain of integral transformations of distributions (2.14), we have the algebraic sequence of computations of characteristics and their Gaussian approximation

$$\ldots \to \Omega_k, \lambda_k, \Psi_k \to \mu_k, \Delta_k, \Phi_k \to \mathrm{P}_k, \sigma_k, \Upsilon_k \to \Omega_{k+1}, \lambda_{k+1}, \Psi_{k+1} \to \ldots, \quad k \geq 0 \tag{3.30}$$

by formulas (3.14), (3.15), (3.21), and (3.28). After the three components of the filter state vector $\mathrm{P}_k(\cdot)$, $\sigma_k(\cdot)$, $\Upsilon_k(\cdot)$ have been obtained, the estimates $\hat{I}_k$, $\hat{X}_k$ are found by formulas (3.16) or (3.17).

3.5. Initial Conditions

The chain of transformations of the Gaussian characteristics (3.30) begins by finding the parameters of the Gaussian approximation (3.10) of the unconditional initial distribution (2.15)

$$\omega_0(i, x, j, y) \approx \Omega_0^{(ij)} N\left(x, y \| \lambda_0^{(ij)}, \mu_0^{(ij)}; \Psi_0^{(ij)}, \Phi_0^{(ij)}, \Delta_0^{(ij)}\right); \tag{3.31}$$

this approximation is similar to (3.10). The meaning of these transformations is as follows:

$$\begin{aligned}
&\Omega_0^{(ij)} = \Pr\left[I_0 = i, J_0 = j\right], \quad \lambda_0^{(ij)} = \mathrm{M}\left[X_0 | I_0 = i, J_0 = j\right], \\
&\mu_0^{(ij)} = \mathrm{M}\left[Y_0 | I_0 = i, J_0 = j\right], \quad \Psi_0^{(ij)} = \mathrm{cov}\left[X_0, X_0 | I_0 = i, J_0 = j\right], \\
&\Delta_0^{(ij)} = \mathrm{cov}\left[X_0, Y_0 | I_0 = i, J_0 = j\right], \quad \Phi_0^{(ij)} = \mathrm{cov}\left[Y_0, Y_0 | I_0 = i, J_0 = j\right]
\end{aligned} \tag{3.32}$$

and they can be found using the Monte-Carlo method.

In order to find analytical approximations of parameters (3.32), we represent them in terms of the initial data. At $k = 0$, we obtain from (3.1) by straightforward integration, taking into account (1.7) and (2.15)

$$\Omega_0^{(ij)} = P_0^{(i)} \int \mathrm{B}_0^{(j|i)}(x) p_0^{(i)}(x) dx.$$

Similarly, we obtain the following expressions, which are similar to (3.19) and (3.24), from (3.6)–(3.9):

$$\lambda_0^{(ij)} = P_0^{(i)} \int x \mathrm{B}_0^{(j|i)}(x) p_0^{(i)}(x) dx \Big/ \Omega_0^{(ij)},$$

$$\mu_0^{(ij)} = P_0^{(i)} \int \nu_0^{(j|i)}(x) p_0^{(i)}(x) dx \Big/ \Omega_0^{(ij)},$$

$$\Psi_0^{(ij)} = P_0^{(i)} \int x x^{\mathrm{T}} \mathrm{B}_0^{(j|i)}(x) p_0^{(i)}(x) dx \Big/ \Omega_0^{(ij)} - \lambda_0^{(ij)} \lambda_0^{(ij)\mathrm{T}},$$

$$\Delta_0^{(ij)} = P_0^{(i)} \int x \nu_0^{(j|i)\mathrm{T}}(x) p_0^{(i)}(x) dx \Big/ \Omega_0^{(ij)} - \lambda_0^{(ij)} \mu_0^{(ij)\mathrm{T}},$$

$$\Phi_0^{(ij)} = P_0^{(i)} \int \mathrm{H}_0^{(j|i)}(x) p_0^{(i)}(x) dx \Big/ \Omega_0^{(ij)} - \mu_0^{(ij)} \mu_0^{(ij)\mathrm{T}},$$

where $\nu(\cdot)$ and $\mathrm{H}(\cdot)$ are the known conditional means of indicator–measurer (3.20).

If the initial state vector X_0 of plant (1.1) is conditionally Gaussian

$$p_0^{(i)}(x) = N\left(x \| m_0^{x|i}; D_0^{x|i}\right), \quad m_0^{x|i} = \mathrm{M}[X_0 | I_0 = i], \quad D_0^{x|i} = \mathrm{cov}[X_0 | I_0 = i], \tag{3.33}$$

then, some of them can be represented, as in (3.21), in terms of the known Gaussian correction functions (3.22) and (3.23):

$$\begin{aligned} \mu_0^{(ij)} &= P_0^{(i)} h_0^{(j|i)}\left(m_0^{x|i}; D_0^{x|i}\right) \Big/ \Omega_0^{(ij)}, \\ \Delta_0^{(ij)} &= P_0^{(i)} G_0^{(j|i)}\left(m_0^{x|i}; D_0^{x|i}\right) \Big/ \Omega_0^{(ij)} - \lambda_0^{(ij)} \mu_0^{(ij)\mathrm{T}}, \\ \Phi_0^{(ij)} &= P_0^{(i)} F_0^{(j|i)}\left(m_0^{x|i}; D_0^{x|i}\right) \Big/ \Omega_0^{(ij)} - \mu_0^{(ij)} \mu_0^{(ij)\mathrm{T}}. \end{aligned} \tag{3.34}$$

The other functions are determined by the algebraic formulas

$$\begin{aligned} \Omega_0^{(ij)} &= P_0^{(i)} \psi_0^{(j|i)}\left(m_0^{x|i}; D_0^{x|i}\right), \\ \lambda_0^{(ij)} &= P_0^{(i)} t_0^{(j|i)}\left(m_0^{x|i}; D_0^{x|i}\right) \Big/ \Omega_0^{(ij)}, \\ \Psi_0^{(ij)} &= P_0^{(i)} \mathrm{X}_0^{(j|i)}\left(m_0^{x|i}; D_0^{x|i}\right) \Big/ \Omega_0^{(ij)} - \lambda_0^{(ij)} \lambda_0^{(ij)\mathrm{T}}, \end{aligned} \tag{3.35}$$

where the Gaussian means of the conditional indication probability $\mathrm{B}_0^{(j|i)}(x)$:

$$\psi_0^{(j|i)}(m, D) = \int \mathrm{B}_0^{(j|i)}(x) N(x \| m; D)\, dx \tag{3.36}$$

and its products with the condition variable

$$t_0^{(j|i)}(m, D) = \int x \mathrm{B}_0^{(j|i)}(x) N(x \| m, D)\, dx, \quad \mathrm{X}_0^{(j|i)}(m, D) = \int x x^{\mathrm{T}} \mathrm{B}_0^{(j|i)}(x) N(x \| m; D)\, dx$$

are used. There are relationships between them, which are similar to (3.23), that represent the two last quantities in terms of (3.36):

$$\begin{aligned} t_0^{(j|i)}(m, D) &= m \psi_0^{(j|i)}(m, D) + D\left(\psi_0^{(j|i)}{}'_m(m, D)\right)^{\mathrm{T}}, \\ \mathrm{X}_0^{(j|i)}(m, D) &= (D - m m^{\mathrm{T}}) \psi_0^{(j|i)}(m, D) + m t_0^{(j|i)\mathrm{T}}(m, D) + t_0^{(j|i)}(m, D) m^{\mathrm{T}} + D\left(\psi_0^{(j|i)}{}''_m(m, D)\right) D. \end{aligned} \tag{3.37}$$

L e m m a 5. Let the initial state X_0 of plant (1.1) be conditionally Gaussian with the probability density (3.33). Then, the parameters of the Gaussian approximation (3.31) of the initial probability density $\omega_0^{(ij)}(\cdot)$ are determined by the algebraic formulas (3.34) and (3.35) with the Gaussian mean functions (3.22), (3.23), (3.36), and (3.37).

