Sequential Monte Carlo methods for filtering of unobservable components of multidimensional diffusion Markov processes

Ellida M. Khazen *

13395 Coppermine Rd. Apartment 410 Herndon VA 20171 USA

Abstract - The problem of filtering of unobservable components $x(t)$ of a multidimensional continuous diffusion Markov process $z(t) = \big(x(t), y(t)\big)$, given the observations of the (multidimensional) process $y(t)$ taken at discrete consecutive times with small time steps, is analytically investigated. On the base of that investigation the new algorithms for simulation of unobservable components, $x(t)$, and the new algorithms of nonlinear filtering with the use of sequential Monte Carlo methods, or particle filters, are developed and suggested. The analytical investigation of observed quadratic variations is also developed. The new closed-form analytical formulae are obtained, which characterize dispersions of deviations of the observed quadratic variations and the accuracy of some estimates for $x(t)$. As an illustrative example, estimation of volatility (for the problems of financial mathematics) is considered. The obtained new algorithms extend the range of applications of sequential Monte Carlo methods, or particle filters, beyond the hidden Markov models and improve their performance.

Keywords - nonlinear filtering; multidimensional diffusion Markov process; particle filters; sequential Monte Carlo methods; simulation; quadratic variation; volatility.

MSC codes: 60G35, 60J60, 62M05, 93E11.

1. Introduction

In the last two decades a great deal of works has been devoted to development and investigation of particle filters, or sequential Monte Carlo algorithms, for filtering of an unobservable process $x(t) \overset{\text{def}}{=} (x_1(t), \ldots, x_m(t))$ given observations of the other process $y(t) \overset{\text{def}}{=} (y_{m+1}(t), \ldots, y_p(t))$, taken at discrete times t_k, $t_0 < t_1 < t_2 < \ldots < t_k < \ldots$, with small time steps $t_k - t_{k-1} \overset{\text{def}}{=} \Delta t_k$ (see, for example, survey [1], collection [2], works [3 - 5], and references wherein). The observations $y(t_k)$ are being obtained consecutively, and an estimate for $x(t_k)$ should be updated at each time t_k. It is assumed that the whole process $z(t) \overset{\text{def}}{=} \big(x(t), y(t)\big) \overset{\text{def}}{=} (z_1(t), \ldots, z_p(t))$ is a Markov process.

*Correspondence to: Ellida M. Khazen, 13395 Coppermine Rd. Apartment 410 Herndon VA 20171 USA

E-mail: ellida_khazen@yahoo.com

In order to solve the problem of filtering of $x(t_n)$, given the sequence of observations $y(t)|_0^n \overset{\text{def}}{=} (y(t_0), y(t_1), \dots, y(t_n))$, with the use of Monte Carlo methods, the samples of the random sequences $x^i(t)|_0^n \overset{\text{def}}{=} (x^i(t_0), x^i(t_1), \dots, x^i(t_n))$ (from the distributions of $x(t_k)$ when $y(t)|_0^k$ and $x(t)|_0^{k-1}$ are given) should have been simulated numerically for $i = 1, \dots, N$.

In general case, first of all, the problem arises: how to obtain an explicit and compact analytical expression (exactly or approximately) for the conditional probability density $P\big(x(t_k)\big|x(t)|_0^{k-1}, y(t)|_0^k\big)$ in order to simulate samples of $x(t_k)$ when $y(t)|_0^k$ and $x(t)|_0^{k-1}$ are given.

Suppose that such sample sequences $x^i(t)|_0^k$ are being simulated, for $i = 1, \dots, N$. The joint probability density for the given observations $y(t)|_0^n$ and the sequence $x^i(t)|_0^n$ could be computed:

$$P_i(n) \overset{\text{def}}{=} P\big(x^i(t)|_0^n, y(t)|_0^n\big) = P_0(x^i(t_0), y(t_0)) \prod_{k=1}^n P(x^i(t_k), y(t_k)|x^i(t_{k-1}), y(t_{k-1})), \qquad (1)$$

where $P(x(t_k), y(t_k)|x(t_{k-1}), y(t_{k-1}))$ represents the transition probability density of the Markov process $(x(t), y(t))$, and $P_0(x_0, y_0)$ is the probability density for the initial joint distribution of the initial value $x(t_0)$ and the initial observation $y(t_0)$.

In a large number of works devoted to particle filters the various algorithms of resampling were introduced in order to obtain the most *a posteriori* probable samples. For the problem of searching for the maximum *a posteriori* probable sample the following algorithm of resampling can be suggested: Introduce the values $W_i(n) \overset{\text{def}}{=} P_i(n)/\sum_{j=1}^N P_j(n)$, for $i = 1, \dots, N$. Then $W_i(n)$ can be considered as a weight of the sample sequence $x^i(t)|_0^n$ (as well as of the sample point $x^i(t_n)$), which characterizes the *a posteriori* probability (or the importance) of the sample sequence $x^i(t)|_0^n$ in comparison with the other samples $x^j(t)|_0^n$ given $y(t)|_0^n$. Due to the Markov property of the process $z(t)$, the recursive formulae for $P_i(n)$ and $W_i(n)$ can be written in general form; we have

$$P_i(n + 1) = P_i(n)P\left(x^i(t_{n+1}), y(t_{n+1})\Big|x^i(t_n), y(t_n)\right). \qquad (2)$$

Then, it is possible to calculate recursively the new values $P_i(n + 1)$ and $W_i(n + 1)$, when the new measurement $y(t_{n+1})$ is obtained and the new sample point $x^i(t_{n+1})$ is augmented to the sequence $x^i(t)|_0^n$, so that $x^i(t)|_0^{n+1} = (x^i(t_0), \dots, x^i(t_n), x^i(t_{n+1}))$. It is easy to see, that it is not necessary to keep in memory all the sequence $x^i(t)|_0^n$, but only the last point $x^i(t_n)$. Then the samples $x^i(t)|_0^n$ with small weights $W_i(n)$ could be deleted, but the 'more important' samples $x^j(t)|_0^n$ should be continued with a few 'offsprings', so that the number of all sample sequences under consideration is equal to N (exactly or approximately). The

point $x^{j^*}(t_n)$ and the sequence $x^{j^*}(t)|_0^n$ that have the maximum weight $W_{j^*}(n)$ correspond to the point $x^j(t_n)$ and to the sample sequence that is maximally *a posteriori* probable given the observations $y(t)|_0^n$, among all the considered sample points $x^j(t_n)$ and sample sequences $x^j(t)|_0^n$. Then the value of $x^{j^*}(t_n)$ can be taken as the sought estimate of $x(t_n)$. On the base of the Laws of Large Numbers, it could be proved that this Monte Carlo estimate converges to the maximum *a posteriori* probability estimate for $x(t_n)$ if N increases.

In most of works the additional assumption is accepted that the process $x(t)$ is a Markov process in itself, and that the conditional probability density $P(y(t_k)|y(t_{k-1}), x(t_k), x(t_{k-1}))$ (for the observable process $y(t)$) can be presented in explicit and simple analytical form. Such cases are often referred to as hidden Markov models. Then, in many works, the sample sequences $x^j(t)|_0^n$ are being simulated as trajectories of the Markov process $x(t)$ in itself, since such a simulation can be easily done. The joint probability density, corresponding to the constructed sample $x^j(t)|_0^n$ and to the observations $y(t)|_0^n$, is equal to

$$P(x^j(t)|_0^n, y(t)|_0^n) =$$
$$P_0\left(x^j(t_0), y(t_0)\right) \prod_{k=1}^n P(x^j(t_k)\left|x^j(t_{k-1})\right) P\left(y(t_k)\left|y(t_{k-1}), x^j(t_k), x^j(t_{k-1})\right.\right) = P_j(n). \quad (3)$$

Then the weights $W_j(n)$ (or the other similar weights) could be easily calculated, and the above procedure of sampling and resampling could be realized, in order to obtain the estimate for $x(t_n)$. In some cases it would be better to simulate samples $x^j(t_k)$ with the use of the conditional probability density $P\left(x(t_k)\left|x(t_{k-1}), y(t_k), y(t_{k-1})\right.\right)$, which includes the observed value $y(t_k)$. But this could have required a large amount of computations if we do not have an explicit and compact analytical expression for that conditional probability density. Meantime, if the process $x(t)$ is being simulated just as a Markov process in itself, the sample sequences $x^j(t)|_0^n$ represent samples from the *a priori* distribution for $x(t_k)$ that can be far apart from the *a posteriori* distribution for $x(t_k)$ given $y(t)|_0^k$. Then resampling with the use of weights and significant increase in the number N are needed in order to find the most *a posteriori* probable values of $x(t_n)$ using the algorithm described above. Nevertheless, for hidden Markov models the particle filters with such a simple simulation of the sample sequences $x^j(t)|_0^n$, with resampling based on consideration of weights, but with large N, have been implemented and proved to be useful in some applications [5, 18].

In general case when $x(t)$ represents only some components of a multidimensional Markov process $(x(t), y(t))$ (and $x(t)$ is not a Markov process in itself), it was not shown in the literature how to simulate sample sequences $x^i(t)|_0^n$ when the values of the other components, $y(t)|_0^n$, are given (i.e. how to simulate sample sequences $x^i(t)|_0^n$ without a formidable amount of computations at each time t_k).

3

Note that the difficulties in obtaining the samples $x^i(t)|_0^n$ that correspond to the distributions $P(x(t_k)|x(t)|_0^{k-1}, y(t)|_0^k)$ have led to the introduction and the use of some 'proposal sampling distributions' (which can be simulated easier) in many works (see, for example, [2]). In the present work the explicit and compact analytical formulae and algorithms for simulation of the sample sequences $x^i(t)|_0^n$, when $y(t)|_0^n$ is given, and for recursive calculations of $P_i(n)$, weights, and estimates are obtained for the general case of a multidimensional continuous diffusion Markov process $(x(t), y(t))$.

For the problem of estimation of a function $f(x(t_n))$ or $\varphi(x(t)|_0^n)$ the new estimates in explicit and closed analytical form are obtained in the following Section 2.2.

Moreover, in the quotients $P_i(n)/\sum_{j=1}^N P_j(n)$ the 'scale' of all the $P_j(n)$ is cancelled, and the information that could characterize the smallness of the *a posteriori* probability of all the generated samples $x^j(t)|_0^n$ is lost. In the present work the new sequential Monte Carlo algorithms are derived that include some tests (developed in Section 2.4) in order to discard the samples of low *a posteriori* probability before the calculation of all the weights is done. The implementation of the suggested tests guarantees that the samples that remain under consideration belong to the domain where the *a posteriori* probability density is localized.

In some important cases of hidden Markov models, in the particle filters that were suggested and studied in the literature, where the samples $x^i(t)|_0^n$ (with $i = 1, \dots, N$) are being simulated as the trajectories of the Markov process $x(t)$ in itself, it is possible that the large part of those N generated samples $x^i(t)|_0^n$ would not be localized in the 'vicinity' of the 'true' realization of the trajectory, $x^{tr}(t)|_0^n$, because they are simulated with the use of the *a priori* transition probability density for $x(t)$, and they could be of low *a posteriori* probability given $y(t)|_0^n$. For example, if the 'level of intensity of noise' in observations $y(t)$ is small, the chance to generate randomly the sample $x(t)|_0^n$ that belongs to the domain where the *a posteriori* probability density (given $y(t)|_0^n$) is concentrated might be as small as the chance to catch randomly a needle in a haystack. Thus, although theoretically in those cases the Monte Carlo estimates converge as $N \to \infty$, the required number N increases if the 'level of intensity of noise' decreases. Meanwhile, a large amount of calculations would be wasted for processing of those samples $x^i(t)|_0^n$ of low *a posteriori* probability. In the present paper, in the following Section 2.6, we shall show (for the problem of nonlinear filtering of a signal) that with the use of the new algorithm (17) of simulation $x^i(t)|_0^n$ given $y(t)|_0^n$ and with some appropriate change of the mathematical model of description of the process of observations (which could be justified in applications), the new algorithm (17) will generate at once the samples $x^i(t)|_0^n$ that are localized in the domain that is *a posteriori* probable given $y(t)|_0^n$, for all $i = 1, \dots, N$.

The special case when the diffusion coefficients of the observable process $y(t)$ depend on unobservable components $x(t)$ is considered. In that case 'observed quadratic variations' of the process $y(t)$ can be introduced (when observations $y(t_k)$ are taken in discrete time), and they contain a lot of information about $x(t)$. In the present paper, the analytical formulae that characterize the observed quadratic variations are obtained in explicit and closed form. On the base of these results, the algorithms of filtering and estimation with the use of the observed quadratic variations are developed.

