SIAM J. APPLIED DYNAMICAL SYSTEMS
Vol. 12, No. 2, pp. 1007–1030 

# Dynamics of Stochastically Blinking Systems. Part I: Finite Time Properties*

Martin Hasler†, Vladimir Belykh‡, and Igor Belykh§

**Abstract.** We consider dynamical systems whose parameters are switched within a discrete set of values at equal time intervals. Similar to the blinking of the eye, switching is fast and occurs stochastically and independently for different time intervals. There are two time scales present in such systems, namely the time scale of the dynamical system and the time scale of the stochastic process. If the stochastic process is much faster, we expect the blinking system to follow the averaged system where the dynamical law is given by the expectation of the stochastic variables. We prove that, with high probability, the trajectories of the two systems stick together for a certain period of time. We give explicit bounds that relate the probability, the switching frequency, the precision, and the length of the time interval to each other. We discover the apparent presence of a soft upper bound for the time interval, beyond which it is almost impossible to keep the two trajectories together. This comes as a surprise in view of the known perturbation analysis results. From a probability theory perspective, our results are obtained by directly deriving large deviation bounds. They are more conservative than those derived by using the action functional approach, but they are explicit in the parameters of the blinking system.

**Key words.** blinking system, stochastic switching, averaging, probability

**AMS subject classifications.** 37H10, 37A50, 37H20, 34D10

**DOI.** 10.1137/120893409

## 1. Introduction.

Dynamical systems that are externally driven by stochastic processes are good models for many physical, biological, and engineering systems. The most obvious examples are those obtained by the presence of noise, but in many other circumstances, e.g., in dynamical networks with randomly present links, such models are also pertinent (see [1, 2] and the references therein). There are two time scales present in such systems, namely the time scale of the dynamical system itself and the time scale of the stochastic process. In this work, we suppose that the stochastic process is much faster than the dynamical system. Hence, we expect the stochastically "blinking" system to behave like the averaged system, where the dynamical law is simply averaged over the driving stochastic variables at each time instant. What this exactly means is a nontrivial problem, and we contribute substantially to its solution in this work.

*Received by the editors October 1, 2012; accepted for publication (in revised form) by B. Sandstede March 29, 2013; published electronically June 20, 2013.
http://www.siam.org/journals/siads/12-2/89340.html

†School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Station 14, 1015 Lausanne, Switzerland (martin.hasler@epfl.ch). This author's work was supported by the Swiss National Science Foundation under grant 200021-112081.

‡Mathematics Department, Volga State Academy, 5 Nesterov st., Nizhny Novgorod, 603 600 Russia (belykh@unn.ac.ru). This author's work was supported by RFFI grant 09-01-00498 and by the Ministry of Education and Science of the Russian Federation under grant 11.G34.31.0039.

§Corresponding author. Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303 (ibelykh@gsu.edu). This author's work was supported by the National Science Foundation under grant DMS-1009744.

There is a long history and a large number of publications on the subject of *averaging*, i.e., on investigating under what conditions the solutions of a time-dependent dynamical system follow the solutions of the time-averaged system. The technique of averaging applied to perturbations of periodic solutions is more than 200 years old [3, 4, 5]. It was also developed for deterministic perturbations of nonperiodic solutions about 60 years ago [6, 7]. More recently, stochastic perturbations of periodic and nonperiodic solutions have caught the attention of the research community [8, 9, 10, 11, 12, 13, 14]. Relevant recent books are [15] for deterministic and [9, 11, 12] for stochastic time-dependence.

In this work, we consider general dynamical systems described by state equations driven by identically distributed independent binary vector-valued random variables. This is a particularly simple case, but it perfectly illustrates the main phenomena pertaining to stochastically driven dynamical systems. It also is a good model for networks and circuits, where the links are switched on and off randomly, i.e., networks that are random in time. The proof of our main theorem, Theorem 3.1, shows that the results can be easily extended to a Markov vector-valued process in discrete or continuous time. This will be discussed in the conclusions. However, what mainly distinguishes our results from previous ones are explicit expressions for bounds on various relevant probabilities. They depend directly on the dynamical system parameters.

In the classical approach by Bogoliubov and Mitropolsky and many other authors (see, for example, [7]) the system

$$\frac{dx}{dt} = F_1(x,t) + \varepsilon F_2(x,t,\varepsilon) \tag{1.1}$$

is considered, where $\varepsilon$ is a small parameter and $F_1, F_2$ are bounded. This system is transformed into a standard form of perturbation analysis,

$$\frac{d\varsigma}{dt} = \varepsilon G(\varsigma,t) + \varepsilon^2 H(\varsigma,t,\varepsilon), \tag{1.2}$$

where $G$ and $H$ are bounded, and the solution of (1.1) can be obtained from the solution of (1.2) (cf. section 1.6 in [15]). It is then shown that the solution of (1.2) is close to the solution of

$$\frac{d\psi}{dt} = \varepsilon G(\psi,t), \tag{1.3}$$

starting from the same initial state, on a time scale of $1/\varepsilon$ (cf. Lemma 1.5.3 in [15], with additional time scaling). Because of the small parameter $\varepsilon$, $\psi$ varies slowly in time, much more slowly than $G$ as a function of time. Therefore, the solution of (1.3) is in turn close to the solution of

$$\frac{d\xi}{dt} = \varepsilon \bar{G}(\xi),$$

starting from the same initial state, where $\bar{G}$ is a certain time-average of $G$ (see Chapter 2 in [15] for periodic time-dependence, and Chapter 4 for general time-dependence with existing time-average). Again, the approximation is valid on a time scale of $1/\varepsilon$.

The system we study in this paper, even when disregarding its stochasticity, is initially not of the form (1.1). The only small parameter is the switching time constant $\tau$. The state equations can be written in the form

$$\frac{d\zeta}{dt} = F\left(\zeta,\ \frac{t}{\tau}\right).$$

Scaling the time variable as $u = t/\tau$ and setting $\eta(u) = \zeta(\tau u)$, we get

$$\frac{d\eta}{du} = \tau F(\eta, u),$$

which is in standard form. Therefore, we can conclude from classical averaging theory that the solution of the system we consider is close to the solution of the time-averaged system if the time-average of the sample path of the stochastic process satisfies certain conditions. Note that this holds on a time scale of $1/\tau$ for $u$, which corresponds to the time scale of 1 for $t$.

More interesting than statements for individual sample paths of the stochastic process are probabilistic properties. It is known that under conditions of ergodicity of the stochastic process the same property as in the deterministic case is valid with high probability. More precisely, on a finite time interval, the solutions of the blinking system remain close to the solutions of the averaged system starting from the same initial state, with a probability that converges to 1 as $\tau \to 0$ [9]. In our case, the stochastic process is just a sequence of independent identically distributed random variables, and so this result is applicable. However, in this paper we shall prove a much stronger result which gives an explicit bound on the distance between the solutions of the blinking and the averaged systems in terms of the system parameters and the length of the time interval. This result will reveal some unexpected phenomena. In the literature, often much more general problems are studied, but the corresponding results are weaker than the results of this paper. As a case in point, in [16] randomly switched dynamical systems are also considered, but the switching is not supposed to be fast with respect to the time scale of the dynamical system, the switching times are random, and, in addition, the system is perturbed by noise. Much effort has to be made to establish the existence and the uniqueness of the solution as a continuous-time Markov process, which is rather trivial in our case. Furthermore, the exponential growth of sample paths (corresponding to the exponential deviation of the solution from the zero solution of the averaged system) has a rate that depends on the bound of the right-hand side of the switched differential equation, whereas in our case it depends on the Lipschitz constant of the averaged system, which is usually much smaller.

Generally speaking, the crucial discipline of probability theory involved in this paper is *large deviations*. It concerns the exponentially fast decay of rare events in stochastic processes. Using the notion of an action functional, it is in principle possible to determine the exact exponential rate of decay of such events. However, in the context of dynamical systems, this involves the solution of a variational problem [10, 11]. We pursue in this and the companion paper a simpler approach. It has the advantage that we can give a lower bound for the probability that the trajectories of the blinking and the averaged system remain within distance $\epsilon$ during a time interval $[0, T]$ that is perfectly explicit in the dynamical systems parameters. The relevant parameters are bounds and Lipschitz constants for the right-hand side (RHS) of the state equations of the blinking and the averaged systems. Except, perhaps, for very special cases, such results are not obtainable by the action functional approach. On the other hand, our techniques do not yield the actual rates of exponential decay of the rare events, but only lower bounds on them.

The work is presented in two closely related companion papers. In Part I we show that, with high probability, the solution of the blinking system follows the solution of the averaged system with a certain precision for a certain finite length of time if it starts from the same initial condition. As a perturbation result this is well known. We add to this explicit bounds that relate the probability, the switching time, the precision, and the length of the time interval to each other. A surprising result that could not be obtained from the perturbation approach is the apparent presence of a soft upper bound on the length of the time interval, beyond which the solution of the blinking system can only follow the solution of the averaged system with good precision for unreasonably fast stochastic driving.

In a companion paper (Part II) [2] we will discuss the asymptotic properties of the solutions of the blinking system as time goes to infinity. The question we will address is whether and how the solutions of the blinking system converge to an attractor of the averaged system. Once again, we will give bounds on the various relevant probabilities that are explicit in the dynamical systems parameters. As mentioned before, the action functional approach of large deviation theory could give the precise exponential decays of the probabilities rather than bounds for them as we derive, but explicit expressions for the precise rates seem out of reach.

