
**CONTROL IN STOCHASTIC SYSTEMS
AND UNDER UNCERTAINTY CONDITIONS**

Optimal Recurrent Nonlinear Filter of a Large Order for Jump Diffusion Markov Signals

E. A. Rudenko

Moscow Aviation Institute (National Research University), Moscow, 125993 Russia

e-mail: rudenkoevg@yandex.ru

Received July 22, 2019; revised September 9, 2019; accepted September 30, 2019

Abstract—In this paper, we consider the problem of the root-mean-square optimal estimation of the current state of a continuous stochastic object of observation exposed to continuous and pulsed random impacts based on the results of discrete measurements of its output at certain clock time points. To obtain real-time estimates using a low-performance computer, a new discrete finite-dimensional filter that provides estimates only at certain clock and possibly inter-cycle time points is proposed. The vector of its state is composed of the last few clock estimates, while the next estimate is sought in the form of its explicit dependence on the last measurement and the previous state of the filter. The number of previous clock estimates to be taken into account can be selected from the condition of a compromise between the required estimation accuracy and the available measurement processing speed. The prediction between measurements is based on the optimal clock and inter-cycle estimates heuristically. The filter synthesis algorithm and methods for constructing its covariance approximations are presented. An example is considered.

DOI: 10.1134/S1064230720010104

INTRODUCTION

The main problem of optimal nonlinear filtering of Markov signals consists in the great complexity of implementing at the required rate over time an *absolutely optimal filter* (AOF) [1–5] that establishes an explicit dependence of the estimate on all the accumulated measurements. It is an infinite-dimensional dynamic (recurrent) system with distributed parameters since its state is determined by the function of the posterior probability density. Therefore, the modern particle filter [6–8] that implements the direct trajectory realization of the AOF requires employing a very powerful computer due to the use of a cumbersome sequential Monte Carlo method. In practice, this forces us to apply suboptimal finite-dimensional estimation algorithms such as the generalized Kalman filter that is the simplest linearized approximation to AOF. However, the accuracy of such approximations may be insufficient, in particular due to the recurrent accumulation of methodological and computational errors over a long period of time [9, 10]. In addition to the vector of the estimate, the need to calculate the matrix of its covariances makes even such covariance approximations to the AOF rather complicated to implement, especially in the case of a significant number of estimated variables [11].

An alternative to the slow AOF is finite-dimensional *optimal structure filters* (OSFs) of various orders, which are generalizations of a conditionally optimal filter of a specified structure and fixed order [3, 12, 13]. Continuously discrete versions of the OSF for Markov continuous (diffusion) signals are given in [14, 15] while the continuous OSF for piecewise continuous (jump diffusion) signals is described in [16]. In a finite-dimensional approach, the nonlinear filtering problem is split into two parts. The most complicated of them consisting in obtaining the filter equations is solved in advance before the appearance of measurements at the filter design stage. As a result, the implementation of the OSF consisting in recurrently obtaining estimates based on incoming measurements requires only the solution (modeling) of these equations. The complexity of this operation is determined by the number of equations and the type of their nonlinearities.

However, the listed continuously discrete OSFs have a number of disadvantages. The accuracy of a *small-order filter* (SOF) having a continuous prediction [14] is limited due to the invariable dimension of its state vector equal to the number of estimated variables. The order of a *finite memory filter* (FMF) and

a piecewise constant prediction [15] can be already chosen since it is a multiple of the dimension of the measurement vector but the estimation of this filter depends only on the last few measurements.

In this paper, we propose a procedure for constructing a sufficiently simple recursive *large-order filter* (LOF) with a piecewise polynomial prediction. Unlike the FMF, its memory time is infinite, while the contradiction between accuracy and complexity can be controlled by choosing the filter order. The estimation vector is still sought as the best function of the last measurement and the filter state vector. However, the latter is now formed from the vectors of several previous estimates, thereby increasing, in comparison with the SOF, the RAM of the filter, which allows us to improve the accuracy by expanding the allowable set of estimates. In this case, the old measurements are not forgotten because information about them is accumulated in the estimates. Besides, the class of diffusion estimated signals extends to the jump diffusion one.

The synthesis of a new LOF is shown to be reduced to finding an advance relevant conditional probability density from a recursive chain of transformations of the prediction-correction type. It is based on the Kolmogorov–Feller integro-differential equation in partial derivatives (prediction) and the Bayes–Stratonovich integral formula (correction). The LOF can be otherwise synthesized numerically by the Monte Carlo method but with the cumbersome construction of a histogram of the desired structural vector-function of the filter on each measuring step depending on a large number of arguments. Therefore, the construction of two traditional numerical-analytical covariance approximations to the LOF is considered. An example of assessing the state of the Van der Pol stochastic oscillator is demonstrated.

1. PROBLEM STATEMENT FOR CONTINUOUS-DISCRETE FILTERING

Let us consider the following hidden Markov model of an observation system.

Let a piecewise-continuous n -dimensional estimated useful signal $X_t \in \mathbb{R}^n$ that varies in time $t \in [0, T]$, is continuous on the right $X_t = X_{t+}$, and has a limit on the left X_{t-} be the state vector of its forming filter. The latter, which is an object of observation, is perturbed both by continuous Gaussian white noise and by pulsed Poisson white noise. It is described by the Ito stochastic differential equation that has the following integral form:

$$X_t = X_0 + \int_0^t a(\tau, X_{\tau-}) d\tau + \int_0^t B(\tau, X_{\tau-}) dW_\tau + \sum_{k=1}^{P_0^t(X_0)} S_{\Theta_k(X_{\Theta_k-})}(X_{\Theta_k-}). \quad (1.1)$$

Here, X_0 is the random initial signal value characterized by the probability density $p_0(x)$, $a(t, x)$ is the bias vector-function, $B(t, x)$ is the diffusion matrix function, W_t is the standard Wiener process, $S_t(X_{t-}) = X_t - X_{t-}$ is a random vector of the hopping amplitude of the random process X_t at time t , which is determined by the conditional probability density $\xi(t, s|x)$, and $\Theta_1, \Theta_2, \dots$ is the Poisson flow of the time points of the hoppings, which depend on the current value of the signal and have a conditional intensity $\mu(t, x)$,

$$\text{Prob}[\Theta_i(X_{\Theta_i-}) - \Theta_{i-1}(X_{\Theta_{i-1}-}) < \tau | X_t = x] = 1 - e^{-\mu(t,x)\tau}, \quad i = 1, 2, \dots,$$

$P_0^t(x)$ is an integer conditional process counting the number of these hoppings over a time period of $[0, t)$,

$$P_0^t(x) = \{j : [\Theta_1(X_{\Theta_1-}), \dots, \Theta_k(X_{\Theta_k-})] \in [0, t) | X_t = x\}$$

with the Poisson distribution law:

$$\text{Prob}[P_0^t(x) = l] = \frac{v^l(t, x)}{l! e^{-v(t, x)}}, \quad l = 0, 1, \dots, \quad v(t, x) = \int_0^t \mu(\tau, x) d\tau.$$

Note that, in Eq. (1.1), the integral over time is understood as the mean square and the integral over the Wiener process W_t is understood as the Ito stochastic integral. The probability density $p(t, x)$ of a piecewise continuous signal X_t is known to satisfy the Kolmogorov–Feller integro-differential equation [2, 17, 18], which is a generalization of the more well-known Fokker–Planck–Kolmogorov differential equation over the class of jump diffusion processes.

