

# Exponential Convergence of Degenerate Hybrid Stochastic Systems with Full Dependence

Svetlana V. Anulova and Alexander Yu. Veretennikov

**Abstract** This research stems from a control problem for a suspension device. For a general class of switching stochastic mechanical systems (including closed-loop control ones), we establish the following: (1) existence and uniqueness of a weak solution and its strong Markov property, (2) mixing property in the form of the local Markov–Dobrushin condition, and (3) exponentially fast convergence to the unique stationary distribution. These results are proved for discontinuous coefficients under nondegenerate disturbances in the force field; for (3) a stability condition is additionally imposed. Linear growth of coefficients is allowed.

## 1 Introduction

Convergence of marginal distributions of (Markov) stochastic systems to a stationary one has been thoroughly studied, and there are classic schemes for proving this property. At the level of ideas, if two facts are established—a version of Doeblin’s condition and recurrence—then this provides convergence. As a version of the former we use the *local Markov–Dobrushin condition*. Quite often it is provided by the nondegeneracy of the Wiener disturbance. However, in this paper we deal with mechanical systems presented by highly degenerate stochastic differential equations of the Langevin–Smoluchowski type driven by a Wiener noise of “smaller

---

S.V. Anulova

Institute for Control Science, Russian Academy of Sciences, Moscow, Russia

e-mail: [anulovas@ipu.rssi.ru](mailto:anulovas@ipu.rssi.ru)

A.Y. Veretennikov (✉)

School of Mathematics, University of Leeds, Leeds, LS2 9JT, UK, and Institute of Information

Transmission Problems, Russian Academy of Sciences, Moscow, Russia

e-mail: [veretenn@maths.leeds.ac.uk](mailto:veretenn@maths.leeds.ac.uk)

dimension". The Wiener term presents a random force in the velocity component equation, while the equation for the state component has, naturally, no Wiener term. Thus, the system is in essence degenerate, and even standard existence and uniqueness results for it require revision.

The recurrence property of a stochastic system may often be reduced to the stability of the corresponding deterministic system (with the removed disturbance term). We establish it in terms of quadratic Lyapunov functions for deterministic systems with switching. The investigation of existence of such functions for general systems is still in progress nowadays; see [7]. Eventually, exponential convergence of marginal distributions in the total variation norm  $\|\cdot\|_{TV}$  will be established.

This work, in fact, stems from the investigation of Campillo and Pardoux into the issue of a vehicle suspension device; see [8, 9].

Stochastic ergodic control—in particular, with expected average in time with infinite horizon as cost functional—proved to be a useful tool for constructing a closed-loop control of a vehicle suspension device; see [8, 12] and references therein. In [3] we have generalized the model of the suspension device to a multi-regime one. That is, we admitted several types of the road surface and assumed that the type of the road surface determines a gear box regime and hence also a working regime of the suspension device. This object may be described by a hybrid system (see [6]) with dynamics of a switching diffusion: position of the device  $X$ , its velocity  $Y$  and the type of the road surface  $V$  (the discrete component). Switchings constitute a Markov chain (see [15, 23, 24]). The novelty in comparison to the earlier works is degenerate diffusion and discontinuous coefficients; the former is due to the nature of the device, while the latter is caused by the control framework—optimal control is never smooth. Similar equations without switching have been studied in [1].

The crucial point in applying the technique of ergodic control is establishing the ergodicity property of the controlled process. Our result in [3] is ergodicity in the sense of Markov processes, the state space type (see [18], [17, Ch.6.3]). Moreover, we have shown that under every (homogeneous) admissible control policy, the distribution of the controlled process converges in time to its limit at an exponential rate. The rate of convergence is uniform over all admissible control policies and locally uniform with respect to initial conditions. We emphasize that *control problems* themselves are *not* addressed in the present paper.

We sketch briefly the contents of [3]—and simultaneously some results from [1] as a partial case—in the next section in order to make intelligible the motivation and reasoning of the present work and the investigation in progress.

The paper consists of Introduction (Sect. 1), Reminder about an earlier background model (Sect. 2), Main Results (Sect. 3) and Proofs of Theorems 1 and 2; Sect. 4 contains the proof of Theorem 3 on just three lines and the proof of Theorem 4 given as a sketch with references.

## 2 Case of Independent Markov Switchings: Reminder

Consider a two-dimensional stochastic differential equation,

$$\begin{aligned} dX_t &= Y_t dt, & X_0 &= x, \\ dY_t &= b(X_t, Y_t, V_t) dt + \sigma(V_t) dW_t, & Y_0 &= y. \end{aligned} \tag{1}$$

Here  $W$  and  $V$  are independent driving processes: the standard Wiener process and a Markov chain, taking values in a finite set  $\mathcal{S} = \{1, 2, \dots, N\}$ ,  $t \geq 0$ . The generator of  $V$  is a matrix  $Q = (q_{ij})_{N \times N}$ , which determines transition probabilities over a small period of time  $\Delta \downarrow 0$ :

$$\mathbf{P}(V_{t+\Delta} = j | V_t = i) = \begin{cases} q_{ij} \Delta + o(\Delta) & \text{if } i \neq j, \\ 1 + q_{ii} \Delta + o(\Delta) & \text{if } i = j. \end{cases}$$

All intensities are positive,  $q_{ij} > 0$  for  $i \neq j$ ; for  $j = i$  the value  $q_{ii}$  is defined as  $q_{ii} = -\sum_{j: j \neq i} q_{ij}$ . All trajectories of  $V$  are right-continuous step functions without accumulations of jumps (recall that  $\mathcal{S}$  is finite and consequently  $\max\{q_{ij}, i, j \in \mathcal{S}, i \neq j\} < \infty$ ).