## 2. Definition of the blinking and the averaged systems.

We consider systems of time-dependent ordinary differential equations of the form

$$\frac{dx}{dt} = F\left(x\left(t\right), s\left(t\right)\right), \qquad x \in \mathbb{R}^N, \quad F : \mathbb{R}^{N+M} \to \mathbb{R}^N, \quad s\left(t\right) \in \{0,1\}^M, \tag{2.1}$$

where the function $s : [0, \infty) \to \{0,1\}^M$ is piecewise constant, taking the constant binary vector value $s^k = (s_1^k, \ldots, s_{\mathrm{M}}^k)$ in the time interval $t \in [(k-1)\,\tau,\ k\tau)$.

We call system (2.1) a *blinking system*, thinking of the piecewise constant binary signal $s_i\left(t\right)$ as a closing and opening eye lid. In this paper, we study the behavior of the solutions of (2.1) starting at $t = 0$ in a finite interval of time. We call the binary vector-valued signal $s(t)$ the *switching signal*, and the sequence of binary vectors $s^k$, $k = 1, 2, \ldots,$ the *switching sequence*, since, in some of the examples we have in mind (see [2] for the examples), each component $s_i^k$ of $s^k$ switches a connection in a circuit or network on ($s_i^k = 1$) or off ($s_i^k = 0$) during the $k$th time interval.

Instead of considering system (2.1) for a particular switching sequence, e.g., a periodic switching sequence, we consider the set of all possible switching sequences simultaneously, supposing that switching is performed probabilistically. More precisely, we suppose that the switching sequences are the instances of the stochastic process $S^k$, $k = 1, 2, \ldots,$ where all random vectors $S^k$ are independent and identically distributed, taking the value $s \in \{0,1\}^M$ with probability $p_s$. Let us remark here that this is the simplest stochastic switching model, which we have chosen because of its direct connection to time-dependent random networks, where $M$ links are independently turned on and off. The independence requirements can be relaxed without changing much in the results and their proofs. The independence of the

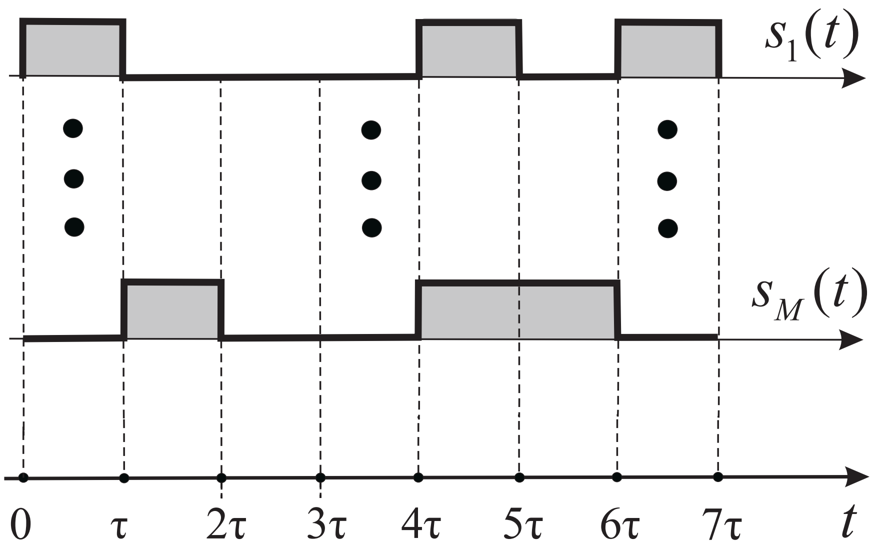


**Figure 1.** *Division of the time axis and type of functions* $\mathbf{s}^k(t)$.

components of the binary vector $S(t)$ at a given time $t$ is never used. Any random variable with a finite number of possible values can be used instead. The independence of the vectors $S(t)$ and $S(t^*)$ for different instants $t$ and $t^*$ can be relaxed to a Markovian dependence with only minor changes that are detailed in the conclusions.

Similarly, we introduce the continuous time stochastic process $S(t)$ defined by

$$S(t) = \sum_{k=1}^{\infty} S^k \, \mathbf{1}_{[(k-1)\tau, k\tau)}(t)$$

and consider simultaneously all solutions of (2.1) by writing

$$\frac{dx}{dt} = F\left(x\left(t\right), S\left(t\right)\right). \tag{2.2}$$

Here, $F\left(x\left(t\right), S\left(t\right)\right)$ is also a stochastic process, and therefore (2.2) represents a system of random differential equations. However, thanks to the fact that $S(t)$ is piecewise constant (cf. Figure 1), we avoid the technical difficulties of stochastic differential equations [9]. Nevertheless, the solutions of (2.2) are also stochastic processes. For simplicity, we shall not distinguish in our notation between the random variable and its instance and denote both by $x(t)$. The context will make it clear which one we mean in each case.

System (2.1) inherently has two time scales, the switching period $\tau$ and the time scale of the dynamics of the nonswitched system, i.e., the system with constant $s$. We limit our attention to the case where switching is fast with respect to the time scale of the nonswitched system dynamics (this statement will be made more precise later). Our intuition tells us that in this case the effect of switching is the same as if the time-dependent system (2.1) were replaced by the averaged system.

Definition 2.1. *The averaged system associated with the time-dependent system* (2.1) *is defined to be*

$$\frac{dx}{dt} = \Phi\left(x\left(t\right)\right), \tag{2.3}$$

*where*

$$\begin{aligned} \Phi(x) &= \mathrm{E}\left(F\left(x,S\right)\right) \\ &= \sum_{s\in\{0,1\}^{\mathrm{M}}} F\left(x,s\right)p_s. \end{aligned} \tag{2.4}$$

In (2.4) we have omitted the upper index $k$ for the switching variables, since in each time interval they have the same probability distribution.

We expect that if we switch fast enough, i.e., if $\tau$ is sufficiently small, then a solution of blinking system (2.1) follows closely the solution of the averaged system (2.3) when starting from the same initial state $x(0)$. Indeed, the following general theorem is a reformulation of the "Statement" in section 3.2 of the book [9].

Theorem 2.2. *Let the following two conditions be satisfied:*

- *for any* $x \in \mathbb{R}^N$ *the expectation* $\mathrm{E}\left\|F\left(x,S\right)\right\| < \infty$,
- $\exists\, L > 0$ *such that for any* $x, \tilde{x} \in \mathbb{R}^N$ *and for any* $s \in \{0,1\}^M$

$$\left\|F\left(x,s\right) - F\left(\tilde{x},s\right)\right\| \le L\left\|x-\tilde{x}\right\|.$$

*Then for any precision* $\varepsilon > 0$ *and time interval length* $T > 0$, *and for any pair of solutions of the blinking and the averaged system, starting from the same initial state* $x\left(0\right) = \xi\left(0\right)$, *the probability that for all times* $t \in [0,T]$

$$\left\|x\left(t\right) - \xi\left(t\right)\right\| \le \varepsilon$$

*converges to* 1 *as the switching time* $\tau \to 0$.

We expect that with decreasing $\tau$ we can take a lower $\varepsilon$ or a higher $T$, or with fixed $\varepsilon$ and $T$ the probability gets closer to 1. However, Theorem 2.2 does not give explicit information on the relation between the three parameters. To be fair, in [9], much more general stochastic processes are considered and many other interesting properties of the solutions are derived. However, these properties are always asymptotic in the limit $\tau \to 0$, whereas we aim at explicit bounds for $\tau$.

Another basic remark is that Theorem 2.2, as well as the theorem we will prove in this paper (Theorem 3.1), never guarantees a uniform approximation of the solution of the averaged system by a solution of the blinking system over the whole time interval $[0,\infty)$. The question of the infinite time interval will be discussed in the companion paper [2]. Briefly, the result is the following. Except in special circumstances, eventually the solutions of the blinking and the averaged systems will always drift apart. However, when a solution of the averaged system converges to an attractor, the solution of the blinking system will in general approach the same attractor. What that exactly means depends on the circumstances. In [2], we will distinguish four different classes of behavior and obtain explicit bounds for each one.

## 3. Basic assumptions and inequalities.

For convenience we make the following, not very restrictive assumptions.

*Hypothesis* 1.

1. The function $F$ that defines the blinking system (2.1) is locally Lipschitz continuous in $x$, the first $N$ arguments, and continuous in $s$, the last $M$ arguments.
2. For any switching signal $s(t)$ and any state $x_0$ there exists exactly one trajectory $x(t)$ of the blinking system with $x(0) = x_0$, defined for $0 \leq t < \infty$. Similarly, there exists a unique trajectory $\xi(t)$ of the averaged system
$$\frac{d\xi}{dt} = \Phi\left(\xi(t)\right),$$
defined for $0 \leq t < \infty$, for a given initial state $\xi(0) = \xi_0$.
3. There is a connected and compact, i.e., closed and bounded, region $\Omega$ in $\mathbb{R}^N$ such that all trajectories of the blinking system and of the averaged system starting in $\Omega$ remain in $\Omega$.

We shall consider only solutions of the blinking and the averaged systems that start in $\Omega$ and therefore remain in $\Omega$. If the blinking and the averaged systems are dissipative for sufficiently large state vectors, then there exists such a region $\Omega$ with the additional property that all solutions reach $\Omega$ in finite time.