Theoretically, the derivative of the process of quadratic variation could be incorporated into the set of observed, known processes. If that derivative process were known, then with the use of the new algorithm of simulation (17), new estimates (26) and algorithms developed in the present paper, the filtering problem would be effectively solved, as shown in Section 2.5. But in practice we cannot assume that this derivative process is straightforwardly observed. However, if additional observations of $y(t)$ are available, namely, $y(t_0 + s\delta t)$, with $\delta t \ll \Delta t$, $s = 1, 2, ...,$ then the estimates $\hat{u}(t_k)$ for the derivative $u(t_k)$ can be obtained and used for the filtering, as shown in Section 2.5.

The similar systems arise if some additional precisely known observations are available, for example, measurements of some function $H(x(t), t)$. In that case the solution of the filtering problem, described in Section 2.5, could be implemented.

The obtained new algorithms improve performance of particle filters, or sequential Monte Carlo methods, and extend the range of their applications beyond the hidden Markov models.

The implementation of particle filters, as it is usual for Monte Carlo methods, requires a lot of repeating computations, but it became accessible and useful in some applications (for example, [5], [18]), since the speed of computations and the capacity of computers have increased dramatically in the last two decades.

2. Derivation of new recursive algorithms for Monte Carlo simulation and filtering of unobservable components of multidimensional diffusion Markov processes.

2.1. Simulation of trajectories of unobservable components. Analytical investigation of a multidimensional diffusion Markov process observed at discrete times.

Consider the multidimensional diffusion Markov process $z(t) \stackrel{\text{def}}{=} \big(x(t), y(t)\big) \stackrel{\text{def}}{=} (z_1(t), ..., z_p(t))$. The components $(z_1(t), ..., z_m(t)) \stackrel{\text{def}}{=} x(t)$ are unobservable, but the other components $(z_{m+1}(t), ..., z_p(t)) \stackrel{\text{def}}{=} y(t)$ are available for observations, which are being taken at discrete times t_k.

A diffusion Markov process with continuous trajectories $z(t)$ can be characterized by its drift and diffusion coefficients:

$$E\left\{\left(z_j(t + \Delta t) - z_j(t)\right) | z(t) = z\right\} = A_j(z, t)\Delta t + o(\Delta t), \quad i, j = 1, \ldots, p, \tag{4}$$

$$E\left\{\left((z_i(t + \Delta t) - z_i(t))\left(z_j(t + \Delta t) - z_j(t)\right) | z(t) = z\right\}\right. = b_{ij}(z, t)\Delta t + o(\Delta t), \tag{5}$$

where Δt is a small time step. Denote $\Delta z(t) \overset{\text{def}}{=} z(t + \Delta t) - z(t) \overset{\text{def}}{=} \left(\Delta z_1(t), \ldots, \Delta z_p(t)\right).$

From the assumption that the trajectories of the process $z(t)$ are continuous functions of t it follows [6, 7] that

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} E\left\{\Delta z_{i_1}(t) \ldots \Delta z_{i_r}(t) | z(t) = z\right\} = 0, \text{ with } r > 2. \tag{6}$$

The matrix $\|b_{ij}\|$ is symmetric and nonnegative definite. Therefore, in general case, that matrix can be represented with the use of its 'square root' in the form $\|b_{ij}\| = \|a_{ij}\|\|a_{ij}\|^T$, where $\|a_{ij}\|^T$ stands for transpose matrix $\|a_{ij}\|$.

Then the process $z(t)$ could be constructed as a solution of the system of stochastic differential equations:

$$dz_i(t) = A_i(z, t)dt + \sum_{j=1}^{p} a_{ij}(z, t)dw_j(t), \quad i = 1, \ldots, p, \quad t > t_0, \tag{7}$$

where $w_j(t)$ are independent Wiener processes. The initial condition at $t = t_0$ is given as $z(t_0) = z_0$, where z_0 is a random variable independent of all $w_j(t)$, with given probability density $P_0(z_0)$. The system (7) should be interpreted as the system of stochastic integral equations:

$$z_i(t) = \int_{t_0}^{t} A_i(z(s), s)ds + \sum_{j=1}^{p} \int_{t_0}^{t} a_{ij}(z(s), s)dw_j(s), \quad i = 1, \ldots, p, \tag{8}$$

with stochastic integrals in the sense of Ito [11]. It is assumed that the drift coefficients $A_i(z, t)$ satisfy the Lipschitz condition

$$\|A_i(z^{(1)}, t) - A_i(z^{(2)}, t)\| \leq K\|z^{(1)} - z^{(2)}\|, \text{ with } K = Const. \tag{9}$$

It is also assumed that the diffusion coefficients $b_{ij}(z, t)$ are continuous and differentiable functions of z and t. The Lipschitz conditions (9) guarantee that trajectories $z(t)$ do not have finite escapes, i.e. $|z(t)|$ does not tend to infinity when t tends to some finite moments of time.

6

The system of integral equations (8) with any given continuous trajectory $w(t) \stackrel{\text{def}}{=} (w_1(t), \dots, w_p(t))$ can be solved with the use of successive approximations, which converge and define the continuous trajectory $z(t)$. Thus, the trajectories of the process $z(t)$ are continuous with the probability 1.

The diffusion Markov process $z(t)$ can be also constructed as the limit solution of the system of finite difference equations:

$$z_i(t_k) - z_i(t_{k-1}) = A_i(z(t_{k-1}), t_{k-1})\Delta t_k + \sum_{j=1}^{p} a_{ij}(z(t_{k-1}), t_{k-1})\eta_j(t_k)\sqrt{\Delta t_k}, \qquad (10)$$

where the random impacts $\eta_j(t_k)$ are independent random variables (for all $j = 1, \dots, p$, $k = 1, 2, \dots$), with $E\eta_j(t_k) = 0$ and $E\big(\eta_j(t_k)\big)^2 = 1$; $\Delta t_k \stackrel{\text{def}}{=} t_k - t_{k-1}$; $i = 1, \dots, p$. In particular, the increments of Wiener processes, $\Delta w_j(t_k) \stackrel{\text{def}}{=} w_j(t_k) - w_j(t_{k-1})$, can be used in the finite difference equations (10) instead of $\eta_j(t_k)\sqrt{\Delta t_k}$:

$$z_i(t_k) - z_i(t_{k-1}) = A_i(z(t_{k-1}), t_{k-1})\Delta t_k + \sum_{j=1}^{p} a_{ij}(z(t_{k-1}), t_{k-1})\Delta w_j(t_k), \qquad (11)$$

Here again it is assumed that the Lipschitz conditions (9) are satisfied, and that the functions $A_i(z, t)$ and $b_{ij}(z, t)$ are differentiable with respect to z. The construction of diffusion Markov processes by the passage to the limit in the scheme of finite difference equations when time steps Δt_k tend to zero was first introduced and investigated by Academician S.N. Bernstein [9, 10] in 1934, 1938 years. The works [9, 10] were published at first in French. They are republished in *Collected works of S.N. Bernstein, vol.4* (in Russian), Publishing House Nauka (Academy of Science of USSR, Moscow, 1964). The results, obtained in these works [9, 10], define analytically all the joint probability distributions for the limit process $z(t)$, for all the random variables $z(t_{i_1})$, $z(t_{i_2}), \dots, z\big(t_{i_r}\big)$, for any r and t_{i_r}. By the Theorem of A.N. Kolmogorov on extension (or continuation) of measures in function space [8], those multidimensional distributions can be continued to the measure on the σ-algebra that corresponds to the process $z(t)$ with continuous time t. The convergence of the linear broken lines, which represent interpolation of the solutions of finite difference stochastic equations (10) or (11), to the continuous trajectories of diffusion Markov processes was also proved in [19]. Thus, the diffusion Markov process $z(t)$ is well defined by the system of finite difference equations (which could be more general, but similar to (10)) and by the passage to the limit from those Markov processes with discrete time [9, 10]. Analytical methods and solutions for some problems of filtering and estimation, based on recursive finite difference stochastic equations, with the passage to the limit (as the time steps tend to zero), similarly to the scheme of S.N. Bernstein [9, 10] were developed and investigated in [13 – 16].

In the following section 2.2, the goal is to obtain the estimates of $x(t_n)$ or $f(x(t_n))$ with the use of Monte Carlo calculation of some integrals that can be interpreted as mathematical expectations. We should notice the following properties that are important in order to achieve that goal. Denote $P_\Delta(z(t_{i_1}), \ldots, z(t_{i_r}))$ the joint probability density for the random process $z_\Delta(t)$ that represents the solution of the finite difference equations (10) or (11) with small time steps $\Delta t_k = \Delta$ (and with piecewise constant or piecewise linear interpolation on the small time intervals $(t_0 + k\Delta, t_0 + (k+1)\Delta)$, $k = 0, 1, 2, \ldots$). Denote $P(z(t_{i_1}), \ldots, z(t_{i_r}))$ the joint probability density for the limit diffusion Markov continuous process $z(t)$ obtained as $\Delta \to 0$. It was established in the works [9, 10] by S.N. Bernstein that $P_\Delta(z(t_{i_1}), \ldots, z(t_{i_r})) \to P(z(t_{i_1}), \ldots, z(t_{i_r}))$ as $\Delta \to 0$. Hence, the contribution to the error of the Monte Carlo estimates of the integrals $\int f(x(t_n))P(x(t)|_0^n, y(t)|_0^n)dx(t)|_0^n$ and $\int P(x(t)|_0^n, y(t)|_0^n \, dx(t)|_0^n$ (which are considered below in (23), (24)), caused by approximation of the limit diffusion process $z(t)$ by its pre-limit finite difference model (11), tends to zero as $\Delta \to 0$. For the scheme (11), it was also proved in [19] (Chapter 10) that trajectories $z_\Delta(t)$ converge in mean square to the trajectories of the limit diffusion process $z(t)$ as $\Delta \to 0$, so that $E\{[z(t) - z_\Delta(t)]^2\} \to 0$. Consequently, $|E\{f(z(t))\} - E\{f(z_\Delta(t))\}| \to 0$, if the function $f(z)$ satisfies Lipschitz condition. Meanwhile, in the problems of filtering the best estimate of $f(x(t_n))$ (by the criterion of minimum mean square error) is $\widehat{f(x(t_n))} = E\{f(x(t_n))|y(t)|_0^n\}$, and, in general case, the value of the mean square error of filtering, $E\left\{[f(x(t_n)) - \widehat{f(x(t_n))}]^2\right\}$, remains finite, positive, limited from below, even if the error of calculation of that conditional expectation tends to zero. Thus, the accuracy of the estimate of $x(t_n)$ or $f(x(t_n))$ could not be noticeably improved even if the accuracy of approximation of the limit diffusion process $z(t)$ by its finite difference model (11) increased as Δ decreased. Therefore, it is justified to use the finite difference model (11) for description and simulation of the considered process $z(t)$, in order to obtain the estimates of $x(t_n)$ or $f(x(t_n))$ that provide solutions to the considered filtering problems (with feasible precision) in case if the observations $y(t_k)$ are being taken in discrete consecutive times t_k with small time steps.

`   In the present work, we are interested to describe analytically the conditional probability density $P(x(t_k)|x(t)|_0^{k-1}, y(t)|_0^k)$. Consider the local increment Δz of the Markov process $z(t)$ when the value $z(t) = z$ is given, with small time step Δt; denote $\Delta z \overset{\text{def}}{=} z(t + \Delta t) - z(t) \overset{\text{def}}{=} (\Delta x, \Delta y)$. We have the relations (4) − (6) for the moments of Δz given $z(t) = z$. Then the characteristic function for the random value Δz can be presented in the form:

$$F(u|z, t) \overset{\text{def}}{=} E\left\{exp\left(i\sum_{k=1}^{p} u_k \, \Delta z_k\right)|z(t) = z\right\} \overset{\text{def}}{=} \int e\,xp\left(i\sum_{k=1}^{p} u_k \, \Delta z_k\right)P(\Delta z|z, t)d\Delta z =$$

$$exp\left\{\Delta t \sum_{k=1}^{p} A_k(z,t)(iu_k) + \frac{1}{2}\Delta t \sum_{k,l=1}^{p} b_{kl}(z,t)(iu_k)(iu_l) + o(\Delta t)\right\}, \tag{12}$$

where $u = (u_1, \ldots, u_p)$ is a real vector. The last equality follows from the known expression for the characteristic function of a random variable with the use of its moments (see, for example, [12]). The inverse transformation provides the representation

$$P(\Delta z|z,t) \stackrel{def}{=} (2\pi)^{-p} \int exp\left(-i\sum_{k=1}^{p} u_k \Delta z_k\right) F(u|z,t)du =$$
$$(2\pi\Delta t)^{-\frac{p}{2}}(Det\|b_{ij}\|)^{-1/2}exp\left\{-\frac{1}{2\Delta t}\sum_{i,j=1}^{p} r_{ij}(z,t)(\Delta z_i - A_i(z,t)\Delta t)(\Delta z_j - A_j(z,t)\Delta t)\right\}, \tag{13}$$

where $\|r_{ij}\| \stackrel{def}{=} \|b_{ij}\|^{-1}$ is the inverse matrix of $\|b_{ij}\|$ (with $1 \le i,j \le p$). The above expressions (12), (13) show that the local increment Δz (in the small, but finite time interval Δt) of the multidimensional diffusion Markov process $z(t)$ with continuous trajectories, when $z(t) = z$ is given, can be considered as a multidimensional Gaussian random variable.