Further,

$$b(x, y, v) = -u(x, y, v)y - \beta x - \gamma(v) \operatorname{sign}(y). \tag{2}$$

Here a function  $u$  (the control policy) is Borel measurable and satisfies  $u \in [u_1, u_2]$  with two constants  $u_1 \leq u_2$ . It is assumed that

$$u_1 > 0, \quad \beta > 0, \quad \min_v \gamma(v) > 0, \quad \min_v \sigma(v) > 0. \tag{3}$$

System (1) describes a mechanical “semi-active” suspension device in a vehicle under external stochastic perturbation forces treated as a white noise. The original model without switching  $V$  was suggested in [8]. In [3] it was extended to various road types by introducing switching.

In [3] the behavior of the stochastic system (1) under a fixed control policy  $u$  was studied, namely, how fast does the system approach its stationary regime. This may be measured by the distance in total variation. Important preliminary results about existence and uniqueness of solutions have been established. We have shown that under our assumptions the stationary regime exists and is unique. It is the discontinuity of  $u$  and the degeneracy of the equation that hinders the derivation of our results directly from the general theory of stochastic differential equations.

In the following theorems (quoted from [3]) we fix the values  $x, y, v$ — initial conditions for the system (1) and for the driving Markov chain. Existence and uniqueness are understood in a weak sense; see [14, Chap. IV, Definitions 1.2 and 1.4].

**Proposition 1 ([3]).** *Under the assumptions (3), the system (1) has a weak solution on  $[0, \infty)$  unique in distribution. The joint process  $(X, Y, V)$  is also unique in distribution, and these distributions form a strong Markov process.*

Denote the marginal distribution of the triple  $(X_t, Y_t, V_t)$  with initial data  $x, y, v$  by  $\mu_t^{x,y,v}$ ,  $t \geq 0$ .

**Proposition 2 ([3]).** *Under the assumptions (3), there exists a stationary probability distribution  $\mu_\infty$  on  $\mathbb{R}^2 \times \mathcal{S}$  and positive constants  $\bar{C}, \bar{c}$  depending on  $\min\{\beta(v), v \in \mathcal{S}\}, \min\{\gamma(v), v \in \mathcal{S}\}, \min\{\sigma(v), v \in \mathcal{S}\}, \max\{\sigma(v), v \in \mathcal{S}\}, \min\{q_{ij}; i, j \in \mathcal{S}, i \neq j\}, \max\{q_{ij}; i, j \in \mathcal{S}, i \neq j\}, u_1, u_2, N$ , such that*

$$\|\mu_t^{x,y,v} - \mu_\infty\|_{TV} \leq \bar{C} \exp(-\bar{c}t)(1 + x^2 + y^2), \quad t \geq 0. \tag{4}$$

The specification of  $\bar{C}, \bar{c}$  assures the rate of convergence to be uniform over all admissible control policies and locally uniform with respect to initial conditions, as stated in the Introduction. Note that although the fixed parameters  $x, y$  and  $v$  (initial values) are not shown in the left-hand side of (4), the measure  $\mu_t$  does depend on them.

### 3 Main Results

#### 3.1 The Model

We want to extend the results of [3] in two directions: (1) to consider general multidimensional mechanical systems and (2) to allow state-dependent switching.

From the theoretical mechanics point of view, we extend the model (1) from the case of one point mass to an ensemble of  $d$  point masses being under the influence of a combined force—the resultant of a force field, friction and interaction.

Let  $d \geq 1$  and consider a system of stochastic differential equations in  $\mathbb{R}^{2d}$ :  
for given  $x^1, x^2 \in \mathbb{R}^d$ , and  $t \geq 0$ ,

$$\begin{aligned} dX_t^1 &= X_t^2 dt, & X_s^1 &= x^1 \in \mathbb{R}^d, \\ dX_t^2 &= b(X_t^1, X_t^2) dt + dW_t, & X_s^2 &= x^2 \in \mathbb{R}^d. \end{aligned} \tag{5}$$

Here  $W$  is a  $d$ -dimensional Wiener process and the drift term  $b$  is a  $d$ -dimensional function. The value  $d > 1$  corresponds to the multi-particle case.

Denote  $X = (X^1, X^2) \in \mathbb{R}^{2d}$ .

Let us now explain what is state-dependent switching. Consider a process  $X_t$ ,  $t \geq 0$ , which is a solution of a stochastic differential equation with coefficients additionally depending on a process  $(V_t, t \geq 0)$  taking values in a finite set  $S = \{1, 2, \dots, N\}$ . The process  $V$  is, informally speaking, a conditional Markov chain: given a “frozen” value of  $X_t = x$ , its generator equals  $Q(x) = (q_{ij}(x))_{N \times N}$ ,  $x \in$

$\mathbb{R}^{2d}$ . Informally, this matrix determines transition probabilities over a small period of time given  $X_t = x$ ,

$$\mathbf{P}(V_{t+\Delta} = j | V_t = i, X_t = x) = \begin{cases} q_{ij}(x)\Delta + o(\Delta), & i \neq j, \\ 1 + q_{ii}(x)\Delta + o(\Delta), & i = j, \end{cases} \tag{6}$$

where  $\Delta \downarrow 0$ . For  $j = i$  the value  $q_{ii}(x)$  is defined as  $q_{ii}(x) := -\sum_{j: j \neq i} q_{ij}(x)$ .