3.6. Equations of the Suboptimal Filter

The relations obtained above allow us to write equations of the Gaussian approximation of the AOF, which is also called NAF. The functions $\eta(\cdot)$, $\tau(\cdot)$, $\Theta(\cdot)$, $h(\cdot)$, $G(\cdot)$, $F(\cdot)$ used in these relations must be found in advance as characteristics of the statistical linearization (3.22) and (3.29) of the corresponding conditional means (3.20) and (3.26). In what follows, $i, l = \overline{1, L}$ and $j, f = \overline{1, M}$, and the random values of deterministic functions at random points are denoted by the same symbols as the functions themselves, e.g., $\mathrm{P}_k^{(i)} = \mathrm{P}_k^{(i|J_k)}(Y_k, \overline{Y}_0^{k-1})$.

T h e o r e m 2. Under the assumptions of Lemma 1, the following equations of the LD NAF hold. First, the initial conditions $\Omega_0^{(ij)}, \lambda_0^{(ij)}, \mu_0^{(ij)}, \Psi_0^{(ij)}, \Delta_0^{(ij)}, \Phi_0^{(ij)}$ are specified, which are found exactly using (3.32) or approximately using (3.34), (3.35) and functions (3.22), (3.36), and (3.37). Next, at each time $k \geq 0$, the following quantities are computed.

1. Characteristics of the conditional predictions of measurement (if $k \neq 0$) using (3.21):

$$\begin{aligned}
\mu_k^{(ij)} &= \sum_\gamma \Omega_k^{(i\gamma)} h_k^{(j|i)}\left(\lambda_k^{(i\gamma)}; \Psi_k^{(i\gamma)}\right) \Big/ \Omega_k^{(ij)}, \\
\Delta_k^{(ij)} &= \sum_\gamma \Omega_k^{(i\gamma)} G_k^{(j|i)}\left(\lambda_k^{(i\gamma)}; \Psi_k^{(i\gamma)}\right) \Big/ \Omega_k^{(ij)} - \lambda_k^{(ij)} \mu_k^{(ij)\mathrm{T}}, \\
\Phi_k^{(ij)} &= \sum_\gamma \Omega_k^{(i\gamma)} F_k^{(j|i)}\left(\lambda_k^{(i\gamma)}; \Psi_k^{(i\gamma)}\right) \Big/ \Omega_k^{(ij)} - \mu_k^{(ij)} \mu_k^{(ij)\mathrm{T}}.
\end{aligned} \tag{3.38}$$

2. Characteristics of the conditional predictions of state and measurement specified by the indicator J_k

$$\Omega_k^{(i)} = \Omega_k^{(iJ_k)}, \quad \lambda_k^{(i)} = \lambda_k^{(iJ_k)}, \quad \mu_k^{(i)} = \mu_k^{(iJ_k)}, \quad \Psi_k^{(i)} = \Psi_k^{(iJ_k)}, \quad \Delta_k^{(i)} = \Delta_k^{(iJ_k)}, \quad \Phi_k^{(i)} = \Phi_k^{(iJ_k)}. \tag{3.39}$$

3. The elements of the current filter state determined by the measurement Y_k by formulas (3.14) and (3.15):

$$\mathrm{P}_k^{(i)} = \Omega_k^{(i)} N\left(Y_k \| \mu_k^{(i)}; \Phi_k^{(i)}\right) \Big/ \sum_i \textit{numerator},$$

$$\sigma_k^{(i)} = \lambda_k^{(i)} + \Delta_k^{(i)} \Phi_k^{(i)\oplus} \left\{Y_k - \mu_k^{(i)}\right\}, \quad \Upsilon_k^{(i)} = \Psi_k^{(i)} - \Delta_k^{(i)} \Phi_k^{(i)\oplus} \Delta_k^{(i)\mathrm{T}};$$

its quadratically simple estimates by formula (3.16)

$$\hat{I}_k \in \operatorname{Arg}\max_{i=\overline{1,L}} \mathrm{P}_k^{(i)}, \quad \hat{X}_k = \sum_i \mathrm{P}_k^{(i)} \sigma_k^{(i)};$$

and its simple multiplicative estimates by formula (3.17)

$$\hat{I}_k \in \operatorname{Arg}\max_{i=\overline{1,L}} \left[\mathrm{P}_k^{(i)} \Big/ \left(2\pi \det \Upsilon_k^{(i)}\right)^{n/2}\right], \quad \hat{X}_k = \sigma_k^{(\hat{I}_k)}.$$

4. Characteristics of possible conditional predictions of the next states by formula (3.28):

$$\begin{aligned}
\Omega_{k+1}^{(lf)} &= \sum_i \mathrm{P}_k^{(i)} \eta_k^{(lf|iJ_k)}\left(Y_k, \sigma_k^{(i)}, \Upsilon_k^{(i)}\right), \\
\lambda_{k+1}^{(lf)} &= \sum_i \mathrm{P}_k^{(i)} \tau_k^{(lf|iJ_k)}\left(Y_k, \sigma_k^{(i)}, \Upsilon_k^{(i)}\right) \Big/ \Omega_{k+1}^{(lf)}, \\
\Psi_{k+1}^{(lf)} &= \sum_i \mathrm{P}_k^{(i)} \mathrm{X}_k^{(lf|iJ_k)}\left(Y_k, \sigma_k^{(i)}, \Upsilon_k^{(i)}\right) \Big/ \Omega_{k+1}^{(lf)} - \lambda_{k+1}^{(lf)} \lambda_{k+1}^{(lf)\mathrm{T}}.
\end{aligned} \tag{3.40}$$

Next, all these computations are repeated by setting $k := k + 1$ and returning to Step 1.

By eliminating the intermediate variables from the equations written above, we obtain a closed system of stochastic difference first-order equations in all the probabilities $\mathrm{P}_k^{(i)}$, expectation vectors $\sigma_k^{(i)}$, and the covariance matrices $\Upsilon_k^{(i)}$, where $i = \overline{1, L}$. The total number of these equations $L(n+1)(n+2)/2$ determines the order of the Gaussian LD AOF. For a monostructured observation system, where $L = 1$ and $\mathrm{P}_k^{(1)} = 1$, the filter state is determined by two moments $\sigma_k^{(1)}$ and $\Upsilon_k^{(1)}$, so that its order is reduced to $n(n+3)/2$ as is the order of the NAF.

3.7. The Case without a Structure Indicator

If measurers (1.2) do not include the indicator, then its equation can be formally considered as the identity $J_k \equiv 1$, so that the probability of indication is degenerated: $\mathrm{B}_k^{(j|i)}(x) = \delta_{j,1}$. This allows us to drop the variables J_k, j, and f in all the preceding expressions. Then, even the basic recurrent formula (2.12) becomes simpler with a lower number of integer arguments of distributions:

$$\rho_{k+1}(l,\chi|y_0^{k+1}) = \frac{\beta_{k+1}(y_{k+1}|l,\chi)\sum_i \int \zeta_k(l,\chi|i,x,y)\rho_k(i,x|y_0^k)dx}{\sum_l \int numerator\ d\chi}.$$

In the case of independent noises, when $\zeta_k(l,\chi|i,x,j,y) = \alpha_k(l,\chi|i,x)$, it coincides with [1, (2.8)].