Using the Theorem on Normal Correlation, we find the following expressions for the first and second moments of the conditional probability distribution of the increments Δx_α (with $(\alpha = 1, \ldots, m)$ provided that the increments Δy_ρ (with $\rho = (m+1), \ldots, p$) and the value $z(t) = z$ are given (see [13] Chapter 3, subsections 3.1.2 and 3.3.1, pages 79-81, 101-106):

$$E\{\Delta x_\alpha|\Delta y, z\} = A_\alpha(z,t)\Delta t + b_{\alpha\sigma}(z,t)c_{\sigma\rho}(z,t)(\Delta y_\rho - A_\rho(z,t)\Delta t) + o(\Delta t), \tag{14}$$

$$E\{\Delta x_\alpha \Delta x_\beta|\Delta y, z\} = (b_{\alpha\beta}(z,t) - b_{\alpha\sigma}(z,t)c_{\sigma\rho}(z,t)b_{\rho\beta}(z,t))\Delta t + o(\Delta t). \tag{15}$$

Hereinafter the summation is being made over repeated indices, with $1 \le \alpha, \beta \le m$, $(m+1) \le \sigma, \rho \le p$; the matrix $\|c_{\sigma\rho}\|$ represents the inverse or pseudo-inverse (Moore - Penrose) matrix of the matrix of the diffusion coefficients of the observable components $y_\rho(t)$, so that $\|c_{\sigma\rho}\| \stackrel{def}{=} \|b_{\sigma\rho}\|^{+}$.

For the probability density of the increments Δy we obtain the following expression:

$$P(\Delta y|z(t) = z) = C(z,t)\,exp\left\{-\frac{1}{2\Delta t}c_{\sigma\rho}(z,t)(\Delta y_\rho - A_\rho(z,t)\Delta t)(\Delta y_\sigma - A_\sigma(z,t)\Delta t)\right\}, \tag{16}$$

where $C(z,t)$ is the normalization factor.

Note, that in the theory of filtering of diffusion Markov processes with observations made in continuous time t, i.e. if $y(s)$ is supposed to be known exactly on the time interval $t_0 \le s \le t$, in such case it is assumed that the diffusion coefficients $b_{\sigma\rho}$, with $(m+1) \le \sigma, \rho \le p$, do not depend on unobservable components $x(t)$. In the contrary case (as it was pointed out by this author in [13], Chapter 3, and [16], Chapter 3) since the diffusion coefficients $b_{\sigma\rho}(x(t), y(t), t)$ could be (at least, theoretically) precisely restored on the base of a single

realization of the observed trajectory $y(s)$ on the small time interval $t - \delta \leq s \leq t + \delta$ (no matter how small the taken value of δ is), the functions $b_{\sigma\rho}(x(t), y(t), t)$ could have been incorporated into the set of observable functions. In the problem at hand, when filtering of the process $x(t)$ should be obtained on the base of observations $y(t_k)$, taken at discrete times t_k with small, but finite time steps Δt_k, that restriction may be cancelled. We shall consider further how to incorporate some estimates of the values of $b_{\sigma\rho}(x(t_k), y(t_k), t_k)$ given $y(t)|_0^k$ in order to improve filtering of $x(t_n)$.

The matrix $\|b_{\alpha\beta} - b_{\alpha\sigma}c_{\sigma\rho}b_{\rho\beta}\|$ is symmetric and nonnegative definite, and it can be presented in the following form:

$$\|b_{\alpha\beta} - b_{\alpha\sigma}c_{\sigma\rho}b_{\rho\beta}\| \stackrel{\text{def}}{=} \|g_{\alpha\beta}\|\|g_{\gamma\delta}\|^T, \text{ with } 1 \leq \alpha, \beta, \gamma, \delta \leq m,$$

Then the samples $x^j(t_k)$, when $x^j(t)|_0^{k-1}$ and $y(t)|_0^k$ are given, can be simulated as

$$x_\alpha^j(t_k) = x_\alpha^j(t_{k-1}) + A_\alpha\left(x^j(t_{k-1}), y(t_{k-1})\right)\Delta t_k + g_{\alpha\beta}(x^j(t_{k-1}), y(t_{k-1}))\xi_\beta^j(k)\sqrt{\Delta t_k} +$$

$$b_{\alpha\sigma}\left(x^j(t_{k-1}), y(t_{k-1})\right)c_{\sigma\rho}(x^j(t_{k-1}), y(t_{k-1}))\left(y_\rho(t_k) - y_\rho(t_{k-1}) - \right.$$

$$A_\rho\left(x^j(t_{k-1}), y(t_{k-1})\right)\Delta t_k\right), \tag{17}$$

where $\xi_\beta^j(k)$ are independent samples from Gaussian standard distribution, with $E\xi_\beta^j(k) = 0$, $E(\xi_\beta^j(k))^2 = 1$ (for all $\beta = 1, \dots, m$, $j = 1, \dots, N$, $k = 1, 2, \dots$); $\Delta t_k = t_k - t_{k-1}$.

We shall denote shortly $b_{\alpha\sigma}\left(x^j(t_{k-1}), y(t_{k-1})\right) \stackrel{\text{def}}{=} b_{\alpha\sigma}^j(t_{k-1})$, $c_{\sigma\rho}\left(x^j(t_{k-1}), y(t_{k-1})\right) \stackrel{\text{def}}{=} c_{\sigma\rho}^j(t_{k-1})$, $A_\sigma\left(x^j(t_{k-1}), y(t_{k-1})\right) \stackrel{\text{def}}{=} A_\sigma^j(t_{k-1})$, $c_{\sigma\rho}(x(t_{k-1}), y(t_{k-1})) \stackrel{\text{def}}{=} c_{\sigma\rho}(t_{k-1})$.

Note, that the above expressions $(14) - (17)$ show that if all the diffusion coefficients $b_{\alpha\rho}$ vanish (with $1 \leq \alpha \leq m$, $(m + 1) \leq \rho \leq p$) and if $A_\alpha(z, t)$, $b_{\alpha\beta}(z, t)$ do not depend on $y(t)$, then $x(t)|_1^n$ will be simulated without the use of $y(t)|_1^n$. Only the initial measurement $y(t_0)$ will be taken into account: the initial sample point $x^j(t_0)$ should be simulated as a random variable with the probability density

$$P_0(x(t_0)|y(t_0) = y_0) = P_0(x(t_0), y_0)/\int P_0(x(t_0), y_0)dx(t_0). \tag{18}$$

This particular case do not correspond to all the hidden Markov models: The process $x(t)$ could be a Markov process in itself, but it is still possible, that some coefficients $b_{\alpha\rho}$ are not equal to zero, for example, in case, if there are some 'white noises', which enter simultaneously into equations for $x(t)$ and into 'random disturbances or random errors' of the observable process $y(t)$.

In general case of a multidimensional continuous Markov diffusion process $(x(t), y(t))$, the influence of the observed data $y(t)|_1^n$ is manifested in the expressions (17), which describe simulation of the random samples $x^j(t_k)$.

The values $P_i(n)$ and $W_i(n)$ can be easily calculated:

$$P_i(n) = \qquad (19)$$

$$P_0\left(x^i(t_0), y(t_0)\right) \prod_{k=1}^{n} P(\Delta x^i(t_k) \mid \Delta y(t_k), x^i(t_{k-1}), y(t_{k-1})) P\left(\Delta y(t_k) \mid x^i(t_{k-1}), y(t_{k-1})\right),$$

where the conditional probability densities for the increments, $\Delta x^i(t_k) \stackrel{\text{def}}{=} x^i(t_k) - x^i(t_{k-1})$, $\Delta y(t_k) \stackrel{\text{def}}{=} y(t_k) - y(t_{k-1})$, are Gaussian densities with moments determined by (14), (15), and (16). And the above procedure of sampling and resampling, which corresponds to a search for the point $x(t_n)$ that provides maximum to the *a posteriori* probability density, could be implemented. In the next sections 2.2 - 2.4 the results of the further analytical investigation and development of the algorithms are presented.

2.2. Estimating of a function $f\left(x(t_n)\right)$.

Consider the problem of estimation of some functions $f\left(x(t_n)\right)$ or $\varphi\left(x((t)|_0^n)\right)$ given $y(t)|_0^n$. The general expression for the estimate can be written in the following form:

$$f(\widehat{x(t_n)}) = E\{f(x(t_n))|y(t)|_0^n\} = \frac{\int f(x(t_n)) P(x(t)|_0^n, y(t)|_0^n) dx(t)|_0^n}{\int P(x(t)|_0^n, y(t)|_0^n) dx(t)|_0^n} \quad . \qquad (20)$$

It is assumed that for the considered functions $f(x)$ or $\varphi(x(t)|_0^n)$ this conditional expectation exists. In case of a Markov process $(x(t_n), y(t_n))$ we can write:

$$P(x(t)|_0^n, y(t)|_0^n) = P_0\left(x(t_0)|y(t_0)\right) P_0\left(y(t_0)\right) \prod_{k=1}^{n} P(x(t_k) \mid y(t_k), y(t_{k-1}), x(t_{k-1})) \times$$

$$\prod_{s=1}^{n} P(y(t_s)|y(t_{s-1}), x(t_{s-1})). \qquad (21)$$

In our case of a multidimensional continuous diffusion Markov process we have derived the explicit analytical expressions (19), (16). Hence, in this case we obtain

$$\prod_{s=1}^{n} P(y(t_s)|y(t_{s-1}), x(t_{s-1})) =$$

11

$$C(x(t)|_0^{n-1}, y(t)|_0^{n-1}) \, exp\left\{-\frac{1}{2}\sum_{s=1}^{n}\frac{1}{\Delta t_s}c_{o\rho}(t_{s-1})[\Delta y_\sigma(t_s)\Delta y_\rho(t_s) - 2\Delta y_\sigma(t_s)A_\rho(t_{s-1})\Delta t_s + \right.$$
$$\left. A_\sigma(t_{s-1})A_\rho(t_{s-1})(\Delta t_s)^2]\right\}, \tag{22}$$

where $C(x(t)|_0^{n-1}, y(t)|_0^{n-1})$ represents the known normalization factor. Denote shortly $C_i(n-1) \overset{\text{def}}{=} C(x^i(t)|_0^{n-1}, y(t)|_0^{n-1})$.

If N samples $x^i(t)|_0^n$, with $i = 1, \dots, N$ and $N \gg 1$, have been independently taken from the distribution

$$P_0\big(x(t_0)\big|y(t_0)\big) \prod_{k=1}^{n} P(x(t_k)\big|y(t_k), y(t_{k-1}), x(t_{k-1})),$$

then, in accordance with basics of Monte Carlo methods, the following integrals can be interpreted as mathematical expectations. Then we obtain

$$\int f(x(t_n))P(x(t)|_0^n, y(t)|_0^n)dx(t)|_0^n \approx P_0(y(t_0)\frac{1}{N}\sum_{i=1}^{N}f(x^i(t_n))\,C_i(n-1)\times$$
$$exp\left\{-\frac{1}{2}\sum_{s=1}^{n}\frac{1}{\Delta t_s}c_{\sigma\rho}^i(t_{s-1})[\Delta y_\sigma(t_s)\Delta y_\rho(t_s) - 2\Delta y_\sigma(t_s)A_\rho^i(t_{s-1})\Delta t_s + \right.$$
$$\left. A_\sigma^i(t_{s-1})A_\rho^i(t_{s-1})(\Delta t_s)^2]\right\}, \tag{23}$$

and similarly

$$\int P(x(t)|_0^n, y(t)|_0^n)dx(t)|_0^n \approx P_0(y(t_0)\frac{1}{N}\sum_{i=1}^{N}C_i(n-1)\times$$
$$exp\left\{-\frac{1}{2}\sum_{s=1}^{n}\frac{1}{\Delta t_s}c_{\sigma\rho}^i(t_{s-1})[\Delta y_\sigma(t_s)\Delta y_\rho(t_s) - 2\Delta y_\sigma(t_s)A_\rho^i(t_{s-1})\Delta t_s + \right.$$
$$\left. A_\sigma^i(t_{s-1})A_\rho^i(t_{s-1})(\Delta t_s)^2]\right\}. \tag{24}$$

Due to the Laws of Large Numbers, the accuracy of those approximate estimates increases as N increases.