Also let instant inertialess measurements of the vector X_t , including incomplete or inaccurate ones, determined by the following formula of the meter:

$$Y_k = c_k(X_{t_k}, V_k), \quad V_k \sim q_k(v), \quad k = \overline{0, K}, \quad (1.2)$$

be performed at known clock time points of

$$t_0 = 0 < t_1 < \dots < t_K \leq T.$$

Here, $Y_k \in \mathbb{R}^m$ is the m -dimensional measurement vector, $c_k(x, v)$ is the meter vector-function, and V_k is the vector of independent discrete white noise with a probability density $q_k(v)$.

Regarding relations (1.1) and (1.2), we make the following natural assumptions.

Assumption 1. Equation (1.1) has a strong solution X_t and it is unique. Sufficient conditions for this are given in [17, 18].

Assumption 2. The random process X_t and its measurements Y_k have finite second moments

$$M\|X_t\|^2 < \infty, \quad M\|Y_k\|^2 < \infty,$$

where M is the mathematical expectation operator and $\|\varepsilon\|^2 = \varepsilon^T \varepsilon$ is the square of the Euclidean norm.

Assumption 3. For simplicity, we will consider all random variables to be absolutely continuous, which allows characterizing them with probability densities.

It is required at each moment of time from any half-interval between two adjacent clock instants of time to find the estimate \hat{X}_t of the vector X_t as a function of all the measurements accumulated by this time:

$$\hat{X}_t = x_{t,k}(Y_0^k), \quad t \in (t_k, t_{k+1}], \quad Y_0^k = (Y_0, Y_1, \dots, Y_k), \quad (1.3)$$

which is optimal in terms of the minimum standard error of the estimate:

$$M[(X_t - \hat{X}_t)^T C_t (X_t - \hat{X}_t)] \rightarrow \min, \quad t \in [0, T]. \quad (1.4)$$

Here, $C_t = C_t^T > 0$ is a symmetric positive definite matrix of weight coefficients. Note that the criterion (1.4) provides the non-bias of the estimate

$$M[\hat{X}_t] = M[X_t],$$

which guarantees the absence of a systematic error in it. Therefore, this criterion serves to minimize the spread of estimates relative to the estimated value.

2. EQUATION SELECTION FOR THE NEW FILTER

We correct the given problem statement, for which we first describe the known result of its solution.

2.1. Constructing an Absolutely Optimal Filter

If no restrictions are imposed on the class of estimates (1.3) determined by the functions $\hat{x}_{t,k}(\cdot)$ of the number of arguments $m(k+1)$ that grow over time, then the optimal function in the sense of (1.4) is the function of the posterior mean:

$$x_{t,k}(y_0^k) = M[X_t | Y_0^k = y_0^k] = \int x \rho_k(t, x | y_0^k) dx, \quad t \in (t_k, t_{k+1}]. \quad (2.1)$$

Here, $\rho_k(\cdot)$ is the posterior probability density, and the integrals here and below are taken over the entire Euclidean space of the corresponding dimension:

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

In this case, the density $\rho_k(\cdot)$ does not depend on the previous estimating functions (2.1) and the corresponding estimates \hat{X}_t and satisfies the Kolmogorov–Feller integro-differential equation in partial derivatives (prediction equation) on each interval between measurements $(t_k, t_{k+1}]$. At the time of the next measurement t_{k+1} , the final section $\rho_k(t_{k+1}^-, x | y_0^k)$ of the solution of this equation according to the Bayes

formula (correction equation) is recalculated into the initial section $\rho_{k+1}(t_{k+1}^+, x|y_0^{k+1})$ for the next interval $(t_{k+1}, t_{k+2}]$. It is usually complicated to implement these calculations at a rate with the arrival of measurements.

Therefore, we abandon the requirement that the estimate be explicitly dependent on all measurements (1.3), which leads to the AOF's complexity indicated above. Note that the search for an explicit dependence of the clock estimates \hat{X}_{t_k} of the diffusion signal X_t similar to (1.3) only on the last few measurements $Y_k, Y_{k-1}, \dots, Y_{k-l}$ led to a much simpler, albeit, less accurate, continuously discrete OSF [15], but its memory time is finite.

2.2. Filter with a Polynomial Prediction

In order to obtain an equally simple discrete OSF but with an infinite memory time, we propose to search for the best estimate Z_k of only the clock value X_{t_k} of the n -dimensional jump diffusion vector X_t at each measuring cycle k . For possible savings in computational costs, it is also possible to evaluate only its certain, most interesting n' -dimensional part $X_t' \in \mathbb{R}^{n'}$, which, without loss of generality, consists of the first $n' \leq n$ components of the vector X_t . Therefore, the criterion of the optimality of the estimate (1.4) is everywhere, at any moment of time, replaced by a similar, but weaker, clock optimality condition:

$$M[(X_{t_k}' - Z_k)^T C_{t_k} (X_{t_k}' - Z_k)] \rightarrow \min, \quad k = 0, 1, \dots \quad (2.2)$$

The estimate Z_k will be sought in the form of an explicit dependence on the last measurement Y_k and on not more than $l \in \mathbb{N}$ previous estimates $Z_{k-l}^{k-1} = [Z_{k-1}, Z_{k-2}, \dots, Z_{k-l}]$, namely,

$$Z_k = f_k(Y_k, Z_{\max(0, k-l)}^{k-1}), \quad k \geq 1, \quad Z_0 = f_0(Y_0), \quad (2.3)$$

determining the best function $f_k(\cdot)$ based on condition (2.2).

We heuristically construct a prediction \hat{X}_t' of the information vector X_t' between measurements using one or several optimal clock estimates Z_k , for example, in the form of an extrapolating polynomial. For example, the prediction may be constant, linear, etc.:

$$\forall t \in [t_k, t_{k+1}): \hat{X}_t' = Z_k, \quad k \geq 0, \quad \text{or} \quad \hat{X}_t' = Z_k + \frac{Z_k - Z_{k-1}}{t_k - t_{k-1}}(t - t_k), \quad k \geq 1, \quad \text{etc.} \quad (2.4)$$

Thus, to evaluate the random process (1.1) from the discrete measurements (1.2), we propose to synthesize a continuously discrete filter with an optimal clock estimate (2.3) and a polynomial prediction (2.4). Unlike the FMF [15], the old measurements are not forgotten here because the information about them is accumulated in the estimates and so the memory time of the new filter as a dynamic system is infinite. A similar filter for discrete Markov signals is described in [19].