The continuity of $F$ implies that $F$ is bounded on $\Omega$, and Definition 2.1 implies the same for $\Phi$. Define constants

$$\begin{aligned} B_F &= \max_{x\in\Omega,\ s\in\{0,1\}^M} \left\|F\left(x,s\right)\right\| < \infty, \\ B_\Phi &= \max_{x\in\Omega} \left\|\Phi\left(x\right)\right\| < \infty, \end{aligned} \tag{3.1}$$

where $||.||$ denotes the Euclidean norm. Similarly, the local Lipschitz continuity implies the existence of the Lipschitz constants $L_F$ and $L_\Phi$ on $\Omega$, which satisfy by definition

$$\begin{aligned} \left\|F\left(x,s\right) - F\left(y,s\right)\right\| &\leq L_F \left\|x-y\right\| \qquad \text{for } x,y\in\Omega,\ s\in\{0,1\}^M, \\ \left\|\Phi\left(\xi\right) - \Phi\left(\eta\right)\right\| &\leq L_\Phi \left\|\xi-\eta\right\| \qquad \text{for } \xi,\eta\in\Omega. \end{aligned} \tag{3.2}$$

It is not difficult to see that

$$B_\Phi \leq B_F, \qquad L_\Phi \leq L_F.$$

We can now formulate the main result of this paper. It implies Theorem 2.2, but in contrast to Theorem 2.2, it gives an explicit lower bound for the probability that the trajectory of the blinking system stays close to the trajectory of the averaged system during a given time interval. Obviously, this probability depends on the length of the time interval, the maximal distance between the two trajectories, the switching time, and the dynamical systems between which the stochastic switching takes place. However, as far as the dynamical systems are concerned, only the bounds and Lipschitz constants introduced in (3.2) are needed, and they can be deduced explicitly from the state equations.

Theorem 3.1. *Suppose that Hypothesis 1 holds. For any precision $\varepsilon > 0$, time interval length $T > 0$, and probability $0 < P_0 < 1$, the solutions of the blinking system and of the averaged system satisfy*

$$\left\|x\left(t\right) - \xi\left(t\right)\right\| \leq \varepsilon \tag{3.3}$$

*for all times* $t \in [0,T]$ *with probability at least* $P_{close} = 1 - P_0$ *if the switching frequency* $\tau$ *is sufficiently small such that*

$$\tau \le \tau_0 = \left(\frac{\varepsilon}{C(T)}\right)^3 \cdot \frac{1}{\ln(2D \cdot TC(T)) - \ln(\varepsilon P_0)} \cdot \frac{1}{DB_F^2}, \tag{3.4}$$

*where* $N$ *is the dimensionality of the blinking system,*

$$C(T) = \left(e^{L_\Phi t} - 1\right)\frac{2}{L_\Phi} + \frac{2}{L_{F\Phi}}, \tag{3.5}$$

*and*

$$\begin{aligned} L_{F\Phi} &= \frac{L_F B_F + L_\Phi B_\Phi}{B_F + B_\Phi}, \\ D &= N\left(L_F B_F + L_\Phi B_\Phi\right). \end{aligned} \tag{3.6}$$

*Proof.* The complete proof of Theorem 3.1 is given in the appendix. Here, we give a short narrative that underlies the main ideas of the proof. For the convenience of the reader, we refer to the corresponding parts of the complete proof.

Let $x(t)$ be a trajectory of the blinking system, and $\xi(t)$ be the solution of the averaged system, defined also for $t \ge 0$, that starts from the same initial state $\xi(0) = x(0)$. To bound the distance between the two solutions probabilistically, we proceed in two steps.

In the first step (cf. Lemma 6.2 in the appendix), we show that in a time interval $[t, t+\Delta t]$ the distance cannot increase more than $\alpha\Delta t$ for $\alpha = B_F + B_\Phi$. This is a rather trivial and conservative bound that always holds. We also prove a more subtle bound, namely that during a time interval of length $\Delta t$ the distance does not increase by more than a factor $1 + L_\Phi \Delta t$ plus a constant $2\lambda\Delta t$. This inequality does not hold always, but only with a probability $1 - P_\lambda$. Here, $\lambda$ is a free parameter, which is chosen later in the proof in a suitable way, much smaller than $\alpha$. Clearly, $P_\lambda$ tends to 1 as $\lambda$ goes to 0, and to 0 as $\lambda$ goes to $\frac{\alpha}{2}$. Furthermore, $P_\lambda$ depends on the ratio $\frac{\Delta t}{\tau}$. $\Delta t$ is also a free parameter that is chosen later in the proof, much larger than $\tau$.

In the second step, we divide the positive time axis into intervals of length $\Delta t$. For each nonnegative integer $q$ we introduce as an auxiliary random variable $Z_q$, the distance between the trajectories of the blinking and the averaged systems at time $t = q\Delta t$. Depending on the switching sequence, the increase from $Z_q$ to $Z_{q+1}$ satisfies the more subtle bound derived in the first step of the proof, or does not satisfy it. In the latter case, only the crude bound can be applied. Combining the bounds on $Z_{q+1} - Z_q$ for $q = 0, 1, \ldots, Q$ leads to a probabilistic bound on $Z_Q$ (cf. Lemmas 6.3 and 6.4 in the appendix) depending on $\Delta t$, $\lambda$, and the number $m$ among the $Q$ increments where the subtle bound does not hold. This step needs a rather detailed analysis of the stochastic process $Z_q$, $q = 0, 1, 2, \ldots$, because the increments $Z_{q+1} - Z_q$ for different $q$ are not independent, and in fact $Z_q$ is not even a Markov process. In fact, the difference between the trajectories of the blinking and the averaged systems, $x(t) - \xi(t)$, is a vector-valued Markov process, but the application of the Euclidean norm to this difference destroys the Markov property.

Applying a union bound for the probability that $m > 0$ (Lemma 6.5 in the appendix), we get a probabilistic bound for $Z_Q$ that still depends on $\Delta t$ and $\lambda$ but is independent of $m$. In fact, it corresponds to $m = 0$. Now, in order to obtain a probabilistic bound on the distance between the trajectories of the blinking and the averaged systems that is valid in the whole time interval $[0, T]$ and not only at the instants $Q\Delta t$, $Q = 0, 1, \ldots,$ we use, to interpolate between those instants, the crude bound that is always valid. Then, we choose the value of $\Delta t$ that gives the smallest bound, which is now proportional to $\lambda$ (Lemma 6.6 in the appendix). The proof is completed by relating $\lambda$ to $\epsilon$. ∎

The following corollary is a reformulation of Theorem 3.1.

Corollary 3.2. *Inequality* (3.4) *in Theorem* 3.1 *can be rearranged to give a bound on the probability* $P_{\text{close}}$ *that the trajectories of the blinking and averaged systems stay* $\varepsilon$*-close during time interval* $T$ *for a given switching frequency* $\tau$*. That is, solving inequality* (3.4) *for* $P_0$ *yields*

$$P_{\text{close}} = 1 - P_0 \geq 1 - \frac{2D \cdot T \cdot C(T)}{\varepsilon} \cdot \exp\left(-\frac{\varepsilon^3}{\tau} \cdot \frac{1}{C(T)^3 DB_F^2}\right).$$

Analyzing more closely the proof of Theorem 3.1 (cf. Lemma 6.6 in the appendix), one obtains the following stronger bound.

Corollary 3.3. *Under the hypotheses of Theorem* 2.2*, for any precision* $\varepsilon > 0$*, time interval length* $T > 0$*, and probability* $0 < P_0 < 1$*, the solutions of the blinking system and of the averaged system satisfy*

$$\|x(t) - \xi(t)\| \leq \frac{C(t)}{C(T)}\varepsilon$$

*for all times* $t \in [0, T]$ *with probability at least* $P_{close} = 1 - P_0$*, if* $\tau \leq \tau_0$*, where* $\tau_0$ *is given by* (3.4) *and*

$$C(t) = \left(e^{L_\Phi t} - 1\right)\frac{2}{L_\Phi} + \frac{2}{L_{F\Phi}}.$$

*Remark* 3.1. The main variation of the bound on the switching time comes from the term $(\frac{\varepsilon}{C(T)})^3$. It can be seen from (3.5) that there are basically two cases to distinguish:

1. $T$ is small, up to $1/L_\Phi$: $C(T)$ varies little, between $\frac{2}{L_{F\Phi}}$ and $\frac{2(e-1)}{L_\Phi} + \frac{2}{L_{F\Phi}}$. Therefore
$$\tau_0 \sim \varepsilon^3.$$
2. $T$ is large with respect to $1/L_\Phi$: In this case, $C(T) \approx e^{L_\Phi T}\frac{2}{L_\Phi}$, and therefore
$$\tau_0 \sim \varepsilon^3 e^{-L_\Phi T}.$$

This fact is illustrated in Figure 2, where for small intervals $[0, T]$ the threshold for $\tau$ varies little, but for large intervals the threshold for $\tau$ drops dramatically. One can interpret $T = \frac{1}{L_\Phi}$ as a soft threshold beyond which the solution of the averaged system is almost impossible to approximate by the blinking system.

## 4. Example: Blinking chaotic system.

In order to test the conjecture that the solutions of the blinking system follow closely the corresponding solutions of the averaged system, we take the hardest case, namely systems with chaotic behavior. Such systems have only unstable solutions, so there is an intrinsic mechanism to drive solutions apart, which should make it difficult for the blinking system to follow the averaged system.