Finally, consider a hybrid SDE system  $(X^1, X^2, V)$  in  $\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{S} = \mathbb{R}^{2d} \times \mathcal{S}$ ,  $d \geq 1$ :

for given  $x^1, x^2 \in \mathbb{R}^d, v \in \mathcal{S}$

$$\begin{aligned} dX_t^1 &= X_t^2 dt, \\ dX_t^2 &= b(X_t^1, X_t^2, V_t) dt + \sigma(V_t) dW_t, \\ t \geq 0, X_0^1 &= x^1 \in \mathbb{R}^d, X_0^2 = x^2 \in \mathbb{R}^d, V_0 = v \in \mathcal{S}. \end{aligned} \tag{7}$$

Here  $W$  is a  $d$ -dimensional Wiener process, the drift term  $b$  is a  $d$ -dimensional function, and  $\sigma(v)$  is a nondegenerate  $d \times d$ -matrix. In order to define this object rigorously, we should describe it through its two-component generator  $L\phi(x, v) =$

$$\begin{aligned} &\left( \frac{\partial \phi}{\partial x^1}(x^1, x^2, v), x^2 \right) + \left( \frac{\partial \phi}{\partial x^2}(x^1, x^2, v), b(x^1, x^2, v) \right) \\ &+ \sum_{i,j=1}^d \sigma_{ij}(v) \frac{\partial^2 \phi}{\partial x_i^2 \partial x_j^2}(x^1, x^2, v) \\ &+ \sum_{j \in \mathcal{S} \setminus v} (\phi(x, j) - \phi(x, v)) q_{vj}(x). \end{aligned} \tag{8}$$

Here generator may be understood in the sense of the martingale problem (see Sect. 5.1 of [4], or [10]); in some papers it is called *extended* generator. This description also makes sense for discontinuous intensities  $q_{ij}$ .

Recall that due to the control origin of the model, no regularity may be assumed about the drift term  $b$ : it is just Borel measurable and of a no more than linear growth.

### 3.2 Standing Assumptions

The following assumptions are standing for the system (7).

The values  $d, N$  are natural numbers; the points of the euclidean space  $\mathbb{R}^{2d}$  are denoted  $x = (x^1, x^2)$  (the first and the last  $d$  coordinates);  $\mathcal{S}$  is the set  $\{1, \dots, N\}$ .

(SA 1) Dimension and measurability:  $b(x, v) : \mathbb{R}^{2d} \times \mathcal{S} \rightarrow \mathbb{R}^d$ ,  $\sigma(v) : \mathcal{S} \rightarrow \mathbb{R}^d \times \mathbb{R}^d$  and  $Q : \mathbb{R}^{2d} \rightarrow \mathbb{R}^N \times \mathbb{R}^N$  are Borel measurable functions.

(SA 2) Nondegenerate diffusion: the matrix  $\sigma(i)\sigma(i)^*$  is nondegenerate for any  $i \in \mathcal{S}$ .

(SA 3) Uniform linear growth: there exists a constant  $C > 0$  such that for all  $x \in \mathbb{R}^{2d}$  and  $v \in \mathcal{S}$ ,

$$|b(x, v)| + \|\sigma(v)\| \leq C(1 + |x|). \quad (9)$$

(SA 4) Intensity bounds: there exist constants  $0 < c_l \leq c_u < \infty$  such that  $c_l \leq q_{ij}(x) \leq c_u$  for all  $x \in \mathbb{R}^{2d}$  and  $i, j \in \mathcal{S}, i \neq j$ ; also,  $q_{ii}$  is defined as  $q_{ii} := -\sum_{j: j \neq i} q_{ij}, i \in \mathcal{S}$ .

### 3.3 Recurrence Assumption

This assumption about a Lyapunov function will be used only in the Sects. 3.6 and 3.7.

(RA 1) There exist a positive definite quadratic function  $\phi : \mathbb{R}^{2d} \rightarrow [0, \infty)$  and positive constants  $c_1, c_2$  such that

$$\left( \frac{\partial \phi}{\partial x^1}(x), x^2 \right) + \left( \frac{\partial \phi}{\partial x^2}(x), b(x, v) \right) \leq -c_1 \phi(x) + c_2 \text{ for all } (x, v) \in \mathbb{R}^{2d} \times \mathcal{S}. \quad (10)$$

The class of systems satisfying (10) is non-empty: indeed, it includes the system (1)–(2) under the assumption (3) and other likewise models. Proposition 2 itself, actually, prompts why we wish to restrict Lyapunov functions to quadratic ones here. Another argument will be given after the Theorem 3.

### 3.4 Weak Existence and Uniqueness

Existence and uniqueness are understood in a weak sense; see [14, Chap. IV, Definitions 1.2 and 1.4].

**Theorem 1.** *Under the assumptions (SA 1)–(SA 4), the system (7) has a weak solution on  $[0, \infty)$  unique in distribution. This solution forms a strong Markov process.*

Existence and uniqueness for the solution of the considered system may be explained as follows. Take a process with no switching (constructed in [19]) and attach to it a random moment, which is a minimum of all stopping times defined

by the switching intensities of transitions to all other discrete states. That is, conditioned on the trajectory of the process, all distributions of these stopping times are “exponential” with corresponding (variable) intensities and independent of each other. Thus, the switched process is constructed up to the first switching. It is clear that its distribution up to the switching moment coincides with that of any solution of the system (7). This construction may be continued further inductively, from one switching moment to the next, and the scheme can be implemented in terms of stochastic differential equations with “rare” jumps—analogs of switchings. Such jumps can be generated with minimal restrictions on jump coefficients—only measurability is required; see [2, 5] and [14, Chap. IV, Sect. 9]. Strong Markov property in [19] was deduced from the Krylov selection method [16], more precisely, due to weak uniqueness. In the present paper, the same idea is helpful and the pasting construction used for establishing existence does preserve the strong Markov property. This procedure will be sketched in the proof of Lemma 2 in the Sect. 4.2.

### 3.5 Local Markov–Dobrushin Condition

This condition describes the following property of the process satisfying (7). For any two initial states at time zero, let us consider two corresponding processes; then, fix some moment of time and compare *marginal* distributions of the processes at this moment; then, they are *non-singular* and in a certain sense even uniformly in initial states.

