ISSN 0005-1179, Automation and Remote Control, 2019, Vol. 80, No. 10, pp. 1755–1782. © Pleiades Publishing, Ltd., 2019.
Russian Text © The Author(s), 2019, published in Avtomatika i Telemekhanika, 2019, No. 10, pp. 3–36.

TOPICAL ISSUE

The Method of Averaged Models for Discrete-Time Adaptive Systems

N. O. Amelina[*,**,a], **O. N. Granichin**[*,**,b], **and A. L. Fradkov**[*,**,c]

[*] *St. Petersburg State University, St. Petersburg, Russia*
[**] *Institute for Problems in Mechanical Engineering, Russian Academy of Sciences, St. Petersburg, Russia*
e-mail: [a]*natalia_amelina@mail.ru,* [b]*o.granichin@spbu.ru,* [c]*fradkov@mail.ru*

Received July 19, 2018
Revised September 6, 2018
Accepted November 8, 2018

Abstract—Dynamical processes in nature and technology are usually described by continuous- or discrete-time dynamical models, which have the form of nonlinear stochastic differential or difference equations. Hence, a topical problem is to develop effective methods for a simpler description of dynamical systems. The main requirement to simplification methods is preserving certain properties of a process under study. One group of such methods is represented by the methods of continuous- or discrete-time averaged models, which are surveyed in this paper. New results for stochastic networked systems are also introduced. As is shown below, the method of averaged models can be used to reduce the analytical complexity of a closed loop stochastic system. The corresponding upper bounds on the mean square distance between the states of an original stochastic system and its approximate averaged model are obtained.

Keywords: dynamical systems, nonlinear stochastic equations, adaptive systems, methods to simplify description, approximate averaged models

DOI: 10.1134/S0005117919100011

1. INTRODUCTION

In most problems of applied mathematics, physics, and engineering, the mathematical model of a process under study is subjected to some simplifications. In modern control theory, the replacement of an original system description by a simpler one is widely used in large-scale systems control, adaptive control, control of stochastic systems and systems with distributed parameters and other problems. Mathematical descriptions of processes in these and many other problems usually involve continuous- or discrete-time dynamical models having the form of differential or difference equations. This fact explains the sustained interest of researchers in the development and justification of effective methods to simplify the dynamical systems. The main requirement to simplification methods is preserving certain properties of a process under study.

For continuous-time systems, the well-known simplification method is based on the separation of fast and slow motions in a system and averaging of the fast component [1–12]. This method, called the averaging method (or principle), was developed for both deterministic and stochastic systems. The averaging effect in such systems occurs due to the trend of the fast component towards a constant value or a rapidly periodic oscillation (under fixed slow variables). The applicability of the averaging method to stochastic differential equations can be justified by the fact that the fast motion (perturbation) has the property of weak dependence: the values of perturbations at remote time instants are “almost” independent.

The averaging effect may also arise in discrete-time systems described by stochastic difference equations. Here the averaging mechanism has a different character, being based on a small increments of the process in one step. Being summed up in large amounts, the effects of these small increments compensate each other. As a result, the process trajectory is likely to be the solution of an averaged difference equation, which can be called *the discrete-time deterministic model* of the original system. As is easily demonstrated, the approximation accuracy of the original system trajectories will not change its order in the case of passing from the discrete-time model to its continuous-time counterpart by letting the sampling interval tend to zero. The obtained ordinary differential equation will be called *the deterministic continuous-time model* of the original system. Together with a deterministic model, which gives a rough approximation to the original system, we can construct a family of *stochastic continuous-time models* described by stochastic differential equations that will better approximate the probability characteristics of the original system.

Thus, instead of solving the analysis or design problem for the original system, we may try to solve a similar problem for its deterministic or stochastic continuous-time counterpart, expecting that the latter would be simpler than the former. This replacement forms the core of *the method of continuous-time models* [13–15], which is surveyed below. This paper has two objectives—to familiarize the readers with the promising (albeit, not novel) analysis method and also to trace its history. For a detailed justification and application of the method of continuous-time models, including a survey of the research up to 1981, see the book [16]. The more recent publications on the subject were briefly overviewed in [17, 18]. We will use *the method of averaged models* as a universal term for the method of discrete-time models and the method of continuous-time models: in both cases, the key element is an averaging procedure of the right-hand sides.

In what follows, the general idea of the method will be clarified by a simple illustrative example—searching for a unique root of a nondecreasing function.

Let $g(x)$ be a given continuously differentiable function. Its root can be found using an iterative procedure of the form

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \alpha g(\hat{\theta}_{k-1}) \tag{1.1}$$

with a fixed sufficiently small coefficient $\alpha > 0$. Here $\hat{\theta}_0$ denotes an initial approximation chosen by the user and $\hat{\theta}_k$ is a current estimate at iteration $k = 1, 2, \ldots$.

If the initial approximation $\hat{\theta}_0$ is sufficiently close to the root θ of the function $g(x)$, this procedure guarantees the convergence of all estimates to θ assuming that $g(x) < 0$ for $x < \theta$, $g(x) > 0$ for $x > \theta$, the derivative of this function is bounded and $g'(x) > 0$ in some neighborhood of the point θ [19]. Generally speaking, this procedure even does not require the differentiability of $g(x)$.

Denote by E the expectation operator. For algorithm (1.1) with noisy measurements of the function $g(x)$, instead of $g(\hat{\theta}_{k-1})$ only the noisy data $Y_k = G(w_k, \hat{\theta}_{k-1})$ can be used, which represent the realizations of some random variables $G(w_1, \hat{\theta}_0), G(w_1, \hat{\theta}_1), \ldots$. The corresponding algorithm to find the root of a function $g(x) = \mathrm{E}G(w, x)$ was suggested in 1951 in the paper [20]; it has the form

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \alpha_k Y_k,$$

with a vanishing sequence of the step parameters $\{\alpha_k\}$ that is chosen by the user so that

$$\alpha_k > 0, \quad \sum_k \alpha_k = \infty, \quad \sum_k \alpha_k^2 < \infty.$$

This algorithm is known as the Robbins–Monro algorithm (RM).

Like in [21], let us demonstrate that the noises with zero mean and bounded variance will not affect the asymptotic behavior of the algorithm as $k \to \infty$. On the one hand, for large values k the step $\alpha_k \to 0$ and the values $\hat{\theta}_k$ have slow variations. On the other, for a sufficiently small $\epsilon > 0$ we define K_k^ϵ so that $\sum_{i=k}^{k+K_k^\epsilon} \alpha_i \approx \epsilon$.

The RM algorithm can be represented as

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \alpha_k g(\hat{\theta}_{k-1}) + \alpha_k \left(g(\hat{\theta}_{k-1}) - Y_k \right).$$

Consequently,

$$\hat{\theta}_{k+K_k^\epsilon} - \hat{\theta}_{k-1} \approx -\epsilon g(\hat{\theta}_{k-1}) + error,$$

where

$$error = \sum_{i=k}^{k+K_k^\epsilon} \alpha_i \left(g(\hat{\theta}_{i-1}) - Y_i \right).$$

Assuming the noises $\{Y_k - g(\hat{\theta}_{k-1})\}_{k=1,2,\dots}$ are a sequence of orthogonal random variables with zero means and a bounded variance $\sigma^2(\hat{\theta}_{k-1})$, we may write the following expression of the error variance:

$$\mathrm{E}\left[\sum_{i=k}^{k+K_k^\epsilon} \alpha_i \Big(y_i - g(\hat{\theta}_{i-1}) \Big) \right]^2 = \sum_{i=n}^{n+N_n^\epsilon} \mathcal{O}(\alpha_i^2) = \mathcal{O}(\epsilon)\alpha_k.$$

At the iterations from the interval $[k, k+K_k^\epsilon)$, for small ϵ and large k the mean change of the parameter is greater than the error. Hence, the asymptotic behavior of the estimates will almost surely coincide with that of some solution to the ordinary differential equation

$$\dot{\theta} = -g(\theta).$$

Under additional restrictions, we may show that $\hat{\theta}_k \to \theta$ with probability 1 if θ is an asymptotically stable equilibrium of this equation.

Let us say a few words about the history of this approach.

The first results to justify the transition from an original discrete-time stochastic system to its simplified averaged model were obtained in the early 1970s independently by several authors. In [22], Meerkov performed the transition to an averaged model of a Markov chain using the Krylov–Bogolyubov averaging method. Like in the first and second Bogolyubov theorems, under the assumption of the asymptotic stability of the averaged model the convergence in probability of the solutions to the original and averaged systems on a finite time interval and also the closeness of their trajectories on infinite time interval were established in the case of independent perturbations.

In January 1973, Derevitskii and Fradkov reported the first results on the construction and justification of deterministic and stochastic continuous-time models at Tsypkin's All-Union School on Adaptive Systems in Agveran [23], also submitting a paper to *Automation and Remote Control*. In June 1973, Fradkov received an invitation from Ya.Z. Tsypkin to present this work at his seminar.[1] At the seminar, Fradkov got acquainted with a young Swede, L. Ljung from Linköping, who had an internship during that period at the Institute of Control Sciences. It turned out that at that time, Ljung had already formed the basic ideas of his results on the study of stochastic approximation processes under dependent perturbations. With the conditions of a decreasing process step, it was possible to closely relate the asymptotic properties of the system and its model.

[1] A note by Fradkov: "At that time, I was still very young, not even a graduate student. This was my first invitation and the first speech to venerable scholars. But a benevolent attitude of Tsypkin won excitement and everything went well; the paper was accepted for publication. So we may say that Tsypkin supported the development of this method even in its cradle. Tsypkin also supported the publication of the book [16], becoming its responsible editor."

The paper by Derevitskii and Fradkov [13] was published in early 1974, while Ljung presented his results at the Budapest conference [24] in the same year. The later paper [15] had a great resonance (781 citations as of April 11, 2019 in the Web of Science database) and was even included into the 25 most influential papers on automatic control of the twentieth century [25]. The authors of [13] were less fortunate, but sometimes this approach is called *the Derevitskii–Fradkov–Ljung* (*DFL*) *scheme* [26]. Interestingly, in the paper [27] published by Polyak and Tsypkin in 1973, a Lyapunov function that establishes the asymptotic stability of the averaged continuous-time model was used for proving the results, but the model itself was not introduced. In a similar manner, in the papers on the averaging method in stochastic systems and its application to the stochastic approximation [28, 29], the original and approximating systems were both continuous-time, while the discrete-time models were not actually introduced and not investigated.

The main text of this survey is organized as follows. In Section 2, a formal description of the method of continuous-time models is given. In Section 3, the results yielding the conditions under which the processes in the original system and its deterministic model are close to each other are summarized. Section 4 is focused on the relation between the asymptotic properties of a system and its deterministic model. Stochastic continuous-time models are described in Section 5. In Section 6, the method of averaged models is extended to hybrid (discrete-continuous) systems. The applications and generalizations of the method as well as related issues are discussed in the remainder of the paper. More specifically, the results on the convergence of gradient algorithms with dependent perturbations are discussed in Section 7. The generalizations to the systems with non-Lipschitz right-hand sides and an implicit entry of the step parameter are presented in Section 8. Section 9 is dedicated to the extension of the method to networked systems. Information on some applications is contained in Section 10. Finally, two illustrative examples are provided in Section 11.

2. METHOD OF CONTINUOUS-TIME MODELS

Consider discrete-time systems described by the stochastic difference equations

$$z_{k+1} = z_k + \alpha_k F(z_k, f_k), \quad k = 0, 1, \ldots, \tag{2.1}$$

where $z_k \in \mathbb{R}^n$ denotes the state vector of the system; $f_k \in \mathbb{R}^m$ is a stationary sequence of random vectors (external perturbations); α_k is the step parameter of the system. Performing the averaging procedure of the right-hand sides of (2.1) with respect to f_k for a fixed vector z_k, we construct an averaged discrete-time system of the form

$$\bar{z}_{k+1} = \bar{z}_k + \alpha_k A(z_k), \quad k = 0, 1, \ldots, \tag{2.2}$$

where $A(z) = \mathrm{E}F(z, f_k)$. (It is assumed that the expectation exists and does not depend on k.) Now, passing to the continuous time, we construct the differential equation of the continuous-time system

$$\frac{dz}{dt} = A(z). \tag{2.3}$$

Equation (2.3) will be called the deterministic or averaged continuous-time model of the original discrete-time system (2.1), while the difference Eq. (2.2) will be called the deterministic or averaged discrete-time model. If the sequence f_k is not stationary but there exists the limit $A(z) = \lim_{k\to\infty} \mathrm{E}F(z, f_k)$ for each z, the continuous-time model (2.3) can still be constructed. Of course, for obtaining rigorous statements the function $A(z)$ must satisfy conditions that will guarantee the existence of a unique solution to (2.3).

Model (2.3) can be applied to study the original system (2.1) because (under certain assumptions presented below) the solutions z_k to system (2.1) with small α_k are close in some sense to the solution $z(t)$ to model (2.3) taken at the time instant

$$t_k = \sum_{i=0}^{k-1} \alpha_i. \tag{2.4}$$

The initial conditions in (2.1) and (2.3) are chosen the same: $z(0) = z_0$. By relation (2.4) a fictitious (equivalent) time is introduced in the discrete-time system (2.1). For studying system (2.1), model (2.3) is applied in two stages as follows:

a) building of the model;

b) analysis of the model (analytical or numerical).