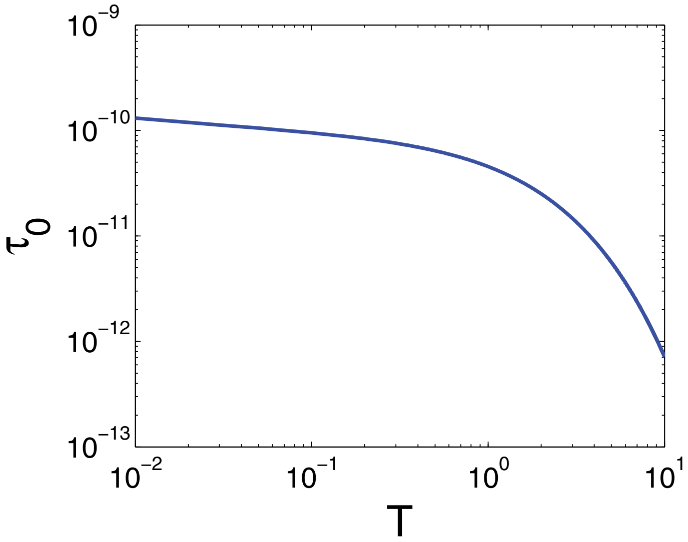


**Figure 2.** *Threshold $\tau_0$ as a function of the length $T$ of the time interval $[0, T]$, calculated via (3.4) Choice of parameters: $\varepsilon = 0.1,\ P_0 = 0.1,\ L_\Phi = 0.1,\ L_{F\Phi} = 0.2,\ D = B_F = 1$. Note the presence of a soft upper bound for the time interval $T$ beyond which the two trajectories can be kept together only by switching almost infinitely quickly.*

We consider the Lorenz system that is described by the following nonlinear system of differential equations:

$$\begin{aligned} \dot{x} &= \sigma\,(y - x), \\ \dot{y} &= rx - y - xz, \\ \dot{z} &= -bz + xy. \end{aligned} \tag{4.1}$$

The customary choice of parameters $\sigma = 10,\ b = 8/3,\ r = 28$ leads to chaotic behavior. We now create a blinking system by switching the parameter $r$ between $r = 28$ and $r = 33$ with probability $p = 0.5$. Hence, the averaged system is again a Lorenz system, with $r = 30.5$.

The behavior of the deterministic Lorenz systems with the three different values is illustrated in Figures 3 and 4. The trajectories immediately drift apart from the common initial state, but they are generally of the same nature.

The comparison between the blinking and averaged systems is done in Figures 5 and 6. Clearly, for a short period the two solutions evolve close to each other. The length of this time interval is almost not affected when $\tau$ decreases by an order of magnitude. The closeness of the two solutions is enhanced when $\tau$ decreases, but by much less than one order of magnitude.

Of course, the trajectory of the blinking system varies with the instance of the switching sequence. In Figure 7, the mean value of the $x$-component of the solution of the blinking system over 100 randomly chosen switching sequences as well as the standard deviation are represented. All solutions start from the same initial condition. Note that after some time the different solutions of the blinking system become uncorrelated, and their mean gets close to 0. Finally, the initial behavior of the solutions of the blinking system depends also on the initial condition. In order to diminish its influence, we have calculated the average distance

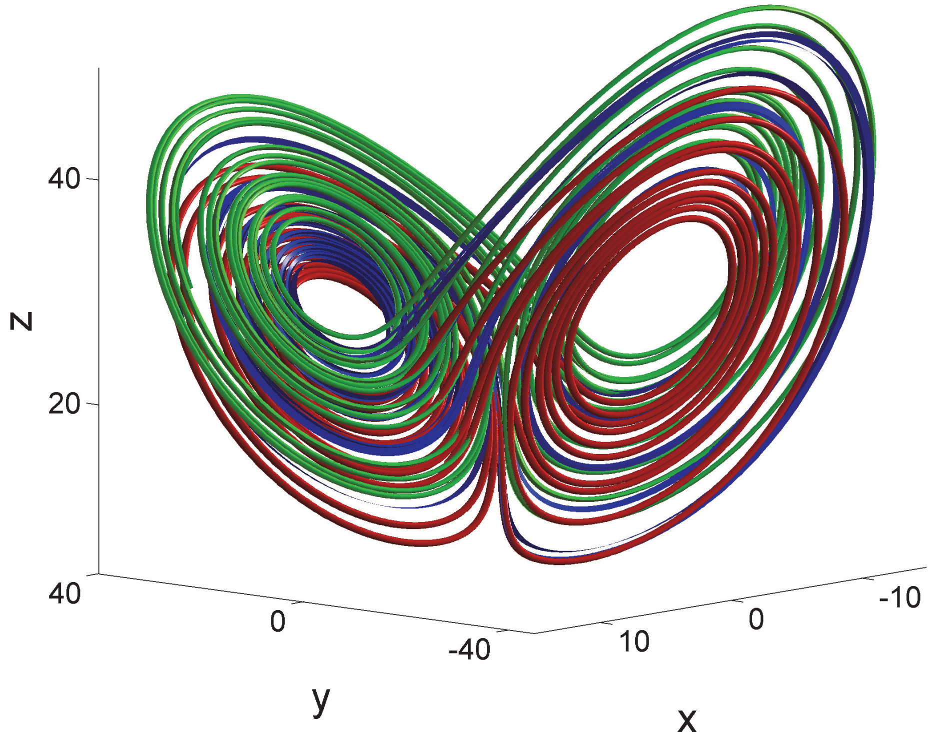


**Figure 3.** *Orbits of the Lorenz system* (4.1) *with* $r = 28$ *(green),* $r = 30.5$ *(blue), and* $r = 33$ *(red), starting from the same initial state that is close to the attractor. The orbits are similar in nature but not coinciding.*

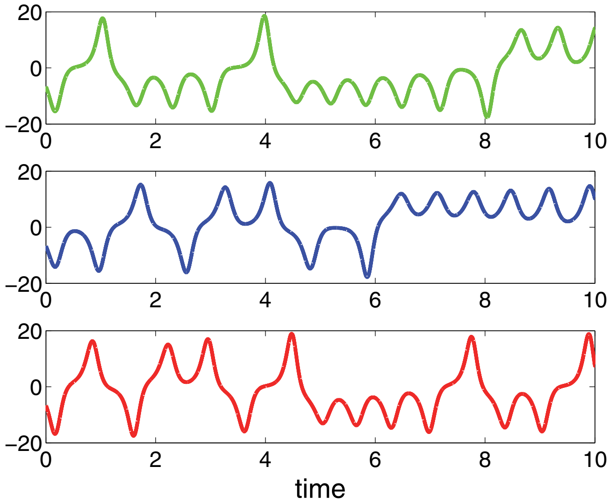


**Figure 4.** $x$*-component of the trajectories of the Lorenz system with* $r = 28$ *(upper panel),* $r = 30.5$ *(middle panel), and* $r = 33$ *(lower panel), starting from the same initial state* $(-6.5763, -11.3758, 16.2133)$*. The trajectories immediately drift apart.*

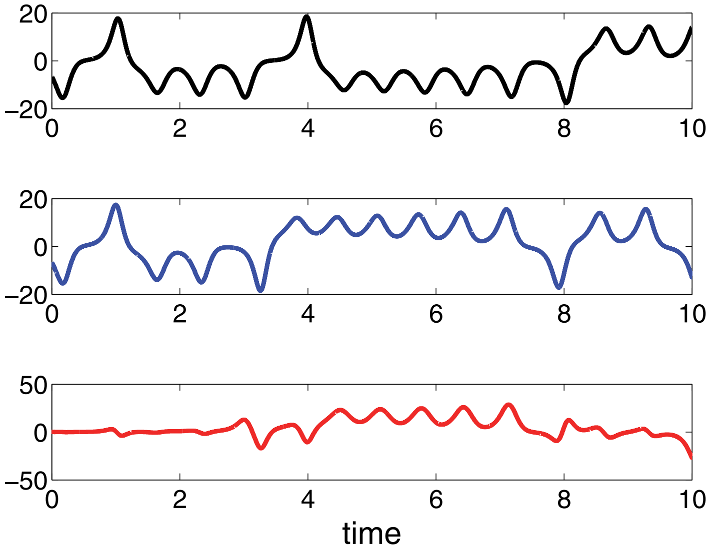


**Figure 5.** *$x$-component of the trajectories of the averaged system (upper panel), the blinking system with $\tau = 0.001$ and $r = 28$ (middle panel), and the difference of the $x$-components of the blinking system and the averaged system (lower panel). Clearly, the two solutions drift apart after some finite time.*

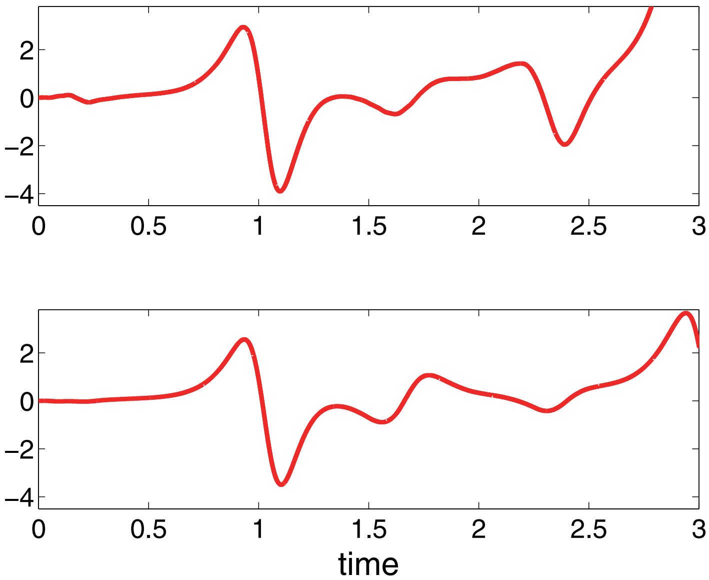


**Figure 6.** *Difference of the $x$-components of the blinking system and the averaged system, with $\tau = 0.001$ (upper panel) and $\tau = 0.0001$ (lower panel). The upper panel represents the same data as the lower panel in Figure 5, but at a higher vertical resolution and for a shorter time interval. Clearly, faster switching reduces the difference between the solutions of the averaged and blinking systems, but it does not substantially increase the interval of time during which the two solutions are close.*