This fact is non-trivial, but a simple “philosophical” background for such non-singularity is Girsanov’s formula. However, the stochastic integral under the exponent with a non-bounded drift along with degeneracy makes the implementation of this idea technically involved. Namely, to make sure that expressions like  $\exp(\int_0^T |b|^2(X_r, V_r))$  are bounded, we will need to consider restricted measures  $\mu^{R'}$  with  $R' < \infty$  instead of simple  $\mu$ ; see the next paragraph.

Let us define the following objects:  $B_R = \{x \in \mathbb{R}^{2d} : |x| < R\}$ ;  $\mu_{s,s+T}(x, v; dydu)$  denotes the transition measure from  $(s, x, v)$  to  $(s + T; dydu)$ ;  $\mu_{s,s+T}^{R'}(x, v; dydu)$  is the restriction of the transition measure  $\mu_{s,s+T}(x, v; dydu)$  to trajectories, whose continuous component does not go beyond the boundary of  $B_{R'}$  on  $[s, s + T]$ ; by definition,  $B_{+\infty} = \mathbb{R}^{2d}$ .

The local Markov–Dobrushin condition, which we need, is formulated for a fixed triple  $T > 0, R > 0, R' \in [R, +\infty]$ :

$$\inf_{s \in [0, \infty)} \inf_{\substack{x, x' \in B_R, \\ v, v' \in \mathbb{S}}} \mu_{s,s+T}^{R'}(x, v) \wedge \mu_{s,s+T}^{R'}(x', v')(B_{R'} \times \mathbb{S}) > 0. \tag{11}$$

Here the minimum  $\mu \wedge \nu$  of two measures  $\mu$  and  $\nu$  is understood in the following way:

$$\mu \wedge \nu(A) := \int_A \left( \frac{d\mu}{d(\mu + \nu)} \wedge \frac{d\nu}{d(\mu + \nu)} \right) (\omega) (\mu + \nu)(d\omega).$$

*Remark 1.* The local Markov–Dobrushin condition, formulated for non-random initial conditions, implies immediately the same statement for distributed initial conditions. Definition (11) suits a nonhomogeneous case; in the homogeneous situation it suffices to take  $s = 0$  and drop  $\inf_{s \in [0, \infty)}$ .

**Theorem 2.** *Under the assumptions (SA 1)–(SA 4), for any  $R > 0$  there exist  $T > 0$  and  $R' > R$  such that the local Markov–Dobrushin condition (11) holds.*

### 3.6 Recurrence

Recurrence of stochastic systems is closely related to stability of deterministic systems and often may be reduced to it, although, in some cases, random perturbations may unexpectedly have a positive effect on the recurrence of the system; see [11, 13, 15]. (It is not unexpectedly that the opposite cases also occur.)

We shall conclude the recurrence from the existence of a quadratic Lyapunov function for our system with the removed stochastic term. It is interesting that the problem of existence of quadratic Lyapunov functions is yet unsolved in full generality even for linear deterministic switching systems.

**Theorem 3.** *Suppose the assumption (RA 1) from the Sect. 3.3 with a function  $\phi$  is fulfilled. Then, there exist positive constants  $c'_1, c'_2$  such that for the generator  $L$  given by (8) the following inequality holds:*

$$L\phi(x, v) \leq -c'_1\phi(x) + c'_2, \quad x \in \mathbb{R}^{2d}, v \in \mathcal{S}. \quad (12)$$

*The constants  $c'_1, c'_2$  depend on a function  $\phi$ , constants  $c_1, c_2$ , a growth constant  $C$  from the inequality (9) and on dimension  $d$ .*

The *proof* follows straightforward, as the second-order term in  $L\phi$  adds a constant to the first-order expression and since  $\lim_{|x| \rightarrow \infty} \phi(x) = +\infty$ . Note that it shows that  $c'_1 = c_1$ .  $\square$

*Remark 2.* One more reason why Lyapunov functions here are restricted to quadratic ones is our concern not to overcomplicate the presentation. Indeed, in the quadratic class, the inequality (12) follows easily, while for a general function, we would need strange additional assumptions; yet, clearly, such a class is wider than only quadratic functions.

Applying Ito’s or Dynkin’s formula to  $\phi(X_t, V_t)$ —the latter being equivalent to the martingale property, at least, for the appropriately stopped process—it is possible to show the following result. Let  $\tau_R := \inf(t \geq s : |X_t| \leq R)$ ,  $R > 0$ .



**Corollary 1.** *Under the assumptions (SA 1)–(SA 4) and (RA 1), there exist  $\alpha > 0$ ,  $R_0 > 0$  and  $C > 0$  such that for any  $R \geq R_0$  (and with  $(X_0, V_0) = (x, v)$ ),*

$$\mathbf{E}_{x,v} \exp(\alpha \tau_R) \leq C(1 + |x|^2), \tag{13}$$

and also

$$\mathbf{E}_{x,v} |X_t|^2 1(t < \tau_R) \leq C|x|^2. \tag{14}$$

This statement admits some modifications: as an example, “for any  $\alpha > 0$  there exist  $R_0, C > 0$  such that for any  $R \geq R_0$  the inequality (13) holds”. The inequality (14) may be also stated without the indicator in the left-hand side (and with a right-hand side as in (13)), but the proof of this version is less elementary and is not necessary for the proof of the Theorem 4 in the next section.