2.3. Filter with a Piecewise Polynomial Prediction

To obtain a more accurate prediction, on the time interval between adjacent measurements $[t_k, t_{k+1}]$, we introduce similarly to [15] additional L inter-cycle points $\tau_k^i, i = \overline{1, L}$, denoting the interval boundaries as τ_k^0 and τ_k^{L+1} :

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

At these points, we will find inter-cycle estimates Z_k^i as their explicit dependences on l clock estimates

$$Z_k^i = g_k^i(Z_{\max(0, k-l+1)}^k), \quad (2.5)$$

determining functions $g_k^i(\cdot)$ from the optimality condition analogous to (2.2):

$$M[(X_{\tau_k^i}' - Z_k^i)^T C_{\tau_k^i} (X_{\tau_k^i}' - Z_k^i)] \rightarrow \min, \quad i = \overline{1, L}, \quad k \geq 1. \quad (2.6)$$

Using these additional estimates Z_k^i , we can construct any of the polynomial predictions of type (2.4) and on each inter-cycle interval separately:

$$\begin{aligned} \forall t \in [\tau_k^i, \tau_k^{i+1}), \quad i = \overline{0, L}: \hat{X}_t^i = Z_k^i, \quad k \geq 0, \\ \text{or} \quad \hat{X}_t^i = Z_k^i + \frac{Z_k^i - Z_k^{i-1}}{\tau_k^i - \tau_k^{i-1}}(t - \tau_k^i), \quad k \geq 1, \quad \text{etc.} \end{aligned} \quad (2.7)$$

3. THE STATE VECTOR AND THE PROPOSED FILTER ORDER

Equation (2.3) linking the current estimate Z_k with the previous ones is a high-order recurrence formula (difference equation). We transform it into a system of first-order equations, for which we collect l of the latest estimates into a block column-vector of a dimension growing initially:

$$U_k = Z_{\max(0, k-l+1)}^k = \begin{cases} Z_0^k, & k = \overline{0, l-1} \text{ (accumulation stage),} \quad \dim Z_0^k = n'(k+1), \\ Z_{k-l+1}^k, & k \geq l \text{ (upgrade stage),} \quad \dim Z_{k-l+1}^k = n'l. \end{cases} \quad (3.1)$$

The process of changing this vector U_k of considered estimates may be written recursively,

$$U_k = s_k(Z_k, U_{k-1}), \quad k \geq 1, \quad U_0 = Z_0, \quad (3.2)$$

using the following vector-functions of estimate accumulation or updating:

$$s_k(Z_k, U_{k-1}) = \begin{cases} \begin{bmatrix} Z_k \\ U_{k-1} \end{bmatrix}, & k = \overline{1, l-1}, \\ \begin{bmatrix} Z_k \\ CU_{k-1} \end{bmatrix}, & k \geq l, \end{cases} \quad C = [E_{(l-1)n'} \quad O_{(l-1)n' \times n'}].$$

Here, C is the matrix of the removal of its last, obsolete block Z_{k-l} from the block column-vector U_{k-1} , while E and O are the unity and zero matrices, respectively. Then, the main filter equation (2.3) and additional expressions for inter-cycle estimates (2.5) take the following form:

$$Z_k = f_k(Y_k, U_{k-1}), \quad k \geq 1, \quad Z_0 = f_0(Y_0), \quad (3.3)$$

$$Z_k^i = g_k^i(U_k), \quad i = \overline{1, L}, \quad k \geq 0. \quad (3.4)$$

Remark 1. Instead of (3.4), we can search for direct dependences of the inter-cycle estimates Z_k^i and on the last measurement Y_k that are similar to (2.3), replacing (2.5) by the relations

$$Z_k^i = f_k^i(Y_k, Z_{\max(0, k-l)}^{k-1}), \quad k \geq 1, \quad Z_0^i = f_0^i(Y_0), \quad i = \overline{1, L}.$$

Given (3.1), these formulas will be similar to (3.3):

$$Z_k^i = f_k^i(Y_k, U_{k-1}), \quad k \geq 1, \quad Z_0^i = f_0^i(Y_0), \quad i = \overline{1, L}.$$

Consequently, the initial input-output filter equations (2.3) and (2.5) are represented by equivalent input-state-output relations (3.2)–(3.4). Indeed, the first-order difference equation (3.2) is the state equation of the filter, while relations (3.3) and (3.4) are the formulas of its outputs. In this case, the state of the filter is determined by the block vector (3.1), and its maximal dimension $p = ln'$, which does not change from the l th step $k = l - 1$, is the filter order. By increasing the coefficient of multiplicity l , this order may be arbitrarily large. Only the filter's output functions $f_k(\cdot)$ and $f_k^i(\cdot)$ are subject to optimization by criterion (2.2), while the function of its state $s_k(\cdot)$ is fixed. As a result, the following statement is proved.

Lemma. *The filter defined by recurrence formulas (2.3) and (2.5) is finite-dimensional with a fixed equation of state (3.2) and optimized formulas of outputs (3.3) and (3.4). Its state vector (3.1) consists of the desired number l of estimates remembered by it, so that the filter order is equal to ln' , where n' is the dimension of the estimation vector.*

Remark 2. Due to the invariance of the state function $s_k(\cdot)$, the proposed estimation algorithm belongs to the type of semi-optimal OSF of an arbitrary order [20]. The new filter differs from the FMF by prin-

ciple replacing past measurements $Y_{\max(0,k-l)}^{k-1}$ with past estimates $Z_{\max(0,k-l)}^{k-1}$. In the particular case when the coefficient of multiplicity $l = 1$ when Eqs. (2.3) and (2.5) take the simplest form

$$Z_k = f_k(Y_k, Z_{k-1}), \quad Z_k^i = f_k^i(Y_k, Z_{k-1}), \quad i = \overline{1, L}, \quad k \geq 1, \quad Z_0 = f_0(Y_0),$$

this filter degenerates into the well-known SOF [15, 21].

Therefore, we give the following definition.

Definition. The recurrent dependence of clock estimates on measurements (2.3), as well as the equivalent pair of the state equation (3.2) and the output formula (3.3), will be termed a filter of large order $p = ln'$ with a polynomial prediction (2.4). These same relations, but with additional formulas for inter-cycle predictions (2.5) or (3.4), will be termed a filter with a piecewise polynomial prediction (2.7).

4. FINDING FILTER OUTPUT FUNCTIONS

Substituting the desired formulas for estimates (2.3) and (2.5) into the corresponding optimality criteria (2.2) and (2.6), we obtain the following result similar to (2.1) based on the well-known theorem on the best mean square regression.

Theorem 1 (on the optimal structure of the LOF). *Under assumptions 1–3, the best functions among all filter output functions measured by Borel exist and are found as the following conditional means:*

$$f_0(y_0) = M[X_0' | y_0] = \int x_0' \rho_0(x_0 | y_0) dx_0, \quad (4.1)$$

$$f_k(y_k, u_{k-1}) = M[X_{t_k}^i | y_k, u_{k-1}] = \int x' \rho_k(t_k^-, x | y_k, u_{k-1}) dx, \quad k \geq 1, \quad (4.2)$$

$$g_k^i(u_k) = M[X_{\tau_k^i}^i | u_k] = \int x' \pi_k(\tau_k^i, x | u_k) dx, \quad i = \overline{1, L}, \quad k \geq 1, \quad (4.3)$$

and their corresponding estimates Z_k and Z_k^i are unbiased.