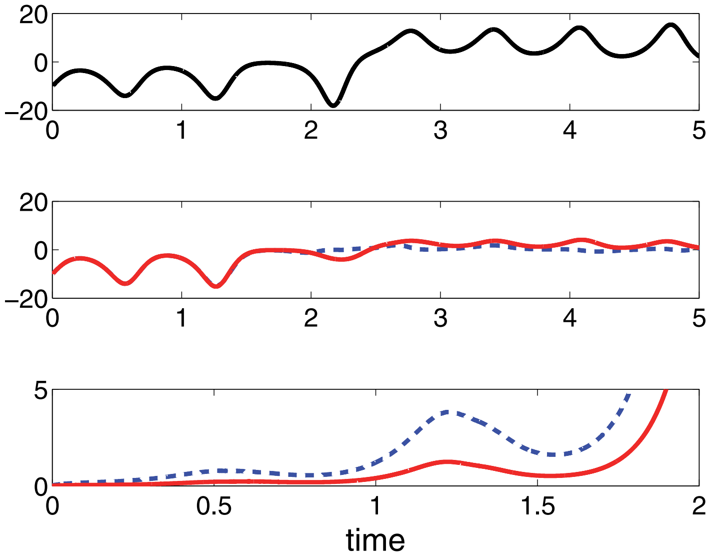


**Figure 7.** *$x$-component of the solution of the averaged system starting from the initial state $[-10.1667, -4.7473, 36.6089]$ (upper panel). Mean value of the $x$-component of 100 solutions of the blinking system starting from the same initial condition, but differing in the switching sequence (middle panel). Standard deviation of the first component of the 100 solutions of the blinking system from the first component of the solution of the averaged system (lower panel). Switching time $\tau = 0.001$ (blue (dashed) curve), switching time $\tau = 0.0001$ (red (solid) curve).*

$\|x(t) - \xi(t)\|$ between the solution of the blinking system and the averaged system for 10 different initial states and for each of 100 switching sequences (see Figure 8). It can be observed that initially the average distance grows slowly and then takes off exponentially, as predicted by the explicit bounds. Afterwards, a saturation effect takes over. The exponential take-off occurs only about two times later when the switching time is decreased by a factor 10.

**5. Conclusions.** We have investigated the finite time properties of the stochastically blinking dynamical systems, previously introduced in our paper [1] in the context of network synchronization. The results of this paper are not constrained by this context and are applicable to any stochastically switched dynamical system in any dynamical regime.

More precisely, we have considered dynamical systems whose parameters are switched within a discrete set of values at regular time intervals. Similar to the blinking of the eye, switching is fast, and it occurs stochastically and independently for different time intervals. There are two time scales present in such systems, namely the time scale of the dynamical system and the time scale of the stochastic process. We have proven that any trajectory of the blinking system closely follows the trajectory of the averaged system, starting from the same initial condition, during a certain lapse of time $T$ if the switching is fast enough. This has already been shown by perturbation theory in a more general context of stochastic dynamical systems [9]. Going beyond the perturbation results, we have given explicit relations among the switching time, the distance between the two trajectories, and the length of the time interval $T$ in the forms of bounds on probabilities.

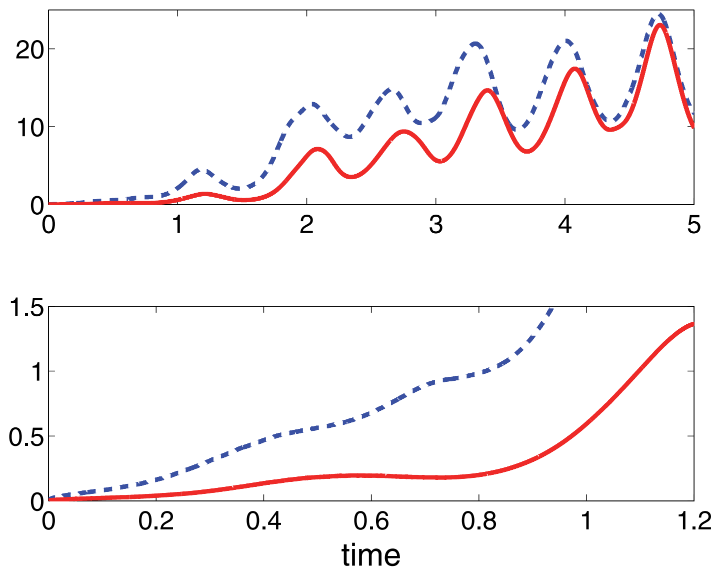


**Figure 8.** *Mean distance of* 1000 *blinking solutions from the averaged solution, for* 10 *different initial states and* 100 *different switching sequences for each. Switching time* $\tau = 0.001$ *(blue (dashed) curve), switching time* $\tau = 0.0001$ *(red (solid) curve). The lower panel is an enlarged version of the upper panel.*

The starting point for the derivation of these bounds is a well-known large deviation bound for sums of bounded independent random variables applied to the switching variables. From this, we have derived probabilistic bounds on the distance between the two trajectories in two steps. First, we have derived the bounds on these probabilities for a small time interval $[t, t + \Delta t]$, conditioned on the state at time $t$. Second, we have combined the bounds for the small time intervals into a bound for the time interval $[0, T]$. The resulting bound on the probability distribution for the distance between the trajectories of the blinking and averaging systems indicates the presence of a soft threshold for the time interval beyond which the two trajectories can be kept together only by switching almost infinitely fast. This comes as a surprise that was not predicted by perturbation theory even though the existence of the soft threshold does not contradict the perturbation theory results. This threshold depends only on the Lipschitz constant of the averaged system: the smaller the Lipschitz constant, the larger the threshold.

More precise bounds that actually contain the precise exponential decay rates could be obtained by applying the action functional approach of large deviation theory [10, 11], but not in the form of explicit expressions in the dynamical system parameters except perhaps for particularly simple systems.

Our method can also be applied when the switching sequences follow a Markov process, instead of being independent binary random vectors. Indeed, there are only two places where the specific nature of the driving stochastic process is used. The first is the probability of large deviation of the sum or integral of random variables from its mean, where we have used Hoeffding's inequality [17]. Similar inequalities can be obtained for Markov processes in discrete or continuous time, and one simply has to substitute the quantities $\mathrm{P}_{W\lambda}$ by the corresponding expressions in all formulas (cf. the appendix). The second place where the nature of the stochastic process is used is when combining the results on the increasing of the distance between the trajectories of the blinking and the averaged systems during an interval of length $\Delta t$ for subsequent intervals on the time axis. Our proofs remain valid as long as the solution $x(t)$ is a vector-valued Markov process. If the driving stochastic process has the Markov property, this is always the case.

In a companion paper (Part II) [2], we study the relation between the trajectories of the averaged and blinking systems when time goes to infinity, using similar methods combined with the Lyapunov function approach. The two trajectories usually cannot stay together, but they may converge to the same attractor.

## 6. Appendix.

In this section, we give the details of the proof for Theorem 3.1. We first prove the preliminary results contained in Lemmas 6.2–6.6 and then arrive at the actual proof of Theorem 3.1.

### 6.1. Preliminaries: Large deviation bounds.

We let $x(t)$ be a trajectory of the blinking system, and $\xi(t)$ be the solution of the averaged system, defined also for $t \geq 0$, that starts from the same initial state $\xi(0) = x(0)$.

By the weak law of large numbers [18], we have for all $x$ and $\lambda > 0$ that

$$P\left\{\left\|\frac{1}{K}\sum_{k=1}^{K}\left(F\left(x,S^k\right)-\Phi\left(x\right)\right)\right\| > \lambda\right\}\underset{K\to\infty}{\to} 0,$$

$$P_\lambda\left(K\right) = \max_{x\in\Omega} P\left\{\left\|\frac{1}{K}\sum_{k=1}^{K} F\left(x,S^k\right)-\Phi\left(x\right)\right\| > \lambda\right\}, \tag{6.1}$$

$$P_{\lambda,i}\left(K\right) = \max_{x\in\Omega} P\left[\left|\frac{1}{K}\sum_{k=1}^{K} F_i\left(x,S^k\right)-\Phi_i\left(x\right)\right| > \lambda\right],$$

where $P$ refers to the probability underlying the stochastic process $S^k$ of switching sequences. These definitions of probabilities $P_\lambda$ and $P_{\lambda,i}$ play an essential role in what follows. Note that $P_\lambda\left(K\right)\underset{K\to\infty}{\to} 0$ and $P_{\lambda,i}\left(K\right)\underset{K\to\infty}{\to} 0$ for any $\lambda > 0$, $i = 1,\ldots,N$, and since $\{S_k\}$ is a stationary stochastic process, for any $k_0 \geq 1$

$$\max_{x\in\Omega} P\left\{\left\|\frac{1}{K}\sum_{k=k_0}^{k_0+K} F\left(x,S^k\right)-\Phi\left(x\right)\right\| > \lambda\right\} = P_\lambda\left(K\right), \tag{6.2}$$

$$\max_{x\in\Omega} P\left\{\left|\frac{1}{K}\sum_{k=k_0}^{k_0+K} F_i\left(x,S^k\right)-\Phi_i\left(x\right)\right| > \lambda\right\} = P_{\lambda,i}\left(K\right).$$

Various bounds, so-called large deviation bounds, can be found in the literature. Since the random variables $F(x,S^k)$ are bounded, we shall use the Hoeffding bound [17]: if $X_k$ is a sequence of independent random variables with $P\left(a_k \leq X_k \leq b_k\right) = 1$ and if $\overline{X} = \frac{1}{K}\sum_{k=1}^{K} X_k$, then

$$P\left(\left|\overline{X}-\mathrm{E}\left(\overline{X}\right)\right|>\lambda\right)\ \le\ 2\exp\left(-\frac{2K^2\lambda^2}{\sum_{k=1}^{K}\left(b_k-a_k\right)^2}\right), \tag{6.3}$$

where $\mathrm{E}\left(\overline{X}\right)$ is the expected value of $\overline{X}$. This leads to the following assertion.