Denote shortly

$$\Upsilon_i(n) = C_i(n-1)\, exp\left\{-\frac{1}{2}\sum_{s=1}^{n}\frac{1}{\Delta t_s}c_{\sigma\rho}^i(t_{s-1})[\Delta y_\sigma(t_s)\Delta y_\rho(t_s) - 2\Delta y_\sigma(t_s)A_\rho^i(t_{s-1})\Delta t_s + \right.$$
$$\left. A_\sigma^i(t_{s-1})A_\rho^i(t_{s-1})(\Delta t_s)^2]\right\}. \tag{25}$$

Then the sought estimate $\widehat{f(x(t_n))}$ can be written in the following form:

$$\widehat{f(x(t_n))} = \sum_{i=1}^{N}f(x^i(t_n))\widetilde{W}_i(n), \text{ where } \widetilde{W}_i(n) = \frac{\Upsilon_i(n)}{\sum_{j=1}^{N}\Upsilon_j(n)}. \tag{26}$$

In most of applications, we can assume that there exists the mathematical expectation of the random value $\big|f(x^i(t_n))\big|\Upsilon_i(n)$, where $y(t)|_0^n$ is fixed and $x^i(t)|_0^n$ represent the results of

independent random simulations with the use of (17); i.e. $E\{|f\left(x^i(t_n)\right)|\Upsilon_i(n)\} < \infty$. This assumption is equivalent to the following: $E\{|f(x(t_n))||y(t)|_0^n\} < \infty$. Denote $\zeta_i = f(x^i(t_n))\Upsilon_i(n)$. The random values ζ_i are independent, and they have one and the same probability distribution, and $E\{|\zeta_i|\} < \infty$. Then the Strong Law of Large Numbers (the Theorem of Kolmogorov) guarantees that the sums (i.e. the arithmetic means $\frac{1}{N}\sum_{i=1}^N \zeta_i$) in right-hand-side of the formulae (23), (24) converge with probability one (i.e. almost surely) to the integrals in left-hand-side if $N \to \infty$.

Note that in case if the diffusion coefficients $b_{\sigma\rho}$ (for $m < \sigma, \rho \le p$) of the observable process $y(t)$ do not depend on unobservable process $x(t)$, then $C_i(n-1)$ and $c_{\sigma\rho}^i(t_{s-1})$ do not depend on i, and cofactors $C_i(n-1)\, exp\left\{-\frac{1}{2}\sum_{s=1}^n \frac{1}{\Delta t_s} c_{\sigma\rho}^i(t_{s-1})\Delta y_\sigma(t_s)\Delta y_\rho(t_s)\right\}$ should be cancelled in the numerator and denominator of the formula (26). In that case we can put:

$$\Upsilon_i(n) = \ exp\left\{-\frac{1}{2}\sum_{s=1}^n c_{\sigma\rho}\,(t_{s-1})[-2A_\rho^i(t_{s-1})\Delta y_\sigma(t_s) + A_\sigma^i(t_{s-1})A_\rho^i(t_{s-1})\Delta t_s]\right\}. \qquad (27)$$

Note that the calculations of the values $Y_i(n)$ (25) or (27) can be implemented recurrently since they are determined by the values of the recursively accumulated sums.

In case of hidden Markov models where the unobservable process $x(t)$ is a Markov process in itself, and the observable process $y(t)$ is described by a stochastic differential equation (SDE):

$$dy(t) = h(x(t),t)dt + \sigma dw(t), \qquad (28)$$

where $h(x,t)$ is a given nonlinear function with respect to x, $\sigma = Const$, and $w(t)$ is a standard Wiener process, the above estimate $\widehat{f(x(t_n))}$ takes the form, which is similar to the estimates for $f\left(x(t_n)\right)$ that were earlier constructed and studied for this particular case in the literature. The novelty of the estimates (23) – (27) and the simulation (17), obtained in the present paper, is that the simulation of unobservable components and estimates $\widehat{f(x(t_n))}$ are obtained in explicit and closed analytical form for the general case of partially observable multidimensional continuous diffusion Markov processes, and the obtained Monte Carlo estimate (26) converges with probability one to the sought posterior expectation of $f(x(t_n))$ given $y(t)|_0^n$, if $N \to \infty$.

Now we shall demonstrate that the Monte Carlo estimates (23) - (26), (27) for $f\left(x(t_n)\right)$ (obtained in the present paper) hold true also in the case if the sample sequences $x^i(t)|_0^n$ are being generated with the use of some branching sampling procedures.

Consider the following branching resampling. As it was already pointed out in the Introduction, it is purposeful to discard the samples $x^i(t)|_0^n$ that have negligibly small weights $W_i(n)$ or $\widetilde{W}_i(n)$, in order to decrease the amount of all the computations.

Suppose that in the end of each time interval $(T_k, T_{k+1}]$ (with $k = 0, 1, 2, \dots$; $T_0 = t_0$; $T_{k+1} = T_k + m_k \Delta t$, and m_k and N_k are given numbers) each sample sequence $x^i(t)|_0^n$ that still existed at the time T_{k+1} is being continued with N_{k+1} 'offsprings'. At the initial moment of time, T_0, there are N_0 independently taken initial sample points $x^i(t_0)$, $i = 1, \dots, N_0$. When $t_n = T_{k+1}$ and $t_{n+1} = T_{k+1} + \Delta t$, sample points $x^{(i,j)}(t_{n+1})$ are being augmented to the sequence $x^i(t)|_0^n$, with $j = 1, \dots, N_{k+1}$, in order to construct the 'offsprings'. The sample points $x^{(i,j)}(t_{n+1})$ are being independently taken from the distribution $P(x(t_{n+1})|x^i(t)|_0^n, y(t)|_0^{n+1})$. Then those sample sequences (with 'offsprings') will be continued until the next time of branching, T_{k+2}, except for some sample sequences that are discarded before T_{k+2} since their weights become 'too small' (for example, smaller than some chosen threshold W_{cr}); and so on. For the simplicity of discussion, we can renumber again all the current sample sequences (that still exist at the time t_n) as $x^i(t)|_0^n$. Consider at first the case when all the sample sequences are being continued without discarding. Then their number is growing, and at the time T_{k+1} it is equal to $N_0 \times N_1 \times \dots \times N_k$ (with $k > 0$).

Consider the estimate

$$\overline{f(x(t_n))} = E\{f(x(t_n))|y(t)|_0^n\},$$

under condition that the above branching sampling procedure is being implemented. Denote $(t_{r(n)}, t_{(r+1)(n)}]$ the time interval (between the moments of branching) that contains t_n, so that $t_{r(n)} = T_{k(n)} < t_n \leq T_{k(n)+1} = t_{(r+1)(n)}$.

From the Markov property of the random process $(x(t), y(t))$ and the factorization (21), (22), and from the 'tower property' of conditional expectations:

$$E\{f(x(t_n))|y(t)|_0^n\} = E\left\{E\{f(x(t_n)|y(t)|_0^n, x(t)|_0^{r(n)}\}\right\} =$$

$$E\left\{E\left\{E\{f(x(t_n)|y(t)|_0^n, x(t)|_0^{r(n)}\}|x(t)|_0^{(r-1)(n)}\right\}\right\} = \cdots,$$

it follows that the formulae (23) -(26), (27) hold true for the Monte Carlo estimate of integral $\overline{f(x(t_n))}$ in the case if the sample sequences $x^i(t)|_0^n$ are generated with the use of the above branching procedure.

We can consider the Monte Carlo estimates for the integrals (23), (24) for each i-th 'tree' of the branching sample sequences, which begins at the point $x^i(t_0)$, with $i = 1, \dots, N_0$. Denote

these random values as $\xi_i(n)$, with $i = 1, \ldots, N_0$. Then the random values $\xi_i(n)$ are independent of each other, and they have one and the same probability distribution, and $E\{|\xi_i|\} < \infty$. Then they obey the Strong Law of Large Numbers, so that $\frac{1}{N_0}\sum_{i=1}^{N_0}\xi_i$ converges with probability one to the sought integral as $N_0 \to \infty$.

But many of the exponential weights $\widetilde{W}_i(n)$ are rapidly decreasing with increase in the number n of time steps. Therefore, it is possible to discard some highly *a posteriori* improbable sample sequences $x^i(t)|_0^n$, which do not provide noticeable contribution into the estimates $\overline{f(x(t_n))}$. The numbers m_k, N_k, and the threshold W_{cr} can be adjusted (in practical implementation for a particular class of applications) in order to decrease the amount of all the calculations and make them feasible, and at the same time to keep the current sample points $x^i(t_n)$ in the domain where the *a posteriori* probability density is localized. In practical implementation, it can be also purposeful to begin branching not at the fixed moment of time T_k but at some current moment of time t_n when the number of all existed sample sequences (or sample points) becomes less than some given fraction of N.

It is worth noting that in the following Section 2.4 the tests (47), (48) are obtained that allow detection and rejection of samples $x^i(t_n)$ of low *a posteriori* probability to be done without the use of the weights $W_i(n)$ or $\widetilde{W}_i(n)$ (for $i = 1, \ldots, N$), i.e. independently for each sample sequence $x^i(t)|_0^n$. Thus, in the above algorithm we can use such tests instead of the comparison of the weight $\widetilde{W}_i(n)$ with the threshold W_{cr}. Then the algorithm described above can be effectively implemented with the use of parallel computing, since the simulation and continuation (the branching resampling) of each sample sequence $x^i(t)|_0^m$ can be done independently of the other samples $x^k(t)|_0^m$ with $k \neq i$. The values $Y_i(n)$ determined by (25) or (27) will be also calculated recurrently and independently for each sample sequence $x^i(t)|_0^n$. Thus, that large amount of calculations could be effectively performed with the use of parallel computing. And only on the other stage all the obtained accumulated values $Y_i(n)$ (with $i = 1, \ldots, N$) will be used for the calculation of the weights $\widetilde{W}_i(n)$, which are needed for the calculation of the sought estimate $\overline{f(x(t_n))}$.

The new Monte Carlo algorithm of estimation of $f\big(x(t_n)\big)$ or $\varphi(x(t)|_0^n)$ (presented above) is derived straightforwardly from Bayes formula (20). The estimates are constructed with the use of the samples $x^i(t_n)$ or $x^i(t)|_0^n$ that are simulated by (17), and their weights $\widetilde{W}_i(n)$ are defined by (23) − (26), (27).

For the particular case of hidden Markov models, another algorithm with random branching resampling for 'particle filtering' of unobservable Markov 'signal' $x(t_n)$ was developed, that is the Sequential Importance Resampling (SIR) algorithm; it is studied in the works [4], [1] (for the processes with discrete time). In that SIR algorithm with random branching resampling, the *a*

15

posteriori probability distribution for $x(t_n)$ is approximated by some 'cloud of random particles' $x^i(t_n)$ with weights $w_i(n) = 1/N$, $i = 1, \dots, N$. With the use of our new algorithm of simulation of unobservable components $x(t_n)$ given $y(t)|_0^n$ (see Section 2.1, (17)) and the new closed form analytical expressions (14) − (17), (23) − (27), that SIR algorithm with random branching can be generalized for the general case of filtering of unobservable components $x(t)$ of a multidimensional diffusion Markov process $(x(t), y(t))$, in the following way.