The number of initial conditions in the equations of the NAF is by a factor of M lower; these are $\Omega_0^{(i)}, \lambda_0^{(i)}, \mu_0^{(i)}, \Psi_0^{(i)}, \Delta_0^{(i)}, \Phi_0^{(i)}$; moreover, in the Gaussian case (2.31), some of them are known: $\Omega_0^{(i)} = P_0^{(i)}$, $\lambda_0^{(i)} = m_0^{x|i}$, and $\Psi_0^{(i)} = D_0^{x|i}$. Instead of (3.38), the characteristics of the conditional measurement predictions are found without summation over the possible values of the indicator variable:

$$\begin{aligned}
&\mu_k^{(i)} = h_k^{(i)}\left(\lambda_k^{(i)};\Psi_k^{(i)}\right),\\
&\Delta_k^{(i)} = \Psi_k^{(i)}\left(h_k^{(j|i)}{}'_m(\lambda_k^{(i)},\Psi_k^{(i)})\right)^{\mathrm{T}},\\
&\Phi_k^{(i)} = F_k^{(i)}\left(\lambda_k^{(i)};\Psi_k^{(i)}\right) - \mu_k^{(i)}\mu_k^{(i)\mathrm{T}},
\end{aligned} \tag{3.41}$$

relations (3.39) are not used any more, and the characteristics of the conditional predictions of the next state are found by the formulas

$$\begin{aligned}
&\Omega_{k+1}^{(l)} = \sum_i \mathrm{P}_k^{(i)}\eta_k^{(l|i)}\left(Y_k,\sigma_k^{(i)},\Upsilon_k^{(i)}\right),\\
&\lambda_{k+1}^{(l)} = \sum_i \mathrm{P}_k^{(i)}\tau_k^{(l|i)}\left(Y_k,\sigma_k^{(i)},\Upsilon_k^{(i)}\right)\Big/\Omega_{k+1}^{(l)},\\
&\Psi_{k+1}^{(l)} = \sum_i \mathrm{P}_k^{(i)}\mathrm{X}_k^{(l|i)}\left(Y_k,\sigma_k^{(i)},\Upsilon_k^{(i)}\right)\Big/\Omega_{k+1}^{(l)} - \lambda_{k+1}^{(l)}\lambda_{k+1}^{(l)\mathrm{T}}
\end{aligned} \tag{3.42}$$

rather than by (3.40).

In addition, the computation of the conditional means both of measurer (3.20)

$$\nu_k^{(i)}(x) = \mathrm{M}\left[b_k^{(i)}(x,W_k)\right], \quad \mathrm{H}_k^{(i)}(x) = \mathrm{M}\left[b_k^{(i)}(x,W_k)\,b_k^{(i)\mathrm{T}}(x,W_k)\right] \tag{3.43}$$

and of the whole observation system (3.26)

$$\begin{aligned}
&\varsigma_k^{(l|i)}(x,y) = \mathrm{M}\left[e_k^{(l|i)}(x,y,V_k^{\triangle},W_k)\right]\Big/\beta_k^{(i)}(y|x),\\
&\varphi_k^{(l|i)}(x,y) = \mathrm{M}\left[a_k^{(i)}(x,V_k)\,e_k^{(l|i)}(x,y,V_k^{\triangle},W_k)\right]\Big/\beta_k^{(i)}(y|x),\\
&\Sigma_k^{(l|i)}(x,y) = \mathrm{M}\left[a_k^{(i)}(x,V_k)a_k^{(i)\mathrm{T}}(x,V_k)\,e_k^{(l|i)}(x,y,V_k^{\triangle},W_k)\right]\Big/\beta_k^{(i)}(y|x)
\end{aligned} \tag{3.44}$$

is simplified; here, the common factor $e(\cdot)$ is $e_k^{(l|i)}(x,y,V_k^{\triangle},W_k) = \delta_{l,\phi_k^{(i)}(x,V_k^{\triangle})}\delta\left(y - b_k^{(i)}(x,W_k)\right)$, and the denominator is equal to the conditional probability density of the measurement

$$\beta_k^{(i)}(y|x) = \mathrm{M}\left[\delta\left(y - b_k^{(i)}(x,W_k)\right)\right].$$

C o r o l l a r y 1. If there is no structure indicator, relations (3.38) and (3.40) take a simpler form (3.41) and (3.42), respectively; to find the structure correction (3.22) and prediction functions (3.29), simpler conditional means (3.43) and (3.44) are used.

Other significant simplifications due to various independences of the white noises in the system can be found in [14]. In that paper, even simpler linearized approximations of the conditional means and, as a consequence, of their Gaussian means are considered, which yields the bank of extended Kalman filters [2]. However, this does not affect the order of such a filter, which remains high.

4. OPTIMAL STRUCTURE FILTER

Now we abandon the complex construction of the AOF by finding the posterior distribution by (2.12) and estimates by (2.4) or (2.5). To obtain simpler dependences (1.3) of estimates on the accumulated measurements that can be implemented in real time, we impose the following recursiveness constraints on these estimates:

$$\hat{X}_k(\bar{Y}_0^k) = f_k\left(\bar{Y}_k, \bar{Y}_{k-1}, \hat{X}_{k-1}(\bar{Y}_0^{k-1})\right), \quad \hat{I}_k(\bar{Y}_0^k) = \iota_k\left(\bar{Y}_k, \bar{Y}_{k-1}, \hat{X}_{k-1}(\bar{Y}_0^{k-1})\right).$$

4.1. Filter Equations

In [10], it was proposed to seek the difference equation for the LD OSF of an order hn, where $h \in \{\overline{1, L}\}$ is the factor multiplying the dimension n of the state vector of the dynamic part of the observation plant (1.1). Below, we consider for simplicity only the fastest such filter—the OSF of a low order n with the multiplicity $h = 1$. The state vector of this filter is the n-dimensional estimate vector $\hat{X}_k$ itself, and the equation of state is sought in the form

$$\hat{X}_k = f_k(\bar{Y}_k, \bar{Y}_{k-1}, \hat{X}_{k-1}), \quad k \geq 1, \quad \hat{X}_0 = f_0(\bar{Y}_0), \tag{4.1}$$

where $\bar{Y}_k = (J_k, Y_k)$ as before, and the estimate of the structure index is given by the formula

$$\hat{I}_k = \iota_k(\bar{Y}_k, \bar{Y}_{k-1}, \hat{X}_{k-1}), \quad k \geq 1, \quad \hat{I}_0 = \iota_0(\bar{Y}_0). \tag{4.2}$$

Both structure functions of this filter $f_k(\cdot)$ and $\iota_k(\cdot)$ are found based on the optimality criterion (1.4), which suggests its name *the optimal structure filter*. Note that filter (4.1) and (4.2) is an LD- generalization of the ordinary (purely dynamic) OSF [11, 12].

4.2. Relation between the Structure Functions and the Estimating Distribution

Substitution of (4.1) and (4.2) into the optimality criterion (1.4) yields the finite-dimensional minimization problem (which is similar to (2.1)) for the *conditional risk* function

$$\Re_k^{(\iota)}(f|\bar{y}_k^*) = \sum_i \int c_k^{(i,\iota)}(x, f)\rho_k(i, x|\hat{x}_{k-1}^*)dx \to \min_{\iota \in \overline{1,L},\, f \in \mathbb{R}^n}, \quad \hat{x}_{k-1}^* = (\bar{y}_{k-1}^k, \hat{x}_{k-1}), \tag{4.3}$$

where $\rho_k(\cdot)$ is now the estimating (truncated posterior) distribution of the extended plant state $\bar{X}_k$. This distribution takes into account possible values of a constant number of conditions $\hat{x}_{k-1}^*$ consisting of the preceding estimate $\hat{x}_{k-1}$ and the last two measurements $\bar{y}_{k-1}^k = (\bar{y}_{k-1}, \bar{y}_k)$. As a result, the structure OSF functions (4.1) and (4.2) are generally found similarly to (2.2) and (2.3) by the formulas

$$\iota_k(\hat{x}_{k-1}^*) \in \operatorname{Arg}\min_{i=\overline{1,L}} \Re_k^{(i)}\left(f_k^{(i)}(\hat{x}_{k-1}^*)|\hat{x}_{k-1}^*\right), \quad f_k(\hat{x}_{k-1}^*) = f_k^{(\iota_k(\hat{x}_{k-1}^*))}(\hat{x}_{k-1}^*); \tag{4.4}$$

in these formulas, each conditional estimation function is determined by

$$f_k^{(i)}(\hat{x}_{k-1}^*) \in \operatorname{Arg}\min_{f \in \mathbb{R}^n} \Re_k^{(i)}(f|\hat{x}_{k-1}^*), \quad i = \overline{1,L}.$$

The quadratically simple and simple multiplicative OSF estimates are also found by (2.4) and (2.5), but with the replacement of the argument $\overline{Y}_0^k$ with $\hat{X}_{k-1}^*$; they also use different not posterior probabilities and densities

$$\mathrm{P}_k^{(i)}(\hat{x}_{k-1}^*) = \int \rho_k(i, x|\hat{x}_{k-1}^*)dx, \quad \rho_k^{(i)}(x|\hat{x}_{k-1}^*) = \rho_k(i, x|\hat{x}_{k-1}^*) \Big/ \mathrm{P}_k^{(i)}(\hat{x}_{k-1}^*). \tag{4.5}$$