We provide a brief sketch of the proof of the Corollary 1 for the reader’s convenience. Dynkin’s formula or, equivalently, the integral form of Ito’s formula applied to the process  $\exp(\alpha t)\phi(X_t)$  by virtue of (12) implies that

$$\begin{aligned} & \mathbf{E}_{x,v} \exp(\alpha(t \wedge \tau_R))\phi(X_{t \wedge \tau_R}) - \phi(x) \\ & + \mathbf{E}_{x,v} \int_0^{t \wedge \tau_R} \exp(\alpha s)(c'_1\phi - c'_2 - \alpha)(X_s) ds \leq 0. \end{aligned} \tag{15}$$

If necessary, this procedure may be accomplished by an appropriate localization. Now, let us choose  $R$  so that

$$\inf_{|x| \geq R} (c'_1\phi - c'_2)(x) \geq 1.$$

Then (13) follows by Fatou’s lemma as  $t \rightarrow \infty$ , at least, if  $\alpha \leq 1$ .

Further, let  $\alpha = 0$ . Then it follows from (15) along with  $c'_1\phi - c'_2 \geq 1 > 0$  that

$$\mathbf{E}_{x,v} \phi(X_t) 1(t < \tau_R) \leq \mathbf{E}_{x,v} \phi(X_{t \wedge \tau_R}) \leq \phi(x), \quad |x| > R,$$

and

$$\mathbf{E}_{x,v} \phi(X_t) 1(t < \tau_R) = 0, \quad |x| \leq R,$$

the latter because  $\tau_R = 0$  for  $|x| \leq R$ . Since quadratic form  $\phi$  is positive definite, this suffices for (14), as required.  $\square$

### 3.7 Exponential Convergence

In this section, for the process  $(X, V)$  satisfying the system (7) with initial values  $(x, v)$ , its marginal distribution at time is denoted by  $\mu_t^{x,v}$ ,  $t \geq 0$ .

There is a routine scheme to deduce exponential—and also many others—convergence in total variation from two facts: (1) “minorization” condition of local Markov–Dobrushin type, here provided by the Theorem 2, and (2) recurrence property, regular returns of the trajectory to a certain set satisfying the “minorization” condition, here provided by the Corollary 1. This scheme is expounded in [20, 21], with the local Markov–Dobrushin condition called differently. Note that the Theorem 3 may also be used directly, without the Corollary 1.

**Theorem 4.** *Suppose the assumption (SA 1)–(SA 4) and (RA 1) are fulfilled. Then there exists a stationary probability distribution  $\mu_\infty$  on  $\mathbb{R}^{2d} \times \mathcal{S}$  and positive constants  $\bar{C}, \bar{c}$ , depending on  $d, C, \phi, c_1, c_2, \min\{\sigma(v), v \in \mathcal{S}\}, \min\{q_{ij}; i, j \in \mathcal{S}, i \neq j\}, \max\{q_{ij}; i, j \in \mathcal{S}, i \neq j\}, N$ , such that*

$$\|\mu_t^{x,v} - \mu_\infty\|_{TV} \leq \bar{C} \exp(-\bar{c}t)(1 + x^2), \quad t \in [0, \infty)$$

(all parameters are described in the Sect. 3.2).

## 4 Proofs of Theorems 1 and 2

### 4.1 Proof of Theorem 1

*Proof.* We shall establish existence on the basis of the paper [5]; another methodological source is the paper [10]. The paper [5] uses the language of stochastic differential equations; thus, we describe our system in such terms. Consider the system (5) with  $b(X^1, X^2) = b(X^1, X^2, v)$  on a stochastic basis  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbf{P})$ , where it has a solution. Let the basis be extended if necessary, and let us add to the system (7) the equation for the discrete component

$$dV_t = \int_{\mathbb{R}^1} K(X_t, V_{t-}, z)N(dt, dz). \tag{16}$$

Here  $N$  is an  $(\mathcal{F}_t)$ -adapted Poisson random measure with a mean (compensator) measure  $ds \times \frac{dz}{z^2}$  independent of the Wiener process. The coefficient  $K$  must be constructed so that it substitutes the intensities  $Q$ : for each  $i \in \mathcal{S}$ , it takes values in  $\{j - i, j \in \mathcal{S}\}$  and

$$\int_{\{z:K(x,v,z)=j-i\}} \frac{dz}{z^2} = q_{ij}, \quad j \in \mathcal{S} \setminus i.$$

We give now a description (slightly non-rigorous, although, hopefully, comprehensible) how to construct  $K$ . For each  $i \in \mathcal{S}$ ,  $K(x, i)$  takes value  $1 - i$  on  $[z_1(x), \infty)$ ; value  $2 - i$  on  $[z_2(x), z_1(x))$ ; ...; value  $-1$  on  $[z_{i-1}(x), z_{i-2}(x))$ ; value  $+1$  on  $[z_{i+1}(x), z_{i-1}(x))$ ; ...; value  $N - i$  on  $[z_N(x), z_{N-1}(x))$ ; value  $0$  on the rest of  $\mathbb{R}^1$ . The points  $z_j, j \in \mathcal{S} \setminus \{i\}$ , are defined by the relations

$$\int_{z_1}^{\infty} \frac{dz}{z^2} = q_{i1}, \int_{z_2}^{z_1} \frac{dz}{z^2} = q_{i2}, \dots, \int_{z_N}^{z_{N-1}} \frac{dz}{z^2} = q_{iN},$$

where the term with index  $ii$  is *excluded*. Proposition 1 of [5] provides existence of a weak solution (its condition 2b is not needed in our case because the jump intensities are bounded and at the moments of jumps, the component  $X$  does not increase). In fact, Proposition 1 of [5] is proved in style of martingale problems, with pasting solutions at the moments of jumps.