In the generalized algorithm that we are proposing the sample sequence $x^i(t)|_0^m$ should be chosen for continuation at the time of branching $T_m = t_{r(m)}$ (and kept up to the next time of branching, T_{m+1}) with probability $p_i(m)$ (which will be determined below) in each of N independent attempts to continue their ensemble, so that the expectation of the random number $K_i(m)$ of its 'offsprings' is equal to $E\left\{K_i(m)|x^1(t_{r(m)}), \dots, x^{K(m)}(t_{r(m)}), y(t)|_0^{r(m)}\right\} = p_i(m)N$, with $i = 1, \dots, K(m)$, where $K(m)$ denotes the number of all the sample points $x^i(t_{r(m)})$ that still existed at the time $t_{r(m)}$. We can suggest that the times of branching $T_{m+1} = T_m + M\Delta t$, with $M \geq 1$, $m = 0, 1, 2, \dots$. (Note, that $M = 1$ in the standard SIR algorithms). The probability $p_i(m)$ depends on all the sample points $x^1(t_m), \dots, x^{K(m)}(t_m)$ or sample sequences $x^i(t)|_0^m$. Such a procedure of random resampling is similar to the standard SIR algorithm, but we have to determine the probability $p_i(m)$ for the general case of multidimensional diffusion Markov processes $(x(t), y(t))$.

For the general case, with the use of analytical expressions (16) − (27), we derive the following expression:

$$p_i(m) = \frac{Q_i(m) \, exp\{-R_i(m)\}}{\sum_{j=1}^{K(m)} Q_j(m) exp\{-R_j(m)\}} \ ,$$

where

$$R_i(m) = \frac{1}{2} \sum_{s=0}^{M-1} \frac{1}{\Delta t} c_{\sigma\rho}^i(t_{r(m-1)+s})[\Delta y_\sigma(t_{r(m-1)+s+1})\Delta y_\rho(t_{r(m-1)+s+1}) -$$

$$2\Delta y_\sigma(t_{r(m-1)+s+1})A_\rho^i(t_{r(m-1)+s})\Delta t + A_\sigma^i(t_{r(m-1)+s})A_\rho^i(t_{r(m-1)+s})(\Delta t)^2] \ ,$$

and $Q_i(m)$ denotes the normalization factor of the probability density $Q_i(m)exp\{-R_i(m)\}$. Here the notations are the same as earlier in Sections 2.1, 2.2.

In the case if the diffusion coefficients $b_{\sigma\rho}$ of the observed process $y(t)$ do not depend on the unobservable process $x(t)$, the above expression can be written in the following more concise form:

$$p_i(m) = \frac{exp\{-V_i(m)\}}{\sum_{j=1}^{K(m)} exp\{-V_j(m)\}},$$

where

$$V_i(m) = \frac{1}{2}\sum_{s=0}^{M-1} c_{\sigma\rho}\left(t_{r(m-1)+s}\right)\left[-2\Delta y_\sigma\left(t_{r(m-1)+s+1}\right)A_\rho^i\left(t_{r(m-1)+s}\right) + A_\sigma^i\left(t_{r(m-1)+s}\right)A_\rho^i\left(t_{r(m-1)+s}\right)\Delta t\right].$$

In case of hidden Markov models (28), which can be considered as a particular case of a multidimensional diffusion Markov process $(x(t), y(t))$, from the above expression follows:

$$p_i(m) = \frac{exp\{-\Gamma_i(m)\}}{\sum_{j=1}^{K(m)} exp\{-\Gamma_j(m)\}},$$

where

$$\Gamma_i(m) = \frac{1}{2\sigma^2}\sum_{s=0}^{M-1}[-2h(x^i\left(t_{r(m-1)+s}\right))\Delta y\left(t_{r(m-1)+s+1}\right) + h^2(x^i\left(t_{r(m-1)+s}\right))\Delta t].$$

For the simplicity of notation, the above formula is written for the case of one-dimensional processes $x(t)$ and $y(t)$, which satisfy (28). The last expression for $p_i(m)$ (for the case $M = 1$) is in agreement with the determination of the probabilities of continuation of the samples $x^i(t_n)$ presented in [4], [1].

With the use of the above procedure of random branching and with our algorithm (17) of simulation of $x(t_n)$ given $y(t)|_0^n$, we can generalize the SIR algorithm of particle filtering with random branching, which was developed for hidden Markov models, to the general case of multidimensional diffusion Markov processes $(x(t), y(t))$.

Thus, in the present paper a few different possible versions of algorithms of particle filters are provided for filtering of unobservable components $x(t_n)$ given $y(t)|_0^n$, which are justified theoretically. The new algorithm developed above, with branching sampling and recurrent calculation of the values (25) or (27), appears to be preferable for implementation with the use of parallel computing. The further practical implementation and comparison for various applications could be achieved in the future works.

2.3. Analytical investigation of observed quadratic variations.

Consider the case when some diffusion coefficients $b_{\sigma\rho}$ of the observable process $y(t)$ depend on the unobservable process $x(t)$. We assume throughout this paper that the diffusion coefficients $b_{ij}(z, t)$ are continuous and differentiable functions with respect to z. For simplicity of notation, consider at first the case when $y(t)$ is one-dimensional. For example, consider the following model that plays an important role in financial mathematics [17]:

$$dx = A(x,t) + \sigma_1(x,t)dw_1(t), \; dy = \left(\mu - \frac{x^2}{2}\right)dt + \sigma_2 x dw_2(t). \tag{29}$$

Here $w_1(t), w_2(t)$ are independent Wiener processes, μ and σ_2 are known constants, and the functions $A(x,t)$ and $\sigma_1(x,t)$ are continuous and differentiable with respect to x. In the famous Black – Scholes models [17], the observable process $y(t)$ in (29) represents the natural logarithm of a stock price, $S(t)$, so that $y(t) \overset{\text{def}}{=} \log S(t)$, while the process $|x(t)|$ corresponds to volatility, which is to be estimated given the path $y(s)$, $t_0 \leq s \leq t$. If the process $y(s)$ were available for observations continuously over the time interval $t_0 \leq s \leq t$, the value $|x(t)|$ could have been restored precisely, at least theoretically.

It is well known that a diffusion Markov process $z(t)$ is not differentiable at any time t. It can be proved that its quadratic variations

$$\sum_{k=1}^{M}[z_i(t_k) - z_i(t_{k-1})][z_j(t_k) - z_j(t_{k-1})] \xrightarrow[max\{\Delta t_k\} \to 0]{} \int_{t-\delta}^{t+\delta} b_{ij}(z(s),s)ds, \tag{30}$$

if time steps Δt_k tend to zero, while $M \to \infty$; here $(t - \delta) = t_0 < t_1 < \cdots < t_k < \cdots < t_M = (t + \delta)$, with given $\delta > 0$, and the value of δ can be arbitrarily small. Note that the relation (30) will be also obtained as a consequence of the new analytical estimates that are developed below.

For the system (29), $b_{22}(x(t), y(t), t) = (\sigma_2)^2 x^2(t)$, and from (30) it follows that the value of $x^2(t)$ could be restored with any desirable accuracy. Then $|x(t)| = +\sqrt{x^2(t)}$, and the filtering problem would be solved exactly.

But the observations $y(t_k)$ are being taken at discrete times t_k with small, but finite time steps Δt_k, so that it is impossible to observe precisely the values of $b_{\sigma\rho}(x(t_k), y(t_k), t_k)$. In general case, we can calculate the 'observed quadratic variations',

$$V_{\sigma\rho}^{(obs)}(t_k) \overset{\text{def}}{=} \sum_{s=0}^{M}[y_\sigma(t_{k-s}) - y_\sigma(t_{k-s-1})][y_\rho(t_{k-s}) - y_\rho(t_{k-s-1})], \tag{31}$$

and consider it as an estimate for $b_{\sigma\rho}(x(t_k), y(t_k), t_k)$. The similar estimates for volatilities were introduced and considered in [17, chapter 15]. It is more convenient to use the recursive averaging instead of the moving averaging (31). Then

$$V_{\sigma\rho}^{(obs)}(t_k) \overset{\text{def}}{=} e^{-\lambda \Delta t_k} V_{\sigma\rho}^{(obs)}(t_{k-1}) + [y_\sigma(t_k) - y_\sigma(t_{k-1})][y_\rho(t_k) - y_\rho(t_{k-1})], \tag{32}$$

with $\lambda > 0$. In the limit, if all Δt_k decreased and tended to zero, we would have obtained

$$V_{\sigma\rho}^{(obs)}(t) \xrightarrow[max\{\Delta t_k\} \to 0]{} \int_{t_0}^{t} e^{-\lambda(t-s)} b_{\sigma\rho}(x(s), y(s), s)ds. \tag{33}$$

18

We can incorporate the observed quadratic variations $V_{\sigma\rho}^{(obs)}(t_k)$ into the set of observed data. We shall show further, in section 2.4, how to use that "observed quadratic variations" in order to reject at once the samples $x^i(t)|_0^k$ that are highly improbable given $y(t)|_0^k$.

The recursive formula (32) implies that

$$V_{\sigma\rho}^{(obs)}(t_k) = \sum_{s=1}^{k} e^{-\lambda(t_k - t_s)} [y_\sigma(t_s) - y_\sigma(t_{s-1})][y_\rho(t_s) - y_\rho(t_{s-1})]. \tag{34}$$

We may assume that the considered realization of the process $z(t)$, $(x(t_0), y(t_0)), (x(t_1), y(t_1)), \dots, (x(t_s), y(t_s)), \dots$, is being obtained consecutively in accordance with finite difference stochastic equations (11). The value of observed quadratic variation $V_{\sigma\rho}^{(obs)}(t_k)$ is being accumulated in parallel, along with that realization of $z(t)$, so that the random increment $\Delta y_\rho(t_s) \stackrel{\text{def}}{=} y_\rho(t_s) - y_\rho(t_{s-1})$ will be produced after $(x(t_{s-1}), y(t_{s-1}))$ is realized. We are interested to describe the properties of the observed quadratic variations $V_{\sigma\rho}^{(obs)}(t_k)$ under condition that the realization of unobservable components, $x(t)|_0^k$, is fixed, although it is unknown, and that the next measurement $y(t_s)$ will arrive after the previous measurement $y(t_{s-1})$ is already given.

The conditional expectation

$$E\{\Delta y_\sigma(t_s)\Delta y_\rho(t_s)|x(t_{s-1}), y(t_{s-1})\} = b_{\sigma\rho}(x(t_{s-1}), y(t_{s-1}), t_{s-1})\Delta t_s .$$

In general case, as it was demonstrated above, the increment $\Delta y = (\Delta y_{m+1}, \dots, \Delta y_p)$ in small time step Δt, given $z(t) = z$, can be described as the Gaussian multidimensional random variable with the probability density (16). The increments $\Delta y(t_k)$ are independent from the past history before t_{k-1} if $z(t_{k-1})$ is given. The following properties of Gaussian distributions will be useful in the sequel: If ξ is a Gaussian random variable with $E\xi = 0$, then $E\xi^4 = 3(E\xi^2)^2$. If $(\xi_1, \xi_2, \xi_3, \xi_4)$ is a 4-dimensional Gaussian random variable with $E\{\xi_i\} = 0$ (for $i = 1, \dots, 4$), then

$$E\{\xi_1\xi_2\xi_3\xi_4\} = E\{\xi_1\xi_2\}E\{\xi_3\xi_4\} + E\{\xi_1\xi_3\}E\{\xi_2\xi_4\} + E\{\xi_1\xi_4\}E\{\xi_2\xi_3\}.$$

In general case of multidimensional process $y(t)$ we obtain

$$E\{(\Delta y_\sigma(t_k)\Delta y_\rho(t_k) - b_{\sigma\rho}(t_k)\Delta t_k)^2|x(t_{k-1}), y(t_{k-1})\} =$$

$$[b_{\sigma\sigma}(x(t_{k-1}), y(t_{k-1}), t_{k-1})\Delta t_k] [b_{\rho\rho}(x(t_{k-1}), y(t_{k-1}), t_{k-1})\Delta t_k] + [b_{\sigma\rho}(x(t_{k-1}), y(t_{k-1}), t_{k-1})\Delta t_k]^2 .$$

Denote shortly $b_{\sigma\rho}(x(t_k), y(t_k), t_k) \stackrel{\text{def}}{=} b_{\sigma\rho}(t_k)$. Introduce the following value:

$$\sum_{s=1}^{k} e^{-\lambda(t_k-t_s)} b_{\sigma\rho}(t_{s-1})(t_s - t_{s-1}) \overset{\text{def}}{=} \widetilde{b_{\sigma\rho}}(t_k).$$ (35)

Formula (35) can be written in recursive form, similar to (32):

$$\widetilde{b_{\sigma\rho}}(t_k) = e^{-\lambda(t_k-t_{k-1})}\widetilde{b_{\sigma\rho}}(t_{k-1}) + b_{\sigma\rho}(t_{k-1})(t_k - t_{k-1}).$$ (36)

Consider deviation

$$\psi_{\sigma\rho}(t_k) \overset{\text{def}}{=} V_{\sigma\rho}^{(obs)}(t_k) - \widetilde{b_{\sigma\rho}}(t_k).$$ (37)

For simplicity of discussion suppose that all the time steps $\Delta t_s = \Delta t$. We obtain: $E\psi_{\sigma\rho}(t_k) = o(\Delta t)$. That notation means that this small value is proportional to a value which decreases faster than Δt if Δt decreases.