4.3. Finding the Estimating Distribution

In order to find the conditional distribution $\rho_k(\cdot)$, we represent it using the Bayes formula as the ratio

$$\rho_k(i, x| j, y, \overline{y}_{k-1}, \hat{x}_{k-1}) = \omega_k(i, x, j, y|\overline{y}_{k-1}, \hat{x}_{k-1}) \Big/ \sum_i \int numerator\ dx, \quad k \ge 1, \tag{4.6}$$

where the numerator $\omega_k(\cdot)$, as in (2.9), is the product

$$\omega_k(i, x, j, y|\overline{y}_{k-1}, \hat{x}_{k-1}) = \beta_k(j, y|i, x)\pi_k(i, x|\overline{y}_{k-1}, \hat{x}_{k-1}). \tag{4.7}$$

Here, the prediction distribution $\pi_k(\cdot)$ can be represented, as (2.11), in terms of the conditional distribution $\xi_{k-1}(\cdot)$ of the random variables corresponding to the preceding time. For the next time point, we have

$$\pi_{k+1}(l, \chi| j, y, \hat{x}) = \sum_i \int \zeta_k(l, \chi|i, x, j, y)\xi_k(i, x| j, y, \hat{x})dx, \tag{4.8}$$

where the conditional distribution $\xi_k(\cdot)$ can be easily represented in terms of the unconditional distribution $q_k(\cdot)$ of the group $\aleph_k = (\overline{X}_k, \overline{Y}_k, \hat{X}_k)$ consisting of five variables as the ratio

$$\xi_k(i, x| j, y, \hat{x}) = q_k(i, x, j, y, \hat{x}) \Big/ \sum_i \int numerator\ dx. \tag{4.9}$$

These random variables form a Markov vector because they are determined by the closed system of difference equations of plant (1.1), measurer (1.2), and filter (4.1) disturbed by white noises. Therefore, there is a recurrent formula for their distribution, which can be written as a chain of relations generalizing the ordinary dynamic case [11]. Indeed, knowing the distribution $q_{k-1}(\cdot)$ for $k \ge 1$ of the preceding group $\aleph_{k-1}$, one can find the distribution of another group $(\overline{X}_k, \overline{Y}_{k-1}, \hat{X}_{k-1})$, which differs from the former group only in the first pair $\overline{X}_k$ by the formula

$$p_k(i_k, x_k, j_{k-1}, y_{k-1}, \hat{x}_{k-1}) = \sum_i \int \zeta_{k-1}(i_k, x_k|i, x, j_{k-1}, y_{k-1})q_{k-1}(i, x, j_{k-1}, y_{k-1}, \hat{x}_{k-1})dx; \tag{4.10}$$

then, the distribution of seven random variables

$$r_k(i_k, x_k, j_k, y_k, j_{k-1}, y_{k-1}, \hat{x}_{k-1}) = \beta_k(j_k, y_k|i_k, x_k)p_k(i_k, x_k, j_{k-1}, y_{k-1}, \hat{x}_{k-1}) \tag{4.11}$$

becomes known. As in (4.6), given this distribution, we easily find the estimating distribution

$$\rho_k(i, x| j_k, y_k, j_{k-1}, y_{k-1}, \hat{x}_{k-1}) = r_k(i, x, j_k, y_k, j_{k-1}, y_{k-1}, \hat{x}_{k-1}) \Big/ \sum_i \int numerator\ dx, \tag{4.12}$$

where $k \ge 1$, and obtain the optimal functions $f_k(\cdot)$ and $\iota_k(\cdot)$ using (4.3) and (4.4). Finally, knowing $f_k(\cdot)$ and $r_k(\cdot)$, we can obtain the following distribution of the Markov vector $\aleph_k$:

$$q_k(i_k, x_k, j_k, y_k, \hat{x}_k) = \sum_j \int \delta[\hat{x}_k - f_k(j_k, y_k, j, y, \hat{x})]r_k(i_k, x_k, j_k, y_k, j, y, \hat{x})dy\, d\hat{x}; \tag{4.13}$$

here, the summation is over all $j = \overline{1, M}$.

The initial condition for the recurrent chain (4.10)–(4.13) is the distribution $q_0(\cdot)$, which is obtained given the known distribution $p_0(\cdot)$ of the initial state of the observation plant (1.1):

$$r_0(i, x, j, y) = \beta_0(j, y|i, x)p_0(i, x), \tag{4.14}$$

$$q_0(i,x,j,y,\hat{x}) = \delta[\hat{x} - f_0(j,y)]r_0(i,x,j,y). \tag{4.15}$$

As a result, the initial estimating distribution $\rho_0(\cdot)$ is also found by (2.13), so that the initial estimates $\hat{I}_0$, $\hat{X}_0$ in the AOF and the OSF are identical.

T h e o r e m 3. The optimal structure functions (4.4) of the OSF (4.1) and (4.2) are found recurrently, along with the required distributions of probabilities using chain (4.6)–(4.15).

4.4. Algorithms for the Filter Synthesis

The relations obtained above make it possible to find the optimal structure functions of the finite-dimensional OSF in advance because the number of arguments of these functions is constant. For this purpose, given the preceding distribution $q_{k-1}(\cdot)$ for each current time $k \geq 1$, we should successively find the following distributions: unconditional $p_k(\cdot)$ and $r_k(\cdot)$ using (4.10) and (4.11) and the estimating distribution $\rho_k(\cdot)$ using (4.12). Then, the structure functions $\iota_k(\cdot)$, $f_k(\cdot)$ should be obtained using (4.3) and (4.4), after which we can determine the new distribution $o_k(\cdot)$ using (4.13). The scheme of these computations is

$$\ldots \to q_{k-1}(\cdot) \to p_k(\cdot) \to r_k(\cdot) \to \rho_k(\cdot) \to f_k(\cdot) \to q_k(\cdot) \to \ldots, \quad k \geq 1. \tag{4.16}$$
$$\searrow \iota_k(\cdot)$$

This sequence of computations begins with finding the initial functions $\iota_0(\cdot)$, $f_0(\cdot)$ and the distribution $q_0(\cdot)$ by formulas (2.13), (4.3), (4.4), (4.14), and (4.15), which can be illustrated by the scheme

$$p_0(\cdot) \to r_0(\cdot) \to \rho_0(\cdot) \to f_0(\cdot) \to q_0(\cdot) \to \ldots \tag{4.17}$$
$$\searrow \iota_0(\cdot)$$

Algorithm (4.16), (4.17) uses only one conditional distribution $\rho_k(\cdot)$; therefore, it is convenient for the numerical synthesis of the OSF, e.g., by the Monte-Carlo method. However, this procedure is fairly complicated because it requires histograms of the unknown structure functions to be constructed. To obtain numerical-analytical approximations of the OSF, it turned out to be more convenient to additionally use other conditional distributions (see [12]) by performing in (4.16) the transition from the joint distribution $q_{k-1}(\cdot)$ to the conditional distribution $\rho_k(\cdot)$ by formulas (4.6)–(4.9) according to the scheme

$$q_{k-1}(\cdot) \to \xi_{k-1}(\cdot) \to \pi_k(\cdot) \to \omega_k(\cdot) \to \rho_k(\cdot). \tag{4.18}$$

Let us make algorithm (4.18) more specific by replacing in it the distributions with probabilities and conditional densities. To this end, we introduce additional notation for groups of variables

$$z_k = (y_k, \hat{x}_k), \quad \bar{z}_k = (j_k, z_k)$$

and replace distributions with the corresponding products as

$$\omega_k(i,x,j,y|\bar{z}_{k-1}) = \Omega_k^{(ij)}(\bar{z}_{k-1})\omega_k^{(ij)}(x,y|\bar{z}_{k-1}), \quad q_k(i,x,j,z) = Q_k^{(ij)}q_k^{(ij)}(x,z),$$

$$\rho_k(i,x|j,y,\bar{z}_{k-1}) = \mathrm{P}_k^{(i|j)}(y,\bar{z}_{k-1})\rho_k^{(i|j)}(x|y,\bar{z}_{k-1}), \quad \xi_k(i,x|j,z) = \Xi_k^{(i|j)}(z)\xi_k^{(i|j)}(x|z).$$