Lemma 6.1.

$$P_\lambda\left(K\right)\ \le\ \sum_{i=1}^{N}P_{\left(\lambda/\sqrt{N}\right),i}\left(K\right)\ \le\ 2N\exp\left(-\frac{K\lambda^2}{2NB_F^2}\right). \tag{6.4}$$

*Proof.*

$$P_\lambda\left(K\right)=\max_{x\in\Omega}P\left(\sum_{i=1}^{N}\left(\frac{1}{K}\sum_{k=1}^{K}F_i\left(x,S^k\right)-\Phi_i\left(x\right)\right)^2>\lambda^2\right). \tag{6.5}$$

The sum over $i$ can be larger than $\lambda^2$ only if at least one term is larger than $\lambda^2/N$. Hence

$$P_\lambda\left(K\right)\le\max_{x\in\Omega}\sum_{i=1}^{N}P\left(\left(\frac{1}{K}\sum_{k=1}^{K}F_i\left(x,S^k\right)\ -\ \Phi_i\left(x\right)\right)^2>\frac{\lambda^2}{N}\right)\le\sum_{i=1}^{N}P_{\left(\lambda/\sqrt{N}\right),i}(K). \tag{6.6}$$

Thus, by (3.1) and Hoeffding's inequality (6.3), we have

$$P_{\lambda,i}\left(K\right)\ \le\ 2\exp\left(-\frac{K\lambda^2}{2B_F^2}\right),$$

which, by substituting into (6.6), yields the second inequality of (6.4). ■

In what follows, this lemma will be used for bounding

$$\left\|\int_0^t\left(F\left(x,s\left(u\right)\right)-\Phi\left(x\right)\right)du\right\|\ =\ \left\|\sum_{k=1}^{K}\tau\left(F\left(x,s^k\right)-\Phi\left(x\right)\right)\right\|,$$

where this equality holds, provided that $t=K\tau$. Actually, we will need to bound $\|\int_t^{t+\Delta t}\left(F\left(x,s\left(u\right)\right)-\Phi\left(x\right)\right)du\|$. Using (6.2), again Lemma 6.1 can be applied, as long as both $t$ and $\Delta t$ are multiples of $\tau$. If this is not the case, then we have to use the more general form

$$\begin{aligned}\int_t^{t+\Delta t}\left(F\left(x,s\left(u\right)\right)-\Phi\left(x\right)\right)du&=\tau\left(1-\frac{t}{\tau}+\left\lfloor\frac{t}{\tau}\right\rfloor\right)\left(F\left(x,s^{\lfloor\frac{t}{\tau}\rfloor}\right)\ -\ \Phi\left(x\right)\right)\\&+\sum_{k=\lfloor\frac{t}{\tau}\rfloor+1}^{k=\lfloor\frac{t+\Delta t}{\tau}\rfloor}\left(F\left(x,s^k\right)-\Phi\left(x\right)\right)+\left(\frac{t+\Delta t}{\tau}-\left\lfloor\frac{t+\Delta t}{\tau}\right\rfloor\right)\left(F\left(x,s^{\lfloor\frac{t+\Delta t}{\tau}\rfloor+1}\right)-\Phi\left(x\right)\right).\end{aligned}$$

Accordingly, we extend the definition of $P_\lambda(K)$ to noninteger $K$ as follows:

$$P_\lambda(K) = \max_{0\le\alpha\le1}\max_{x\in\Omega} P\Bigg(\frac{1}{K}\Bigg\|(1-\alpha)F\left(x,S^1\right) + \sum_{k=2}^{\lfloor K+\alpha\rfloor} F\left(x,S^k\right)$$
$$+(K+\alpha-\lfloor K+\alpha\rfloor)F\left(x,S^{\lfloor K+\alpha\rfloor+1}\right) - K\Phi(x)\Bigg\| > \lambda\Bigg). \tag{6.7}$$

It is not difficult to see that the bounds (6.4) remain valid and that

$$P\left(\left\|\int_t^{t+\Delta t}\left(F\left(x,s\left(u\right)\right)-\Phi\left(x\right)\right)du\right\| > \Delta t\cdot\lambda\right) \le P_\lambda\left(\frac{\Delta t}{\tau}\right) \tag{6.8}$$

for any $t\ge0,\ \Delta t\ge0,\ \lambda>0,\ x\in\Omega$.

**6.2. Important lemmas.** In the following, we prove that the trajectory of the blinking system and the trajectory of the averaged system stay close together for a certain interval of time. We investigate in detail the relation between the switching time $\tau$, the precision $\varepsilon$, the length of the time interval $T$, and the probability that a solution of the blinking system approximates with precision $\varepsilon$ during a time interval of length $T$ the solution of the averaged system starting from the same initial state.

We proceed in two steps. First, we derive the bounds on these probabilities for a small time interval $[t,t+\Delta t]$, conditioned on the state at time $t$. Second, we combine the bounds for the small time intervals into a bound for the time interval $[0,T]$. That is, in the first step, we prove that for some constant $\alpha>0$

$$\|x(t+\Delta t)-\xi(t+\Delta t)\| \le \|x(t)-\xi(t)\| + \alpha\Delta t.$$

In addition, we prove that the following inequality holds for a set of switching sequences with probability at least $1-P_\lambda\left(\frac{\Delta t}{\tau}\right)$, where $P_\lambda$ is given by (6.1) and (6.7):

$$\|x(t+\Delta t)-\xi(t+\Delta t)\| \le (1+L_\Phi\Delta t)\,\|x(t)-\xi(t)\| + 2\lambda\Delta t.$$

The probability $P_\lambda$ goes to zero as the argument goes to infinity, which is the case when the switching time goes to zero. In addition, typically $2\lambda$ is much smaller than $\alpha>0$.

In the second step, we prove that during a time $T$ we have the inequality

$$\|x(t)-\xi(t)\| \le \left[\left(e^{L_\Phi t}-1\right)\frac{2}{L_\Phi} + \frac{2\alpha}{L_{F\Phi}}\right]\cdot\lambda, \tag{6.9}$$

with a probability that converges to 1 as the switching time goes to zero, where the constant $L_{F\Phi}$ has a value between $L_F$ and $L_\Phi$. By an appropriate choice of $\lambda$, the LHS of (6.9) can be made arbitrarily small during an arbitrarily long time interval with a probability arbitrarily close to 1 if the switching time is small enough, as claimed by Theorem 2.2. However, the dependence of the time interval on the switching time $\tau$ is far from uniform.

*First step*: We prove the following.

Lemma 6.2.

1. *For any* $t, \Delta t \geq 0$

$$\|x(t+\Delta t) - \xi(t+\Delta t)\| \leq \|x(t) - \xi(t)\| + \alpha\,\Delta t, \text{ where} \tag{6.10}$$

$$\alpha = (B_F + B_\Phi). \tag{6.11}$$

2. *For any* $\lambda > 0$*, for* $t \geq 0$*, for any value of* $x(t)$ *(which depends on the switching sequence), and for any* $\Delta t$ *satisfying*

$$0 \leq \Delta t \leq \frac{2\lambda}{L_F B_F + L_\Phi B_\Phi}, \tag{6.12}$$

*the probability that*

$$\|x(t+\Delta t) - \xi(t+\Delta t)\| \leq (1 + L_\Phi \Delta t)\|x(t) - \xi(t)\| + 2\lambda\Delta t \tag{6.13}$$

*holds is larger than* $1 - P_\lambda\left(\frac{\Delta t}{\tau}\right)$ *(cf. (6.1)).*

*Proof.*

1. For any $t,\ \Delta t \geq 0$

$$\begin{aligned}
&\|x(t+\Delta t) - \xi(t+\Delta t)\| \\
&\leq \left\|x(t) + \int_t^{t+\Delta t} F(x(u), s(u))\,du - \xi(t) - \int_t^{t+\Delta t} \Phi(\xi(u))\,du\right\| \\
&\leq \|x(t) - \xi(t)\| + \int_t^{t+\Delta t} \|F(x(u), s(u))\|\,du + \int_t^{t+\Delta t} \|\Phi(\xi(u))\|\,du \\
&\leq \|x(t) - \xi(t)\| + (B_F + B_\Phi)\,\Delta t,
\end{aligned}$$

where we have used (3.1).

2.

$$\begin{aligned}
&\|\ x(t+\Delta t) - \xi(t+\Delta t)\ \| \\
&\leq \left\|x(t) + \int_t^{t+\Delta t} F(x(u), s(u))\,du - \xi(t) - \int_t^{t+\Delta t} \Phi(\xi(u))\,du\right\| \\
&\leq \|x(t) - \xi(t)\| + \left\|\int_t^{t+\Delta t} [F(x(u), s(u)) - F(x(t), s(u))]\,du\right\| \\
&+ \left\|\int_t^{t+\Delta t} [F(x(t), s(u)) - \Phi(x(t))]\,du\right\| + \left\|\int_t^{t+\Delta t} [\Phi(x(t)) - \Phi(\xi(t))]\,du\right\| \\
&\qquad + \left\|\int_t^{t+\Delta t} [\Phi(\xi(u)) - \Phi(\xi(t))]\,du\right\|.
\end{aligned} \tag{6.14}$$

The fourth term on the RHS of (6.14) can be bounded by applying (3.2):

$$\begin{aligned}
\left\|\int_t^{t+\Delta t} [\Phi(x(t)) - \Phi(\xi(t))]\ du\right\| &= \|\Phi(x(t)) - \Phi(\xi(t))\|\,\Delta t \\
&\leq L_\Phi \|x(t) - \xi(t)\|\,\Delta t.
\end{aligned}$$

The second term can be bounded using both (3.1) and (3.2):

$$
\begin{aligned}
\left\| \int_t^{t+\Delta t} \left[ F\left(x\left(u\right), s\left(u\right)\right) - F\left(x\left(t\right), s\left(u\right)\right)\right] du \right\| &\le L_F \int_t^{t+\Delta t} \left\| x\left(u\right) - x\left(t\right)\right\| du \\
&= L_F \int_t^{t+\Delta t} \left\| \int_t^u F\left(x\left(v\right), s\left(v\right)\right) dv \right\| du \\
&\le L_F B_F \int_t^{t+\Delta t} \int_t^u dv\, du \;=\; L_F B_F \frac{\left(\Delta t\right)^2}{2}. \qquad (6.15)
\end{aligned}
$$

Similarly, for the fifth term,

$$
\left\| \int_t^{t+\Delta t} \left[ \Phi\left(\xi\left(u\right)\right) \;-\; \Phi\left(\xi\left(t\right)\right)\right] du \right\| \le L_\Phi B_\Phi \frac{\left(\Delta t\right)^2}{2}. \qquad (6.16)
$$

Adding (6.15) and (6.16) and applying (6.12) yields the bound $\lambda \Delta t$.