The increments $\Delta y_\rho(t_s)$ and $\Delta y_\rho(t_q)$ with $t_q < t_s$ are independent when $(x(t_{s-1}), y(t_{s-1}))$ is given. We obtain that the dispersion of the deviation $\psi_{\sigma\rho}(t_k)$ (37) (under condition that $x(t)|_0^k$ is fixed) is equal to

$$Var\{\psi_{\sigma\rho}(t_k)\} \overset{\text{def}}{=} E\left\{\left(\psi_{\sigma\rho}(t_k)\right)^2 |x(t)|_0^k\right\} =$$

$$\sum_{s=1}^{k} e^{-2\lambda(t_k-t_s)}E\left\{\left(b_{\sigma\sigma}(t_{s-i})b_{\rho\rho}(t_{s-1}) + [b_{\sigma\rho}(t_{s-1})]^2\right)|x(t)|_0^{s-1}\right\}(t_s - t_{s-1})^2.$$ (38)

In case if the diffusion coefficients of the observable process, $b_{\sigma\rho}(z,t),$ are the functions of x and t only, so that $b_{\sigma\rho}(z(t),t) = b_{\sigma\rho}(x(t),t)$ (with $(m+1) \leq \sigma, \rho \leq p$) we obtain

$$Var\{\psi_{\sigma\rho}(t_k)\} = \sum_{s=1}^{k} e^{-2\lambda(t_k-t_s)}\left(b_{\sigma\sigma}(t_{s-1})b_{\rho\rho}(t_{s-1}) + [b_{\sigma\rho}(t_{s-1})]^2\right)(t_s - t_{s-1})^2 \overset{\text{def}}{=}$$

$$\widetilde{d_{\sigma\rho}}(x(t)|_0^k, t_k) \overset{\text{def}}{=} \widetilde{d_{\sigma\rho}}(t_k).$$ (39)

The value (39) also can be calculated recursively, similarly to (36). Note, that the value $\widetilde{d_{\sigma\rho}}(t_k)$ is a value proportional to Δt. If $\sigma = \rho$, the above expression takes more concise form:

$$\widetilde{d_{\rho\rho}}(t_k) = 2\sum_{s=1}^{k} e^{-2\lambda(t_k-t_s)} [b_{\rho\rho}(t_{s-1})]^2(t_s - t_{s-1})^2.$$ (40)

In general case, if $b_{\sigma\rho}(z(t),t)$ depend on $y(t)$, we can find the estimates $E\{b_{\sigma\rho}(t_s)|x(t)|_0^{s-1}\} < K_{\sigma\rho} < \infty$, where $K_{\sigma\rho}$ may be determined either as some functions of $x(t)|_0^{s-1}$ or as some constants $K_{\sigma\rho} \leq K_1$. Then for the variance (38) we obtain the following estimates:

$$Var\{\psi_{\sigma\rho}(t_k)\} < 2\sum_{s=1}^{k} e^{-2\lambda(t_k-t_s)}(K_{\sigma\rho}(x(t)|_0^{s-1}))^2(\Delta t_s)^2,$$

$$Var\{\psi_{\sigma\rho}(t_k)\} < (K_1)^2 (\tfrac{1}{\lambda}) \Delta t. \tag{41}$$

The process $z(t)$ is considered on the finite time interval $t_0 \leq s \leq T$, so that the values of $K_{\sigma\rho}(x(t)|_0^{s-1})$ are limited. Thus, the variance of the deviation (37) is a value proportional to Δt, and if $\Delta t \to 0$ the observed quadratic variation tends to the limit (33). We shall demonstrate below, that the probability distribution for the deviation $\psi_{\rho\rho}$ or $\psi_{\sigma\rho}$ tends to the Gaussian one as $\Delta t \to 0$. It follows that the deviations converge to zero if $\Delta t \to 0$.

Besides, we obtained new analytical formulae (38) - (41), which describe the dispersion of the deviation (37) of the observed quadratic variation when observations are taken in discrete times, with small time steps.

For the system (28), the general analytical formulae (38) - (40), obtained in the present paper, provide characterization of the accuracy of the estimate of volatility $|x(t_k)|$, which is constructed with the use of the observed quadratic variation. If this accuracy satisfies requirements, the problem of estimation is solved. In that case the value of the given time step Δt can be considered as 'small enough'.

In general case, if the value of $x(t_k)$ cannot be uniquely recovered only on the base of observed quadratic variations, the further filtering may be needed.

Finally, we are going to demonstrate that the probability density of the random value $\psi_{\rho\rho}(t_k)$ can be approximately described as a Gaussian one. The deviation $\psi_{\rho\rho}(t_k)$ contains the sum of squares of the increments $\Delta y_\rho(t_s)$, which are independent of the past history before t_{s-1} given $y(t_{s-1})$, $x(t_{s-1})$. The increments $\Delta y_\rho(t_s)$ (given $y(t_{s-1})$, $x(t_{s-1})$) can be described as the Gaussian random variables with probability density (16). The sum of squares of Gaussian random variables is not a Gaussian random variable. However, the probability density for the random deviation $\psi_{\rho\rho}(t_k)$ can be approximately described as the Gaussian probability density. Consider the case when $b_{\sigma\rho}(z(t),t)$ (with $(m+1) \leq \rho, \sigma \leq p$) are the functions of $x(t)$ and t only. Then the variance of deviation $\psi_{\rho\rho}(t_k)$ is equal to $\widetilde{d_{\rho\rho}}(t_k)$ (40), which represents a value proportional to Δt (here the time step Δt can be small, but Δt is finite).

Note, that for a Gaussian random variable ξ with $E\xi = 0$ the following relations hold true: $E\{\xi^{2r}\} = Q(r)(E\{\xi^2\})^r$, where $Q(r)$ is a constant, and for the odd moments $E\{\xi^{2r+1}\} = 0$, with $r = 1, 2, 3, \dots$. Those relations can be useful in estimating of the moments of $\Delta y_\rho(t_s)$ when $(y(t_{s-1}), x(t_{s-1}))$ is given. Then the consideration of the higher moments of the deviation proves that $E(\psi_{\rho\rho}(t_k))^q = o(\Delta t)$, if $q > 2$. Then the characteristic function of the random deviation $\psi_{\rho\rho}(t_k)$ takes the form similar to (12), which implies the Gaussian expression for its inverse transformation, similar to (13). Hence, the probability density of that

deviation can be approximately described as a Gaussian probability density. The smaller Δt the higher the precision of that approximation, although the deviation itself tends to zero if $\Delta t \rightarrow 0$. But in practice the time step should not be chosen too small since in our mathematical model $y(t)$ is considered as a diffusion random process, which is not differentiable at any time t. That is not the case in practice where we have a smooth trajectory $\tilde{y}(t)$, and the quadratic variation of $\tilde{y}(t)$ is equal to zero. Diffusion approximation can be accepted when the time steps $\Delta t_s \geq \tau_{cor}$, where τ_{cor} represents the characteristic time span of decay of correlation, so that correlation between the random values $\Delta \tilde{y}(t_s)$ and $\Delta \tilde{y}(t_{s-1})$ (given $x(t_{s-1}), \tilde{y}(t_{s-1})$) would be negligibly small [13 - 16].

2.4. Detection and rejection of highly a posteriori improbable samples

In the process of resampling, the *a posteriori* improbable samples $x(t)_0^n$ are being rejected when their weights become small. Besides, it is possible to discard some highly *a posteriori* improbable samples before all the weights $W_i(n)$ or $\widetilde{W}_i(n)$ are calculated. It can be done with the use of the following tests that are based on analysis of the obtained analytical expressions (19), (21), (22).

The measurements $y(t)|_0^k$ are being obtained under condition that there is an unobservable realization of the process $x(t)$, $x^{tr}(t)|_0^k$, which is unknown. (Here the superscript "tr" stands for "true"). The first and the second moments of the random variables $\Delta y_\rho(t_s)$ given $x^{tr}(t)|_0^k$ are equal to

$$E\{\Delta y_\rho(t_s)|x^{tr}(t_{s-1}), y(t_{s-1})\} = A_\rho(x^{tr}(t_{s-1}), y(t_{s-1}), t_{s-1})\Delta t_s \overset{\text{def}}{=} A_\rho^{tr}(t_{s-1})\Delta t_s, \qquad (42)$$

$$E\{\Delta y_\sigma(t_s)\Delta y_\rho(t_s)|x^{tr}(t_{s-1}), y(t_{s-1})\} = b_{\sigma\rho}(x^{tr}(t_{s-1}), y(t_{s-1}), t_{s-1})\Delta t_s + o(\Delta t_s) \overset{\text{def}}{=} b_{\sigma\rho}^{tr}(t_{s-1})\Delta t_s + o(\Delta t_s).$$

For simplicity of discussion we can assume that all the time steps are equal to Δt.

Consider at first the case when the diffusion coefficients $b_{\sigma\rho}$ of the observable process $y(t)$ do not depend on $x(t)$; then $b_{\sigma\rho}^i = b_{\sigma\rho}^{tr} = b_{\sigma\rho}$.

The value $P_i(k)$ (19) contains the following cofactor:

$$C \exp\left\{-\frac{1}{2}\sum_{s=1}^{k}\frac{1}{\Delta t_s}c_{o\rho}(t_{s-1})[\Delta y_\sigma(t_s)\Delta y_\rho(t_s) - 2\Delta y_\sigma(t_s)A_\rho^i(t_{s-1})\Delta t_s + \right.$$

$$\left. A_\sigma^i(t_{s-1})A_\rho^i(t_{s-1})(\Delta t_s)^2]\right\}, \qquad (43)$$

where the normalization factor C does not depend on $x^i(t)|_0^k$ in that case.

Consider the hypothesis H_i that the sample $x^i(t)|_0^k$ is situated in the vicinity of the true realization $x^{tr}(t)|_0^k$, and the 'closeness' means that $|A_\rho^i(t_{s-1}) - A_\rho^{tr}(t_{s-1})| \leq D_\rho(x^i(t_{s-1}), y(t_{s-1}), t_{s-1}) \overset{\text{def}}{=} D_\rho^i(t_{s-1})$. Here the function $D_\rho(x, y, t)$ can be determined on the basis of some preliminary analysis of the considered dynamical system, for example, we can put $D_\rho(x, y, t) = \left|\frac{\partial A_\rho(x,y,t)}{\partial x_\alpha}\right| R_\alpha$, where R_α are constants, $1 \leq \alpha \leq m$. In order to distinguish between two hypotheses, H_i and its negation $\overline{H_i}$, consider the following value:

$$Q^i(t_k) \overset{\text{def}}{=} \sum_{s=1}^k c_{\sigma\rho}(t_{s-1}) A_\rho^i(t_{s-1})\Delta y_\sigma(t_s) . \tag{44}$$

Note, that $Q^i(t_k)$ can be recursively calculated. Denote

$$G^i(t_k) \overset{\text{def}}{=} \sum_{s=1}^k c_{\sigma\rho}(t_{s-1}) A_\sigma^i(t_{s-1})A_\rho^i(t_{s-1})\Delta t_s. \tag{45}$$

The larger the difference $|Q^i(t_k) - G^i(t_k)|$ the farther the sample $x^i(t)|_0^k$ can be from $x^{tr}(t)|_0^k$ Denote $\widetilde{\Delta y_\sigma}(t_s) = \Delta y_\sigma(t_s) - A_\sigma^{tr}(t_{s-1})\Delta t_s$, $U^i(t_k) \overset{\text{def}}{=} \sum_{s=1}^k c_{\sigma\rho}(t_{s-1}) A_\rho^i(t_{s-1})\widetilde{\Delta y_\sigma}(t_s)$.