Then, we easily obtain from (4.6) the following analogs of relations (3.3) and (3.4):

$$\mathrm{P}_k^{(i|j)}(y,\bar{z}) = \Omega_k^{(ij)}(\bar{z})\int\omega_k^{(ij)}(x,y|\bar{z})dx \Big/ \sum_i numerator, \tag{4.19}$$

$$\rho_k^{(i|j)}(x|y,\bar{z}) = \omega_k^{(ij)}(x,y|\bar{z}) \Big/ \int numerator\ dx; \tag{4.20}$$

they differ only in that the growing group of k arguments $v_k = (\bar{y}_0, \ldots, \bar{y}_{k-1})$ is replaced with the constant group of variables $\bar{z}_{k-1} = (\bar{y}_{k-1}, \hat{x}_{k-1})$. The probability $\Omega_k^{(ij)}(\cdot)$ and the conditional density $\omega_k^{(ij)}(\cdot)$ appearing in these relations are found using expressions implied by (4.7):

$$\Omega_k^{(ij)}(\bar{z}) = \int \mathrm{B}_k^{(j|i)}(x)\pi_k(i,x|\bar{z})dx, \quad \omega_k^{(ij)}(x,y|\bar{z}) = \beta_k(j,y|i,x)\,\pi_k(i,x|\bar{z}) \big/ \Omega_k^{(ij)}(\bar{z}).$$

In turn, formula (4.8) represents the next prediction distribution $\pi_{k+1}(\cdot)$ in terms of the conditional distribution $\xi_k(\cdot)$. For its probability and density, we obtain the following formulas for the transformation of the joint probability $Q_k^{(ij)}$ and density $q_k^{(ij)}(\cdot)$ to the conditional ones

$$\Xi_k^{(i|j)}(z) = Q_k^{(ij)} \int q_k^{(ij)}(x,z)dx \Big/ \sum_i numerator, \tag{4.21}$$

$$\xi_k^{(i|j)}(x|z) = q_k^{(ij)}(x,z) \Big/ \int numerator\ dx; \tag{4.22}$$

these formulas are similar to (4.19) and (4.20).

5. THE GAUSSIAN OPTIMAL STRUCTURE FILTER

To simplify the procedure for the synthesis of the exact OSF described above and make it use only algebraic relations, we approximate two probability densities.

5.1. Approximation of the Conditional Probability Density

First, we approximate the joint conditional density in (4.20) by a Gaussian one:

$$\omega_k^{(ij)}(x,y|\bar{z}) \approx N\left(x,y \| \lambda_k^{(ij)}(\bar{z}), \mu_k^{(ij)}(\bar{z}); \Psi_k^{(ij)}(\bar{z}), \Phi_k^{(ij)}(\bar{z}), \Delta_k^{(ij)}(\bar{z})\right). \tag{5.1}$$

It differs from (3.5) only in the replacement of the condition v_k with $\bar{z}_{k-1}$, while the relations between their parameters, which are similar to (3.6)–(3.9), remain the same. Then, accurate to the form of this condition, Lemma 1 holds with relations (3.14)–(3.17). Therefore, the estimating probability and density (4.5) have the analytical approximations

$$\mathrm{P}_k^{(i|j)}(y,\bar{z}) \approx \Omega_k^{(ij)}(\bar{z}) N\left(y \| \mu_k^{(ij)}(\bar{z}); \Phi_k^{(ij)}(\bar{z})\right) \Big/ \sum_i numerator, \tag{5.2}$$

$$\rho_k^{(i|j)}(x|y,\bar{z}) \approx N\left(x \| \sigma_k^{(i|j)}(y,\bar{z}), \Upsilon_k^{(i|j)}(\bar{z})\right);$$

and the parameters of the density are determined by the formulas

$$\begin{aligned} \sigma_k^{(i|j)}(y,\bar{z}) &= \lambda_k^{(ij)}(\bar{z}) + \Delta_k^{(ij)}(\bar{z}) \Phi_k^{(ij)\oplus}(\bar{z})\{y - \mu_k^{(ij)}(\bar{z})\}, \\ \Upsilon_k^{(i|j)}(\bar{z}) &= \Psi_k^{(ij)}(\bar{z}) - \Delta_k^{(ij)}(\bar{z}) \Phi_k^{(ij)\oplus}(\bar{z}) \Delta_k^{(ij)\mathrm{T}}(\bar{z}). \end{aligned} \tag{5.3}$$

Estimates of the Gaussian OSF are also found using (3.16) or (3.17) but with the argument $\bar{Y}_0^{k-1}$ replaced with $\bar{Z}_{k-1} = (\bar{Y}_{k-1}, \hat{X}_{k-1})$, so that now we have in these formulas

$$\mathrm{P}_k^{(i)} = \mathrm{P}_k^{(i|J_k)}(Y_k, \bar{Z}_{k-1}), \quad \hat{X}_k^{(i)} \approx \sigma_k^{(i|J_k)}(Y_k, \bar{Z}_{k-1}), \quad \Upsilon_k^{(i)} = \Upsilon_k^{(i|J_k)}(\bar{Z}_{k-1}). \tag{5.4}$$

According to (4.7), the prediction distribution $\pi_k(\cdot)$ is marginal with respect to the distribution $\omega_k(\cdot)$ and also in the case of the OSF similarly to (2.10); hence the poly-Gaussian property of the prediction density is preserved accurate to its condition:

$$\pi_k^{(i)}(x|\bar{z}) \approx \sum_j \Omega_k^{(ij)}(\bar{z}) N\left(x \| \lambda_k^{(ij)}(\bar{z}); \Psi_k^{(ij)}(\bar{z})\right) \Big/ \Pi_k^{(i)}(\bar{z}).$$

Therefore, Lemma 2 with expressions (3.21) also holds with the same difference, so that the Gaussian means (3.22) and (3.23) also allow us to represent the measurement prediction functions in terms of the state prediction functions by the formulas

$$\begin{aligned} \mu_k^{(ij)}(\bar{z}) &\approx \sum_\gamma \Omega_k^{(i\gamma)}(\bar{z}) h_k^{(j|i)}\left(\lambda_k^{(i\gamma)}(\bar{z}); \Psi_k^{(i\gamma)}(\bar{z})\right) \Big/ \Omega_k^{(ij)}(\bar{z}), \\ \Delta_k^{(ij)}(\bar{z}) &\approx \sum_\gamma \Omega_k^{(i\gamma)}(\bar{z}) G_k^{(j|i)}\left(\lambda_k^{(i\gamma)}(\bar{z}); \Psi_k^{(i\gamma)}(\bar{z})\right) \Big/ \Omega_k^{(ij)}(\bar{z}) - \lambda_k^{(ij)}(\bar{z}) \mu_k^{(ij)\mathrm{T}}(\bar{z}), \\ \Phi_k^{(ij)}(\bar{z}) &\approx \sum_\gamma \Omega_k^{(i\gamma)}(\bar{z}) F_k^{(j|i)}\left(\lambda_k^{(i\gamma)}(\bar{z}); \Psi_k^{(i\gamma)}(\bar{z})\right) \Big/ \Omega_k^{(ij)}(\bar{z}) - \mu_k^{(ij)}(\bar{z}) \mu_k^{(ij)\mathrm{T}}(\bar{z}). \end{aligned} \tag{5.5}$$

Here, similarly to (3.24), we have the following representations for the state prediction functions:

$$\begin{aligned}
\Omega_{k+1}^{(lf)}(\bar{z}_k) &= \int \mathrm{B}_{k+1}^{(f|l)}(\chi)\pi_{k+1}(l,\chi|\bar{z}_k)d\chi, \quad \bar{z}_k = (j_k, y_k, \hat{x}_k),\\
\lambda_{k+1}^{(lf)}(\bar{z}_k) &= \int \chi \mathrm{B}_{k+1}^{(f|l)}(\chi)\pi_{k+1}(l,\chi|\bar{z}_k)d\chi \Big/ \Omega_{k+1}^{(lf)}(\bar{z}_k),\\
\Psi_{k+1}^{(lf)}(\bar{z}_k) &= \int \chi\chi^{\mathrm{T}} \mathrm{B}_{k+1}^{(f|l)}(\chi)\pi_{k+1}(l,\chi|\bar{z}_k)d\chi \Big/ \Omega_{k+1}^{(lf)}(\bar{z}_k) - \lambda_{k+1}^{(lf)}(\bar{z}_k)\lambda_{k+1}^{(lf)\mathrm{T}}(\bar{z}_k).
\end{aligned} \tag{5.6}$$

L e m m a 6. If the Gaussian approximation of the conditional probability density (5.1) holds, then elements (5.4) of the OSF estimates (3.16) or (3.17) are represented by the algebraic formulas (5.2) and (5.5) in terms of the state prediction functions (5.6).