However, these relationships only express the desired functions in terms of three conditional probability densities: initial $\rho_0(x_0 | y_0)$, corrective $\rho_k(t_k^-, x | y_k, u_{k-1})$, and predictive $\pi_k(t, x | u_k)$, $t \in [t_k, t_{k+1})$. Using the Bayes formula, we find the first of them by the well-known density $p_0(x)$ of the initial state of object (1.1),

$$\rho_0(x_0 | y_0) = \beta_0(y_0 | x_0) p_0(x_0) / \int \text{numerator} dx_0, \quad (4.4)$$

and represent each correcting density $\rho_k(x_{t_k} | y_k, u_{k-1})$ through the final cross section at $t = t_k^-$ of the previous predictive $\pi_{k-1}(t, x | u_{k-1})$, which is valid only on the interval $t \in [t_{k-1}, t_k)$:

$$\rho_k(t_k^-, x | y_k, u_{k-1}) = \beta_k(y_k | x) \pi_{k-1}(t_k^-, x | u_{k-1}) / \int \text{numerator} dx, \quad k \geq 1. \quad (4.5)$$

In the last two relations, $\beta_k(y_k | x_k)$ is the likelihood function obtained from the measuring formula (1.2),

$$\beta_k(y_k | x_k) = \int \delta[y_k - c_k(x_k, v)] q_k(v) dv, \quad (4.6)$$

where $\delta(\cdot)$ is the Dirac function, and the symbol *numerator* hereinafter designates the numerator of the corresponding fraction. Similarly, the predicted density $\pi_k(t, x | u_k)$ can be expressed in terms of the joint probability density $r_k(t, x, u_k)$ of random states of the observation object X_t and filter U_k :

$$\pi_k(t, x | u_k) = r_k(t, x, u_k) / \int \text{numerator} dx, \quad t \in [t_k, t_{k+1}), \quad k \geq 0. \quad (4.7)$$

Finally, using the Markov property of the jump diffusion process X_t determined by the Ito equation (1.1), each of the probability densities $r_k(\cdot)$ on the corresponding time interval between measurements can be shown [15] to satisfy the well-known Kolmogorov–Feller equation:

$$\frac{\partial r_k(t, x, u_k)}{\partial t} = K_x[r_k(t, x, u_k)] + F_x[r_k(t, x, u_k)], \quad t \in [t_k, t_{k+1}), \quad k \geq 0. \quad (4.8)$$

Here, K_x and F_x are two direct generating process operators: the Fokker–Planck–Kolmogorov differential operator

$$K_x[r(t, x, u)] = -\nabla_x^T[a(t, x)r(t, x, u)] + 0.5\text{tr}[\nabla_x \nabla_x^T (B(t, x)B^T(t, x)r(t, x, u))]$$

with the vector ∇_x of the gradient with respect to x and the trace operator of the matrix tr , as well as the Feller integral operator

$$F_x[r(t, x, u)] = -\mu(t, x)r(t, x, u) + \int \mu(t, x-s)r(t, x-s, z)\xi(t, s|x-s)ds.$$

Note that, if the bias function $a(t, x)$ is non-differentiable with respect to the variable x once and the diffusion functions $B(t, x)$ is non-differentiable with respect to the variable x twice, the solution to Eq. (4.8) is understood in the generalized sense by Bubnov–Galerkin [22, 23].

Based on the relations between random variables $U_0 = Z_0$ and $Z_0 = f_0(Y_0)$ given in (3.2) and (3.3) and the known properties of probability densities, the initial condition for equation (4.8) on the first interval $t \in [t_0, t_1)$ is the probability density:

$$r_0(t_0, x, u_0) = p_0(x) \int \delta[u_0 - f_0(y_0)]\beta_0(y_0|x)dy_0. \quad (4.9)$$

Here, the initial estimate function $f_0(y_0)$ is already known from (4.2) and (4.4).

Similarly, the initial condition $r_k(t_k, x, u_k)$ for Eq. (4.8) is found on each of the subsequent inter-cycle time intervals $t \in [t_k, t_{k+1})$, $k \geq 1$. It is related to the final value $r_{k-1}(t_k^-, x, u_{k-1})$ of the solution to this equation on the previous interval $t \in [t_{k-1}, t_k)$ due to the change in the filter state vector from U_{k-1} to U_k at the time instant $t = t_k$ by formula (3.2). Indeed, as a result of the emergence of a new measurement Y_k and obtaining, according to (3.3), the next estimate of Z_k using the found function $f_k(y_k, u_{k-1})$, the latter is added to the vector U_{k-1} with the possible removal of the outdated estimate Z_{k-l} from it at the update stage. Since $U_k = Z_0^k$ at the estimate accumulation stage and $U_k = Z_{k-l+1}^k$ at the update stage, the sections of probability densities $r_k(\cdot)$ and $r_{k-1}(\cdot)$ at the same time t_k are related by one of the following two relations:

$$\begin{aligned} r_k(t_k, x, z_0^k) &= r_{k-1}(t_k^-, x, z_0^{k-1}) \int \delta[z_k - f_k(y_k, z_0^{k-1})]\beta_k(y_k|x)dy_k, & k = \overline{1, l-1}, \\ r_k(t_k, x, z_{k-l+1}^k) &= \iint r_{k-1}(t_k^-, x, z_{k-l}^{k-1})\delta[z_k - f_k(y_k, z_{k-l}^{k-1})]\beta_k(y_k|x)dy_k dz_{k-l}, & k \geq l. \end{aligned} \quad (4.10)$$

Here, the function $f_k(\cdot)$ is known from (4.2), (4.5), and (4.7), while integration over the variable z_{k-l} in the second of these expressions corresponds to the removal of its block Z_{k-l} from the vector $U_{k-1} = Z_{k-l}^{k-1}$.

5. ALGORITHM FOR PRECISE FILTER SYNTHESIS

The relations given above allow determining the following sequence of actions. They are performed in advance before the measurement results appear and, therefore, can be implemented using a rather powerful computer. The latter favorably distinguishes this filter and other OSFs from the classical AOF.

The first method for finding the optimal filter clock functions (4.2) is to alternate the process of solving the Kolmogorov–Feller equation (4.8) on the next inter-cycle time interval $t \in [t_k, t_{k+1})$, $k = 0, 1, \dots$ with a suitable clock conversion (4.10) of the final value of the obtained probability density $r_k(t_{k+1}^-, x, u_k)$ to the initial condition $r_{k+1}(t_{k+1}, x, u_{k+1})$ for solving the same equation on the next interval. To perform such a conversion, a clock function $f_{k+1}(y_{k+1}, u_k)$ is required. Therefore, each solution $r_k(\cdot)$ to this equation obtained on the k th interval by formula (4.7) is converted into a predictive probability density $\pi_k(\cdot)$, the final section of which, according to (4.5), is converted into a correction density $\rho_{k+1}(\cdot)$. The latter allows finding by (4.2) a function $f_{k+1}(\cdot)$ that, firstly, is the result of the filter's synthesis at a specified clock cycle and, secondly, allows performing the aforementioned clock conversion of $r_k(\cdot)$ to $r_{k+1}(\cdot)$ in accordance with (4.10). The initial condition for solving Eq. (4.8) on the interval $t \in [t_0, t_1)$ is the probability density $r_0(\cdot)$ immediately obtained from (4.1), (4.4), and (4.9). At the same time, each inter-cycle cross section of the predictive density $\pi_k(\cdot)$ at time τ_k^i by (4.3) also allows finding the function $g_k^i(u_k)$ of calculating an additional estimate Z_k^i , which serves to construct a more accurate piecewise polynomial prediction (2.7).

The second numerical method for filter synthesis is to obtain the almost exact optimal functions of the conditional means (4.1)–(4.3) due to these calculations in advance using the Monte Carlo method applied step-by-step in time. For this purpose, it is necessary to perform multiple statistical simulations of the equations of object (1.1), meter (1.2), and filter (3.2) and (3.3) at each measuring step k in order to calculate sufficiently large packages of realizations of the random variables $\{X_{t_k^-}\}$, $\{Y_k\}$, and $\{U_{k-1}\}$. In this case, the Ito equation can be integrated using one of the well-known stochastic difference schemes of the Runge–Kutta type, for example, the Euler–Maruyama method. Then, the k th processing of the simulation results should be performed by finding a histogram of the optimal function of the conditional mean $f_k(y_k, u_{k-1})$ and smoothing it out using any of the known methods. Knowledge of this function allows continuing the process of filter modeling for the next clock cycle. At the same time, if it is necessary, the inter-cycle values of the signal $\{X_{t_k^-}\}$ should be fixed for similarly obtaining additional functions $g_k^i(u_k)$.