Consider the variance of the random value $U^i(t_k)$ when the realization $x^{tr}(t)|_0^k$ is given. We obtain

$$Var\{U^i(t_k)\} = \sum_{s=1}^k E\{c_{\sigma\rho}(t_{s-1})b_{\sigma\eta}(t_{s-1})c_{\eta\tau}(t_{s-1})A_\rho^i(t_{s-1})A_\tau^i(t_{s-1})|x^{tr}(t)|_0^s\}\Delta t_s .$$

If the matrix $\|b_{\sigma\rho}\|$ is not degenerated, then $b_{\sigma\eta}c_{\eta\tau} = \delta_{\sigma\tau}$, where $\delta_{\sigma\tau}$ is Kronecker symbol, $\delta_{\sigma\tau} = 1$ if $\sigma = \tau$ and $\delta_{\sigma\tau} = 0$ if $\sigma \neq \tau$. In that case, the above expression for $Var\{U^i(t_k)\}$ can be written in more concise form:

$$Var\{U^i(t_k)\} = \sum_{s=1}^k E\{c_{\sigma\rho}(t_{s-1})A_\rho^i(t_{s-1})A_\sigma^i(t_{s-1})|x^{tr}(t)|_0^s\}\Delta t_s . \tag{46}$$

In order to reject the samples $x(t)|_0^k$, which are highly improbable given $y(t)|_0^k$, we can use approximately the following test. If the considered dynamic system is stable, the following value can be used as an approximate estimate for the above expression (46), if the hypothesis H_i is true and the measurements $y(t)|_0^k$ are obtained:

$$\widetilde{Var}\{U^i(t_k)\} = \sum_{s=1}^k c_{\sigma\rho}(t_{s-1})A_\rho^i(t_{s-1})A_\sigma^i(t_{s-1})\Delta t_s.$$

The probability of the following event:

$$|Q^i(t_k) - G^i(t_k)| > \left[\sum_{s=1}^k c_{\sigma\rho}(t_{s-1})A_\rho^i(t_{s-1})D_\sigma^i(t_{s-1})\Delta t_s + 3\sqrt{\widetilde{Var}\{U^i(t_k)\}}\right] \tag{47}$$

is small under condition that H_i is true. Hence, if the inequality (47) is satisfied for the considered sample $x^i(t)|_0^k$, the hypothesis H_i should be rejected, and that sample $x^i(t)|_0^k$

should be discarded. If all the considered sample sequences are discarded, the new sample points $x(t_k)$ should be generated as initial points, i.e. as samples from the initial distribution $P_0(x(t_k)|y(t_k))$ (18). For the samples $x^i(t)|_0^k$ that remain under consideration the values $P_i(k)$ and $W_i(k)$ or $\widetilde{W}_i(k)$ will be recursively computed. The samples that were not discarded can be continued with a few 'offsprings'. Due to the test (47) the samples $x^i(t)|_0^k$ that are highly *a posteriori* improbable (when the sequence of observations $y(t)|_0^k$ is given) will be rejected at once, before the weights $W_j(k)$ or $\widetilde{W}_i(k)$ are computed.

Consider the case when some diffusion coefficients $b_{\sigma\rho}$ *depend on* $x(t)$. In some cases, as it was shown above, the unobservable value of $x(t)$ can be restored with the use of the observed quadratic variations. In other cases, if $x(t)$ cannot be uniquely recovered, it is purposeful to use observed quadratic variations $V_{\rho\rho}^{(obs)}(t_k)$ in order to reject the samples $x^i(t)|_0^k$ that are highly *a posteriori* improbable. That test is similar to the above test (47). Assume that if hypothesis H_i is true, the sample $x^i(t)|_0^k$ is 'close' to the true realization $x^{tr}(t)|_0^k$, and the 'closeness' means that

$$\left|b_{\rho\rho}(x^i(t_s),y(t_s),t_s) - b_{\rho\rho}(x^{tr}(t_s),y(t_s),t_s)\right| \le B_\rho(x^i(t_s),y(t_s),t_s) \overset{\text{def}}{=} B_\rho^i(t_s).$$

Here the function $B_\rho(x,y,t)$ can be determined as $B_\rho(x,y,t) = \left|\frac{\partial b_{\rho\rho}(x,y,t)}{\partial x_\alpha}\right| R_\alpha$, where R_α are constants. Denote shortly (similarly to (35), (36), (38),(39)):

$$\widetilde{b_{\sigma\rho}^i}(t_k) \overset{\text{def}}{=} \widetilde{b_{\sigma\rho}}(x^i(t)|_0^k,y(t)|_0^k,t_k)\,, \quad \widetilde{d_{\sigma\rho}^i}(t_k) \overset{\text{def}}{=} \widetilde{d_{\sigma\rho}}(x^i(t)|_0^k,y(t)|_0^k,t_k)\,,$$

$$\widetilde{B_\rho^i}(t_k) \overset{\text{def}}{=} \widetilde{B_\rho}(x^i(t)|_0^k,y(t)|_0^k,t_k) \overset{\text{def}}{=} \sum_{s=1}^k e^{-\lambda(t_k-t_s)} B_\rho(x^i(t_s),y(t_s),t_s)(t_s - t_{s-1}),$$

$$\widetilde{B_\rho^i}(t_k) = e^{-\lambda(t_k-t_{k-1})}\widetilde{B_\rho^i}(t_{k-1}) + B_\rho^i(t_k)(t_k - t_{k-1}).$$

Consider the case when $b_{\sigma\rho}(z(t),t) = b_{\sigma\rho}(x(t),t)$ (with $(m+1) \le \sigma,\rho \le p$). Then the probability of the following event

$$\left|V_{\rho\rho}^{(obs)}(t_k) - \widetilde{b_{\rho\rho}^i}(t_k)\right| > \left[\widetilde{B_\rho^i}(t_k) + 3\sqrt{\widetilde{d_{\rho\rho}^i}(t_k)}\right] \tag{48}$$

is small if the hypothesis H_i is true. Hence, if the inequality (48) is satisfied at least for one number ρ, with $(m+1) \le \rho \le p$, then the hypothesis H_i should be rejected, and the sample $x^i(t)|_0^k$ should be discarded at once, before the calculation of all the weights $W_j(k)$ or $\widetilde{W}_j(k)$ is done.

In case if $b_{\sigma\rho}(z(t),t) = b_{\sigma\rho}(x(t),y(t),t)$, with $E\{(b_{\sigma\rho})^2|x^{tr}(t)|_0^k\} \le K_1^2 = Const$, the following test can be used instead of (48):

$$\left| V_{\rho\rho}^{(obs)}(t_k) - \widetilde{b_{\rho\rho}^i}(t_k) \right| > \left[\widetilde{B_\rho^i}(t_k) + 3K_1 \sqrt{\frac{1}{\lambda} \Delta t} \right]. \tag{49}$$

2.5. Incorporation of the derivative of the process of quadratic variation into the set of observed data. Systems with some additional precise observations

In case of continuous time of observation, the process of quadratic variation can be included into the set of observed processes, as it was demonstrated in Section 2.3.

For simplicity of notation, consider the system of one-dimensional processes $x(t)$, $y(t)$:

$$dx = m(x,t)dt + \sigma_1(x,t)dw_1(t), \tag{50}$$

$$dy = h(x,t)dt + \sigma_2(x,t)dw_2(t), \tag{51}$$

where $w_1(t)$, $w_2(t)$ are independent Wiener processes, the functions $m(x,t)$, $h(x,t)$ satisfy Lipschitz condition, and the functions $h(x,t), \sigma_1(x,t)$, $\sigma_2(x,t)$ are twice differentiable with respect to x. The process

$$V(t) = \int_{t_0}^t e^{-\alpha(t-s)} \sigma_2^2(x(s),s)ds \tag{52}$$

can be considered theoretically as an observed process, given the trajectory $y(s)$ for $t_0 \leq s \leq t$. Here $\alpha = Const \geq 0$. The process $V(t)$ obeys the ordinary differential equation

$$\frac{dV}{dt} = -\alpha V + \sigma_2^2(x(t),t).$$

For continuous time of observation and multidimensional diffusion processes $(x(t), y(t))$, in some cases the possible values of $x(t)$ are precisely determined by the equalities similar to (52) (and, consequently, $x(t)$ can be restored precisely, at least theoretically) , and in other cases the trajectory $x(s)$ must be localized in some 'layer', where those equalities are satisfied, given the observed path of the process of quadratic variation, $V(s)$, for $t_0 \leq s \leq t$.

Theoretically, it is possible to consider the derivative $\frac{dV(t)}{dt} \stackrel{\text{def}}{=} u(t)$ as an observed process, and incorporate it into the set of observed components: $(y(t), u(t))$. For the system (50), (51), the process $u(t)$ satisfies the following stochastic differential equation (in the sense of Ito):

$$du = -\alpha u\, dt + 2\sigma_2(x,t)\frac{\partial \sigma_2}{\partial x}[m(x,t)dt + \sigma_1(x,t)dw_1(t)] + \left[\sigma_2 \frac{\partial^2 \sigma_2}{\partial x^2} + \left(\frac{\partial \sigma_2}{\partial x}\right)^2\right]\sigma_1^2(x,t)dt.$$

$$\tag{53}$$

The above SDE (53) could be augmented to the system of SDE (50) – (51) for $x(t)$, $y(t)$.

But in practice we cannot assume that the process $u(t)$ (or, equivalently, the process $\sigma_2^2(x(t),t)$) is straightforwardly observed and known precisely. Practically, the process $V(t_k)$ is not precisely given. Even in case if the deviations of 'observed quadratic variation' $V^{(obs)}(t_k)$ from $V(t_k)$ are small, in obtaining the estimates for derivative, $\frac{dV(t)}{dt}|_{t=t_k} \overset{\text{def}}{=} u(t_k)$, the sought values could be inevitably corrupted by errors.

However, if additional observations of $y(t)$ are available, namely, $y(t_0 + s\delta t)$, with $\delta t \ll \Delta t$, $s = 1, 2, ...$, then the estimates $\hat{u}(t_k)$ for the values of $u(t_k)$ can be improved. Suppose that there is a special digital device that takes measurements of $y(t'_s)$ at times $t'_s = t_0 + s\delta t$, with $t_{k-1} < t'_s \leq t_k$, and use them in order to calculate the estimate $\hat{u}(t_k)$. For the moment, consider the case if we had $\hat{u}(t_k) = u(t_k)$. Then the problem of filtering (based on observations $y(t_k)$, $u(t_k)$ taken at discrete times t_k, with small time steps Δt_k) could be solved effectively with the use of the new algorithm of simulation (17) and the estimates (26) and algorithms suggested in the present paper. For the system (50), (51), (53) we obtain in (17): $b_{11} - b_{1\sigma}c_{\sigma\rho}b_{\rho 1} \equiv 0$. That means that the sample paths $x^i(t)|_1^n$ would be deterministically continued given the observations $u(t)|_1^n$ and the initial values $x^i(t_1)$, with $i = 1, ..., N$. It is purposeful to choose the probability distribution of the initial random samples $x^i(t_1)$ in such way that $E\{\sigma_2^2(x^i(t_1), t_1)\} = u(t_1)$, with some random scattering. Meanwhile, the weights $\widetilde{W}_i(n)$ are being defined by (25), (26) with the use of observations $y(t)|_1^n$. Thus, the *a posteriori* distribution is being determined not only by $u(t)|_1^n$ but also by $y(t)|_1^n$. Then, as it was shown in Section 2.2, the estimate (26) converges with probability one (due to the Strong Law of Large Number, as $N \to \infty$) to the sought *a posteriori* expectation that provides solution to the filtering problem. That algorithm could be implemented with the use of parallel computing.

The random errors of the obtained estimates $\hat{u}(t_k)$ can be taken into account in the following way. The random deviations $\hat{u}(t_k) - u(t_k) \overset{\text{def}}{=} \varepsilon(t_k)$ can be approximately described as independent of each other (for different k) Gaussian random values. Then $\hat{u}(t_k) = u(t_k) + \varepsilon(t_k)$ can be considered as measurements (taken at discrete times t_k) of the process $\hat{u}(t)$ that satisfies the following SDE:

$$d\hat{u} = -\alpha\hat{u}\, dt + 2\sigma_2(x,t)\frac{\partial\sigma_2}{\partial x}[m(x,t)dt + \sigma_1(x,t)dw_1(t)] + \left[\sigma_2\frac{\partial^2\sigma_2}{\partial x^2} + \left(\frac{\partial\sigma_2}{\partial x}\right)^2\right]\sigma_1^2(x,t)dt +$$
$$\sigma_3 dw_3(t), \tag{54}$$

where $w_3(t)$ is an independent Wiener process, and $\sigma_3^2(t_k)$ is proportional to the variance of $\varepsilon(t_k)$, i.e. $E\{\varepsilon^2(t_k)\}$. Thus, σ_3 can be determined as some function of x, t. Then the solution of the filtering problem for the system (50), (51), (54) can be obtained in the similar way as above.