For example, the possible limit points of the process z_k can be approximated by solving the equation $A(z) = 0$. In this case, the analytical solution (if exists) will be simpler and numerical solution will require less computing effort. In particular, a stable limit point can be determined by numerical integration of (2.3) with a constant or increasing step. This approach turns out to be "more profitable" than the modeling of the original system (2.1), where often $\alpha_k \to 0$ as $k \to \infty$. If the problem is to estimate the time instant when the trajectory z_k will reach a given neighborhood of the equilibrium state, as a rough approximation we may take the corresponding time for the solution $z(t_k)$ to the deterministic system (2.3), etc. Note that in adaptation problems, the vector z_k usually consists of tunable coefficients, i.e., the current estimates of the coefficients of the controlled system or some "perfect" controller, and system (2.1) describes the adaptation algorithm itself.

The type of results proving the applicability of the method of continuous-time models actually depends on the properties of the original system that have to be studied. If we are interested in the behavior of the system on a finite or infinite time interval (the dynamics analysis problem), the estimates of the closeness of trajectories (2.1) and (2.3) will be needed. If the subject of research is the asymptotic properties of the system (stability, dissipativity, etc.), we will need the conditions under which such properties of system (2.1) follow from the similar properties of model (2.3).

The complexity to justify the method depends on the nature of the perturbations f_k. In the simplest case, the vectors f_k are independent and identically distributed. This situation occurs, e.g., in the problems of identification and adaptive control of static systems. If the controlled system is dynamical, the vectors f_k (system state vectors) become dependent. To justify the method in this case, we have to assume that the dependence between f_k and f_l is weakened with increasing $k - l$. The conditions of dependence weakening (the mixing conditions) have become widespread in the limit theorems of probability theory [30]. In particular, they hold if f_k are generated by the linear difference equation

$$f_{k+1} = G f_k + B\eta_k, \tag{2.5}$$

where η_k represents a set of independent identically distributed random vectors and the matrix G is stable (meaning that all its eigenvalues are less than 1 by absolute value). This situation occurs in identification of linear dynamical systems. In terms of the averaging method [1, 12], we may say that in system (2.1), (2.5) the components of the vector z_k (tunable coefficients) are slow variables while the components f_k (the phase coordinates of the system) are fast variables.

Even more complex problems arise when the dynamics of the fast motion may depend on the slow variables. For example, let the perturbations be described by the following equation instead of (2.5):

$$f_{k+1} = G(z_k) f_k + B(z_k)\eta_k. \tag{2.6}$$

This case is typical for adaptive control problems in which the behavior of a controlled system is determined by the values of the tunable parameters z_k. For building a continuous-time model, we have to consider the fast motion equation with the "frozen" slow variables

$$f_{k+1}(z) = G(z) f_k(z) + B(z)\eta_k \tag{2.7}$$

and let $A(z) = \lim_{k\to\infty} \mathrm{E}F\left(z, f_k(z)\right)$ for model (2.3). In this case, the right-hand side of (2.3) will be defined only for those z for which the limit exists. The existence of the limit and also the weak dependence of the vectors $f_k(z)$ are guaranteed, e.g., if the vectors η_k are independent and identically distributed while the matrix $G(z)$ is stable. Note that the averaging procedure of the fast motions of system (2.1), (2.6) leads to the simplified model (2.3) of smaller order. With this approach, system (2.1), (2.6) is treated as a singularly perturbed one with respect to (2.3); see [2, 5–7].

However, in a number of cases, the separation of fast and slow motions in the system (hence, order reduction for model (2.3)) is incorrect. For example, consider the adaptive stabilization problem of a linear continuous-time system

$$\frac{dx}{dt} = Ax + bu(t) + \eta(t) \tag{2.8}$$

using a discrete-time adaptive controller

$$\begin{cases} u(t) = u_k = \theta_k^{\mathrm{T}} x_k, \quad t_k \leqslant t < t_{k+1} \\ \theta_{k+1} = \theta_k + \alpha_k \Phi(\theta_k, x_k). \end{cases} \tag{2.9}$$

Here $x(t)$ denotes the state vector of the system, $x(k) = x(t_k)$; u_k and θ_k are the values of control and tunable coefficients on a time interval $[t_k, t_{k+1}]$; $\eta(t)$ is a random perturbation such that $\mathrm{E}\eta(t) = 0$; $t_k = kh$, $k = 0, 1, \ldots,$ indicate the time instants when the values u_k and θ_k are corrected. If the time interval $h = t_{k+1} - t_k$ between successive correction time instants is short, the vector $x(t)$ will have a small change during this interval. Therefore, reducing the system's dynamical equation to the discrete form (2.6), we obtain an almost identity matrix $G(z_k)$. The degree of dependence between the vectors x_k and x_l will grow with decreasing h, even in the case of a stable matrix A. System (2.8), (2.9) can be considered singularly perturbed only if α_k are small compared to h. This requirement leads to an unreasonable slowdown in the adaptation process, making the design scheme of a continuous-time model unacceptable.

In the described problems, a continuous-time model of the same order as the original system [14] can be constructed. For example, let the perturbation $\eta(t)$ (2.8) be centered, $\mathrm{E}\eta(t) = 0$, and also let the step α_k in the adaptation algorithm be proportional to the sampling interval h, i.e., $\alpha_k = \overline{\alpha} h$. Then, as a continuous-time model we may adopt the system

$$\begin{cases} \dfrac{dx}{dt} = Ax + bu, \quad u = \theta^{\mathrm{T}} x \\ \dfrac{d\theta}{dt} = \overline{\alpha}\Phi(\theta, x). \end{cases} \tag{2.10}$$

Note that the original initial system (2.8), (2.9) is regularly perturbed with respect to the former system. In this case, no motion separation into the fast and slow components takes place in (2.8), (2.9).

Now, let us formulate the results justifying the applicability of the method of continuous-time models. As it has been already mentioned, these results are divided into two groups, namely, (1) the connection between the asymptotic properties of the original system and its model and (2) the closeness of statistical characteristics of the original and model processes.

3. CLOSENESS OF TRAJECTORIES OF ORIGINAL AND MODEL SYSTEMS

Here are some estimates for the approximation accuracy of the trajectories of the original system (2.1) with the trajectories of the deterministic model (2.3). The first statement refers to the case of independent stationary perturbations f_k; the right-hand side of model (2.3) should satisfy the global Lipschitz condition. Let $b(z) = \mathrm{E}\|F(z, f_k) - A(z)\|^2$. Denote by z_k the solution to system (2.1) and by $z(t_k)$ the solution to system (2.3) with the same initial condition, i.e., $z(t_0) = z_0$.

Theorem 1 [13]. *Let the vectors f_k be independent and identically distributed and also let the following conditions hold:*

$$\|A(z) - A(z')\| < L_1 \|z - z'\|, \tag{3.1}$$

$$b(z) \leqslant L_2 \left(1 + \|z\|^2\right). \tag{3.2}$$

Then there exist values $C_1 > 0$ and $C_2 > 0$ such that, given $0 \leqslant \alpha_k \leqslant \alpha$, $k = 0, 1, \ldots, N-1$,

$$\mathrm{E} \max_{1 \leqslant t_k \leqslant t_N} \|z_k - z(t_k)\|^2 \leqslant C_1 e^{C_2 t_N} \alpha. \tag{3.3}$$

The statements close to Theorem 1 also appeared in [22, 31]. In the fundamental paper [31], the convergence in probability of the trajectories of (2.1) and (2.3) was established; in [22], the probability that the trajectory of (2.1) will leave a ε-neighborhood of the solution $\overline{z}_k$ to the discrete-time deterministic model

$$\overline{z}_{k+1} = \overline{z}_k + \alpha_k A(\overline{z}_k) \tag{3.4}$$

was estimated.

Theorem 1 proves the closeness of the trajectories of (2.1) and (2.3) in the mean-square sense uniformly on a finite time interval. In other words, for a fixed time instant t_N the mean-square distance between the trajectories of the original and model systems has an order of the square root of the maximum step α. For higher values t_N, the accuracy of estimation will rapidly decrease, which is an essential fact. A similar result holds for weakly dependent perturbations f_k. The following definition gives a precise meaning to the concept of weak dependence.

Definition 1 [30]. Let f_k, $k = 0, 1, \ldots,$ be a random process and $\mathcal{F}_k^l$ be a σ-algebra generated by the values f_r, $k \leqslant r \leqslant l$. The process f_k is said to satisfy the strong mixing condition if

$$\sup_k \sup_{\xi, \eta} |\mathrm{E}\xi\eta - \mathrm{E}\xi\mathrm{E}\eta| = \zeta_r \xrightarrow[r \to \infty]{} 0, \tag{3.5}$$

where ξ, is a random variable measurable[2] with respect to $\mathcal{F}_0^k$ and η is a random variable measurable with respect to $\mathcal{F}_{k+r}^\infty$ such that $|\xi| \leqslant 1$ and $|\eta| \leqslant 1$ with probability 1. The function ζ_r is called *the mixing coefficient.*

Theorem 2. *Let the vectors f_k satisfy the strong mixing condition, $\mathrm{E}\|F(z, f_k)\|^8 \leqslant C_R < \infty$ for $\|z\| < R$ and*

$$\sum_{k=1}^{\infty} k \zeta_k^{1/2} < \infty. \tag{3.6}$$

Assume there exists a value $L > 0$ and a domain $\Omega \subset \mathbb{R}^m$ such that $f_R \in \Omega$ with probability 1 and

$$\|F(z, f) - F(z', f)\| \leqslant L \|z - z'\| \tag{3.7}$$

for all $f \in \Omega$.

[2] The measurability of ξ with respect to the σ-algebra $\mathcal{F}_k^l$ means that the values of ξ are completely determined by the values f_r for $k \leqslant r \leqslant l$.

Then there exist $C_3 > 0$ and $C_4 > 0$ such that

$$\mathrm{E}\max_{0\leqslant t_k\leqslant t_N}\|z_k - z(t_k)\|^2 \leqslant C_3 e^{C_4 t_N}\alpha \tag{3.8}$$

for $0 \leqslant \alpha_k \leqslant \alpha$.

The proof of Theorem 2 is given in the Appendix. Note that condition (3.6) holds, e.g., if f_k are generated by Eq. (2.5) with a stable matrix G and a bounded f_k. In this case, the mixing coefficient of the process ζ_k decreases exponentially: $\zeta_k \leqslant C_\zeta q^k$, where $C_\zeta > 0$ and $0 < q < 1$.

If model (2.3) has a unique globally exponentially stable equilibrium state z_*, the closeness of the trajectories can be estimated on the infinite interval. As is well-known [32], a sufficient[3] condition for the global exponential stability of a point z_* is the existence of a twice continuously differentiable function $V(z)$ that satisfies for some positive κ_1, κ_2 and κ_3 the inequalities

$$\dot{V} \leqslant -\kappa_1 V(z), \tag{3.9}$$

$$\|\nabla^2 V(z)\| \leqslant \kappa_2, \qquad V(z) \geqslant \kappa_3\|z - z_*\|^2, \tag{3.10}$$

where $\dot{V}(z) = \nabla V(z)^{\mathrm{T}} A(z)$ denotes the derivative of $V(z)$ along the trajectories of model (2.3) and $\nabla^2 V(z)$ is the Hessian matrix.

Theorem 3. *Assume the conditions of Theorem 1 hold and, moreover, there exists a function $V(z)$ satisfying* (3.9), (3.10). *Then for some $C_5 > 0$ and $\beta > 0$ the following inequalities are true:*

$$\mathrm{E}\|z_k - z(t_k)\|^2 \leqslant C_5\alpha^\beta, \quad k = 0, 1, \ldots. \tag{3.11}$$

A more complicated situation will occur if model (2.3) has several stable equilibria $z_*^{(1)}, \ldots, z_*^{(r)}$. In this case, the trajectories of the original system may have transitions between the domains of attraction of the points $z_*^{(1)}, \ldots, z_*^{(r)}$, which take place "contrary to" the average motion. Such transitions are described by the theorems on large deviations [11, 12]. The probability of transition in finite time is positive (albeit, very small). Therefore, on the infinite time interval transitions will occur with probability 1.

For the sake of simplicity, let $\alpha_k \equiv \alpha > 0$. Using the methods from [12, 30], under mild additional assumptions on the processes z_k and f_k it can be established that, for small α, the transitions of the trajectories of (2.1) between the neighborhoods of $z_*^{(i)}$ are described by a Markov chain with the transition rates $p_{ij} = \exp(-v_{ij}/\alpha)$ (the accuracy is within the logarithmic equivalence). The values $v_{ij} > 0$ are determined by solving a variational problem. Obviously, if $v_{ij} > v_{i'j'}$, $p_{ij} > p_{i'j'} \to 0$ as $\alpha \to 0$. Therefore, for small α the first exit from the neighborhood of $z_*^{(i)}$ with an overwhelming probability will occur into a neighborhood of $z_*^{(i_1)}$, where the number $i_1 = l(i)$ is found from the condition $v_{il(i)} = \min_{k\neq l} v_{ik}$. From the neighborhood of $z_*^{(i_1)}$ the trajectory will evolve into a neighborhood of $z_*^{(i_2)}$, where $i_2 = l(i_1)$, and so on until the states repeat. Thus, all points $z_*^{(i)}$ will be divided into cycles. The transitions between these cycles will occur even rarer than the rare transitions within them. As a result, the cycles of rank 2 will be formed, etc.