However, both the described methods for the precise synthesis of the optimal filter are rather technically complicated. Therefore, we further consider the construction of two rather simple numerical-analytical approximations to the proposed filter. These approximations take into account only the covariance of some random variables losing the estimation accuracy due to this, while the form of their functions can be represented by analytical expressions. It only remains for us to numerically obtain the parameters of these expressions.

6. GAUSSIAN APPROXIMATION TO THE FILTER

We use the procedure described in [15] for constructing this well-known approximation to the FMF that, as noted above, differs from the proposed LOF only in the form of the state vector U_k .

We approximate the numerators of the fractions of the Bayes formulas (4.4), (4.5), and (4.7), namely, one conditional and two unconditional probability densities:

$$\eta_k(t_k^-, x, y_k | u_{k-1}) = \beta_k(y_k | x) \pi_{k-1}(t_k^-, x | u_{k-1}), \quad r_{k-1}(t_k^-, x, u_{k-1}), \quad r_0(x_0, y_0) = \beta_0(y_0 | x_0) p_0(x_0)$$

by the normal density $N(\cdot | m, D)$ with the corresponding conditional or unconditional parameters, such as the mean m and covariance D . Then, by the property of the Gaussian distribution law, we obtain that the conditional densities $\rho_k(t_k^-, x | y_k, u_{k-1})$, $\pi_k(t, x | u_k)$, and $\rho_0(x_0 | y_0)$ determined by these fractions are also almost Gaussian, while their parameters, among which the desired conditional means (4.1)–(4.3) are present, are expressed in terms of m and D by the normal correlation theorem. Therefore, the optimal filter output formulas (3.3) obtain approximations that are only linear in the measurements of Y_k and Y_0 . Furthermore, due to the following relation between the densities

$$\pi_{k-1}(t_k^-, x | u_{k-1}) = \int \eta_k(t_k^-, x, y_k | u_{k-1}) dy_k$$

and type (4.6) of the likelihood function $\beta_k(y_k | x)$, three of the five parameters of the Gaussian approximation of the conditional density $\eta_k(t_k^-, x, y_k | u_{k-1})$ can be expressed by its two other parameters using the measuring function $c_k(x, v)$ (1.2). As a result, similarly to [15], we obtain this algorithm for computing clock and inter-cycle estimates.

Theorem 2 (on the Gaussian LOF equations). *If two conditional means of the meter (1.2)*

$$v_k(x) = M[c_k(x, V_k)] \triangleq \int c_k(x, v) q_k(v) dv, \\ \Pi_k(x) = M[c_k(x, V_k) c_k^T(x, V_k)]$$

have Gaussian moments (the characteristics of their statistical linearization according to Kazakov)

$$h_k(m, D) = M_N^{m, D}[v_k(X)] \triangleq \int v_k(x) N(x | m, D) dx, \quad G_k(m, D) = \frac{\partial h_k(m, D)}{\partial m}, \\ F_k(m, D) = M_N^{m, D}[\Pi_k(X)] - h_k(m, D) h_k^T(m, D), \quad (6.1)$$

then the following suboptimal filter equations hold:

$$\begin{aligned} Z_0 &= H_0 Y_0 + e_0, \quad U_k = Z_{\max(0, k-l+1)}^k, \quad \Lambda_k = \Gamma_k U_{k-1} + \kappa_k, \quad k \geq 1, \\ Z_k &= \Lambda_k' + T_k' G_k^T (\Lambda_k, T_k) F_k^{\oplus} (\Lambda_k, T_k) [Y_k - h_k (\Lambda_k, T_k)], \quad k \geq 1, \\ Z_k^i &= \Gamma_k^i U_k + \kappa_k^i, \quad i = \overline{1, L}, \quad k \geq 1. \end{aligned} \quad (6.2)$$

Here, Λ_k is the clock prediction vector, Λ_k' and T_k' are the matrices from the first n' rows of the n -row matrices Λ_k and T_k , \oplus is the Moore–Penrose matrix's pseudoinverse symbol, H_0 , and e_0 are the initial filter parameters, Γ_k , κ_k , and T_k are the clock parameters, and Γ_k^i and κ_k^i are the inter-cycle parameters. In this case, all these parameters are expressed through the first two moments m and D of the well-known random variables by the following formulas:

$$\begin{aligned} H_0 &= D_{00}^{x'y} (D_0^y)^{\oplus}, \quad e_0 = m_0^{x'} - H_0 m_0^y, \\ \Gamma_k &= D_{\bar{t}_k, k-1}^{x,u} (D_{k-1}^u)^{\oplus}, \quad \kappa_k = m_{\bar{t}_k}^x - \Gamma_k m_{k-1}^u, \quad T_k = D_{\bar{t}_k}^x - \Gamma_k (D_{\bar{t}_k, k-1}^{x,u})^T, \\ \Gamma_k^i &= D_{\bar{t}_k, k-1}^{x',u} (D_k^u)^{\oplus}, \quad \kappa_k^i = m_{\bar{t}_k}^{x'} - \Gamma_k^i m_k^u, \quad i = \overline{1, L}. \end{aligned} \quad (6.3)$$

Note that these parameters are also now determined easily by the Monte Carlo method by processing the results of the multiple cycle-by-cycle statistical modeling of the equations of object (1.1), meter (1.2), and filter (6.2) without plotting histograms.

7. LINEARIZED APPROXIMATION TO THE FILTER

Unfortunately, the integral procedure for finding the characteristics of the statistical linearization of nonlinearities (6.1) that is necessary for constructing a Gaussian filter is rather complicated, while they do not always exist in real estimation problems [24]. Therefore, another well-known approximation, which is much simpler and, therefore, more widely used, is applicable. Similarly to [15], we can easily obtain such a statement.

Corollary (on the linearized LOF correction functions). If the nonlinearity $c_k(x, v)$ of meter (1.2) is differentiable with respect to both arguments and the Taylor formula at the point (Λ_k, m_k^v) , where Λ_k is the prediction of the vector $X_{\bar{t}_k}$ and m_k^v is the average value of the interference V_k ,

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

is valid for it, then the functions of the Gaussian correction (6.1) are approximated by the following expressions:

$$\begin{aligned} h_k(m, D) &\approx c_k(m, m_k^v), \quad G_k(m, D) \approx C_k^x(m), \\ F_k(m, D) &\approx C_k^x(m) D C_k^{xT}(m) + C_k(m) R_k C_k^{vT}(m). \end{aligned} \quad (7.1)$$

Here, $C_k^x(x) = \partial c_k(x, v) / \partial x|_{v=m_k^v}$ and $C_k^v(x) = \partial c_k(x, v) / \partial v|_{v=m_k^v}$ are the sections of the Jacobi matrices of partial derivatives, and $R_k = \text{cov}[V_k, V_k]$ is the covariance matrix of the interference measurement.

As a result, the linearized LOF equations also have form (6.2), but with easily obtained functions (7.1), while its parameters are still calculated by formulas (6.3).

Remark 3. The disadvantages of the two considered numerical-analytical approximations to the OSF such as linearized (differentiability of nonlinearities and a small deviation from the linearization point) and Gaussian (the difficulty of analytically obtaining Gaussian moments of nonlinearities and the possibility of their non-existence) may be overcome by using purely numerical approximations occupying an intermediate position between them, such as cubature [25] and unscented or sigma-point [9, 26, 27] approximations.