5.2. Approximation of the Unconditional Probability Density

Finally, we find the algebraic dependences of functions (5.6) on their common argument $\bar{z}_k$. Now, the prediction distribution $\pi_{k+1}(\cdot)$, in contrast to (2.11), is determined by (4.8). Substitute it into (5.6) and use Lemma 3 to obtain for the state prediction functions the following representations with the conditional means (3.26):

$$\begin{aligned}
\Omega_{k+1}^{(lf)}(\bar{z}) &= \sum_i \int \varsigma_k^{(lf|ij)}(x,y)\xi_k(i,x|j,y,\hat{x})dx, \quad \bar{z} = (j,y,\hat{x}),\\
\lambda_{k+1}^{(lf)}(\bar{z}) &= \sum_i \int \varphi_k^{(lf|ij)}(x,y)\xi_k(i,x|j,y,\hat{x})dx \Big/ \Omega_{k+1}^{(lf)}(\bar{z}),\\
\Psi_{k+1}^{(lf)}(\bar{z}) &= \sum_i \int \Sigma_k^{(lf|ij)}(x,y)\xi_k(i,x|j,y,\hat{x})dx \Big/ \Omega_{k+1}^{(lf)}(\bar{z}) - \lambda_{k+1}^{(lf)}(\bar{z})\lambda_{k+1}^{(lf)\mathrm{T}}(\bar{z});
\end{aligned} \tag{5.7}$$

however, with the conditional distribution (4.9) whose factors have the form (4.21) and (4.22).

Therefore, in order to obtain algebraic dependences from (5.7), we also approximate the unconditional density by the Gaussian one as

$$q_k^{(ij)}(x,z) \approx N\left(x,z \| m_k^{x|ij}, m_k^{z|ij}; D_k^{x|ij}, D_k^{z|ij}, D_{kk}^{xz|ij}\right); \tag{5.8}$$

as the proximity measure of these two densities, we also choose the equality of all their expectations and covariances

$$m_k^{u|ij} = \mathrm{M}[U_k|I_k = i, J_k = j], \quad D_{kk}^{u\sigma|ij} = \mathrm{cov}[U_k, \Sigma_k|I_k = i, J_k = j], \quad D_k^{u|ij} = D_{kk}^{uu|ij}.$$

Then, the conditional probability (4.21) has the fractional Gaussian approximation

$$\Xi_k^{(i|j)}(z) \approx Q_k^{(ij)} N(z\|m_k^{z|ij}; D_k^{z|ij}) \Big/ \sum_i numerator, \tag{5.9}$$

and the conditional density (4.22) is also approximated by the Gaussian one by the same normal correlation theorem as

$$\xi_k^{(i|j)}(x|z) \approx N\left(x \| u_k^{(i|j)}(z); T_k^{(i|j)}\right). \tag{5.10}$$

The expectation of this approximation is a linear function of its condition

$$u_k^{(i|j)}(z) = \Gamma_k^{(i|j)} z + \kappa_k^{(i|j)}, \quad z = (y^{\mathrm{T}}\hat{x}^{\mathrm{T}})^{\mathrm{T}}, \tag{5.11}$$

and the covariance is independent of the expectation; along with the parameters of function (5.11), it is determined by the formulas

$$\Gamma_k^{(i|j)} = D_{kk}^{xz|ij}\left(D_k^{z|ij}\right)^{\oplus}, \quad \kappa_k^{(i|j)} = m_k^{x|ij} - \Gamma_k^{(i|j)} m_k^{z|ij}, \quad T_k^{(i|j)} = D_k^{x|ij} - \Gamma_k^{(i|j)}\left(D_{kk}^{xz|ij}\right)^{\mathrm{T}}. \tag{5.12}$$

Due to (5.10), the conditional distribution (4.9) has the approximation

$$\xi_k(i,x|j,z) \approx \Xi_k^{(i|j)}(z) N\left(x \| u_k^{(i|j)}(z); T_k^{(i|j)}\right).$$

Substitute it into (5.7) similarly to (3.25) to find the algebraic expressions

$$\begin{aligned}
\Omega_{k+1}^{(lf)}(j,z) &\approx \sum_i \Xi_k^{(i|j)}(z)\eta_k^{(lf|ij)}\left(y,u_k^{(i|j)}(z);T_k^{(i|j)}\right),\\
\lambda_{k+1}^{(lf)}(j,z) &\approx \sum_i \Xi_k^{(i|j)}(z)\tau_k^{(lf|ij)}\left(y,u_k^{(i|j)}(z);T_k^{(i|j)}\right)\Big/\Omega_{k+1}^{(lf)}(j,z),\\
\Psi_{k+1}^{(lf)}(j,z) &\approx \sum_i \Xi_k^{(i|j)}(z)\mathrm{X}_k^{(lf|ij)}\left(y,u_k^{(i|j)}(z);T_k^{(i|j)}\right)\Big/\Omega_{k+1}^{(lf)}(j,z) - \lambda_{k+1}^{(lf)}(j,z)\lambda_{k+1}^{(lf)\mathrm{T}}(j,z)
\end{aligned} \tag{5.13}$$

with the known Gaussian means (3.29).

L e m m a 7. If the Gaussian approximation of the unconditional probability density (5.8) holds, then the state prediction functions are analytically expressed by (5.13) in terms of characteristics (5.9) and (5.11) of the preceding measurement Y_k and estimate $\hat{X}_k$, which are found using the numerical parameters of the filter $Q_k^{(ij)}$, $m_k^{z|ij}$, $D_k^{z|ij}$, and (5.12).

5.3. Equations of the Gaussian Filter

We summarize the above reasoning in the following theorem.

T h e o r e m 4. If the approximations of the probability densities (5.1) and (5.8) are valid, then we have the following equations of the Gaussian approximation of the OSF (these equations are written as a chain of formulas below). Upon specifying the same initial conditions $\Omega_0^{(ij)}, \lambda_0^{(ij)}, \mu_0^{(ij)}, \Psi_0^{(ij)}, \Delta_0^{(ij)}, \Phi_0^{(ij)}$ as in the NAF, the following quantities are computed for each $k \geq 0$.