4. CONNECTION BETWEEN ASYMPTOTIC PROPERTIES OF ORIGINAL SYSTEM AND ITS DETERMINISTIC MODEL

Consider the asymptotic stability and dissipativity[4] of system (2.1) (in the mean-square sense, or almost surely). The connection between the asymptotic properties of system (2.1) and model (2.3)

[3] And necessary, under some assumptions.

[4] In this paper, the dissipativity of a system is understood in the sense of Levinson, i.e., as the ultimate boundedness and convergence of its trajectories to a bounded set D as $t \to \infty$.

depends on fulfillment of the condition $\alpha_k \to 0$ as $k \to \infty$. If $\alpha_k \not\to 0$ (e.g., $\alpha_k \equiv \alpha$), even with an exponentially stable model (2.3) with a single limit point, generally the original system (2.1) will only be mean-square dissipative.[5] At the same time, in accordance with [34], the limit variance of z_k vanishes as $\alpha \to 0$. If model (2.3) is dissipative (under the hypotheses of Theorem 3, inequality (3.9) is replaced by $\dot V(z) \leqslant -\kappa_1 V(z) + \kappa_0$), for small α system (2.1) will also be dissipative, and the limit variance of z_k will not vanish as $\alpha \to 0$.

The properties of systems (2.1) and (2.3) have a closer connection to each other as $\alpha_k \to 0$. In this case, the trajectories of (2.1) and (2.3), roughly speaking, converge to each other as $k \to \infty$ and system (2.1) turns out to be stable if model (2.3) is stable. We give a precise formulation of this fact in the case where the dynamics of fast motion does not depend on the slow variables.

Theorem 4. *Let there exist a domain $\Omega \subset \mathbb{R}^n$, a twice continuously differentiable function $V(z)$ for $z \in \Omega$, an integer number $p \geqslant 1$ and a number $\varepsilon : 0 < \varepsilon < 1$ that satisfy the following conditions:*

1. *For all $z \in \Omega$, $f \subset R^m$, the function $F(z, f)$ is continuously differentiable in z.*

2. *$V(z) \geqslant 0$, $\dot V(z) \leqslant 0$ for $z \in \Omega$, where $\dot V(z)$ denotes the derivative of the function $V(z)$ along the trajectories of the continuous-time model* (2.3).

3. *For almost all trajectories z_k of system* (2.1), *there exists a sequence $k_s \to \infty$ such that $z_{k_s} \in \overline{\Omega}$, where $\overline{\Omega}$ is some bounded closed subset of Ω.*

4. *Values α_k in* (2.1) *satisfy the conditions*

$$\alpha_k \geqslant \alpha_{k+1}, \quad \sum_{k=0}^{\infty} \alpha_k = \infty, \tag{4.1}$$

$$\sum_{k=0}^{\infty} \alpha_k^{2p} < \infty. \tag{4.2}$$

5. *The sequence f_k is almost surely bounded and satisfies the strong mixing condition, while the mixing coefficient ζ_k satisfies the relation*

$$\sum_{k-1}^{\infty} \alpha^{p-1} \zeta_k^{\varepsilon} < \infty. \tag{4.3}$$

Then the trajectories of (2.1) *with probability* 1 *either converge to a set $\Omega_0 = \{z \in \Omega : \dot V(z) = 0\}$, or have a limit point on the boundary of Ω. The latter option disappears if the set Ω_0 is bounded and closed.*

Note that in the case of the linear system (2.1) with the perturbations f_k described by a Markov chain, the connection between the stability of system (2.1) and its continuous-time model (2.3) was studied in [33]. Some analogs of Theorem 4 for the stochastic approximation-type algorithms and independent f_k were described in [27, 34, 35]. Results similar to Theorem 4 were obtained in [15, 36]; in [15], the case with the equation of fast motions depending on the slow variables was also considered.

5. STOCHASTIC CONTINUOUS-TIME MODELS

Up to this point, we have considered a deterministic continuous-time model of the stochastic system (2.1) described by the ordinary differential Eqs. (2.3). In accordance with estimate (3.3), the deviations between the trajectories of system (2.1) and model (2.3) have an order of $\sqrt{\alpha_k}$, where

[5] Of course, for separate classes of processes, the original system (2.1) can be stable as well; see [33, 34]. For example, (2.1) with a sufficiently small value α is mean-square asymptotically stable if $b(z) \leqslant L_2\|z - z_*\|^2$ (the case of multiplicative perturbations).

α_k is the step parameter of (2.1). These deviations are generated by random fluctuations of the trajectory of (2.1) with respect to its systematic component. It is natural to try to build a more accurate model of the continuous-time system (2.1) that will take into account the random fluctuations. With such a model, we may expect to obtain more accurate quantitative characteristics of the original process that cannot be yielded by deterministic models (e.g., using a deterministic model it is impossible to estimate the steady-state error of the linear adaptive algorithm with a constant step under noisy conditions [13]). The models that take into account the randomness will be called *continuous-time stochastic models or stochastic models.*

A convenient tool to describe and study stochastic continuous-time models is the theory of stochastic differential equations [10]. We will use the stochastic differential equations in the Itô form

$$dz(t) = a(z,t)dt + D(z,t)dW(t), \tag{5.1}$$

where $z(t) \in \mathbb{R}^n$; $a(\cdot)$ is an n-dimensional vector known as the drift coefficient; $D(\cdot)$ is a matrix of dimensions $n \times r$ known as the diffusion coefficient; $dW(t)$ is a stochastic differential (Itô differential) of an n-dimensional Wiener process $W(t)$ with independent components.

As is well-known, under the conditions of smoothness and limited growth of the functions $a(\cdot)$ and $D(\cdot)$, Eq. (5.1) has a unique solution $z(t)$ that is a diffusion Markov process. A process $z(t)$ is determined by its local vector-matrix characteristics: the first and second conditional moments of the increments $\Delta z(t) = z(t+\Delta t) - z(t)$. These characteristics have the form

$$\mathrm{E}\{\Delta z(t)|z(t) = z\} = a(z,t)\Delta t + O(\Delta t^{3/2}), \tag{5.2}$$

$$\mathrm{E}\{\Delta z(t)\Delta z(t)^{\mathrm{T}}|z(t)\} = D(z,t)D(z,t)^{\mathrm{T}}\Delta t + O(\Delta t^{3/2}), \tag{5.3}$$

where $O(\Delta t^3)$ denotes a value tending to zero not slower than Δt^3 as $\Delta t \to 0$. Obviously, the deterministic model (2.3) can be written as (5.1) by choosing $a(z,t) = A(z)$ and $D(z,t) = 0$. Let $\overline{z}(t)$ be the solution to the deterministic model (2.3) with the initial condition $\overline{z}(0) = z_0$. Then the increment $\overline{z}(t)$ in the time $\alpha_k = t_{k+1} - t_k$ for small α_k is close to the conditional mean increase of the original process:

$$\mathrm{E}\{z_{k+1} - z_k|z_k = z\} = \alpha_k \mathrm{E}F(z, f_k) = \alpha_k A(z) \approx z(t_{k+1}) - z(t_k)$$

for $\overline{z}(t_k) = z$. In a similar way,

$$\mathrm{E}\left\{(z_{k+1} - z_k)(z_{k+1} - z_k)^{\mathrm{T}}|z_k = z\right\} \approx [\overline{z}(t_{k+1}) - \overline{z}(t_k)][\overline{z}(t_{k+1}) - z(t_k)]^{\mathrm{T}},$$

and in both cases the approximate equalities are true with an accuracy of order α_k^2. The simplest version of the stochastic model is obtained from (5.1) by choosing

$$a(z,t) = a(t) = A(\overline{z}(t)), \quad D(z,t) = D(t) = \sqrt{\alpha_k} \times \sqrt{B(\overline{z}(t))} \tag{5.4}$$

for $t_k \leqslant t \leqslant t_{k+1}$, where $B(z) = Eh(z, f_k)h(z, f_k)^{\mathrm{T}}$, $h(z, f_k) = F(z, f_k) - A(z)$; $\sqrt{B(\cdot)}$ denotes the square root of a symmetric nonnegative matrix $B(\cdot)$; $\overline{z}(t)$ is the solution to (2.3) with the initial condition $\overline{z}(t_0) = z_0$. Coefficients (5.4) are calculated along the trajectory of the deterministic model. It can be demonstrated that in such a stochastic model, the second conditional moments of the increments of its trajectories are close to the corresponding characteristics of the original system (2.1) with an accuracy of order α_k^3, i.e., with a greater accuracy than for the deterministic model.

Denote by $\tilde{z}(t)$ the solution to Eqs. (5.1), (5.4). Then we can write the approximate equality

$$\tilde{z}(t_{k+1}) - \tilde{z}(t_k) \approx \alpha_k A(\overline{z}(t_k)) + \sqrt{\alpha_k B(\overline{z}(t_k))}\eta_k, \tag{5.5}$$

where $\eta_k = W(t_{k+1}) - W(t_k)$ give the increments of the Wiener process that are independent Gaussian random vectors with the covariance matrix $\mathrm{E}\eta_k\eta_k^{\mathrm{T}} = \alpha_k I_r$. Relation (5.5) is used, e.g., for the statistical modeling of system (5.1), (5.4). From (5.5) it can be observed that the replacement of the original system (2.1) by the stochastic model (5.1), (5.4) leads to a "normalization" of fluctuations: the actual fluctuations are replaced by their Gaussian counterparts with the same covariance matrix.

Despite the simple design procedure, the stochastic model (5.1), (5.4) is inconvenient for applications: for obtaining analytical or numerical estimates of the solutions to (5.1), (5.4), first we have to find the solutions to the deterministic model (2.3). This drawback can be eliminated by choosing

$$a(z,t) = A(z) \quad \text{and} \quad D(z,t) = \sqrt{\alpha_k B(z)} \tag{5.6}$$

for $t_k \leqslant t \leqslant t_{k+1}$. It is easy to show that model (5.1), (5.6) approximates the local characteristics of initial system (2.1) with the same accuracy as model (5.1), (5.4). However, instead of (5.5) the trajectories of (5.1), (5.6) satisfy the relation

$$\tilde{z}(t_{k+1}) - \tilde{z}(t_k) \approx \alpha_k A(\tilde{z}(t_k)) + \sqrt{\alpha_k B(\tilde{z}(t_k))}\eta_k, \tag{5.7}$$

where $\tilde{z}(t_k)$ is the solution to Eq. (5.1), (5.6). If we adopt (5.7), there is no need to find the trajectories of the deterministic model (2.3). With system (2.1) replaced by its stochastic model, the analytical and numerical study becomes easier. This is due to the partial averaging and normalization of the right-hand sides of the original system equations. Some examples of analytical estimation using the stochastic model were presented in [13, 37, 38].

As it has been mentioned, the approximation accuracy of the local characteristics of system (2.1) by the stochastic model (5.1), (5.6) has an order of α_k^2 in the first moments and an order of α_k^3 in the second moments. It turns out that we may construct a family of stochastic models with an arbitrarily high order of approximation. For example, a model of the accuracy with an order of α_k^3, both in the first and second moments, can be obtained by changing the drift coefficient of model (5.1), (5.6). The modified stochastic model is described by the equation

$$d\tilde{z} = \left[I_n - \frac{\alpha_k}{2}\frac{\partial A(\tilde{z})}{\partial \tilde{z}}\right] A(\tilde{z})dt + \sqrt{\alpha_k B(\tilde{z})}dW(t) \tag{5.8}$$

for $t_k \leqslant t < t_{k+1}$. However, note that more accurate models are less convenient to use, since the complexity of model equations is rapidly increasing with their accuracy.

6. METHOD OF AVERAGED MODELS FOR HYBRID SYSTEMS

In modern computer-controlled systems, e.g., cyber physical systems, control or state estimation algorithms of a physical plant may operate in discrete time while the plant itself is operating in continuous time. In this case, the original system is hybrid (discrete-continuous). It seems natural to adopt simplified (averaged) models for the analysis and design of such systems, too. In particular, during the design procedure of a discrete-time adaptive system using its continuous-time simplified model, the problem is to preserve the properties of a continuous-time system during the discretization of the control and adaptation algorithms. One method to solve this problem was developed in [13, 16]. This method can be employed to establish that, if the simplified continuous-time model of the initial system has the property of exponential dissipativity, the original discrete-time system has the limiting dissipativity property as $\epsilon_k \to 0$ and the estimate of the limit set is close to the corresponding estimate of the continuous-time model.

Let a controlled system be described by a continuous-time state-space model of the form

$$\dot{x} = F(x, \theta, t) + \varphi(t), \tag{6.1}$$

where $x \in \mathbb{R}^n$ denotes the state vector; $\theta \in \mathbb{R}^m$ gives the control vector (the vector of tunable parameters); a vector function $F(\cdot)$ is defined for all $x \in \mathbb{R}^n$, $\theta \in R^m$ and $t \geqslant 0$ as well as has the properties of piecewise continuity in t and continuous differentiability with respect to x and θ; finally, $\varphi(t)$ is a bounded vector function of perturbations.