The third term is bounded, according to (6.8) with probability larger than $1 - P_\lambda\left(\frac{\Delta t}{\tau}\right)$, by

$$
\left\| \int_t^{t+\Delta t} \left[ F\left(x\left(t\right), s\left(u\right)\right) \;-\; \Phi\left(x\left(t\right)\right)\right] du \right\| \le \Delta t \lambda.
$$

Collecting all these inequalities leads to (6.13). ■

*Second step*: For any choice of the initial state $x(0) = \xi(0)$ in the region $\Omega$, the constants $\lambda > 0$ and $\tau > 0$, and the constant $\Delta t$ satisfying (6.12), we consider the following sequence of random variables on the probability space of switching sequences:

$$
Z_q\left(s\right) \;=\; \left\| x\left(q\Delta t\right) - \xi\left(q\Delta t\right)\right\|, \qquad q = 0, 1, 2, \ldots. \qquad (6.17)
$$

The dependence of $Z_q(s)$ on the switching sequence $s$ on the RHS of (6.17) is through the solution $x(q\Delta t)$ of the blinking system. For simplicity, we shall suppress the argument $s$ of $Z_q$. Note that the stochastic process $Z_q$ is not a Markov process, even though $x\left(q\Delta t\right)$ is a vector-valued Markov process. In fact, the norm destroys a large part of the information contained in the state vector. However, according to (6.10) we have the deterministic bound

$$
Z_q \;\le\; Z_{q-1} + \alpha \Delta t \quad \text{for } \; q = 1, 2, \ldots,
$$

and according to (6.10) the bound

$$
Z_q \le \left(1 + L_\Phi \Delta t\right) Z_{q-1} + 2\lambda \Delta t, \qquad (6.18)
$$

which holds with probability larger than $1 - P_\lambda\left(\frac{\Delta t}{\tau}\right)$ for $q = 1, 2, \ldots$. Actually, the second part of Lemma 6.2 implies the stronger statement that the following conditional probability has the uniform bound

(6.19)

$$
P - \text{a.s.}\left(Z_q \le \left(1 + L_\Phi \Delta t\right) Z_{q-1} + 2\lambda \Delta t \;\left| x\left(\left(q-1\right)\Delta t\right)\right.\right) > 1 - P_\lambda\left(\frac{\Delta t}{\tau}\right) \text{ for } q = 1, 2, \ldots.
$$

Note that we have used the Markov property of $(x(q\Delta t))_{q\in N}$. It is convenient to introduce the auxiliary random variable

$$\theta_q = \begin{cases} 1 & \text{if } Z_q \le (1 + L_\Phi \Delta t) Z_{q-1} + 2\lambda\Delta t, \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 6.3. *Let $\sigma = (\sigma_1, \ldots, \sigma_Q) \in \{0,1\}^Q$ be a binary vector of length $Q$, and $m = Q - \sum_{q=1}^{Q} \sigma_q$ the number of zeros in this vector. Then*

$$P(\theta_Q = \sigma_Q, \ldots, \theta_1 = \sigma_1) \le \left[ P_\lambda \left( \frac{\Delta t}{\tau} \right) \right]^m. \tag{6.20}$$

*Proof.* If $\sigma_Q = 1$, we simply use

$$P(\theta_Q = 1,\ \theta_{Q-1} = \sigma_{Q-1}, \ldots, \theta_1 = \sigma_1) \le P(\theta_{Q-1} = \sigma_{Q-1}, \ldots, \theta_1 = \sigma_1). \tag{6.21}$$

If $\sigma_Q = 0$, we write

$$\begin{aligned} &P(\theta_Q = 0,\ \theta_{Q-1} = \sigma_{Q-1}, \ldots, \theta_1 = \sigma_1) \\ &\quad = \sum_{x((Q-1)\Delta t)} P(\theta_Q = 0,\ \theta_{Q-1} = \sigma_{Q-1}, \ldots, \theta_1 = \sigma_1, x((Q-1)\Delta t)) \\ &\quad = \sum_{x((Q-1)\Delta t)} P(\theta_Q = 0 \,|\, \theta_{Q-1} = \sigma_{Q-1}, \ldots, \theta_1 = \sigma_1, x((Q-1)\Delta t)) \\ &\qquad\qquad \times P(\theta_{Q-1} = \sigma_{Q-1}, \ldots, \theta_1 = \sigma_1, x((Q-1)\Delta t)), \end{aligned}$$

where the summation is over (the finite number of) all possible values of $x((Q-1)\Delta t)$. Since $\theta_Q$ depends only on $x(Q\Delta t)$ and $x((Q-1)\Delta t)$, and since $x(q\Delta t)$, $q = 1, 2, \ldots$, is a vector-valued Markov process, we get

$$\begin{aligned} &P(\theta_Q = 0,\ \theta_{Q-1} = \sigma_{Q-1}, \ldots, \theta_1 = \sigma_1) \\ &= \sum_{x((Q-1)\Delta t)} P(\theta_Q = 0 \,|\, x((Q-1)\Delta t))\, P(\theta_{Q-1} = \sigma_{Q-1}, \ldots, \theta_1 = \sigma_1, x((Q-1)\Delta t)), \end{aligned}$$

and applying the complement of (6.19),

$$P(\theta_Q = 0 \,|\, x((Q-1)\Delta t)) \le P_\lambda \left( \frac{\Delta t}{\tau} \right),$$

we obtain

$$\begin{aligned} &P(\theta_Q = \sigma_Q, \ldots, \theta_1 = \sigma_1) \\ &\quad \le P_\lambda \left( \frac{\Delta t}{\tau} \right) \sum_{x((Q-1)\Delta t)} P(\theta_{Q-1} = \sigma_{Q-1}, \ldots, \theta_1 = \sigma_1, x((Q-1)\Delta t)) \\ &\quad = P_\lambda \left( \frac{\Delta t}{\tau} \right) \cdot P(\theta_{Q-1} = \sigma_{Q-1}, \ldots, \theta_1 = \sigma_1). \end{aligned} \tag{6.22}$$

Recursively applying (6.21) or (6.22), depending on the value of the corresponding $\sigma_q$, proves the lemma. ∎

Lemma 6.4. *The following inequality holds true:*

$$Z_Q \leq \left[e^{L_\Phi(Q-m)\Delta t} - 1\right] \cdot \frac{2\lambda}{L_\Phi} + e^{L_\Phi(Q-m)\Delta t} m\alpha\Delta t \quad P(\cdot \,|\, \theta_Q = \sigma_Q, \ldots, \theta_1 = \sigma_1) \; - \; a.s., \tag{6.23}$$

*where as before* $m = Q - \sum_{q=1}^{Q} \sigma_q$ *is the number of zeros in the binary vector* $\sigma = (\sigma_1, \ldots, \sigma_Q)$.

*Proof.*

1. $m = 0$: Applying (6.18) recursively, we get

$$Z_Q \leq \sum_{q=1}^{Q} (1 + L_\Phi \Delta t)^{Q-q}\, 2\lambda\Delta t = \left[(1 + L_\Phi\, \Delta t)^Q - 1\right] \frac{2\lambda}{L_\Phi}$$
$$\leq \left[e^{L_\Phi Q\, \Delta t} - 1\right] \frac{2\lambda}{L_\Phi}.$$

2. $m > 0$: Let $\sigma_{j_1} = \sigma_{j_2} = \cdots = \sigma_{j_m} = 0$ for $1 \leq j_1 < j_2 < \cdots < j_m \leq Q$, whereas for all other $j$, $\sigma_j = 1$. Then

$$\begin{aligned} Z_Q \leq &\sum_{q=j_m+1}^{Q} (1 + L_\Phi\Delta t)^{Q-q}\, 2\lambda\Delta t + (1 + L_\Phi \Delta t)^{Q-j_m}\, \alpha\Delta t \\ &+ \sum_{q=j_{m-1}+1}^{j_m-1} (1 + L_\Phi\Delta t)^{Q-q-1}\, 2\lambda\Delta t + \cdots + \sum_{q=j_1+1}^{j_2-1} (1 + L_\Phi\Delta t)^{Q-q-m+1}\, 2\lambda\Delta t \\ &+ (1 + L_\Phi\Delta t)^{Q-j_1-m+1}\, \alpha\Delta t + \sum_{q=1}^{j_1-1} (1 + L_\Phi\Delta t)^{Q-q-m}\, 2\lambda\Delta t \\ \leq &\left[(1 + L_\Phi\Delta t)^{Q-m} - 1\right] \frac{2\lambda}{L_\Phi} + \Big[(1 + L_\Phi\Delta t)^{Q-j_1-m+1} \\ &+ (1 + L_\Phi\Delta t)^{Q-j_2-m+2} + \cdots + (1 + L_\Phi\Delta t)^{Q-j_m}\Big]\, \alpha\Delta t. \end{aligned} \tag{6.24}$$