1. Characteristics of possible conditional measurement predictions (if $k \neq 0$) using (5.5):

$$\begin{aligned}
\mu_k^{(ij)} &= \sum_\gamma \Omega_k^{(i\gamma)} h_k^{(j|i)}\left(\lambda_k^{(i\gamma)};\Psi_k^{(i\gamma)}\right)\Big/\Omega_k^{(ij)},\\
\Delta_k^{(ij)} &= \sum_\gamma \Omega_k^{(i\gamma)} G_k^{(j|i)}\left(\lambda_k^{(i\gamma)};\Psi_k^{(i\gamma)}\right)\Big/\Omega_k^{(ij)} - \lambda_k^{(ij)}\mu_k^{(ij)\mathrm{T}},\\
\Phi_k^{(ij)} &= \sum_\gamma \Omega_k^{(i\gamma)} F_k^{(j|i)}\left(\lambda_k^{(i\gamma)};\Psi_k^{(i\gamma)}\right)\Big/\Omega_k^{(ij)} - \mu_k^{(ij)}\mu_k^{(ij)\mathrm{T}}.
\end{aligned} \tag{5.14}$$

2. Characteristics of the conditional state and measurement predictions specified by the indicator J_k

$$\Omega_k^{(i)} = \Omega_k^{(iJ_k)},\quad \lambda_k^{(i)} = \lambda_k^{(iJ_k)},\quad \mu_k^{(i)} = \mu_k^{(iJ_k)},\quad \Psi_k^{(i)} = \Psi_k^{(iJ_k)},\quad \Delta_k^{(i)} = \Delta_k^{(iJ_k)},\quad \Phi_k^{(i)} = \Phi_k^{(iJ_k)},$$

and the numerical filter parameters

$$\Gamma_k^{(i)} = \Gamma_k^{(i|J_k)},\quad \kappa_k^{(i)} = \kappa_k^{(i|J_k)},\quad T_k^{(i)} = T_k^{(i|J_k)},\quad Q_k^{(i)} = Q_k^{(i|J_k)},\quad m_k^{z|i} = m_k^{z|iJ_k},\quad D_k^{z|i} = D_k^{z|iJ_k}.$$

3. Parameters of the Gaussian approximation of the estimating distribution, which are determined by the measurement Y_k, using (5.2) and (5.3):

$$\mathrm{P}_k^{(i)} = \Omega_k^{(i)} N\left(Y_k \| \mu_k^{(i)};\Phi_k^{(i)}\right)\Big/\sum_i \textit{numerator},$$

$$\sigma_k^{(i)} = \lambda_k^{(i)} + \Delta_k^{(i)}\Phi_k^{(i)\oplus}\left\{Y_k - \mu_k^{(i)}\right\},\quad \Upsilon_k^{(i)} = \Psi_k^{(i)} - \Delta_k^{(i)}\Phi_k^{(i)\oplus}\Delta_k^{(i)\mathrm{T}};$$

quadratically simple estimates using (3.16)

$$\hat{I}_k \in \operatorname{Arg}\max_{i=\overline{1,L}} \mathrm{P}_k^{(i)},\quad \hat{X}_k = \sum_i \mathrm{P}_k^{(i)}\sigma_k^{(i)};$$

simple multiplicative estimates using (3.17)

$$\hat{I}_k \in \operatorname{Arg}\max_{i=\overline{1,L}}\left[\mathrm{P}_k^{(i)}\Big/\left(2\pi\det\Upsilon_k^{(i)}\right)^{n/2}\right],\quad \hat{X}_k = \sigma_k^{(\hat{I}_k)};$$

and the characteristics of the state vector $\hat{X}_k$ modified using the parameters of filter (5.12) and the repeated account for the measurement Y_k by formulas (5.9) and (5.11):

$$\Xi_k^{(i)} = Q_k^{(i)} N(Z_k \| m_k^{z|i}; D_k^{z|i}) \Big/ \sum_i numerator, \qquad Z_k = \begin{bmatrix} Y_k \\ \hat{X}_k \end{bmatrix}. \qquad (5.15)$$
$$u_k^{(i)} = \Gamma_k^{(i)} Z_k + \kappa_k^{(i)},$$

4. Characteristics of possible conditional predictions of the next state using (5.13):

$$\Omega_{k+1}^{(lf)} = \sum_i \Xi_k^{(i)} \eta_k^{(lf|iJ_k)} \left(Y_k, u_k^{(i)}; T_k^{(i)}\right),$$
$$\lambda_{k+1}^{(lf)} = \sum_i \Xi_k^{(i)} \tau_k^{(lf|iJ_k)} \left(Y_k, u_k^{(i)}; T_k^{(i)}\right) \Big/ \Omega_{k+1}^{(lf)}, \qquad (5.16)$$
$$\Psi_{k+1}^{(lf)} = \sum_i \Xi_k^{(i)} \mathrm{X}_k^{(lf|iJ_k)} \left(Y_k, u_k^{(i)}; T_k^{(i)}\right) \Big/ \Omega_{k+1}^{(lf)} - \lambda_{k+1}^{(lf)} \lambda_{k+1}^{(lf)\mathrm{T}}.$$

5. Set $k := k + 1$ and go to Step 1.

5.4. Finding the Numerical Parameters

All the parameters of the *Gaussian OSF* (GOSF) are numerical characteristics of the random state of the plant $\bar{X}_k = (I_k, X_k)$, its measurer $\bar{Y}_k = (J_k, Y_k)$, and the filter state $\hat{X}_k$ in the form of the probability $Q_k^{(ij)}$ of their discrete components I_k, J_k, the first conditional moments $m_k^{z|ij}$, and $D_k^{z|ij}$ of their absolutely continuous components; alternatively; they can be represented in terms of such moments by (5.12). Therefore, all these characteristics can be easily determined using sufficiently large samples of these random variables using direct statistical simulation of the observation system (1.1) and (1.2), and the filter (5.14)–(5.16) itself, in the same way as this is done for the monostructure OSF in [12] and for the two-step LD OSF in [14]. Simultaneously, the sample value of the optimality criterion (1.4) can be easily obtained at each k, thus analyzing the GOSF accuracy.

C o r o l l a r y 2. An algorithm for the computation of the GOSF parameters using the Monte-Carlo method can be described by the scheme

$$\begin{array}{ccccccccccccccc} \{\bar{V}_{k-1}\} & & \{\bar{W}_k\} & & & & & & & & & & & & \\ \downarrow & & \downarrow & & & & & & & & & & & & \\ \begin{Bmatrix} I_k \\ X_k \end{Bmatrix} & \to & \begin{Bmatrix} J_k \\ Y_k \end{Bmatrix} & \to & \begin{Bmatrix} \mathrm{P}_k^{(i)} \\ \sigma_k^{(i)} \\ \Upsilon_k^{(i)} \end{Bmatrix} & \to & \begin{Bmatrix} \hat{I}_k \\ \hat{X}_k \end{Bmatrix} & \to & \begin{bmatrix} Q_k^{(ij)} \\ m_k^{e|ij} \\ D_k^{e|ij} \end{bmatrix} & \to & \begin{bmatrix} \Gamma_k^{(i|j)} \\ \kappa_k^{(i|j)} \\ T_k^{(i|j)} \end{bmatrix} & \to & \begin{Bmatrix} \Xi_k^{(i)} \\ u_k^{(i)} \end{Bmatrix} & \to & \begin{Bmatrix} \Omega_{k+1}^{(lf)} \\ \lambda_{k+1}^{(lf)} \\ \Psi_{k+1}^{(lf)} \end{Bmatrix} \to \begin{Bmatrix} \mu_{k+1}^{(lf)} \\ \Delta_{k+1}^{(lf)} \\ \Phi_{k+1}^{(lf)} \end{Bmatrix}. \\ \uparrow & & & & & & & & & & & & & & \\ \uparrow & & & & \uparrow & & & & & & & & & & \downarrow \\ \longleftarrow & & & & \longleftarrow & & k := k+1, & & l := i, & & f := j, & & j := J_k & & \end{array}$$

Here, the braces {·} denote sets of impelementations of the corresponding random variables, the square brackets denote sets of deterministic parameters in the form of probabilities $Q_k^{(ij)} = \Pr\left[I_k = i, J_k = j\right]$, conditional means $m_k^{e|ij}$, the covariance $D_k^{e|ij}$ of the vector $E_k = (X_k, Y_k, \hat{X}_k)$, and parameters (5.12) obtained from these realizations. All these computations begin with the generation of the plant initial state $\bar{X}_0 = (I_0, X_0)$ based on the given distribution $p_0(i, x)$.

6. COMPARISON OF THE GAUSSIAN FILTERS

The difference of the GOSF equations presented in Subsection 5.3 from the equations of the NAF (3.38)–(3.40) is as follows:

the second part of Step 2—the presence of six numerical parameters $\Gamma_k^{(i|j)}$, ..., $D_k^{z|ij}$;

the second part of Step 3—the new relations (5.15);

step 4—the use in the state prediction equations (5.16) of two new random variables $\Xi_k^{(i)}$ and $u_k^{(i)}$ obtained in (5.15), along with the deterministic parameter $T_k^{(i)}$ instead of the three variables of the NAF $\mathrm{P}_k^{(i)}$, $\sigma_k^{(i)}$, and $\Upsilon_k^{(i)}$ in expressions similar to (3.40), respectively.