Let the control vector $\theta(t)$ be updated at sampling time instants $t_k = k\epsilon_k$, $k = 0, 1, 2, \ldots$, $\epsilon_k > 0$, as follows:

$$\theta(t) = \begin{cases} \theta(t_k), & t_k \leqslant t < t_{k+1} \\ \theta_k + \epsilon_k \Phi(x_k, \theta_k, t_k), & t = t_{k+1}, \end{cases} \tag{6.2}$$

where $x_k = x(t_k)$ and $\theta_k = \theta(t_k)$. It is assumed that the function $\Phi(\cdot)$ satisfies the same regularity conditions as $F(\cdot)$. Recall that we are interested in analyzing stability or dissipativity (the boundedness of solutions) of system (6.1), (6.2) for small $\epsilon_k > 0$. Hence, it seems natural to choose a rough approximation (simplified) model as the continuous-time one with neglected perturbations:

$$\begin{aligned} \dot{x} &= F(x, \theta, t), \\ \dot{\theta} &= \Phi(x, \theta, t). \end{aligned} \tag{6.3}$$

Then the following question immediately arises: is it possible to judge about the stability or similar properties of system (6.1), (6.2) by the corresponding properties of its simplified model (6.3)? The answer is provided by the theorem [14]; see below and also Theorem 3.13 in [16]. For formulating this result, we should introduce the concepts of exponential dissipativity and ultimate dissipativity as $\epsilon_k \to 0$.

The exponential dissipativity of a system $\dot{z} = F(z, t)$, $z \in \mathbb{R}^n$, is the existence of a smooth function $V(z)$, a vector $z_* \in \mathbb{R}^n$ and values $\zeta > 0$ and $\beta_i > 0$, $i \in 1, 2, 3, 4$, such that

$$\begin{gathered} V(z,t) \leqslant -\zeta V(z,t) + \beta_1, \quad \|\nabla^2 V(z,t)\| \leqslant \beta_2, \\ \beta_3 \|z - z_*\| \leqslant \|\nabla_z V(z,t)\| \leqslant \beta_4 \|z - z_*\|. \end{gathered} \tag{6.4}$$

As is easily seen, under condition (6.4) all trajectories of the system satisfy the inequality $\|z(t)\| \leqslant K_1 e^{-\zeta t} + K_2$ with some K_1 and K_2, thereby converging to a certain bounded set in the state space.

Let $\epsilon = \sup_{k \geqslant 0} \epsilon_k$ and $\sum_{k=1}^{\infty} \epsilon_k = \infty$. System (6.1), (6.2) is called *ultimately dissipative* as $\epsilon \to 0$ if in the phase space X of the system there exist a family of sets $D_0(\epsilon)$ with the property $\cup_{\epsilon > 0} D_0(\epsilon) = X$, a bounded set D_∞ and a value $\epsilon_* > 0$ such that all trajectories evolving from the set $D_0(\epsilon)$ will converge into the set D_∞ as $t \to \infty$.

Theorem 5. *Let $F(x, \theta, t)$ be Lipschitz continuous in x uniformly in x and θ in any bounded set and also let model* (6.3) *be exponentially dissipative. Then the original system* (6.1), (6.2) *is ultimately dissipative as $\epsilon_k \to 0$.*

An advantage of Theorem 5 over the well-known results on the analysis of sampled-data nonlinear systems (see references in [39]) is that the right-hand sides of the system model not necessarily satisfy the global Lipschitz condition.

For a special case in which the limit set is a singleton, in 1990 a refined result was established by Dragan and Khalanai—the conditions under which exponential stability is preserved after discretization [40]. Consider a continuous-time controlled system $\dot{x} = f(x, u)$, $y = s(x)$, where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $y \in \mathbb{R}^l$, with a dynamical controller $\dot{w} = g(w, y)$, $u = U(w, y)$, where $w \in \mathbb{R}^p$ and $p > 0$. The closed loop control system is described the differential equations

$$\dot{x} = f(x, U(w, s(x))), \qquad \dot{w} = g(w, s(x)). \tag{6.5}$$

Also consider the continuous-time system with the discrete controller in the closed loop:

$$\dot{x} = f(x(t), u(t)), \ \ u(t) = u_k, \ \ w(t) = w_k, \ \ t_k \leqslant t < t_{k+1}, \tag{6.6}$$

$$u_k = U(w_k, s(x(t_k))), \ \ w_{k+1} = w_k + hg(w_k, s(x(t_k))), \tag{6.7}$$

where $t_k = kh$, $h > 0$, $j = 0, 1, 2, \ldots$.

Theorem 6 [40]. *Let the functions f, g, s and U be locally Lipschitz and $(\bar{x}, \bar{w})$ be the equilibrium of system (6.5), i.e., $f(\bar{x}, U(\bar{w}, s(\bar{x}))) = 0$, $g(\bar{w}, s(\bar{x})) = 0$. Assume there exists a domain $\mathcal{D}$ containing the point $(\bar{x}, \bar{w})$ and also constants $\alpha > 0$ and $\beta \geqslant 1$ such that, if $(x(0), w(0)) \in \mathcal{D}$, for $t \geqslant 0$ the solution $(x(t), w(t))$ to system (6.5) satisfies the inequality*

$$||x(t) - \bar{x}|| + ||w(t) - \bar{w}|| \leqslant \beta e^{-\alpha t}(||x(0) - \bar{x}|| + ||w(0) - \bar{w}||). \tag{6.8}$$

Let $\mathcal{B}_r = \{(x, w) : ||x - \bar{x}|| + ||w - \bar{w}|| \leqslant r\}$, where $r > 0$ is such that $\mathcal{B}_r \subset \mathcal{D}$. Let $L > 0$ be the Lipschitz constant for all functions f, g, s and U on the compact set $\mathcal{B}_r$.

Then there exists a value $h > 0$ that depends on α, β, L and $\tilde{\alpha} > 0$, $\tilde{\beta} \geqslant 1$, such that $(\tilde{x}(t), \tilde{w}(t))$, $t \geqslant 0$, is the solution to system (6.6), (6.7); moreover, if $(\tilde{x}(0), \tilde{w}(0)) \in \mathcal{B}_{r/2\beta}$,

$$||\tilde{x}(t) - \bar{x}|| + ||\tilde{w}(t) - \bar{w}|| \leqslant \tilde{\beta} e^{-\tilde{\alpha} t}(||\tilde{x}(0) - \bar{x}|| + ||\tilde{w}(0) - \bar{w}||). \tag{6.9}$$

7. CONVERGENCE OF GRADIENT ALGORITHMS WITH DEPENDENT INPUTS

The averaging method can be effectively applied to analyze the differential-difference stochastic systems (6.1), (6.2). We formulate a corresponding result for a practically important case in which Eq. (6.2) is described by the gradient-type algorithm, i.e., takes the form

$$\theta_{k+1} = \theta_k + \alpha_k \nabla_\theta Q(\theta_k, f_k), \tag{7.1}$$

where $Q(\cdot)$ is a goal function that characterizes the achievement of an adaptation goal.

An important condition for standard theorems on the convergence of gradient algorithms [19] is the independence of the input actions x_k. This condition complicates the use of standard theorems in control problems for dynamical systems, because the vectors x_k in system (6.1), (6.2) are stochastically dependent. With the method of continuous-time models, the convergence of gradient algorithms can be established even in the case of dependent actions.

However, completely abandoning the condition of independence is not possible. We have to require that the sequence of inputs f_k satisfy the so-called *strong mixing condition*; see Definition 1 in Section 3.

A wide class of random processes satisfying the strong mixing consists of the Markov processes generated by the stable stochastic differential and difference equations [30, 41]. In particular, this applies to the process $f_k \in \mathbb{R}^n$ generated by the linear Eq. (2.5), and moreover the mixing coefficient of process (2.5) is exponentially decreasing: $\zeta_r \leqslant c\rho^r$, where $c > 0$ and $0 < \rho < 1$.

Theorem 7. *Let the following assumptions hold.*

1. *For each θ, there exists $\lim_{k\to\infty} \mathrm{E}Q(\theta, f_k) = J(\theta)$, where the function $J(\theta)$ is continuously differentiable, $J(\theta) \to +\infty$ as $\|\theta\| \to \infty$ and the set $\Omega_0 = \{\theta : \nabla J(\theta) = 0\}$ is bounded.*

2. *There exist values $L > 0$ and $\kappa > 0$ and a domain D such that $\|\nabla_\theta Q(\theta, f)\| \leqslant \kappa$,*

$$\|\nabla_\theta Q(\theta, f) - \nabla_\theta Q(\theta', f')\| \leqslant L\|\theta - \theta'\| \tag{7.2}$$

for each $f \in D$ and $f' \in D$ with probability 1.

3. *The steps α_k of algorithm (7.1) with some $p \geqslant 2$ satisfy the conditions*

$$\alpha_{k+1} \leqslant \alpha_k, \quad \sum_{k=0}^{\infty} \alpha_k = +\infty, \quad \sum_{k=0}^{\infty} \alpha_k^p < \infty; \tag{7.3}$$

4. *The random process f_k satisfies the strong mixing condition with the mixing coefficient ζ_k and*

$$\sum_{k=0}^{\infty} k^{p-1} \zeta_k^{1/2} < \infty. \tag{7.4}$$

5. *Almost each trajectory θ_k contains a bounded subsequence θ_{k_s}, $s = 1, 2, \ldots$.*

Then $\theta_k \to \Omega_0$ almost surely as $k \to \infty$.

Moreover, it can be shown that if all points of Ω_0 are isolated, $\theta_k \to^{\text{a.s.}} \theta_*$, where θ_* is one of the local minima of the function $J(\theta)$. (Hereinafter, $\to^{\text{a.s.}}$ denotes the almost sure convergence, i.e., the convergence with probability 1.)

The conclusion of Theorem 7 follows from a more general result presented in [16], where the conditions under which the stability of the original stochastic differential-difference Eq. (6.1), (6.2) is guaranteed by the stability of its continuous-time model of the form $\theta^{\cdot} = -\nabla J(\theta)$ with algorithm (7.1) were established. To prove the stability of the model, the Lyapunov function $V(\theta) = J(\theta)$ should be used.

8. FURTHER RESULTS

Let us describe in brief some generalizations of the results from Sections 3 and 4, which will enlarge the applicability of the method of continuous-time models. First of all, note that Theorems 1–4 can be easily extended to the case in which, instead of (2.1), the original system is described by the more general equation

$$z_{k+1} = \Phi(z_k, f_k, \alpha_k), \tag{8.1}$$

where[6] $\mathrm{E}\Phi(z, f_k, \alpha) = z + \alpha A(z) + \alpha^2 a(z, \alpha)$. For example, the statement of Theorem 1 for system (8.1) will remain in force [42] under conditions (3.1), (3.2) and the inequality $\|a(z, \alpha)\| \leqslant L_3(1 + \|z\|)$. Note that for system (8.1) the right-hand side of model (2.3) can be defined by the relation $A(z) = \lim_{\alpha\to 0} \alpha^{-1}[\mathrm{E}\Phi(z, f_k, \alpha) - z]$.

In a number of problems (in particular, in control problems for dynamical systems), there arise systems of type (2.1) or (8.1) for which the right-hand side of model (2.3) does not satisfy the global Lipschitz condition (3.1). An example of such a system is the continuous-discrete system (2.8), (2.9). (Its equations can be transformed to the discrete form (8.1).) The right-hand sides $A(z)$ of the continuous-time model (2.10) contain the terms $\theta^{\mathrm{T}} x$ and, regardless of the type of adaptation algorithm, have a quadratic order of growth as $\|z\| \to \infty$ because $z = \mathrm{col}(x, \theta)$. The Lipschitz

[6] For brevity, only the stationary sequences of f_k will be considered below.

condition (3.1) for model (2.10) is satisfied only locally, i.e., in each bounded domain $\|z\| < r$ the constant L_1 in (3.1) takes a specific value: $L_1 = L_1(r)$. As is well-known, the solutions to differential and difference equations with rapidly (superlinearly) growing right-hand sides can "go to infinity" in a finite time. To exclude this possibility, additional requirements have to be imposed while formulating the statements that justify the method of continuous-time models for locally Lipschitz systems.

However, even under these conditions it turns out that the degree of smallness of the steps α_k with which the original system can be replaced by the model may depend on the value of the initial conditions $\|z_0\|$. This means that the trajectories of (2.1) and (2.3) are close to each other if $\alpha_k \leqslant \alpha < \overline{\alpha}(\|z_0\|)$, where $\overline{\alpha}(r)$ is a decreasing scalar function such that $\overline{\alpha}(r) \to \infty$ as $r \to \infty$. In addition, while studying the asymptotic properties of the original system, we have to consider not the dissipativity of (2.1) but its ultimate dissipativity as $\alpha_k \to 0$.