8. COMPARISON OF THE NEW AND CLASSIC COVARIANCE FILTERS

To clarify the fundamental differences of the suboptimal LOF (6.2) and FMF [15] from a similar approximation to a continuously discrete AOF, we give the equations of the latter. In this case, for sim-

plicity, we restrict ourselves to considering the problem of estimating a purely diffusion signal, assuming that there are no hoppings in the process X_t , so that $X_{t-} = X_t$. Moreover, let the LOF estimate the entire vector X_t , i.e., $n' = n$, as a result of which its estimates Z_k and Z_k^i can be designated as \hat{X}_{t_k} and $\hat{X}_{t_k}^i$.

The linearized version of the AOF is well known as an *extended Kalman filter* (EKF), while its Gaussian version is not as well-known as a *normal approximation filter* (NAF). Their equations also differ only in the form of structural functions. Any of them builds a prediction between measurements $t \in [t_k, t_{k+1})$ continuously, integrating an autonomous (closed) system of ordinary differential equations for the estimate $\hat{X}_t = M[X_t|Y_0^k]$ and the matrix $P_t = \text{cov}[X_t|Y_0^k]$ of the posterior covariances of the estimated vector [3, p. 354]:

$$\begin{cases} \frac{d}{dt} \hat{X}_t = \tau(t, \hat{X}_t, P_t), \\ \frac{d}{dt} P_t = A(t, \hat{X}_t, P_t)P_t + P_t A^\top(t, \hat{X}_t, P_t) + \Theta(t, \hat{X}_t, P_t), \end{cases} \quad t \in [t_k, t_{k+1}), \quad k = 0, 1, \dots \quad (8.1)$$

Then, at each clock time point t_k , the final values of the solutions $\hat{X}_{t_k}^-$ and $P_{t_k}^-$ of system (8.1) on the previous interval $t \in [t_{k-1}, t_k)$ are adjusted by measuring Y_k in the initial conditions $\hat{X}_{t_k}^-$ and $P_{t_k}^-$ for the next interval $t \in [t_k, t_{k+1})$ using the following formulas [3, p. 458]:

$$\begin{cases} \hat{X}_{t_k} = \hat{X}_{t_k}^- + H_k[Y_k - h_k(\hat{X}_{t_k}^-, P_{t_k}^-)], \\ P_{t_k} = P_{t_k}^- - H_k G_k(\hat{X}_{t_k}^-, P_{t_k}^-) P_{t_k}^-. \end{cases} \quad H_k = P_{t_k}^- G_k^\top(\hat{X}_{t_k}^-, P_{t_k}^-) F_k^\oplus(\hat{X}_{t_k}^-, P_{t_k}^-), \quad k = 0, 1, \dots \quad (8.2)$$

Here, the initial values are deterministic:

$$\hat{X}_{t_0}^- = m_0^x, \quad P_{t_0}^- = D_0^x.$$

Thus, in addition to the same correction functions $h_k(\cdot)$, $G_k(\cdot)$, and $F_k(\cdot)$ as in the LOF, three prediction functions $\tau(\cdot)$, $A(\cdot)$, and $\Theta(\cdot)$ are also needed now. In the Gaussian approximation for the NAF, they are also found as, similar to (6.1), the characteristics of statistical linearization but of the bias and diffusion functions of the object's equation (1.1):

$$\begin{aligned} \tau(t, m, D) &= M_N^{m, D}[a(t, X)], & A(t, m, D) &= \frac{\partial \tau(t, m, D)}{\partial m} \\ \Theta(t, m, D) &= M_N^{m, D}[B(t, X)B^\top(t, X)], \end{aligned} \quad (8.3)$$

while, in the linearized approximation, for the EKF as in (7.1), they have a simpler form:

$$\begin{aligned} \tau(t, m, D) &= a(t, m), & A(t, m, D) &= \frac{\partial a(t, m)}{\partial m} \\ \Theta(t, m, D) &= B(t, m)B^\top(t, m), \end{aligned} \quad (8.4)$$

However, the system of differential equations (8.1), the second of which is an Riccati-type equation, has to be solved numerically with a sufficiently small step while ensuring the symmetry and positive definiteness of the covariance matrix, which is not at all simple [9, 10]. An n -dimensional vector of the estimate \hat{X}_t and various elements of the $n \times n$ -matrix of the covariance P_{t_k} form a rather large state vector of the filter, so that its order is equal to $n(n+3)/2$. In this case, here it is not possible to save by evaluating only part of the state variables.

Comparing the equations of the new filter (6.2) with the well-known (8.1) and (8.2), we note the following fundamental advantages of approximate OSFs.

1. Clock estimates of suboptimal filters are calculated using the same (common to them) correction function

$$\tilde{f}_k(\xi, \Psi, y) = \xi + \Psi G_k^\top(\xi, \Psi) F_k^\oplus(\xi, \Psi)[y - h_k(\xi, \Psi)],$$

but with respect to different values of its arguments:

$$\hat{X}_{t_k}^{\text{AOF}} = \tilde{f}_k(\hat{X}_{t_k}^-, P_{t_k}^-, Y_k), \quad \hat{X}_{t_k}^{\text{LOF}} = \tilde{f}_k(\Lambda_k, T_k, Y_k).$$

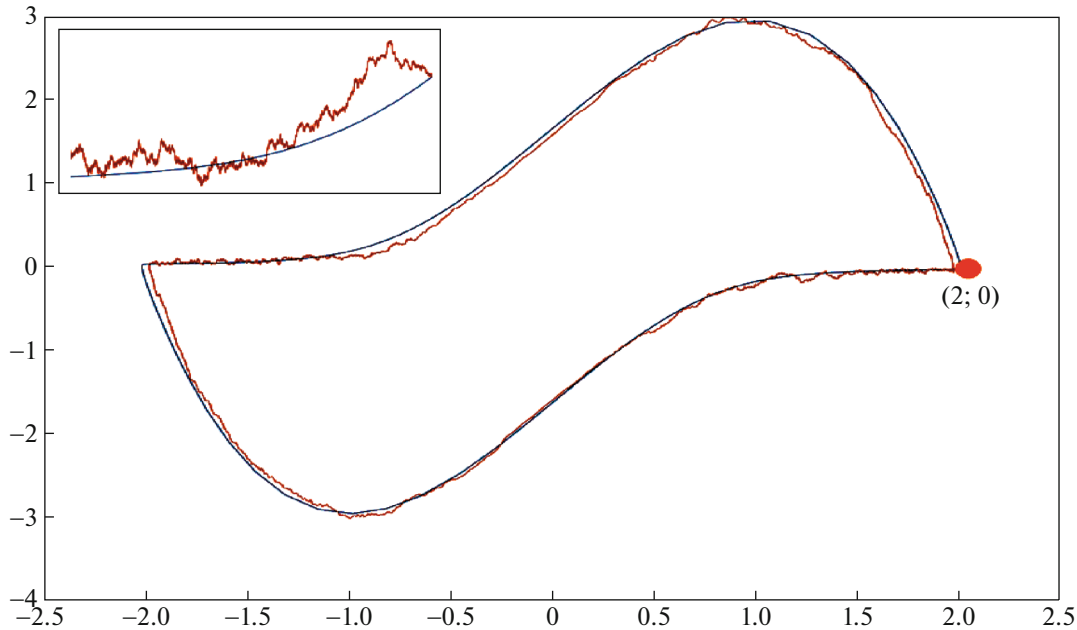


Fig. 1. Limit cycles of the Van der Pol oscillator on the phase plane $(X_{1,t}, X_{2,t})$.