To prove uniqueness, we use Lemma 2 of [5]. It uses a solution  $(\tilde{X}_t, t \geq 0)$  of the equation (5) *without switching*. Given this trajectory, the first switching moment  $\tau$  of the solution  $(X, V)$  of the system (7)–(16) has the following distribution (note that on  $[0, \tau)$  the trajectories of  $X$  and  $\tilde{X}$  coincide by construction):

**Lemma 1.** *Under the assumptions (SA 1)–(SA 4), given the trajectory  $\tilde{X}_t, t \in [0, \infty)$ , of the solution of (5) with  $b(X^1, X^2) = b(X^1, X^2, V_0)$ , the conditional probability of the event  $\{\tau > r\}$  equals*

$$\exp \left\{ - \int_0^r \sum_{j \in \mathcal{S} \setminus v} q_{vj}(\tilde{X}_u) du \right\}.$$

Finally, we give a sketch of the proof of strong Markov property. The solution of system (5) does possess a strong Markov property; see [1] and [19]. This entails a strong Markov property of the switched process (7). To prove it, adopt the method of [2, Sect. 4], where nondegenerate diffusions are considered. Instead of making sequentially infinite number of switchings, let us limit ourselves to the first  $k$  switchings and make no further ones. The result is a distribution on the space of trajectories (both continuous and discrete). Let us take an arbitrary stopping time  $\tau$  and calculate the conditional distribution of this distribution given  $\mathcal{F}_\tau$ , restricted to  $t \in [\tau, \infty)$ . It equals the distribution of the process with initial conditions  $(\tau, X_\tau, V_\tau)$  switched finitely many times—so many times how many out of the first  $k$  switchings took place after the moment  $\tau$ . With  $k \rightarrow \infty$  the proof is completed—the limiting conditional distribution is again that of a switching process, and due to its uniqueness, this provides the strong Markov property.  $\square$

### 4.2 Proof of Theorem 2

*Proof.* Here we shall explain how to deduce Theorem 2 for a general set  $\mathcal{S}$  from the statement of this Theorem 2 for  $\mathcal{S} = \{1\}$ . Recall that it suffices to take  $s = 0$  in (11).

Let us fix initial state  $(x, v)$ ; denote by  ${}^v X_u, u \in [0, \infty)$ , the corresponding solution of the equation (5) (without switching) with  $b(X^1, X^2) \equiv b(X^1, X^2, v)$ . In some cases it will be convenient to use a more sophisticated notation  $({}^v X_u^{0,x}, u \geq 0)$  for the same process where  $x$  is the initial data at 0. Respectively,  $({}^v X_u^{r,x'}, u \geq r)$  signifies a solution of the equation (5) with  $b(X^1, X^2) = b(X^1, X^2, v)$  on the interval  $[r, \infty)$  with initial value  $x'$  at  $r$ . Let us inspect what occurs on time interval  $[0, T]$ . The discrete component  $V$  is a point process with compensator intensities lying between the given lower and upper bounds. This implies that the probability that  $V_T$  equals 1 is bounded away from zero uniformly in all  $x, v$ .

We are now going to give a rigorous explanation of this fact, although its implementation may look a bit more complicated than it actually is, due to the inevitably involved notations. To simplify the latter a little bit, denote  ${}^1 \tilde{X}_u^{(r)} := {}^1 X_u^{r, {}^v X_r^{0,x}}, u \geq r$  (this will be used only in this subsection); recall that  $({}^v X_u^{0,x}, u \geq 0)$  is a process *without switching* and emphasize that likewise without switching is the process  $({}^1 \tilde{X}_u^{(r)}, u \geq r)$ . Then for  $v \neq 1$  the nonconditional probability of the event  $\{V_T \text{ equals } 1\}$  is greater than or equal to the expectation of

$$\int_0^T \left( \exp\left\{ - \int_0^r \sum_{j \in \mathcal{S} \setminus v} q_{vj}({}^v X_u) du \right\} \times \exp\left\{ - \int_r^T \sum_{j \in \mathcal{S} \setminus 1} q_{1j}({}^1 \tilde{X}_u^{(r)}) du \right\} \right) q_{v1}({}^v X_r) dr,$$

or (notations  $p$  and  $q$  are defined below), equivalently, of

$$\int_0^T p(r)q(r)q_{v1}({}^v X_r)dr. \tag{17}$$

Here the conditional probability that the discrete component remains at state  $v$  on the time interval  $[0, r)$  given  $({}^v X_u, 0 \leq u < r)$  reads

$$p(r) = \exp\left\{ - \int_0^r \sum_{j \in \mathcal{S} \setminus v} q_{vj}({}^v X_u) du \right\};$$

the conditional probability that the discrete component jumps from state  $v$  to state 1 on the time interval  $[r, r + dr)$  given  $({}^v X_u, 0 \leq u < r)$  equals

$$q_{v1}({}^v X_r)dr;$$

the conditional probability that the discrete component remains at state 1 on the time interval  $[r + dr, T]$  given  $({}^v X_u, 0 \leq u < r)$  may be presented as

$$q(r) = \exp\left\{-\int_r^T \sum_{j \in \mathcal{S} \setminus 1} q_{1j}({}^1 \tilde{X}_u^{(r)}) du\right\}.$$

Due to the assumptions on all intensities  $q_{ij}$ , we get a lower bound  $(\exp(-cT))$ , the proof of which is based on the Lemma 1. Integration with respect to  $r$  in (17) is a complete probability formula, a rigorous justification for which may be given, for example, as in [22].

For  $v = 1$  it is even easier to obtain a desired lower bound by virtue of the same Lemma 1, as in this case the probability in question is greater than or equal to the expectation of

$$\exp\left\{-\int_0^T \sum_{j \in \mathcal{S} \setminus 1} q_{1j}({}^1 X_u) du\right\}.$$

Further, fix  $R > 0$  and assume that  $x \in B_R$ . Copying the reasoning of [19] and [1], we obtain that there exists  $R' \in (R, +\infty)$  such that the continuous component of the trajectory does not go beyond the boundary of  $B_{R'}$  on  $[0, T]$  with probability almost 1, and this is uniformly in initial conditions, belonging to  $B_R \times \mathcal{S}$ . Combining these two facts, we conclude that the probability that both events take place is bounded away from zero uniformly in  $(x, v) \in B_R \times \mathcal{S}$ .