In the paper [14], it was shown that system (2.1) is ultimately dissipative as $\alpha_k \to 0$; moreover, in the case of independent f_k and locally Lipschitz $A(z)$, inequality (3.3) holds for $\alpha_k \leqslant \alpha < \overline{\alpha}(\|z_0\|)$ if the perturbations in (2.1) are bounded (more precisely, if $\|F(z, f_k) - A(z)\| \leqslant C$) and model (2.3) is exponentially dissipative (more precisely, if there exists a function $V(z) \geqslant 0$ satisfying (3.10) and the inequality $\dot{V}(z) \leqslant -\kappa_1 V(z) + \kappa_0$.) Similar results hold for systems (8.1) as well as for weakly dependent f_k.

In [43, 44], consideration was given to the continuous-discrete systems

$$dx(t) = A_1(x(t), \theta(t_k))dt + \sigma D(x(t), \theta(t_k))dW(t), \tag{8.2}$$

$$\theta(t) = \begin{cases} \theta(t_k) & \text{for} \quad t_k \leqslant t < t_{k+1} \\ \theta(t_k) + hF(x(t_k), \theta(k), \eta_k) & \text{for} \quad t = t_{k+1}, \end{cases} \tag{8.3}$$

where $W(t)$ denotes a Wiener process; η_k are independent; $t_k = kh$, $h > 0$; the function $D(x, \theta)$ is bounded, while the functions $A_1(x, \theta)$ and $A_2(x, \theta) = \mathrm{E}F(x, \theta, \eta_k)$ are globally Lipschitz. The stochastic differential Eq. (8.2) was understood in the Itô sense on each interval $[t_k, t_{k+1}]$. It was demonstrated that for $\sigma^2 \leqslant L_4 h$ the solution $z(t) = \mathrm{col}(x(t), \theta(t))$ to system (8.2), (8.3) satisfies the inequality

$$\mathrm{E}\|z(t) - \hat{z}(t)\|^2 \leqslant C_6 e^{C_7 T} h,$$

where $\hat{z}(t) = \mathrm{col}(\hat{x}(t), \hat{\theta}(t))$ is the solution to the model equations

$$\begin{cases} d\hat{x}/dt = A_1(\hat{x}, \hat{\theta}) \\ d\hat{\theta}/dt = A_2(\hat{x}, \hat{\theta}) \end{cases}$$

given $\hat{z}(0) = z(0)$. In [43], for systems (8.2) and (8.3) a continuous-time stochastic model of form (5.1) and (5.6) was also built.

Finally, the main results of Sections 3–5 remain in force for the time-varying model (2.3) (see [45]) as well as for model (2.3) with a stable manifold of equilibria on which the right-hand side of $A(z)$ has a discontinuity [46].

9. CONTINUOUS-TIME MODELS FOR NETWORKS

In this section, we consider averaged and continuous-time models for networked systems.

New classes of problems such as control of synchronization, consensus, group behavior of agents, etc. for networked control systems have become popular in the two recent decades. One of the important problems is the analysis of multiagent stochastic systems with nonlinear dynamics, uncertainty in the topology and/or in the control protocol, noisy and delayed information about the agents' states. In such networked systems, it is reasonable to consider the approximate consensus problem. In [47], this problem was deeply investigated. Such problems arise in the analysis and control of production networks, multiprocessor, sensor, wireless or multicomputer networks and other systems with very many agents. The stochastic systems of this type can be studied using the method of averaged models. This method was adopted to analyze different classes of information systems (e.g., see [48–50]) and to reduce the analytical complexity of the closed loop stochastic systems. In [47, 51, 52], upper bounds on the mean-square distance between the states of the original stochastic system and its approximate averaged model were derived for different cases (e.g., the cases with or without measurement delays). The upper bounds were used to obtain conditions for reaching an approximate consensus.

In what follows, we will formulate some results on the method of averaged models for networked systems.

Let a network be described by a set of agents $N = \{1, 2, \ldots, n\}$, in which each agent i interacts with his neighbors from the neighbors set $N^i = \{j : (j, i) \in E\}$. The dynamical topology of the network is modeled by a sequence of digraphs $\{(N, E_t)\}_{t\geqslant 0}$, where $E_t \subset E$ evolves with time. Denote by A_t the corresponding adjacency matrices.

Assume the time-varying value $x^i_t \in \mathbb{R}$ reflects the state of each agent $i \in N$ at a time instant $t \in [0, T]$. His dynamics in discrete time obey the equation

$$x^i_{t+1} = x^i_t + f^i(x^i_t, u^i_t), \quad t = 0, 1, 2, \ldots, T. \tag{9.1}$$

To form his control, each agent uses his state (possibly, with measurement noises)

$$y^{i,i}_t = x^i_t + w^{i,i}_t. \tag{9.2}$$

If $N^i_t \neq \emptyset$, each agent also uses the noisy measurements of the neighbors' states, which may have delays:

$$y^{i,j}_t = x^j_{t-d^{i,j}_t} + w^{i,j}_t, \quad j \in N^i_t, \tag{9.3}$$

where $w^{i,i}_t, w^{i,j}_t$ denotes the noise; $0 \leqslant d^{i,j}_t \leqslant \bar{d}$ is an integer-valued delay; $\bar{d} \geqslant 0$ gives the maximum possible delay.

In [47], the following control protocol was considered, which is called *the local voting protocol*:

$$u^i_t = \alpha_t \sum_{j\in \overline{N}^i_t} b^{i,j}_t (y^{i,j}_t - y^{i,i}_t), \tag{9.4}$$

where $\alpha_t > 0$ is the step of the protocol; $b^{i,j}_t > 0 \ \ \forall j \in \bar{N}^i_t$. Let $b^{i,j}_t = 0$ for other pairs (i, j) and define $B_t = [b^{i,j}_t]$ as the control protocol matrix.

First, we study the case without delays in measurements ($\bar{d} = 0$) [53, 54].

The agents' dynamics can be written in the vector-matrix form

$$\bar{x}_{t+1} = \bar{x}_t + F(\alpha_t, \bar{x}_t, \bar{w}_t), \tag{9.5}$$

where $F(\alpha_t, \bar{x}_t, \bar{w}_t)$ is the n-dimensional vector

$$F(\alpha_t, \bar{x}_t, \bar{w}_t) = \begin{pmatrix} \cdots \\ f^i\left(x^i_t, \alpha_t \sum\limits_{j\in \bar{N}^i_t} b^{i,j}_t \left((x^j_t - x^i_t) + (w^{i,j}_t - w^{i,i}_t)\right)\right) \\ \cdots \end{pmatrix}. \tag{9.6}$$

To analyze the behavior of the stochastic system with a specific choice of the coefficients (parameters), the protocol adopts *the method of averaged models* in the form [13], which was also applied in [47, 53–55]. A feature of using this method for networked systems is that the asymptotic stability condition of the continuous-time model (2.3) is violated in the networked systems, thereby being replaced by partial stability conditions. Such cases can be studied with the theorem [17] presented below. Prior to it, consider a relevant definition.

Definition 2. Let Ω, Ω_0 and $\Omega \subseteq \Omega_0$ be closed subsets of $\mathbb{R}^n$ and also let Ω be consisting of the equilibria of system (2.3). The set Ω is said to be Ω_0-pointwise stable if it is Lyapunov stable and any solution evolving from Ω_0 will converge to a point from Ω as $t \to \infty$.

Theorem 8. *Let the Lipschitz and growth conditions*

$$\|A(z) - A(z')\| \leqslant L_1\|z - z'\|, \;\; b(z) \leqslant L_2(1 + \|z\|^2), \tag{9.7}$$

where $b(z) = \mathrm{E}\|F(z, f_k) - A(z)\|^2$, *be satisfied. Assume there exists a smooth mapping* $h : \mathbb{R}^n \to \mathbb{R}^l$ *and a bounded set* $\Omega_0 \subseteq \mathbb{R}^n$ *such that* $\mathrm{rank}\, \frac{\partial h}{\partial z} = l$ *for* $z \in \Omega = \{z \in \Omega_0 : h(z) = 0\}$ *and also assume the set* Ω *is* Ω_0*-pointwise stable. In addition, assume there exist a twice continuously differentiable function* $V(z)$ *and positive values* $\varkappa_1, \varkappa_2$ *and* $\varkappa_3$ *such that*

$$\dot V(z) \leqslant -\varkappa_1 V(z), \tag{9.8}$$

$$\left|\frac{\partial^2 V(z)}{\partial z^{(i)} \partial z^{(j)}}\right| \leqslant \varkappa_3, \quad V(z) \geqslant \varkappa_2||h(z)||^2. \tag{9.9}$$

Then there exist values $\bar\alpha > 0$, $K_2 > 0$ *and* $0 < \zeta < 1$ *such that, for* $0 \leqslant \alpha_k \leqslant \alpha < \bar\alpha$,

$$\mathrm{E}||y_k - y(t_k)||^2 \leqslant K_2\alpha^\zeta, \quad k = 1, 2, \ldots, \tag{9.10}$$

where $y_k = h(z_k), \; y(t_k) = h(z(t_k))$.

Theorem 8 imposes an upper mean-square bound on the distance between the current state and the limit manifold $\Omega = \{z \in \Omega_0 : h(z) = 0\}$.

In the case under consideration, the application of the method of averaged models consists in an approximate replacement of the original stochastic difference Eq. (9.5) by the ordinary differential equation

$$\frac{d\bar x}{d\tau} = R(\alpha, \bar x), \tag{9.11}$$

where

$$R(\alpha, \bar x) = R\begin{pmatrix} & x^1 \\ \alpha, & \vdots \\ & x^n \end{pmatrix} = \begin{pmatrix} \cdots \\ \frac{1}{\alpha} f^i(x^i, \zeta s^i(\bar x)) \\ \cdots \end{pmatrix}, \tag{9.12}$$

$$s^i(\bar x) = \sum_{j \in N^i_{\max}} a^{i,j}_{\max}(x^j - x^i) = -d^i(A_{\max})x^i + \sum_{j=1}^n a^{i,j}_{\max} x^j, \; i \in N,$$

where $A_{\max}$ is the adjacency matrix of dimensions $\bar n \times \bar n$, $\bar n = n \times (\bar d + 1)$.

In accordance with [13], the trajectories $\{\bar x_t\}$ of (9.5), (9.6) and $\{\bar x(\tau_t)\}$ of (9.11), (9.12) are close to each other on a finite time interval. Hereinafter, $\tau_t = \alpha_0 + \alpha_1 + \ldots + \alpha_{t-1}$.

Let the following conditions hold.

B1. $\forall i \in N$ the functions $f^i(x,u)$ are Lipschitz in x and u, $|f^i(x,u)-f^i(x',u')| \leqslant L_1(L_x|x-x'|+|u-u'|)$, and for any fixed x the function $f^i(x,\cdot)$ is such that $\mathrm{E}_x f^i(x,u) = f^i(x, \mathrm{E}_x u)$. The latter part of condition **B1** is satisfied if the system is almost affine in the control variable.

From the Lipschitz condition it follows that the rate of growth is bounded: $|f^i(x,u)|^2 \leqslant L_2(L_c + L_x|x|^2 + |u|^2)$.

B2. **a**) $\forall i \in N$ and $j \in N^i_{\max}$ the noises $w_t^{i,j}$ are centered, independent and have a bounded variance $\mathrm{E}(w_t^{i,j})^2 \leqslant \sigma_w^2$.

b) $\forall i \in N$ and $j \in N^i_{\max}$ variable arcs (j,i) in the graph $\mathcal{G}_{A_t}$ occur in a random and independent way.

c) $\forall i \in N$ and $j \in N^i_{\max}$ the weights $b_t^{i,j}$ in protocol (9.4) are independent random variables with $b^{i,j} = \mathrm{E} b_t^{i,j}$, $\sigma_b^{i,j} = E(b_t^{i,j} - b^{i,j})^2 < \infty$.

The next theorem gives upper bounds on the mean-square distance between the original system and its averaged continuous-time model.

Theorem 9. *Let conditions* **B1** *and* **B2a**–**B2c** *hold and also* $\forall i \in N$ *let the function* $f^i(x,u)$ *be smooth in* u, $f^i(x,0) = 0$ *for any* x *and* $0 < \zeta_t \leqslant \bar{\zeta}$. *Then there exists a value* $\tilde{\alpha}$ *such that, for* $\bar{\alpha} < \tilde{\alpha}$,

$$\mathrm{E} \max_{0 \leqslant \tau_t \leqslant \tau_{\max}} ||\bar{x}_t - \bar{x}(\tau_t)||^2 \leqslant C_1 e^{C_2 \tau_{\max}} \bar{\alpha}, \tag{9.13}$$

where $C_1 > 0$, $C_2 > 0$ *and* $\bar{\alpha} > 0$ *are some constants.*

In [47, 51], the case with measurement delays was also considered and upper estimates of the mean-square distance between the original system and its averaged discrete-time model were derived. Note that a still open problem is to extend the presented results to the discontinuous models that arise in economic games and pattern recognition; some special cases were studied in [46, 56].