The comparison of the relations in Subsections 3.6 and 5.3 shows that the considerable reduction in the computer memory used by the GOSF compared with the NAF (reduction in the filter order) does not materially speed up the measurement processing because the estimates are computed by almost identical formulas that involve the same structure correction functions (3.22) and (3.23), and the prediction functions (3.29). However, the use of the preliminary determined numerical parameters (5.12) in the GOSF equations allows one to better adjust it to the problem under consideration, as was the case with the monostructure GOSF in [12].

An example of comparing the NAF with the GOSF and with its two-step version for LD systems, which illustrates this fact, can be found in [14]. In that paper, the Gaussian approximation of another LD low order filter—*two-step OSF (2OSF)* [13], which also produces the optimal predictions $\tilde{I}_k$ and $\tilde{X}_k$—was obtained. The Gaussian 2OSF turned out to be simpler than the NAF and the GOSF because its equations do not require the computation of the covariances of the prediction errors $\Psi_{k+1}^{(lf)}$, which are computed in (3.40) and (5.16) using a complex analytical matrix function $\mathrm{X}_k^{(lf\,|\,i\,j)}(\cdot)$ defined in (3.29). The random covariance $\Psi_{k+1}^{(lf)}$ in the arguments of the 2OSF functions is replaced with its deterministic parameter.

Note that, in the absence of the structure indicator, the equations of the Gaussian OSF and 2OSF are simplified not only as NAF equations (see Subsection 3.7) but the number of their numerical parameters is reduced by a factor of M. Indeed, instead of $\Gamma_k^{(i|j)}$ and $D_k^{z|ij}$ in this case, only $\Gamma_k^{(i)}$ and $D_k^{z|i}$ remain.

CONCLUSIONS

A generalization of the two known methods for the optimal estimation of the state of an LD observation system with discrete time for the case when the measurement errors depend on the plant disturbances is proposed. Constructive Gaussian approximations of the corresponding filters are found, and the comparison of these algorithms is performed.

First, the classical recurrent formula for the posterior probability distribution of the plant state is generalized (Theorem 1), and the ways of its use in real time are analyzed. Properties of the Gaussian approximation of the posterior distribution are established (Lemmas 1–5) and used to find equations of the bank of NAFs (Theorem 2), which, however, have a large order. A simpler version of the NAF is also considered for the case when there is no structure indicator (Corollary 1).

Next, the synthesis procedure of the structure of a finite-dimensional filter with the order equal to the dimension of the state vector is described. An algorithm for the integral recurrent computation of the filter's structure functions is obtained (Theorem 3), which is reduced to a form that is convenient for finding their analytical-numerical approximations. Based on the two proposed Gaussian approximations (Lemmas 6 and 7), an approximation of the OSF (Theorem 4) is constructed, which is similar to the NAF, and an algorithm for the numerical computation of the parameters of this approximation by the Monte-Carlo method is described (Corollary 2). Finally, similarities between the Gaussian OSF and the NAF and their fundamental differences are analyzed.

The results provide a foundation for the construction of more accurate approximations of the AOF, OSF, and 2OSF. However, the use of segments of the Gauss-like Gram–Charlier and Edgeworth series (which account for higher moments) not only increases the complexity of the structure functions of the corresponding approximations of the AOF but also drastically increases its order compared with the NAF [8]; however, it is difficult to obtain a good approximation of the posterior distribution, especially a polymodal one, in this way. The poly-Gaussian approximation is more promising due to its polymodality, and the order of such an approximation of the AOF is equal to the product of NAF order with the number of Gaussian terms. The application of the two-moment parametric approximation [4] transforms the AOF into another suboptimal filter of the same order as that of the NAF but with different prediction and correction functions. In order to obtain them, it is sufficient to replace in (3.5) the joint Gaussian probability density $N(x,y\|m,D)$ with the multidimensional two-parametric probability density $B(x,y\|a,b)$ and then use its marginal $\bar{B}(x\|\bar{a},\bar{b})$, $\tilde{B}(y\|\tilde{a},\tilde{b})$ and conditional $\hat{B}(x\|\hat{a}(y),\hat{b}(y))$ modifications. As a result, all the Gaussian means (the statistical linearization coefficients) such as (3.22) automatically turn into two parametric means of the type $h_k^{(j|i)}(\hat{a},\hat{b}) = \int \nu_k^{(j|i)}(x)\hat{B}(x\|\hat{a},\hat{b})dx$.

The orders of all similar approximations of the OSF and 2OSF remain equal to the number of components of the state vector to be estimated; only the form of the structure functions and the number of the numerical parameters of the suboptimal filter vary. A fundamentally different method of improving the accuracy of the OSF and 2OSF, which proved to be effective, is the construction of Pugachev's conditionally optimal modifications of their Gaussian or linearized approximations [11, 17]. They differ in terms of the additional correction parameters introduced in the suboptimal equations already obtained; these parameters have the form of amplification matrices and offset vectors, which are also optimized. As a result, they are computed along with the basic parameters (5.12) using similar formulas by the Monte-Carlo method.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Science and Education (project no. 1.1191.2014K) and by the Russian Foundation for Basic Research (project no. 15-08-01902-a).

REFERENCES

1. A. Nemura and E. Klekis, *Estimation of State and Parameters of Systems: Systems with Jump-like Varying Properties* (Mokslas, Vilnus, 1988) [in Russian].
2. S. Ya. Zhuk, "Joint filtering of mixed Markov discrete time processes," Izv. Vyssh. Uchebn. Zaved., Radioelektron., No. 1, 33–39 (1988).
3. V. I. Tikhonov and V. N. Kharisov, *Statistical Analysis and Synthesis of Radio Engineering Devices and Systems* (Radio Svyaz', Moscow, 1991) [in Russian].
4. V. A. Bukhalev, *Recognition, Estimation and Control in Systems with Random Jump Structure* (Nauka, Moscow, 1996) [in Russian].
5. A. Doucet, N. Freitas, and N. Gordon, *Sequential Monte-Carlo Methods in Practice* (Springer, New York, 2001).
6. *Markovian Estimation Theory in Radio Engineering*, Ed. by M. S. Yarlykov (Radiotekhnika, Moscow, 2004) [in Russian].
7. A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering* (Springer, New York, 2009).
8. I. N. Sinitsyn, *Kalman and Pugachev Filters* (Logos, Moscow, 2007) [in Russian].
9. A. V. Bosov and A. R. Pankov, "Conditionally minimax filtering of process in system with switching observation channels," Autom. Remote Control **56**, 835 (1995).
10. E. A. Rudenko, "Recurrent finite algorithms for operating mode recognition and state estimation of intelligent management systems," Inform., Ser. Avtomatiz. Proektir., Nos. 2–3, 79–91 (1992).
11. E. A. Rudenko, "Optimal structure of discrete nonlinear filters of low order," Autom. Remote Control **60**, 1261 (1999).
12. E. A. Rudenko, "Optimal discrete nonlinear filters of the object's order and their gaussian approximations," Autom. Remote Control **71**, 320 (2010).
13. E. A. Rudenko, "An optimal discrete nonlinear logical-dynamical filter-predictor," J. Comput. Syst. Sci. Int. **52**, 354 (2013).
14. E. A. Rudenko, "Analytical-numerical approximations of the optimal recurrent logical–dynamical low order filter-predictor," J. Comput. Syst. Sci. Int. **54**, 691 (2015).
15. A. P. Trifonov and Yu. S. Shinakov, *Joint Discrimination of Signals and Estimation of their Parameters against an Interference Background* (Radio Svyaz', Moscow, 1986) [in Russian].
16. A. N. Shiryaev, *Probability* (Nauka, Moscow, 1980) [in Russian].
17. A. V. Panteleev, E. A. Rudenko, and A. S. Bortakovskii, *Nonlinear Control Systems: Description, Analysis and Synthesis* (Vuzovsk. Kniga, Moscow, 2008) [in Russian].

Translated by A. Klimontovich