The conditional distribution of  $(X, V)$  admits on these events a useful estimate. Take  $T' > 0$  and consider the distribution of  $(X, V)$  on  $[T, T + T']$ , conditioned on the past time  $[0, T]$  history. This distribution is minorized by the distribution of the solution of system (5) with initial conditions  $T, X_T$ , with a positive constant multiplier, which follows from the calculus in [1]—more precisely, from the proofs of the proofs of Lemmas 3 and 4 from [1]—accomplished by the Lemma 2 and its Corollaries. This suffices for the local Markov–Dobrushin condition (11). Note that in [1, 19] the initial conditions are assumed non-random; however, the Remark 1 removes this restriction.

To realize this plan, for  $x \in \mathbb{R}^{2d}$ ,  $v \in \mathcal{S}$ , let us define the following:

$\mu(x, v; dXdV)$ —the distribution of the solution of the system (7) on  $[0, \infty)$  with initial conditions  $x, v$ ;

${}^1\mu(x; dX)$ —the distribution of the solution on  $[0, \infty)$  of the system (5) with  $b(X^1, X^2) = b(X^1, X^2, 1)$  and initial data  $x$ .

**Lemma 2.** *Under the assumptions (SA 1)–(SA 4), for any  $T > 0$  there exists*

$$c(c_u, T) = e^{-T(N-1)c_u}$$

*such that for any  $x \in \mathbb{R}^{2d}$  and any event  $A$  defined through the trajectory of  $X$  on the time interval  $[0, T]$ , the following inequality holds:*

$$\mu(x, 1; A \times \{V|_{[0,T]} \equiv 1\}) \geq c(c_u, T) \times {}^1\mu(x; A).$$

*Proof.* Let us fix  $x$  and construct on a stochastic basis the following objects:

- (1) A process  ${}^1X = ({}^1X(t), t \in [0, \infty))$ —a solution of the system (5) with  $b(X^1, X^2) = b(X^1, X^2, 1)$  and initial data  $x$ .
- (2) A switching process  $(X, V)$  satisfying the system (7) with initial conditions  $(x, 1)$  in such a way that they coincide up to the first switching. For this purpose take a process  ${}^1X$  satisfying (5) and a random variable  $\tau$  such that for  $t \geq 0$ , the probability of the event  $\{\tau \leq t\}$  given  $({}^1X_u, 0 \leq u < t)$  according to the Lemma 1 equals

$$1 - \exp\left\{-\int_0^t \sum_{j \in S \setminus 1} q_{1j}({}^1X_u) du\right\}$$

(this can be done on the product space  $\times [0, \infty)$ ). The moment  $\tau$  is the moment of the first switching, and the value of  $V_\tau = j$  is chosen proportionally to  $q_{1j}, j \in S \setminus 1$ . At the moment  $\tau$  the switched process  $X$  satisfying (7) acquires the corresponding conditional probability  $\mu_{\tau, X_\tau, V_\tau}$ , while the process  ${}^1X$  satisfying (5) develops further in its dynamics. It is easy to see that the probability of the event  $\{\tau > T\}$  conditioned on a trajectory of  ${}^1X$  is uniformly bounded away from zero on the space of all trajectories: it is greater than or equal to  $c(c_u, T)$ . For any event  $A$  on time interval  $[0, T]$ , the probability that the switched process  $X$  lies in  $A$  is greater than or equal to the probability of the event  $\{\text{the process } {}^1X \text{ lies in } A \text{ and the switched process } X \text{ coincides with } {}^1X\}$ , which, in turn, is greater than or equal to  $c(c_u, T) \times \{\text{the probability that } {}^1X \in A\}$ . Thus, it is also greater than or equal to  $c(c_u, T) \times \{\text{the probability that } {}^1X \in A\}$ . □

**Corollary 2.** *Under the assumptions (SA 1)–(SA 4), for any  $T > 0$  there exists a constant*

$$c(c_u, T) = e^{-T(N-1)c_u}$$

*such that for any  $s \in [0, \infty)$ ,  $x \in \mathbb{R}^{2d}$  and any event  $A$  defined through trajectories of  $X, V$  on the time interval  $[0, T]$ , the following inequality holds:*

$$\mu(x, 1; A) \geq c(c_u, T) \times {}^1\mu(x, A \cap \{V|_{[0,T]} \equiv 1\}).$$

*Proof.* Indeed,  $\mu(x, 1; A) \geq \mu(x, 1; A \cap \{V|_{[0,T]} \equiv 1\})$ . □

**Corollary 3.** *Under the assumptions (SA 1)–(SA 4), for any  $T > 0$  there exists a constant*

$$c(c_u, T) = e^{-T(N-1)c_u},$$

such that for any initial condition  $x$ , a measurable  $\Gamma \subseteq \mathbb{R}^{2d} \times \mathcal{S}$  and  $R > 0$  the following inequality holds:

$$\begin{aligned} & \mu(x, 1; (X, V)_T \in \Gamma, X_u \in B_R, u \in [0, T]) \\ & \geq c(c_u, T)^1 \mu(X_T \in \Gamma \cap \{V = 1\}, X_u \in B_R, u \in [0, T]). \end{aligned}$$

This completes the proof of the Theorem 2. □

## 5 Conclusion

We have proved for highly degenerate stochastic mechanical hybrid systems under quite general conditions (discontinuity and linear growth of coefficients and the Wiener process perturbations) the following properties:

- Existence and uniqueness theorem and a strong Markov property for solutions of such systems
- A local mixing property in the Markov–Dobrushin form for these solutions
- Exponential stochastic stability in total variation metric for solutions of such systems under the additional assumption (10)

The work of the first author was supported by the RFBR grant 13-08-01096. Both authors are grateful to ICMS Edinburgh for support during the programme “Mixing and averaging in stochastic processes” (20.01.2013–09.02.2013). For the second author, various versions of this work were written also while visiting the trimester programme “Stochastic Dynamics in Economics and Finance” at Hausdorff Research Institute for Mathematics (HIM) Bonn, Nikolai Pertsov White Sea Biological Station of Moscow State University (WSBS) and Bielefeld University; also, his work was supported by the RFBR grant 13-01-12447 ofi\_m2. The authors are thankful to the reviewer for very useful comments.