10. APPLICATIONS OF CONTINUOUS-TIME MODELS

Continuous-time models were adopted to analyze algorithms used for identification [13, 16, 21, 36, 57–65], optimization [16, 21, 66–68], random search [37, 42], filtering [57, 58, 69–71], adaptive control [14, 16, 40, 43, 44, 63, 72–80], calculation of eigenvalues for random matrices [81, 82]; calculation of saddle points in games [22, 83]; decentralized resource allocation [84], automata games [22], queueing systems [85], image recognition [13, 15, 46, 63]; learning of neural networks [86], etc. A series of recent publications have opened up new areas connected with networks: convergence analysis of learning algorithms to control the coverage of mobile sensing agents in stochastic networks [87], distributed learning and cooperative control in multiagent systems [88, 89], distributed topology control in wireless networks [90], etc.

Another class of problems is associated with the randomization of systems. In adaptive control, while "tuning" to minimize a certain loss function $f(x)$, there is an additional opportunity to compensate the systematic noises of observations (not only the "good statistical ones") using randomization in the feedback channel [91–93]. In randomized versions of the stochastic approximation method, the current estimate $\hat{\theta}_{n-1}$ is supplied to the input of the algorithm along with some centered independent randomized perturbation Δ_n, and $\bar{Y}_n = \Delta_n G(w_n, \hat{\theta}_{n-1} + \Delta_n)$ is taken as the observations. In [94], also the consistency of such estimates and their asymptotic properties were analyzed using the described approach.

11. EXAMPLES

Here are some examples that illustrate the use of continuous-time models.

Example 1. Consider the identification problem of a linear dynamical system described by the equation

$$y_{k+1} + \sum_{i=1}^{n} a_i y_{k+1-i} = \sum_{i=1}^{n} b_i u_{k+1-i} + v_{k-i}, \tag{11.1}$$

where u_k and y_k denote the system's input and output that are measured at a time instant k; v_{k+1} is an unmeasurable perturbation. All values in (11.1) are scalar. The sequences u_k and v_k, $k = 0, 1, \ldots$, will be considered random and stationary. In addition, assume system (11.1) is stable, i.e., all zeros of the polynomial $a(\lambda) = \lambda^n + a_1\lambda^{n-1} + \ldots + a_n$ are greater than 1 by absolute value. Let the estimates $\widehat{a}_{i,k}$ and $\widehat{b}_{i,k}$ of the system parameters a_i and b_i be constructed using the gradient-type algorithm [95]:

$$\begin{aligned} \widehat{a}_{i,k+1} &= \widehat{a}_{i,k} - \alpha_k \delta_{k+1} y_{k+1-i}, \quad i = 1, \ldots n, \\ \widehat{b}_{i,k+1} &= \widehat{b}_{i,k} - \alpha_k \delta_{k+1} u_{k+1-i}, \quad i = 1, \ldots n, \\ \delta_{k+1} &= y_{k+1} + \sum_{i}^{n} \widehat{a}_{i,k} y_{k+1-i} - \sum_{i}^{n} \widehat{b}_{i,k} u_{k+1-i}, \end{aligned} \tag{11.2}$$

where $\alpha_k \geqslant 0$, $k = 0, 1 \ldots$. To study system (11.1), (11.2), we will construct its continuous-time deterministic model. With the notations

$$\begin{gathered} z_k = col(\widehat{a}_{1,k}, \ldots, \widehat{a}_{n,k}, -\widehat{b}_{1,k}, \ldots, -\widehat{b}_{n,k}), \\ z_* = col(a_1, \ldots, a_n, -b_1, \ldots, b_n), \\ g_k = col(y_k, \ldots, y_{k+1-n}, u_k, \ldots, u_{k+1-n}), \end{gathered}$$

algorithm (11.2) can be written as

$$z_{k+1} = z_k + \alpha_k \left[(z_k - z_*)^{\mathrm{T}} g_k + v_{k+1} \right] g_k. \tag{11.3}$$

Performing the averaging procedure of the right-hand side of (11.3) with a fixed vector $z_k = z$, we obtain the deterministic model (2.3) described by the linear differential equation

$$dz/dt = -P(z - z_*) + q, \tag{11.4}$$

where $P = \lim_{k\to\infty} g_k g_k^{\mathrm{T}}$ and $q = \lim_{k\to\infty} v_{k+1} g_k$. These limits exist due to the stability of system (11.1). Obviously, any solution to (11.4) tends to z_* as $t \to \infty$ if

$$P > 0, \quad q = 0. \tag{11.5}$$

From this point onwards, assume $\mathrm{E}v_k = 0$, $\mathrm{E}v_k^2 > 0$, $\mathrm{E}v_k^8 < \infty$, $\mathrm{E}u_k^8 < \infty$ and the sequences u_k and v_k consist of independent random variables and are mutually independent. Then relations (11.5) hold, and the vectors $f_k = \mathrm{col}(g_k, v_{k+1})$ satisfy the hypotheses of Theorem 2 (due to the stability of system (11.1)). It follows from Theorem 2 that, for small α_k, the vectors z_k with a high probability will be close in terms of the mean-square distance to the vectors $z(t_k) = z_* + e^{-p\sum_{i=0}^{k-1}\alpha_i}(z_0 - z_*)$ uniformly on any finite interval $[0, t_N]$. Hence, a variety of approximate estimates for the convergence rate of algorithm (11.2) can be obtained, and their accuracy will be higher for smaller "steps" α_k.

Now, let $\alpha_k \to 0$ so that conditions (4.1), (4.2) are satisfied while the vectors u_k and v_k are almost surely bounded. Then the hypotheses of Theorem 4 hold; in accordance with this theorem, $z_k \to z_*$ almost surely, i.e., the estimates z_k are strongly consistent.

Example 2. Consider the adaptive control problem for a continuous-time system with the transfer function $W(p) = \kappa/(\tau p + 1)$. This system is described by the equation

$$\tau\dot{y} + y = \kappa u. \tag{11.6}$$

The control action $u(t)$ is generated by a discrete-time controller and is considered piecewise constant, i.e., $u(t) = u_k$ for $kh \leqslant t < (k+1)h$, where $h > 0$ denotes a sampling interval. For a high quality of control, the value h should be chosen as small as possible. The behavior of the system with sufficient accuracy can be described by the difference equations

$$y_{k+1} = (1 - h/\tau)y_k + (\kappa h/\tau)u_k, \tag{11.7}$$

where $y_k = (kh)$. Assume the goal of control is to track a given reference signal $r_k = r(kh)$ by the system output y_{k+1}, where $r(t)$ is some bounded piecewise continuous function. As is easily seen, the goal is achieved in one step if the control algorithm has the form

$$u_k^* = \frac{h-\tau}{\kappa h}y_k + \frac{\tau}{\kappa h}r_k. \tag{11.8}$$

However, algorithm (11.8) with decreasing h evidently leads to unlimited values of the control action, which is unacceptable in practice. The reason is a too "strong" choice of the objective condition, which neglects the inertia of the real continuous-time system (11.6).

Let the goal of control be "softer":

$$y_{k+1} = g_0 y_k + g_1 r_k, \tag{11.9}$$

where the coefficients g_0 and g_1 are tuned so that Eq. (11.9) results from the discretization of the dynamical equation of the continuous-time reference model

$$\tau_*\dot{y} + y = r(t). \tag{11.10}$$

The above requirement makes the coefficients g_0 and g_1 dependent on the sampling interval h; for h, this dependence is described by the relation

$$g_0 = 1 - h/\tau_*, \quad g_1 = h/\tau_*. \tag{11.11}$$

The "perfect" control algorithm that achieves goal (11.9) has the form

$$u_k^* = \frac{\tau_* - \tau}{\kappa\tau_*}y_k + \frac{\tau}{\kappa\tau_*}r_k. \tag{11.12}$$

As a matter of fact, this algorithm is independent of h. For adaptive controller design, the direct approach [96] will be used. Replacing the unknown coefficients in (11.12) by the tunable coefficients θ_1 and θ_2, we choose the real control algorithm

$$u_k = \theta_{1k}y_k + \theta_{2k}r_k. \tag{11.13}$$

With the notation $\theta_k = col(\theta_{1k}, \theta_{2k})$ and $z_k = col(y_k, r_k)$, algorithm (11.13) can be written as

$$u_k = \theta_k^{\mathrm{T}} z_k. \tag{11.14}$$

To tune the vector θ_k, we choose the gradient-type algorithm [96]

$$\theta_{k+1} = \theta_k - \alpha_k\delta_{k+1}z_k, \tag{11.15}$$

where $\delta_{k+1} = y_{k+1} - g_0 y_k - g_1 r_k$ is the current residual in the goal of control (11.9). From (11.11) it follows that, for small h, $\delta_{k+1} = y_{k+1} - y_k + h(y_k(\tau_* - r_k/\tau_*)) \approx (\tau_* \dot{y} + y - r(t))h/\tau_*$. Now, we may pass to continuous time, constructing a continuous-time model of system (11.7), (11.14), (11.15) in a similar way as (2.10). Assuming $\alpha_k = \alpha h$ for simplicity, we write the following equation of this continuous-time deterministic model:

$$\begin{aligned} \tau\dot{y} + y &= \kappa u, \quad u = \theta_1 y + \theta_2 r, \\ d\theta_1/dt &= -\alpha(\tau_*\dot{y} + y - r(t))y/\tau_*, \\ d\theta_2/dt &= -\alpha(\tau_*\dot{y} + y - r(t))r(t)/\tau_*. \end{aligned} \tag{11.16}$$

Equations (11.16) describe a continuous-time adaptive system whose adaptation algorithm is the same as in [97]. As was demonstrated in [97], the goal of control $\lim_{t\to\infty} |y(t) - r(t)| = 0$ is achieved in system (11.16).

Note that model (11.16) is on the border of stability; for judging the properties of the original system (11.7), (11.14), (11.15) by the properties of model (11.16), the adaptation algorithm has to be regularized (roughened), e.g., by introducing a negative feedback [98]. After a regularization process, system (11.16) becomes exponentially dissipative. Hence, it follows (see Section 6) that the original system (11.7), (11.14), (11.15) will be ultimately dissipative as $h \to 0$, and its trajectories for small h will be close to the trajectories of (11.16). A similar result will hold for the system affected by weakly bounded perturbations with small variance [16, 44].

A disadvantage of system (11.16) is the need to measure the highest derivative of the system's output. For a high order of the system, similar arguments allow to establish that the adaptive controller (11.16) is nonimplementable in practice. This conclusion remains valid for the original adaptive controller (11.14), (11.15) with a small sampling interval h.

This example shows that the method of continuous-time models can be used for making recommendations on adaptive control design, simplifying the analysis of its dynamics and also for identifying some features of the original system that stay in the background while the system is viewed as a purely discrete one.

12. CONCLUSIONS

The method of (continuous- and discrete-time) averaged models discussed in this paper can be used for the analysis and design of discrete-time stochastic systems. A natural field of its application is the systems with separate fast and slow motions. For the problems of analysis, the method of averaged models allows simplifying an original system, both in analytical and numerical study. For the design problem of a discrete-time system with given properties, the method of continuous-time models may employ a rich arsenal of design tools for continuous-time systems. In both cases, the resulting solutions are approximate but their accuracy grows with decreasing the sampling interval. It may seem that as the power of computing devices is rapidly increasing, the relevance of the simplified methods to study dynamical systems will drop. However, the complexity of modern problems faced by researchers and engineers is often increasing even faster. From the authors' view, taking into account the ubiquitous distribution of networked systems and growing dimensions of the dynamical systems under study, the relevance of approaches based on simplifications will not fall but raise.

Note that the method of continuous-time models has long been used in engineering practice at the first (preliminary) stages of design. The results obtained so far (see Sections 3–6) well justify this method for the analysis and design of complex nonlinear stochastic discrete-time systems, in particular, discrete-time adaptive control systems. The results presented in this survey cover a wide class of systems arising in applications.

However, the method of averaged models is an approximate asymptotic method. Therefore, the theorems proving its applicability should be understood, in accordance with Ljung [15], not in their literal sense but rather as a "moral support" for replacing an original system by its deterministic or stochastic continuous-time model.

APPENDIX

Proof of Theorem 2. The proof will be preceded by two preliminary results as follows.

Lemma A.1. *Let $\xi_k, k = 1, 2, \ldots,$ be a random process with values in $\mathbb{R}^n$, that satisfies the strong mixing condition with a mixing coefficient ζ_k, and*

$$\mathrm{E}\xi_k = 0, \quad \mathrm{E}\|\xi_k\|^{2p+\delta} \leqslant C < \infty, \tag{A.1}$$

$$\sum_{k=1}^{\infty} k^{p-1}\zeta_k^{\delta/(2p+\delta)} < \infty \tag{A.2}$$

for some $C > 0$, $\delta > 0$ and an integer number $p \geqslant 2$.