The RHS of (6.24) is largest for the smallest $j_1, \ldots, j_m$, i.e., for $j_k = k$, and we obtain the bound that is independent of $j_k$:

$$Z_Q \leq \left[e^{L_\Phi(Q-m)\Delta t} - 1\right] \frac{2\lambda}{L_\Phi} + e^{L_\Phi(Q-m)\Delta t} m\alpha\Delta t. \qquad \blacksquare$$

Lemma 6.5.

$$P\left(\sum_{q=1}^{Q} \theta_q < Q\right) \leq \left(1 + P_\lambda\left(\frac{\Delta t}{\tau}\right)\right)^Q - 1 \leq \exp\left(Q P_\lambda\left(\frac{\Delta t}{\tau}\right)\right) - 1. \tag{6.25}$$

*Proof.* According to (6.20),

$$P\left(\sum_{q=1}^{Q} \theta_q < Q\right) \leq \sum_{m=1}^{Q} \binom{Q}{m} \left[P_\lambda\left(\frac{\Delta t}{\tau}\right)\right]^m = \sum_{m=0}^{Q} \binom{Q}{m} \left[P_\lambda\left(\frac{\Delta t}{\tau}\right)\right]^m - 1$$
$$\leq \left(1 + P_\lambda\left(\frac{\Delta t}{\tau}\right)\right)^Q - 1 \leq e^{Q P_\lambda\left(\frac{\Delta t}{\tau}\right)} - 1. \qquad \blacksquare$$

Lemma 6.6. *For arbitrary $\tau > 0$, $T > 0$, and $\lambda > 0$, the probability $P_{\tau,T,\lambda}$ that*

$$\|x(t) - \xi(t)\| > \left(\left[e^{L_\Phi t} - 1\right] \frac{2}{L_\Phi} + \frac{2}{L_{F\Phi}}\right)\lambda$$

*for at least one value of $t$ with $0 \le t \le T$ is bounded by*

$$P_{\tau,T,\lambda} \le \exp\left(\frac{T}{\Delta t} P_\lambda\left(\frac{\Delta t}{\tau}\right)\right) - 1,$$

*where*

$$\Delta t = \frac{2\lambda}{L_F B_F + L_\Phi B_\Phi}, \qquad L_{F\Phi} = \frac{L_F B_F + L_\Phi B_\Phi}{B_F + B_\Phi}. \tag{6.26}$$

*Proof.* Let

$$Q = \left\lfloor \frac{T}{\Delta t} \right\rfloor,$$

where $\lfloor . \rfloor$ denotes the integral part. Consider a switching sequence $s$ such that

$$\theta_1(s) = \theta_2(s) = \cdots = \theta_Q(s) = 1.$$

Then, according to (6.23) for $q = 1, \dots, Q$

$$Z_q(s) \le \left[e^{L_\Phi q \Delta t} - 1\right] \cdot \frac{2\lambda}{L_\Phi};$$

i.e.,

$$\|x(t) - \xi(t)\| \le \left[e^{L_\Phi t} - 1\right] \cdot \frac{2\lambda}{L_\Phi} \quad \text{for } t = 0, \Delta t, 2\Delta t, \dots, Q\Delta t.$$

For the other values of $t$ in the interval $[0, T]$, using (6.10), we get

$$\begin{aligned}
\|x(t) - \xi(t)\| &= \left\| x\left(\left\lfloor \frac{t}{\Delta t} \right\rfloor \Delta t\right) - \xi\left(\left\lfloor \frac{t}{\Delta t} \right\rfloor \Delta t\right) \right\| + \alpha \cdot \left(t - \left\lfloor \frac{t}{\Delta t} \right\rfloor \Delta t\right) \\
&\le \left[\exp\left(L_\Phi \left\lfloor \frac{t}{\Delta t} \right\rfloor \Delta t\right) - 1\right] \cdot \frac{2\lambda}{L_\Phi} + \alpha \Delta t \\
&\le \left[e^{L_\Phi t} - 1\right] \cdot \frac{2\lambda}{L_\Phi} + \alpha \Delta t.
\end{aligned}$$

Setting $\Delta t$ to the upper limit in (6.12) and substituting the value (6.11) for $\alpha$, we get

$$\|x(t) - \xi(t)\| \le \left(\left[e^{L_\Phi t} - 1\right] \frac{2}{L_\Phi} + \frac{2}{L_{F\Phi}}\right)\lambda. \tag{6.27}$$

This implies that for any switching sequence $s$ such that the inequality (6.27) is violated for at least one value of $t$ in the interval $[0, T]$, we must have

$$\sum_{q=1}^{Q} \theta_q(s) < Q.$$

Applying now (6.25), we obtain

$$P_{\tau,T,\lambda} \le P\left(\sum_{q=1}^{Q} \theta_q(s) < Q\right) \le \exp\left(Q P_\lambda\left(\frac{\Delta t}{\tau}\right)\right) - 1 \le \exp\left(\frac{T}{\Delta t} P_\lambda\left(\frac{\Delta t}{\tau}\right)\right) - 1. \qquad \blacksquare$$

**6.3. Proof of Theorem 3.1.** Having derived the lemmas, we are ready to give the proof of the main result, formulated in Theorem 3.1.

*Proof.* Choose

$$\lambda = \frac{\varepsilon}{C(T)}. \tag{6.28}$$

Then, by Lemma 6.6 inequality (3.3) holds for all times $t \in [0, T]$ with probability at least $1 - \left(e^{\frac{T}{\Delta t} P_\lambda(\frac{\Delta t}{\tau})} - 1\right)$, where $\Delta t$ is given by (6.26). According to Lemma 6.1,

$$\frac{T}{\Delta t} P_\lambda \left(\frac{\Delta t}{\tau}\right) \leq 2 \frac{T}{\Delta t} N \exp\left(-\frac{\Delta t}{\tau} \frac{\lambda^2}{2NB_F^2}\right). \tag{6.29}$$

For a given $0 < P_0 < 1$ choose $\tau_0$ such that the RHS of (6.29) equals $P_0/2$. Explicitly,

$$\tau_0 = \left(\frac{\varepsilon}{C(T)}\right)^3 \cdot \frac{1}{\ln(2D \cdot TC(T)) - \ln(\varepsilon P_0)} \cdot \frac{1}{DB_F^2},$$

where we have used (6.28), (3.6), and (6.26). It also implies that $\frac{T}{\Delta t} P_\lambda(\frac{\Delta t}{\tau_0}) < \frac{1}{2}$, and therefore for $\tau \leq \tau_0$

$$\exp\left(\frac{T}{\Delta t} P_\lambda \left(\frac{\Delta t}{\tau}\right)\right) - 1 \leq \frac{T}{\Delta t} P_\lambda \left(\frac{\Delta t}{\tau}\right) \exp\left(\frac{T}{\Delta t} P_\lambda \left(\frac{\Delta t}{\tau}\right)\right) \leq 2 \frac{T}{\Delta t} P_\lambda \left(\frac{\Delta t}{\tau}\right) \leq P_0.$$

Therefore, inequality (3.3) holds for all times $t \in [0, T]$ with probability at least $P_{close} = 1 - P_0$ if the switching time satisfies $\tau \leq \tau_0$. ∎

## REFERENCES


[1] I. Belykh, V. Belykh, and M. Hasler, *Blinking model and synchronization in small-world networks with a time-varying coupling*, Phys. D, 195 (2004), pp. 188–206.

[2] M. Hasler, V. Belykh, and I. Belykh, *Dynamics of stochastically blinking systems. Part* II: *Asymptotic properties*, SIAM J. Appl. Dyn. Syst., 12 (2013), pp. 1031–1084.

[3] J.-L. Lagrange, *Mécanique Analytique*, Albert Blanchard, Paris, 1788.

[4] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste, Vol.* I, Gauthiers-Villars, Paris, 1892.

[5] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste, Vol.* II, Gauthiers-Villars, Paris, 1893.

[6] N. N. Krylov and N. N. Bogoliubov, *Introduction to Nonlinear Mechanics*, Izd. AN UkSSR, Kiev, 1937 (in Russian).

[7] N. N. Bogoliubov and Yu. A. Mitropolsky, *Asymptotic Methods in the Theory of Nonlinear Oscillations*, Gordon and Breach, New York, 1961.

[8] C. Beck, *Ergodic properties of a kicked damped particle*, Comm. Math. Phys., 130 (1990), pp. 51–60.

[9] A. Skorokhod, F. Hoppensteadt, and H. Salehi, *Random Perturbation Methods*, Springer-Verlag, New York, 2002.

[10] M. I. Freidlin and A. D. Wentzell, *Diffusion processes on graphs and the averaging principle*, Ann. Probab., 21 (1993), pp. 2215–2245.

[11] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer-Verlag, New York, 1998.

[12] Yu. Kifer, *Large Deviations and Adiabatic Transitions for Dynamical Systems and Markov Processes in Fully Coupled Averaging*, Mem. Amer. Math. Soc. 201, AMS, Providence, RI, 2009.

[13] R. Z. Khaśminskii, *A limit theorem for the solutions of differential equations with random right-hand sides*, Theory Probab. Appl., 11 (1966), pp. 390–406.

[14] M. A. Pinsky, *Instability of the harmonic oscillator with small noise*, SIAM J. Appl. Math., 46 (1986), pp. 451–463.
[15] J. Sanders, F. Verhulst, and J. Murdock, *Averaging Methods in Nonlinear Dynamical Systems*, Springer, New York, 2007.
[16] X. Mao and C. Yuan, *Stochastic Differential Equations with Markovian Switching*, Imperial College Press, London, 2006.
[17] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc., 58 (1963), pp. 13–30.
[18] M. Loève, *Probability Theory* 1, 4th ed., Springer-Verlag, Berlin, 1977.