Here, the clock AOF prediction $\hat{X}_{t_k}^-$ obtained by integrating the system of Eqs. (8.1) is replaced by the OSF prediction vector Λ_k , which is easily found by the state vector of this filter $U_{k-1} = \hat{X}_{t_{\max(0, k-t)}}^{t_{k-1}}$ as $\Lambda_k = \Gamma_k U_{k-1} + \kappa_k$ using the precalculated parameters Γ_k and κ_k , and thus it is unbiased. Besides, the clock matrix $P_{t_k}^- = \text{cov}[X_{t_k}^- | Y_0^{k-1}]$ obtained from the same system (8.1) is replaced in the OSF by the predetermined determinate matrix $T_k = \text{cov}[X_{t_k}^- - \Lambda_k]$, so that it is also not necessary to solve a complex equation of the Riccati type.

2. Over the time interval between measurements, the AOF prediction is constructed by solving a complex system of differential equations (8.1), while the LOF prediction is determined by a simple extrapolation (2.4) or (2.7) of the clock \hat{X}_{t_k} or inter-cycle $\hat{X}_{t_k}^i$ estimates. The latter are calculated as $\hat{X}_{t_k}^i = \Gamma_k^i U_k + \kappa_k^i$ and are unbiased.

The performed analysis allows drawing the following conclusion.

Statement (on comparing approximations to LOF and AOF). The approximate OSFs do not need to integrate the system of equations of approximate AOFs for the prediction vector and its covariance matrix (8.1) and, therefore it is not necessary to find three prediction functions according to (8.3) or (8.4). Instead, it is necessary to preliminarily find from (6.3) and use the clock parameters Γ_k, κ_k, T_k and inter-cycle parameters Γ_k^i, κ_k^i , which compensate for the lack of information about the covariance of the estimation error P_t .

9. AN EXAMPLE OF THE NUMERICAL COMPARISON OF SUBOPTIMAL FILTERS¹

To analyze the accuracy of the approximations to the AOF and the OSF, we consider the problem of estimating a two-dimensional state vector of a stochastic version of a nonlinear Van der Pol oscillator. The corresponding system of Ito equations (1.1), which is given here in a non-strict Langevin form, has the following form:

$$\begin{cases} \dot{X}_{1,t} = X_{2,t}, & \omega = 0.1\pi, \quad \alpha = 2, \quad \beta = 1, \quad X_{1,0} \sim N(x_1 \parallel 2, 0.15), \\ \dot{X}_{2,t} = -\omega^2 X_{1,t} + \alpha X_{2,t}(1 - \beta X_{1,t}^2) + X_{1,t} \dot{W}_t, & X_{2,0} \sim N(x_2 \parallel 0, 0.15), \end{cases}$$

¹ The calculations were performed by a student, A.A. Rick, under the supervision of the author.

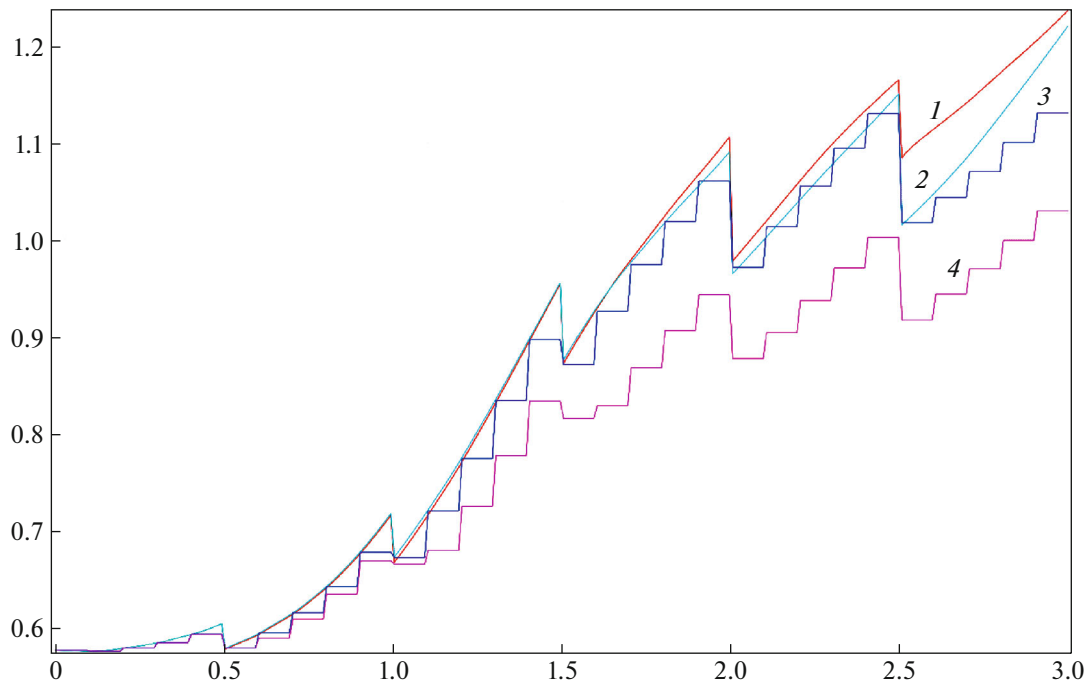


Fig. 2. Time-dependent standard deviations of errors in estimating the variable $X_{1,t}$ of the Van der Pol oscillator by various linearized filters: (1) L-AOF, (2) L-SOF with a continuous prediction, (3) L-LOF of multiplicity $l = 1$ with piecewise constant predictions, and (4) L-FMF of multiplicity $l = 4$ with piecewise constant predictions

where \dot{W}_t is Gaussian white noise. Figure 1 shows the phase plane $(X_{1,t}, X_{2,t})$ with two limit cycles of this oscillator, perturbed (at $\dot{W}_t \neq 0$) and unperturbed (at $\dot{W}_t = 0$), that form the corresponding hysteresis loops.

Let both state variables of this oscillator be measured with additive errors, so that the discrete meter (1.2) is two-dimensional and non-linear in $X_{1,t}$:

$$\begin{cases} Y_{1,k} = X_{1,t_k}^2 + V_{1,k}, & V_{1,k} \sim N(v_1 \parallel 0, 1), \\ Y_{2,k} = X_{2,t_k} + V_{2,k}, & V_{2,k} \sim N(v_2 \parallel 0, 1). \end{cases}$$

The accuracy of the compared filters was determined by the Monte Carlo method with averaging over a sample of 800 realizations. Statistical modeling of the differential equations of the oscillator and the filter with continuous prediction was carried out using the method of the Runge–Kutta type on the interval $[0, 3]$ with an integration step of $\Delta t_{\text{int}} = 0.01$. The time interval was taken to be $\Delta t_{\text{meas}} = 0.5$ between measurements and $\Delta t_{\text{prediction}} = 0.1$ between piecewise-constant discrete predictions. The structural functions of the linearized and Gaussian filters for this example were found in [15].

Figure 2 shows the plots of the *standard deviations* (RMS) of the errors Se_1 for estimating the variable $X_{1,t}$, by several *linearized* (L) continuously-discrete filters with various types of predictions. Four filters are compared:

1. L-AOF (EKF) (8.1), (8.2), and (8.4), the prediction of which is necessarily continuous;
2. L-SOF with a continuous prediction [14];
3. L-LOF of multiplicity $l = 1$ with piecewise constant predictions (2.7) at $L = 4$;
4. L-FMF of multiplicity $l = 4$ [15] with the same piecewise constant predictions.