Then there exists a value $C_p > 0$ such that, for any integer number $N \geqslant 0$ and any numerical sequence $\alpha_1, \ldots, \alpha_N$, the following inequality holds:

$$\mathrm{E}\left\{\max_{1\leqslant k\leqslant N}\left\|\sum_{i=1}^{k}\alpha_i\xi_i\right\|^2 \leqslant C_p\left(\sum_{k=1}^{N}\alpha_k^2\right)^p\right\}. \tag{A.3}$$

Corollary. *Under the hypotheses of Lemma A.1,*

$$\mathrm{E}\left\{\max_{1\leqslant k\leqslant N}\left\|\sum_{i=1}^{k}\alpha_i\xi_i\right\|^2 \leqslant \sqrt{C_2}\sum_{k=1}^{N}\alpha_k^2\right\}. \tag{A.4}$$

Lemma A.2 (discrete version of the Bellman–Grönwall inequality). *Let a sequence of values $\mu_k \geqslant 0$, $k = 1, \ldots, N$, satisfy the inequalities $\mu_k \leqslant \rho_1 + \rho_2\sum_{i=0}^{k-1}\alpha_i\mu_i$, $k = 1, \ldots, N$, where $\rho_1 \geqslant 0$, $\rho_2 \geqslant 0$ and $\alpha_i \geqslant 0$. Then*

$$\mu_k \leqslant \rho_1 \exp\left(\rho_2\sum_{i=0}^{k-1}\alpha_i\right), \quad k = 1, \ldots, N. \tag{A.5}$$

For proving this theorem, first we estimate the value $||z_k - \bar{z}_k||^2$, where $\bar{z}_k$ is the solution to the discrete-time deterministic model (3.4) with $\bar{z}_0 = z_0$. From (2.1), (3.4) we have

$$z_k - \bar{z}_k = \sum_{i=0}^{k-1}\alpha_i[F(z_i, f_i) - F(\bar{z}_i, f_i)] + \sum_{i=0}^{k-1}\alpha_i[F(\bar{z}_i, f_i) - A(\bar{z}_i)]. \tag{A.6}$$

Let $\mu_k = \max_{1\leqslant i\leqslant k}||z_i - \bar{z}_i||$. Calculating the maximum over $i = 1, \ldots, k$, first in the right- and then in the left-hand sides of (A.6), due to condition (3.7) we obtain

$$\mu_k \leqslant L\sum_{i=0}^{k-1}\alpha_i\mu_i + \max_{0\leqslant k\leqslant N-1}\left\|\sum_{i=0}^{k-1}\alpha_i[F(\bar{z}_i, f_i) - A(\bar{z}_i)]\right\|. \tag{A.7}$$

Applying Lemma (A.2), raising both sides of (A.7) to the square and passing to the expectations, we get

$$\mathrm{E}\mu_N^2 \leqslant e^{2Lt_N}\kappa_N, \tag{A.8}$$

where $\kappa_N = \mathrm{E}\max_{1\leqslant i\leqslant k}\left\|\sum_{i=0}^{k-1}\alpha_i[F(\bar{z}_i, f_i) - A(\bar{z}_i)]\right\|^2$. For estimating the value κ_N, we note that the random process $\varphi_i = F(\bar{z}_i, f_i) - A(\bar{z}_i)$ satisfies the strong mixing condition with the same mixing coefficient as the process f_i [30]. Since for a finite t_N the vectors $\bar{z}_1, \ldots, \bar{z}_N$ are bounded in the aggregate, the hypotheses of Lemma A.1 are satisfied for $\xi_i = \varphi_i$, $p = 2$ and $\delta = 4$. From (A.4) we get

$$\kappa_N \leqslant \sqrt{C_2}\sum_{k=0}^{N-1}\alpha_k^2 \leqslant \sqrt{C_2}t_N\alpha. \tag{A.9}$$

Now, we observe that

$$\mathrm{E}\left\{\max_{0\leqslant t_k\leqslant t_N}||z_k - z(t_k)||^2\right\} \leqslant 2\mathrm{E}\mu_N^2 + 2\nu_N^2, \tag{A.10}$$

where $\nu_N = \max_{1\leqslant k\leqslant N}||\bar{z}_k - z(t_k)||$. The value ν_N can be considered the error of solving Eq. (2.3) by the Euler method on the interval $[0, t_N]$. For ν_N, standard calculations yield the estimate

$$\nu_N \leqslant ||A(z_0)||e^{2Lt_N}\left[e^{L\alpha} - 1\right]L^{-1} \tag{A.11}$$

(see, for example, [13]). Comparing (A.8)–(A.11), we finally deduce inequality (3.8). The proof of Theorem 2 is complete.

Remark. Using the results of [99], it can be demonstrated that inequality (3.8) with α replaced by $\alpha^{1-\varepsilon}$ with an arbitrary value $\varepsilon > 0$ will hold if only the fourth moments of the values $F(\bar{z}_i, f_i)$ are assumed to be bounded in the hypotheses of Theorem 2.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research, projects nos. 17-08-01728, 19-03-00375. The results on the analysis of continuous-discrete and networked systems in Sections 8–11 were obtained at the Institute for Problems in Mechanical Engineering, Russian Academy of Sciences, under the support of the Russian Science Foundation, project no. 16-19-00057-P.

REFERENCES

1. Bogolyubov, N.N. and Mitropol'skii, Yu.A., *Asimptoticheskie metody v teorii nelineinykh kolebanii* (Asymptotical Methods in Theory of Nonlinear Oscillations), Moscow: Gostekhizdat, 1955.
2. Volosov, V.M., Averaging in Systems of Ordinary Differential Equations, *Russ. Math. Surveys*, 1962, vol. 17, no. 6, pp. 1–126.
3. Mishchenko, E.F. and Rozov, N.Kh., *Differentsial'nye uravneniya s malym parametrom i relaksatsionye kolebaniya* (Differential Equations with Small Parameter and Relaxation Oscillations, Moscow: Nauka, 1975, vol. 1.
4. Gerashchenko, E.I. and Gerashchenko, S.M., *Metod razdeleniya dvizhenii i optimizatsiya nelineinykh sistem* (The Method of Motions Separation and Optimization of Nonlinear Systems), Moscow: Nauka, 1975.
5. O'Malley, R., *Introduction to Singular Perturbations*, New York: Academic, 1974.
6. Kokotovic, P.V., O'Malley, R., and Sannuti, P., Singular Perturbations and Order Reduction in Control Theory—An Overview, *Automatica*, 1976, vol. 12, no. 2. P. 123–132.
7. Young, K.-K., Kokotovic, P., and Utkin, V., A Singular Perturbation Analysis of High-Gain Feedback Systems, *IEEE Trans. Autom. Control*, 1977, vol. 22, no. 6, pp. 931–938.

8. Khas'minskii, R.Z., On Random Processes Defined by Differential Equations with Small Parameter, *Teor. Veroyatn. Primen.*, 1966, vol. 11, no. 2, pp. 240–259.

9. Khas'minskii, R.Z., On the Averaging Principle for Itô Stochastic Differential Equations, *Kibern.*, 1968, vol. 4, no. 3, pp. 260–279.

10. Gikhman, I.I. and Skorokhod, A.V., *Stokhasticheskie differentsial'nye uravneniya* (Stochastic Differential Equations), Kiev: Naukova Dumka, 1968.

11. Freidlin, M.I., The Averaging Principle and Theorems on Large Deviations, *Russ. Math. Surveys*, 1978, vol. 33, no. 5, pp. 117–176.

12. Wentzell, A.D. and Freidlin, M.I., *Fluktuatsii v dinamicheskikh sistemakh pod deistviem malykh sluchainykh vozmushchenii* (Fluctuations in Dynamical Systems Effected by Small Random Perturbations), Moscow: Nauka, 1979.

13. Derevitskii, D.P. and Fradkov, A.L., Two Models Analyzing the Dynamics of Adaptation Algorithms, *Autom. Remote Control*, 1974, vol. 35, no. 1, pp. 67–75.

14. Derevitskii, D. and Fradkov, A., Investigation of Discrete-Time Adaptive Control Systems Using Continuous-Time Models, *Izv. Akad. Nauk SSSR, Tekhn. Kibern.*, 1975, vol. 5, pp. 93–99.

15. Ljung, L., Analysis of Recursive Stochastic Algorithms, *IEEE Trans. Autom. Control*, 1977, vol. 22, no. 4, pp. 551–575.

16. Derevitskii, D.P. and Fradkov, A.L., *Prikladnaya teoriya diskretnykh adaptivnykh sistem upravleniya* (Applied Theory of Discrete-Time Adaptive Control Systems), Moscow: Nauka, 1981.

17. Fradkov, A., Continuous-Time Averaged Models of Discrete-Time Stochastic Systems: Survey and Open Problems, *Proc. 2011 50 IEEE Conf. on Decision and Control and Eur. Control Conf. (CDC-ECC)*, Orlando, Florida, USA, 2011, pp. 2076–2081.

18. Fradkov, A.L., Averaged Continuous-Time Models in Identification and Control, *Proc. Eur. Control Conf.*, Strasbourg, 2014, pp. 2822–2826.

19. Polyak, B.T., *Vvedenie v optimizatsiyu*, Moscow: Nauka, 1983. Translated under the title *Introduction to Optimization*, Translations Series in Mathematics and Engineering, New York: Optimization Software, 1987.

20. Robbins, H. and Monro, S., A Stochastic Approximation Method, *Ann. Math. Stat.*, 1951, vol. 22, pp. 400–407.

21. Kushner, H.J. and Clark, D.S., *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, New York: Springer-Verlag, 2012.

22. Meerkov, S.M., On Simplification of the Description of Slow Markovian Roaming, *Autom. Remote Control*, 1972, vol. 33, no. 3, pp. 404–414.

23. *Informatsionnye materialy: Kibernetika, tom 68* (Information Materials: Cybernetics, vol. 68), 1973.

24. Ljung, L., Convergence of Recursive Stochastic Algorithms, *Proc. IFAC Sympos. on Stochastic Control*, 1974, pp. 551–575.

25. Basar, E.T., *Control Theory: Twenty-Five Seminal Papers (1932–1981)*, New York: Wiley–IEEE Press, 2001.

26. Gerencsér, L., A Representation Theorem for the Error of Recursive Estimators, *SIAM J. Control Optim.*, 2006, vol. 44, no. 6, pp. 2123–2188.

27. Polyak, B.T. and Tsypkin, Ya.Z., Pseudogradient Algorithms of Adaptation and Learning, *Autom. Remote Control*, 1973, vol. 34, no. 3, pp. 45–68.

28. Nevel'son, M.B. and Khas'minskii, R.Z., Continuous Procedures of Stochastic Approximation, *Probl. Inf. Transm.*, 1971, vol. 7, no. 2, pp. 139–148.

29. Khas'minskii, R.Z., The Averaging Principle for Stochastic Differential Equations, *Probl. Inf. Transm.*, 1968, vol. 4, no. 2, pp. 68–69.

30. Ibragimov, I.A. and Linnik, Yu.V., *Nezavisimye i statsionarno svyazannye velichiny* (Independent and Stationary-Bound Variables), Moscow: Nauka, 1965.

31. Bernstein, S.N., *Stokhasticheskie uravneniya v konechnykh raznostyakh i stokhasticheskie differentsial'nye uravneniya* (Stochastic Finite-Difference Equations and Stochastic Differential Equations), Moscow: Akad. Nauk SSSR, 1964, vol. 4, pp. 484–542.

32. Krasovskii, N., *Stability of Motion*, Stanford: Stanford Univ. Press, 1963.

33. Akhmetkaliev, T., On the Relationship between the Stability of Stochastic Difference and Differential Equations, *Differ. Uravn.*, 1965, no. 8, pp. 1016–1026.

34. Polyak, B.T., Convergence and Rate of Convergence of Recursive Stochastic Algorithms. I, *Autom. Remote Control*, 1976, vol. 37, pp. 537–542.

35. Nevel'son, M.B. and Khas'minskii, R.Z., *Stokhasticheskaya approksimatsiya i rekurrentnoe otsenivanie* (Stochastic Approximation and Recursive Estimation), Moscow: Nauka, 1972.

36. Kushner, H., Convergence of Recursive Adaptive and Identification Procedures via Weak Convergence Theory, *IEEE Trans. Autom. Control*, 1977, vol. 22, no. 6, pp. 921–930.

37. Derevitskii, D. and Fradkov, A., Application of the Theory of Markov Processes to Analyze the Dynamics of Adaptation Algorithms, *Avtom. Vychisl. Tekh.*, 1974, no. 2, pp. 39–48.

38. Derevitskii, D. and Fradkov, A., Analysis of the Dynamics of Some Adaptation Algorithms, in *Voprosy kibernetiki. Adaptivnye sistemy* (Problems of Cybernetics. Adaptive Systems), Moscow: Akad. Nauk SSSR, 1974, pp. 79–84.

39. Nešić, D., Teel, A.R., and Kokotović, P., Sufficient Conditions for Stabilization of Sampled-Data Nonlinear Systems via Discrete-Time Approximations, *Syst. Control Lett.*, 1999, vol. 38, no. 4, pp. 259–270.

40. Dragan, V. and Khalanai, A., Preserving Exponential Stability In Discrete-Time Control Systems with Adaptive Stabilization, *Sib. Math. J.*, 1990, vol. 31, no. 6, pp. 1046–1050.

41. Braverman, E. and Pyatnitskii, Yu.S., Passing of Random Signal Through Absolutely Stable Systems, *Autom. Remote Control*, 1971, vol. 32, no. 2, pp. 202–206.

42. Derevitskii, D.P., Ripa, K.K., and Fradkov, A.L., Investigation of the Dynamics of Certain Random-search Algorithms, *Probl. Sluchainogo Poiska*, 1975, no. 4, pp. 32–47.