## References

1. Abourashchi, N., Veretennikov, A. Yu.: On stochastic averaging and mixing. *Theory of Stoch. Processes.* **16**(1), 111–130 (2010)
2. Anulova, S., Pragarauskas, H.: Weak Markov solutions of stochastic equations. *Lith. Math. J.* **17**(3), 141–155 (1978)
3. Anulova, S., Veretennikov, A.: On ergodic properties of degenerate hybrid stochastic control systems. In: 49th IEEE Conference on Decision and Control (CDC), 2010, Issue Date: 15–17 Dec. 2010, pp. 2292–2297. Location: Atlanta, GA (2010) doi:10.1109/CDC.2010.5717412
4. Arapostathis, A., Borkar, V.S., Ghosh, M.K.: Ergodic control of diffusion processes. *Encyclopedia of Mathematics and its Applications* 143. xvi, 323 p. Cambridge University Press, Cambridge. (2012)

5. Athreya, K., Kliemann, W., Koch, G.: On sequential construction of solutions of stochastic differential equations with jump terms. *Syst. Control Lett.* **10**, 141–146 (1988)
6. Bujorianu, M., Lygeros, J.: Toward a general theory of stochastic hybrid systems. In: Blom, H., Lygeros, J. (eds.) *Stochastic Hybrid Systems*, pp. 3–30. Springer, Berlin (2006) <http://www.springerlink.com/index/n62r24355788p6h7.pdf>
7. Buyukkoroglu, T., Esen, O., Dzhaferov, V.: Common Lyapunov functions for some special classes of stable systems. *IEEE Trans. Automat Contr.* **56**, 1963–1967 (2011)
8. Campillo, F.: Optimal ergodic control for a class of nonlinear stochastic systems—Application to semi-active vehicle suspensions. In: *Proceedings of the 28th IEEE Conference on Decision Control*, pp. 1190–1195. Tampa (1989)
9. Campillo, F., Pardoux, E.: Numerical methods in ergodic optimal stochastic control and application. In: Karatzas, I., Ocone, D. (eds.) *Applied Stochastic Analysis*, pp. 59–73. Springer, Berlin (1992) <http://www.springerlink.com/index/y13p561655383ht7.pdf>
10. Davis, M.H.A.: Piecewise-Deterministic Markov Processes: A General class of non-diffusion stochastic models. *J. Royal Stat. Soc. Ser. B (Methodological)* **46**(3), 353–388 (1984)
11. Deng, F., Luo, Q., Mao, X., Pang, S.: Noise suppresses or expresses exponential growth. *Syst. Control Lett.* **57**(3), 262–270 (2008)
12. Guglielmino, E., Sireteanu, T., Stammers, C., Ghita, G., Giuclea, M.: *Semi-active Suspension Control: Improved Vehicle Ride and Road Friendliness*. Springer, Berlin (2008)
13. Hu, G., Liu, M., Mao, X., Song, M.: Noise expresses exponential growth under regime switching Systems. *Syst. Control Lett.* **58**(9), 691–699 (2009)
14. Ikeda, N., Watanabe, S.: *Stochastic Differential Equations and Diffusion Processes*. 2nd edn. North-Holland Mathematical Library, 24. xvi, 555 p. Amsterdam etc.: North-Holland; Tokyo: Kodansha Ltd. (1989)
15. Khasminskii, R., Zhu, C., Yin, G.: Stability of regime-switching diffusions. *Stoch. Processes Appl.* **117**(8), 1037–1051 (2007)
16. Krylov, N.V.: On the selection of a Markov process from a system of processes and the construction of quasi-diffusion processes. *Math. USSR, Izv.* **7**, 691–709 (1973)
17. Nummelin, E.: *General irreducible Markov chains and non-negative operators*. Cambridge Tracts in Mathematics, 83. XI, 156 p. Cambridge etc.: Cambridge University Press. (1984)
18. Shur, M.: Ergodic theorems for Markov processes. In: Prohorov, Ju.V.(ed.) *Probability and Mathematical Statistics: Encyclopedia*, pp. 825–827. Moscow, The Big Russian Encyclopedia, 911 p. (1999) ISBN 5-85270-265-X
19. Veretennikov, A.: On weak solution of an SDE. In: 10th Vilnius conference on Probability Theory and Mathematical Statistics, 28.06-02.07.2010, Vilnius, Lithuania, Abstracts of Communications, vol. 287, TEV Publishers, Vilnius (2010)
20. Veretennikov, A. Yu.: Bounds for the mixing rate in the theory of stochastic equations. *Theory Probab. Appl.* **32**, 273–281 (1987)
21. Veretennikov, A. Yu.: On Polynomial mixing and convergence rate for stochastic difference and differential equations. *Theory Probab. Appl.* **44**(2), 361–374 (2000)
22. Veretennikov, A. Yu., Zverkina, G.A.: Simple proof of Dynkin’s formula for single-server systems and polynomial convergence rates. arXiv:1306.2359v1 [math.PR]
23. Wu, L.: Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stoch. Processes Appl.* **91**(2), 205–238 (2001)
24. Yuan, C., Mao, X.: Asymptotic stability in distribution of stochastic differential equations with Markovian switching. *Stoch. Processes Appl.* **103**(2), 277–291 (2003)