Note that the systems that are similar to (50), (51), (53) arise if the measurements of some given function $H(x(t), t)$ are known without errors. Here the function $H(x, t)$ is supposed to

be twice differentiable with respect to x. Then the solution of the filtering problem (given observations of $y(t)$ and $H(x(t), t)$ at discrete times $t_k = t_0 + k\Delta t$, with small time step Δt) can be obtained with the use of the algorithm of simulation (17) and the estimates (26), similarly to the solution described above for the system (50), (51), (53).

2.6. The problem of nonlinear filtering of a signal. Change of the mathematical model that describes the process of observations

For simplicity of notations, consider one-dimensional processes $x(t)$, $y(t)$. In most of works devoted to nonlinear filtering problems the following hidden Markov model is accepted for description of a signal process $x(t)$ and an observed process $y(t)$:

$$dx = m(x,t)dt + \sigma_1(x,t)dw_1(t), \tag{55}$$

$$dy = h(x(t),t)dt + \sigma_2 dw_2(t), \tag{56}$$

where $w_1(t)$, $w_2(t)$ are independent Wiener processes, $\sigma_2 \equiv Const$, the functions $m(x,t)$, $h(x,t)$, $\sigma_1(x,t)$ satisfy Lipschitz condition, and the functions $h(x,t)$, $\sigma_1(x,t)$ are twice differentiable with respect to x.

In many applications, it would be possible to describe an observed process $Y(t)$ as follows:

$$Y(t) = h(x(t),t) + u(t), \tag{57}$$

where $u(t)$ represents the errors of measurements. The 'noise process' $u(t)$ can be considered as a Gaussian Markov process, which can be described by the following stochastic differential equation:

$$du = -\beta u dt + \sigma_3 dw_3(t), \tag{58}$$

where β and σ_3 are constant, and $w_3(t)$ is an independent Wiener process. Then $E\{u\} = 0$, the variance $E\{u^2\} = \sigma_3^2/2\beta$, and the correlation function $E\{u(t)u(t+\tau)\} = (\sigma_3^2/2\beta)e^{-\beta\tau}$.

Denote $z_1(t) \overset{\text{def}}{=} x(t)$, $z_2(t) \overset{\text{def}}{=} u(t)$, $z_3(t) \overset{\text{def}}{=} Y(t)$.

Then the model (55), (57), (58) of the signal process $x(t)$ and the observed process $Y(t)$ can be described as follows:

$$dz_1 = m(z_1,t)dt + \sigma_1(z_1,t)dw_1(t), \quad dz_2 = -\beta z_2 dt + \sigma_3 dw_3(t),$$

$$dz_3 = \frac{\partial h(z_1,t)}{\partial t}dt + \frac{\partial h}{\partial z_1}[m(z_1,t)dt + \sigma_1(z_1,t)dw_1] + \frac{1}{2}\frac{\partial^2 h}{\partial z_1^2}\sigma_1^2(z_1,t)dt - \beta z_2 dt + \sigma_3 dw_3. \tag{59}$$

Here $z_3(t)$ represents the observed process, and $z_1(t)$, $z_2(t)$ are unobservable components of the process $z(t) \overset{\text{def}}{=} (z_1(t), z_2(t), z_3(t))$. It is easy to see that the diffusion coefficients

27

$$b_{11} = \sigma_1^2(z_1, t), \ b_{13} = \frac{\partial h}{\partial z_1} \sigma_1^2(z_1, t), \ b_{12} = 0, \ b_{22} = \sigma_3^2, \ b_{23} = \sigma_3^2, \ b_{33} = \left(\frac{\partial h}{\partial z_1} \sigma_1(z_1, t)\right)^2 + \sigma_3^2.$$

Consider the matrix $\|r_{\alpha\beta}\| = \|b_{\alpha\beta} - b_{\alpha\sigma} c_{\sigma\rho} b_{\rho\beta}\|$, which was introduced in (15): This is the variance-covariance matrix for the conditional probability density of the increments $\Delta z_1 = z_1(t + \Delta t) - z_1(t)$, $\Delta z_2 = z_2(t + \Delta t) - z_2(t)$, given $z(t)$ and the increment $\Delta z_3 = z_3(t + \Delta t) - z_3(t)$. For the system (59) we find: $r_{11} = \sigma_1^2 \left[1 - \frac{1}{1 + \sigma_3^2 / \left(\frac{\partial h}{\partial z_1} \sigma_1\right)^2} \right],$

$$r_{12} = -\sigma_3^2 \sigma_1^2 \frac{\partial h}{\partial z_1} \left[\left(\frac{\partial h}{\partial z_1} \sigma_1\right)^2 + \sigma_3^2\right]^{-1}, \ r_{22} = \sigma_3^2 \left[1 - \sigma_3^2 \left[\left(\frac{\partial h}{\partial z_1} \sigma_1\right)^2 + \sigma_3^2\right]^{-1} \right], \text{ and } Det\|r_{\alpha\beta}\| \equiv 0.$$

The case when σ_3 is small corresponds to the small errors of measurements. All the entries of the matrix $\|r_{\alpha\beta}\|$ (shown above) is small if σ_3 is small. That means that most of the samples $(z_1(t), z_2(t))|_0^n$ (given $z_3(t)|_0^n$) simulated with the use of the algorithm (17) continue to be localized in the domain where the *a posteriori* probability density is concentrated. Besides, the test (48) can be implemented (unless the function $h(x, t)$ is linear with respect to x and $\sigma_1 \equiv Const$) in order to discard *a posteriori* improbable samples with the use of 'observed quadratic variation' (for the observations made in discrete time) of the observed process $z_{33}(t)$.

Thus, although the dimensionality of the unobservable process increased for the model of description (59) in comparison with the model (55), (56), the problem of nonlinear filtering (with observations made in discrete time) will be solved effectively with the use of the new Monte Carlo estimates (23) - (27) and with the use of the new algorithm of simulation (17), derived in the present paper.

Note that the model (59) is a hidden Markov model where unobservable process $(z_1(t), z_2(t))$ is a Markov process in itself. But if the samples $(z_1(t), z_2(t))|_0^n$ were simulated as trajectories of the Markov process $(z_1(t), z_2(t))$ in itself (as it is suggested in particle filters, developed and studied in the literature), then, in case of the low 'level of intensity of noise' (i.e. if σ_3 is small) the large part of generated samples would be localized in the domain of low *a posteriori* probability (and that part is growing when σ_3 decreases). Then, in order to obtain some feasible accuracy of filtering, the number N should be increased significantly if the value of σ_3 decreased. Meanwhile, the large amount of calculations would be wasted in vain for processing of the samples of low *a posteriori* probability. Thus, this example (the system (59)) also proves the advantage of the new algorithms, developed in the present paper, in comparison with the known algorithms for hidden Markov models.

3. Conclusion

In this work the analytical investigation of multidimensional diffusion Markov processes considered at discrete times t_k, with small time steps Δt_k, is developed. The analytical formulae in closed form for the conditional probability density for increments $\Delta x = x(t + \Delta t) - x(t)$ of unobservable components $x(t)$ given increments $\Delta y = y(t + \Delta t) - y(t)$ of the observable components $y(t)$ and for the probability density of Δy, given $(x(t), y(t))$, are obtained. On the base of this investigation the new algorithms for simulation of unobservable components, $x(t_k)$, of a diffusion Markov process given the measurements of the other components $y(t)|_0^k$ and the new algorithms of nonlinear filtering and estimation with the use of sequential Monte Carlo methods, or particle filters, are derived and suggested.

The analytical investigation of observed quadratic variations is also developed. If diffusion coefficients of the observed components, $y(t)$, depend on unobserved components, $x(t)$, then the observed quadratic variations can be used for estimation of unknown values of $x(t)$. The new analytical formulae, (38) – (40), obtained in this work, characterize the accuracy of such estimates. Incorporation of the observed quadratic variations into the set of observed data is investigated.

Besides the use of the weights $\widetilde{W}_i(n)$ or $W_i(n)$, some new tests for detection and rejection of highly *a posteriori* improbable samples $x(t)|_0^k$ are derived. Realization of such tests provides opportunity for implementation of the new suggested particle filtering algorithm (with branching sampling and recurrent accumulation of the values (25) or (27)) with the use of parallel computing.

The obtained new algorithms and estimates extend the range of applications of sequential Monte Carlo methods, or particle filters, beyond the hidden Markov models, and improve their performance.

References

1. D. Crisan, A. Doucet. A survey on convergence results on particle filtering methods for practitioners. *IEEE Trans. Signal Process.* 2002; **50**: 736 – 746.
2. A.Doucet, N. De Freitas, N. Gordon, editors. *Sequential Monte Carlo methods in practice.* Statistics for Engineering and Information Science. Springer – Verlag, New York, 2001.
3. A. Doucet, S. Godsill, C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* 2000; **10**: 197 – 208.

4. P. Del Moral. Measure-valued processes and interacting particle systems. Application to nonlinear filtering problems. *Ann. Appl. Probab.*1998; **8**: 438 – 495.

5. H. Carvalho, P. Del Moral, A. Monin, and G. Salut. Optimal Nonlinear filtering in GPS/INS integration. *IEEE Trans. Aerospace Electr. Syst.* 1997; **39**: 835 – 850.

6. A.N. Kolmogorov. Uber die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.* 1931, Bd. 104, S. 413-458. Russian translation in "Uspekhi Mat. Nauk", 1938, issue 5, pages 5-41. Republished in *Collected works of A.N. Kolmogorov*, *The Theory of Probability and Mathematical Statistics*, Moscow Publishing House: Nauka, 1986 (in Russian).

7. A.N. Kolmogorov. Zur Theorie der stetigen zufalligen Prozesse. *Math. Ann.* 1933, Bd. 108, S. 149-160. Republished in *Collected works of A.N. Kolmogorov*, *The Theory of Probability and Mathematical Statistics*, Moscow Publishing House: Nauka, 1986 (in Russian).

8. A.N. Kolmogorov. *Foundations of the Theory of Probability.* Published at first in German as *"Grundbegriffe der Wahrscheinlichkeitsrechnung"* in 1933. Russian editions in 1936, 1974. English edition: Chelsea, New York, 1950.

9. S.N. Bernstein. Principes de la theorie des equations differentielles stochastiques. *Proceeding of the Phys.-Mat. Steklov Institute* 1934; **5** : 95 -124. Republished in *Collected works of S.N. Bernstein*, vol.**4**. Moscow Publishing House: Nauka, 1964 (in Russian).

10. S.N. Bernstein. Equations differentielles stochastiques. *"Actualites Sci. Ind."* 1938; **738**: 5 - 31. Conf. internat. Sci. math. Univ. Geneve. Theorie des probabilites, **V**. Les fonctions aleatoires. Republished in *Collected works of S.N. Bernstein,* vol.**4.** Moscow Publishing House: Nauka, 1964 (in Russian).

11. J.L. Doob. *Stochastic processes*. New York–John Wiley & Sons, London–Chapman & Hall, 1953.

12. B.V. Gnedenko. *The Theory of Probability.* Mir Publishers, Moscow, 1976.

13. E.M. Khazen. *Methods of optimal statistical decisions, optimal control, and stochastic differential equations.* Xlibris Publ. House: USA, 2009.

14. E.M. Khazen. *Methods of optimal statistical decisions and optimal control problems.* Soviet Radio Publ. House: Moscow, 1968 (in Russian).

15. E.M. Khazen. On stochastic differential equations for the *a posteriori* probability distribution in problems of adaptive filtering and signal detection. *Automation and Remote Control* 1971**; 32**: 1776 - 1782.

16. B.N. Petrov, G.M. Ulanov, S.V. Ulyanov, E.M. Khazen. *Information Theory in Processes of Optimal Control and Organization*. Chapters 2, 3, 5 by E.M. Khazen, p. 47-66, 67-102, 193-243. Moscow Publishing House: Nauka, 1977 (in Russian).

17. J.C. Hull. *Options, Futures, & Other derivatives.* Prentice Hall, Inc., NJ, 2000, 1997, 1993, 1989 (four editions).

18. S. Thrun, D. Fox, W. Burgard, F. Dellaert. Robust Monte Carlo Localization for Mobile Robots. *Artificial Intelligence* 2001; **128:** 99 -141.

19. H. Kushner, *Introduction to stochastic control.* N.Y.: Holt, Rinchart, and Winston, Inc., 1971