43. Andrievskii, B.R., Blajkin, A.T., Derevitskii, D.P., and Fradkov, A.L., A Method to Investigate the Dynamics of Digital Adaptive Control Systems for Aerial Vehicles, in *Upravlenie v prostranstve* (Spatial Control), Moscow: Nauka, 1976, vol. 1, pp. 149–153.

44. Derevitskii, D.P., Synthesis of a Discrete Stochastic Adaptive Stabilization System Using a Continuous-Time Model, *Avtom. Vychisl. Tekh.*, 1975, no. 6, pp. 50–52.

45. Andrievskii, B.R., Blajkin, A.T., Derevitskii, D.P., and Fradkov, A.L., A Synthesis Method of Discrete-Time Adaptive Stabilization Systems for a Stochastic Plant, *Referaty dokladov VII Vsesoyuznogo soveshchaniya po problemam upravleniya* (Abstracts of papers of the VII All-Union Meeting on Control Problems), 1977, pp. 102–105.

46. Levitan, V.D., Analysis of the Dynamics of Discrete-Time Adaptation Processes with a Discontinuous Stochastic Model, in *Voprosy kibernetiki. Adaptivnye sistemy upravleniya* (Problems of Cybernetics. Adaptive Control Systems), Moscow: Akad. Nauk SSSR, 1977, pp. 127–129.

47. Amelina, N., Fradkov, A., Jiang, Y., and Vergados, D.J., Approximate Consensus in Stochastic Networks with Application to Load Balancing, *IEEE Trans. Inform. Theory*, 2015, vol. 61, no. 4, pp. 1739–1752.

48. Pezeshki-Esfahani, H. and Heunis, A.J., Strong Diffusion Approximations for Recursive Stochastic Algorithms, *IEEE Trans. Inform. Theory*, 1997, vol. 43, no. 2, pp. 512–523.

49. Weiss, A. and Mitra, D., Digital Adaptive Filters: Conditions for Convergence, Rates of Convergence, Effects of Noise and Errors Arising from the Implementation, *IEEE Trans. Inform. Theory*, 1979, vol. 25, no. 6, pp. 637–652.

50. Benveniste, A., Métivier, M., and Priouret, P., *Adaptive Algorithms and Stochastic Approximations*, Berlin: Springer-Verlag, 2012.

51. Amelina, N.O. and Fradkov, A.L., Approximate Consensus in the Dynamic Stochastic Network with Incomplete Information and Measurement Delays, *Autom. Remote Control*, 2012, vol. 73, no. 11, pp. 1765–1783.

52. Amelina, N. and Fradkov, A., Approximate Consensus in Multi-Agent Nonlinear Stochastic Systems, *Proc. Europ. Control Conf. (ECC)*, 2014, pp. 2833–2838.

53. Amelina, N., Fradkov, A., and Amelin, K., Approximate Consensus in Multi-Agent Stochastic Systems with Switched Topology and Noise, *Proc. IEEE Multiconf. on Systems and Control (MSC'2012)*, Dubrovnik, Croatia, 2012, pp. 445–450.

54. Amelin, K., Amelina, N., Granichin, O., and Granichina, O., Multi-Agent Stochastic Systems with Switched Topology and Noise, *Proc. 13 ACIS Int. Conf. on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Kyoto, Japan, 2012, pp. 438–443.

55. Amelina, N., Granichin, O., and Kornivetc, A., Local Voting Protocol in Decentralized Load Balancing Problem with Switched Topology, Noise, and Delays, *Proc. IEEE 52 Ann. Conf. on Decision and Control (CDC)*, 2013, pp. 4613–4618.

56. Gorodeisky, Z., Deterministic Approximation of Best-Response Dynamics for the Matching Pennies Game, *Game. Econom. Behav.*, 2009, vol. 66, no. 1, pp. 191–201.

57. Benveniste, A., Design of Adaptive Algorithms for the Tracking of Time-Varying Systems, *Int. J. Adapt. Control Signal Process.*, 1987, vol. 1, no. 1, pp. 3–29.

58. Benveniste, A., Priouret, P., and Metivier, M., *Algorithmes adaptatifs et approximations stochastiques: théorie et applications à l'identification, au traitement du signal et à la reconnaissance des formes*, Paris: Masson, 1987.

59. Benveniste, A. and Ruget, G., A Measure of the Tracking Capability of Recursive Stochastic Algorithms with Constant Gains, *IEEE Trans. Autom. Control*, 1982, vol. 27, no. 3, pp. 639–649.

60. Dugard, L. and Landau, I., Recursive Output Error Identification Algorithms Theory and Evaluation, *Automatica*, 1980, vol. 16, no. 5, pp. 443–462.

61. Kulchitskii, O., The New Method of Dynamical Stochastic Uncertain Systems Investigation, *Preprints of Int. Conf. "Stochastic Optimization,"* Part 1, 1984, pp. 127–129.

62. Kushner, H.J., *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory*, Cambridge: MIT Press, 1984, vol. 6.

63. Ljung, L., On Positive Real Transfer Functions and the Convergence of Some Recursive Schemes, *IEEE Trans. Autom. Control*, 1977, vol. 22, no. 4, pp. 539–551.

64. Ljung, L. and Söderström, T., *Theory and Practice of Recursive Identification*, Cambridge: MIT Press, 1983.

65. Solo, V., The Convergence of AML, *IEEE Trans. Autom. Control*, 1979, vol. 24, no. 6, pp. 958–962.

66. Bozin, A. and Zarrop, M., Self Tuning Optimizer-Convergence and Robustness Properties, *Proc. 1st Eur. Control Conf.*, 1991, pp. 672–677.

67. Ermol'ev, Yu.M. and Kaniovskii, Yu.M., Asymptotic Properties of Some Stochastic Programming Methods with Constant Step, *Zh. Vych. Mat. Mat. Fiz.*, 1979, vol. 19, no. 2, pp. 356–366.

68. Coito, F.J. and Lemos, J.M., Adaptive Optimization with Constraints: Convergence and Oscillatory Behaviour, in *Pattern Recogn. Image Anal.*, New York: Springer, 2005, pp. 19–26.

69. Kulchitskii, O.Yu., The Method of Convergence Analysis of Adaptive Filtration Algorithms, *Probl. Inf. Transm.*, 1985, vol. 21, no. 4, pp. 285–297.

70. Kushner, H.J. and Shwartz, A., Weak Convergence and Asymptotic Properties of Adaptive Filters with Constant Gains, *IEEE Trans. Inform. Theory*, 1984, vol. 30, no. 2, pp. 177–182.

71. Metivier, M. and Priouret, P., Applications of a Kushner and Clark Lemma to General Classes of Stochastic Algorithms, *IEEE Trans. Inform. Theory*, 1984, vol. 30, no. 2, pp. 140–151.

72. Derevitskii, D.P., The Method of Investigation of the Dynamics of Discrete-Time Adaptive Control Systems of Dynamical Objects, in *Voprosy kibernetiki. Adaptivnye sistemy* (Problems of Cybernetics. Adaptive Systems), Moscow: Akad. Nauk SSSR, 1976, pp. 64–72.

73. Andrievsky, B., Blazhkin, A., Derevitsky, D., and Fradkov, A., The Method of Investigation of Dynamics of Digital Adaptive Systems of Flight Control, *Proc. 6th IFAC Sympos. on Automatic Control in the Space*, Yerevan, 1974.

74. Derevitskii, D.P. and Fradkov, A.L., The Method of Continuous-Time Models in the Theory of Discrete-Time Adaptive Systems, in *Voprosy kibernetiki. Zadachi i metody adaptivnogo upravleniya* (Problems of Cybernetics. Problems and Methods of Adaptive Control), Moscow: Akad. Nauk SSSR, 1980, pp. 75–98.

75. Derevitskii, D. and Fradkov, A., Averaging Method for Discrete-Time Stochastic Systems and Its Application in Adaptive Control, *Preprints Int. Conf. "Stochastic Optimization,"* vol. 1, pp. 74–76.

76. Fradkov, A.L., *Adaptivnoe upravlenie slozhnymi sistemami* (Adaptive Control of Complex Systems), Moscow: Nauka, 1990.

77. Kulchitskii, O.Y., Algorithms of Stochastic Approximation in Adaptation Loop of Linear Dynamical System. I, II, *Autom. Remote Control*, 1983, vol. 44, no. 9, pp. 1189–1203; 1984, vol. 45, no. 3, pp. 104–113.

78. Mosca, E., Zappa, G., and Lemos, J.M., Robustness of Multipredictor Adaptive Regulators: Musmar, *Automatica*, 1989, vol. 25, no. 4, pp. 521–529.

79. Mosca, E., Lemos, J.M., Mendonca, T., and Nistri, P., Input Variance Constrained Adaptive Control and Singularly Perturbed ODE's, *Proc. 1st ECC*, Grenoble, 1991, pp. 2176–2180.

80. Tsykunov, A.M., *Adaptivnoe upravlenie ob"ektami s posledeistviem* (Adaptive Control of Plants with Aftereffect), Moscow: Nauka, 1984.

81. Oja, E. and Karhunen, J., On Stochastic Approximation of the Eigenvectors and Eigenvalues of the Expectation of a Random Matrix, *J. Math. Anal. Appl.*, 1985, vol. 106, no. 1, pp. 69–84.

82. Zhdanov, A.I., Recursive Estimation of Minimum Eigenvalues of Information Matrices, *Autom. Remote Control*, 1987, vol. 48, no. 4, part 1, pp. 443–451.

83. Stankovic, M.S., Johansson, K.H., and Stipanović, D.M., Distributed Seeking of Nash Equilibria in Mobile Sensor Networks, *Proc. 49th IEEE Conf. on Decision and Control (CDC)*, 2010, pp. 5598–5603.

84. Astanovskaya, N.V., Varshavskii, V.I., Revako, V.M., and Fradkov, A.L., The Method of Continuous-Time Models in the Problem of Decentralized Resource Allocation, in *Primenenie metodov sluchainogo poiska v SAPR* (Application of Random Search in CAD), Tallin: Valgus, 1979.

85. Borovkov, A.A., *Asimptoticheskie metody v teorii massovogo obsluzhivaniya* (Asymptotic Methods in Queueing Theory), Moscow: Nauka, 1979.

86. Kuan, C.-M. and Hornik, K., Convergence of Learning Algorithms with Constant Learning Rates, *IEEE Trans. Neural Networks*, 1991, vol. 2, no. 5, pp. 484–489.

87. Choi, J. and Horowitz, R., Learning Coverage Control of Mobile Sensing Agents in One-Dimensional Stochastic Environments, *IEEE Trans. Autom. Control*, 2010, vol. 55, no. 3, pp. 804–809.

88. Choi, J., Oh, S., and Horowitz, R., Distributed Learning and Cooperative Control for Multi-Agent Systems, *Automatica*, 2009, vol. 45, no. 12, pp. 2802–2814.

89. Huang, M. and Manton, J., Coordination and Consensus of Networked Agents with Noisy Measurements: Stochastic Algorithms and Asymptotic Behavior, *SIAM J. Control Optim.*, 2009, vol. 48, no. 1, pp. 134–161.

90. Borkar, V.S. and Manjunath, D., Distributed Topology Control of Wireless Networks, *Wireless Networks*, 2008, vol. 14, no. 5, pp. 671–682.

91. Granichin, O. and Fomin, V., Adaptive Control Using Test Signals in the Feedback Channel, *Autom. Remote Control*, 1986, vol. 47, no. 2, part 2, pp. 238–248.

92. Granichin, O., Volkovich, V., and Toledano-Kitai, D., *Randomized Algorithms in Automatic Control and Data Mining*, Berlin: Springer-Verlag, 2015.

93. Amelin, K. and Granichin, O., Randomized Control Strategies under Arbitrary External Noise, *IEEE Trans. Autom. Control*, 2016, vol. 61, no. 5, pp. 1328–1333.

94. Spall, J.C., *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, vol. 64, Hoboken: Wiley-Interscience, 2003.

95. Tsypkin, Ya.Z., *Adaptatsiya i obuchenie v avtomaticheskikh sistemakh* (Adaptation and Learning in Automatic Systems), Moscow: Nauka, 1968.

96. Yakubovich, V.A., The Method of Recurrent Objective Inequalities in the Theory of Adaptive Systems, in *Voprosy kibernetiki. Adaptivnye sistemy upravleniya* (Problems of Cybernetics. Adaptive Control Systems), Moscow: Akad. Nauk SSSR, 1976, pp. 32–63.

97. Andrievskii, B.R. and Fradkov, A.L., Analysis of the Dynamics of an Adaptive Control Algorithm for a Linear Dynamical Plant, in *Voprosy kibernetiki. Adaptivnye sistemy upravleniya* (Problems of Cybernetics. Adaptive Control Systems), Moscow: Akad. Nauk SSSR, 1976, pp. 99–103.

98. Fradkov, A.L., Speed-Gradient Scheme and Its Application in Adaptive Control Problems, *Autom. Remote Control*, 1980, vol. 40, no. 9, pp. 1333–1342.

99. Yoshihara, K., Moment Inequalities for Mixing Sequences, *Kodai Math. J.*, 1978, vol. 1, no. 2, pp. 316–328.

This paper was recommended for publication by A.V. Nazin, a member of the Editorial Board