As can be seen in Fig. 2, the accuracy of the L-AOF, which is of order $p = 2 \times (2 + 3)/2 = 5$ in this example, over time loses increasing to both L-FMPs of order $p = 2$ and, moreover, to 4-fold L-FMF of order $p = 4 \times 2 = 8$.

The implementation time of the new filters in comparison with the classic L-AOF is also better:

—for an L-SOF with a continuous prediction, it is smaller by a factor of 1.6, since, unlike L-AOF, the latter does not integrate the three differential equations for covariances,

—for L-SOF and L-FMF with piecewise constant predictions, it is even smaller by a factor of 2.8 because they do not integrate any differential equations, even for prediction, while the large order of L-FMF barely affected the estimation rate.

Thus, this example both confirms the conclusion formulated in Section 8 about the computational advantages of the linearized LOF and FMF and even demonstrates some of their superiority in accuracy over EKF.

CONCLUSIONS

The synthesis method for a fast continuously discrete nonlinear filter with infinite memory that produces estimates only at discrete time points, based on which it is possible to construct a prediction of the desired type, is proposed. This filter can be implemented in real time on a low-performance computer and has the highest accuracy in its class of simple filters, which remember the last few clock estimates of the information part of the state vector of a continuous object, which is also exposed to pulsed impacts.

Formulas for the optimal output functions of the new filter are obtained. For the corresponding probability densities, a chain of integro-differential equations and clock conversion formulas for its solution is found. A method for calculating the filter's structural functions by the Monte Carlo method is described. A Gaussian approximation to the proposed filter and its linearized simplification are also constructed. A comparative analysis of these approximations with similar ones for the classical filter are performed. The accuracy and estimation rate of the linearized filter is illustrated by a two-dimensional example, which confirms the theoretical conclusions.

FUNDING

This work was financially supported by the Russian Federal Property Fund, project no. 17-08-00530-a.

REFERENCES

1. A. H. Jazwinski, *Stochastic Processes and Filtering Theory* (Academic, New York, London, 1970).
2. M. S. Yarlykov and M. A. Mironov, *Markov Theory of Random Process Estimation* (Radio Svyaz', Moscow, 1993) [in Russian].
3. I. N. Sinitsyn, *Kalman and Pugachev Filters* (Logos, Moscow, 2007) [in Russian].
4. A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering* (Springer, New York, 2009).
5. K. A. Rybakov, *Statistical Methods of Analysis and Filtering in Continuous Stochastic Systems* (MAI, Moscow, 2017) [in Russian].
6. Z. Chen, "Bayesian filtering: from Kalman filters to particle filters, and beyond," Technical Report (Adaptive Syst. Lab., McMaster Univ., Hamilton, ON, Canada, 2003).
7. A. Budhiraja, L. Chen, and C. Lee, "A survey of numerical methods for nonlinear filtering problems," *Phys. D (Amsterdam, Neth.)* **230**, 27–36 (2007).
8. K. A. Rybakov, "Solving the nonlinear problems of estimation for navigation data processing using continuous particle filter," *Gyrosc. Navig.* **10**, 27–34 (2019).
9. M. V. Kulikova and G. Yu. Kulikov, "Numerical methods for nonlinear filtering of signals and measurements," *Vychisl. Tekhnol.* **21** (4), 64–98 (2016).
10. M. V. Kulikova, Yu. V. Tsyganova, "Numerically stable Kalman filter implementations for estimating linear pairwise Markov models in the presence of Gaussian noise," *Vychisl. Tekhnol.* **22** (3), 45–60 (2017).
11. D. A. Kozorez, M. N. Krasil'shchikov, D. M. Kruzhkov, and K. I. Sypalo, "A solution of the navigation problem for autonomous insertion of payload into a geostationary orbit using a low-thrust engine," *J. Comput. Syst. Sci. Int.* **54**, 104 (2015).
12. V. S. Pugachev, I. N. Sinitsyn, and V. I. Shin, "Conditionally optimal discrete filtering of processes in continuously discrete stochastic systems," *Dokl. Akad. Nauk SSSR* **289**, 297–301 (1986).
13. J. R. Raol and N. K. Sinha, "A nonlinear filter for estimation of state of a continuous-time systems with discrete measurements," in *Proceedings of the 2nd IFAC Symposium on Stochastic Control, Vilnius, USSR, 1986*, pp. 211–216.
14. E. A. Rudenko, "Optimal continuous discrete nonlinear filter of small order," in *Proceedings of the 10th International Conference on System Identification and Control Problems (SICPRO-15)* (IPU RAN, Moscow, 2015), pp. 1335–1349.
15. E. A. Rudenko, "Optimal continuous-discrete nonlinear finite memory filter with a discrete predictions," *J. Comput. Syst. Sci. Int.* **55**, 878 (2016).

16. E. A. Rudenko, “Continuous finite-dimensional locally optimal filtering of jump diffusions,” *J. Comput. Syst. Sci. Int.* **57**, 505 (2018).
17. I. I. Gikhman and A. V. Skorokhod, *Introduction to the Theory of Random Processes* (Nauka, Moscow, 1977) [in Russian].
18. V. S. Korolyuk, N. I. Portenko, A. V. Skorokhod, and A. F. Turbin, *Handbook of Probability Theory and Mathematical Statistics* (Nauka, Moscow, 1985) [in Russian].
19. E. A. Rudenko, “Methods and algorithms for optimal finite-dimensional nonlinear filtering of random Markov sequences,” in *Proceedings of the 31st Conference in Memory of N. N. Ostryakov RMPKU-18* (Elektroprigor, St. Petersburg, 2018), pp. 133–145.
20. E. A. Rudenko, “A optimal discrete nonlinear arbitrary-order filter,” *J. Comput. Syst. Sci. Int.* **49**, 548 (2010).
21. E. A. Rudenko, “Optimal structure of discrete nonlinear filters of small order,” *Avtom. Telemekh.*, No. 9, 58–71 (1999).
22. O. A. Ladyzhenskaya, V. A. Solonnikov, and N. N. Ural'tseva, *Parabolic Linear and Quasilinear Equations* (Nauka, Moscow, 1967) [in Russian].
23. A. V. Panteleev, E. A. Rudenko, and A. S. Bortakovskii, *Nonlinear Control Systems: Description, Analysis and Synthesis* (Vuzovsk. Kniga, Moscow, 2008) [in Russian].
24. E. A. Rudenko, “Autonomous estimation of the descent vehicle trajectory by recurrent Gaussian filters,” *Izv. Akad. Nauk, Teor. Syst. Upravl.*, No. 5, 9–29 (2018).
25. I. Arasaratnam and S. Haykin, “Cubature Kalman filters,” *IEEE Trans. Automat. Control.* **54**, 1254–1269 (2009).
26. S. Julier and J. Uhlmann, “Unscented filtering and nonlinear estimation,” *Proc. IEEE* **92**, 401–422 (2004).
27. I. A. Kudryavtseva, “Performance analysis of extended Kalman filter (EKF), the uncented Kalman filter (UKF) and unscented Particle Filter (UPF),” *Nauch. Vestn. MGTU GA*, No. 224, 43–52 (2016).

Translated by A. Ivanov