EXPONENTIAL ERGODICITY OF THE SOLUTIONS TO SDE'S WITH A JUMP NOISE

ALEXEY M.KULIK

Abstract. The mild sufficient conditions for exponential ergodicity of a Markov process, defined as the solution to SDE with a jump noise, are given. These conditions include three principal claims: recurrence condition **R**, topological irreducibility condition **S** and non-degeneracy condition **N**, the latter formulated in the terms of a certain random subspace of $\mathbb{R}^m$, associated with the initial equation. The examples are given, showing that, in general, none of three principal claims can be removed without losing ergodicity of the process. The key point in the approach, developed in the paper, is that the *local Doeblin condition* can be derived from **N** and **S** via the stratification method and criterium for the convergence in variations of the family of induced measures on $\mathbb{R}^m$.

Introduction

In this paper, we study ergodic properties of a Markov process X in $\mathbb{R}^m$, given by an SDE

$$dX(t) = a(X(t))dt + \int_{\|u\|\leq 1} c(X(t-),u)\tilde{\nu}(dt,du) + \int_{\|u\|>1} c(X(t-),u)\nu(dt,du). \tag{0.1}$$

Here, ν is a Poisson point measure, $\tilde{\nu}$ is correspondent compensated measure, and coefficients a, c are supposed to satisfy usual conditions sufficient for the strong solution of (0.1) to exist and be unique. Our aim is to give sufficient conditions for exponential ergodicity of (0.1), that impose as weak restrictions on the Lévy measure of the noise, as it is possible.

There exists two well developed methods to treat the ergodicity problem for the discrete time Markov processes, valued in a locally compact phase space. The first one is based on the *coupling technique* (see detailed overview in [14]), the second one uses the notions of *T-chain* and *petite sets* (see [21], [6] and references therein). These methods can be naturally extended to continuous time case either by making a procedure of time discretization (like it was made for the solutions to SDE's with jumps in the recent paper [20]), or by straightforward use of the coupling technique in a continuous time settings (see [27],[28] for such kind of a technique for diffusion processes). Typically, in the methods mentioned above, the two principal features should be provided:

– recurrence outside some large ball;

– regularity of the transition probability in some bounded domain.

The first feature can be provided in a quite standard way via an appropriate version of the Lyapunov criterium (condition **R** in Theorem 1.1 below). The second one is more intrinsic, and requires some accuracy in the choice both of the concrete terms, in which such feature is formulated, and of the conditions on the process, sufficient for such feature to hold true. We deal with the form of the regularity feature, that is usually called the *local Doeblin condition*, and is formulated below.

LD. For every $R > 0$, there exists $T = T(R) > 0$ such that

$$\inf_{\|x\|,\|y\|\leq R} \int_{\mathbb{R}^m} [P_x^T \wedge P_y^T](dz) > 0,$$

2000 *Mathematics Subject Classification.* Primary 60J25; Secondary 60H07.

Key words and phrases. β-mixing coefficients, local Doeblin condition, admissible time-stretching transformations, stratification method, convergence in variation of induced measures.

where $P_x^t(\cdot)\equiv P(X(t)\in\cdot|X(0)=x)$, and, for any two probability measures $\mu,\varkappa$,

$$[\mu\wedge\varkappa](dz)\stackrel{df}{=}\min\left[\frac{d\mu}{d(\mu+\varkappa)}(z),\frac{d\varkappa}{d(\mu+\varkappa)}(z)\right](\mu+\varkappa)(dz).$$

The non-trivial question is what is the proper form of the conditions on the coefficients a,c of the equation (0.1) and the Lévy measure of the noise, sufficient for the local Doeblin condition to hold true. In a diffusion settings, standard strong ellipticity (or, more general, Hörmander type) non-degeneracy conditions on the coefficients provide that the transition probability of the process possesses smooth density w.r.t. Lebesgue measure, and thus **LD** holds true ([27],[28]). In a jump noise case, one can proceed analogously and claim the transition probability of the solution to (0.1) to possess a locally bounded density (exactly this claim was used as a basic assumption in the recent paper [20]). However, in the latter case such kind of a claim appears to be too restrictive; let us discuss this question in more details. Consider, for simplicity, equation (0.1) with $c(x,u)=u$, i.e. a following non-linear analogue of the Ornstein-Uhlenbeck equation:

$$dX(t)=a(X(t))dt+dU_t, \tag{0.2}$$

where $U_t=\int_0^t\int_{\|u\|\leq1}c(u)u\tilde{\nu}(ds,du)+\int_0^t\int_{\|u\|>1}c(u)\nu(ds,du)$ is a Lévy process. There exist two methods to provide the process defined by (0.2) to possess a bounded (moreover, belonging to the class C^∞) transition probability density. The first one was initially proposed by J.Bismut (see [3],[2],[19],[15]), the second one – by J.Picard (see [24],[13]). Both these methods require, among others, the following condition on the Lévy measure Π of the process U to hold true:

$$\exists\rho\in(0,2):\quad \varepsilon^{-\rho}\int_{\|u\|\leq\varepsilon}\|u\|^2\Pi(du)\to\infty,\quad \varepsilon\to0+. \tag{0.3}$$

This limitation is not a formal one. It is known (see [18], Theorem 1.4), that if

$$\liminf_{\varepsilon\to0+}\left[\varepsilon^2\ln\left(\frac{1}{\varepsilon}\right)\right]^{-1}\sup_{\|v\|=1}\int_{\mathbb{R}^m}[|(u,v)|\wedge\varepsilon]^2\Pi(du)=0, \tag{0.4}$$

then the transition probability density, if exists, does not belong to any $L_{p,loc}(\mathbb{R}^m), p>1$, and therefore is not locally bounded (note that (0.4) implies that (0.3) fails). One can say that when the intensity of the jump noise is "sparse near zero" in a sense of (0.4), the behavior of the density essentially differs from the diffusion one, and the density either does not exist or is essentially irregular.

On the other hand, let us formulate a corollary of the general ergodicity result, given in Theorems 1.1,1.3 below.

Proposition 0.1. *Let $m=1$, suppose that $a(\cdot)$ is locally Lipschitz on $\mathbb{R}$ and $\lim\sup_{|x|\to+\infty}\frac{a(x)}{x}<0$. Suppose that the Lévy measure Π of the process U satisfies the following conditions:*

(i) there exists $q>0$: $\int_{|u|>1}|u|^q\Pi(du)<+\infty$;

(ii) $\Pi(\mathbb{R}\backslash\{0\})\neq0$.

Then the solution to (0.1) is exponentially ergodic, i.e. its invariant distribution μ_{inv} exists and is unique, and, for some positive constant C,

$$\forall x\in\mathbb{R},\quad \|P_x^t-\mu_{inv}\|_{var}=O(\exp[-\mathsf{C}t]),\quad t\to+\infty. \tag{0.5}$$

In this statement, the non-degeneracy condition (ii) on the jump noise, obviously, is the weakest possible one: if it fails, then (0.2) is an ODE, and (0.5) fails also. We can conclude, that the proper conditions on the jump noise, sufficient to provide exponential ergodicity of the process, defined by (0.2), are much milder than the conditions that should be imposed in order to provide that this process possesses regular (locally bounded or even locally L_p-integrable) transition probability density.

Our way to prove the local Doeblin condition for the solution to (0.1) strongly relies on the finite-dimensional criterium for the convergence in variations of the family of induced measures on $\mathbb{R}^m$. This criterium was obtained in [1] (the case $m=1$ was treated in [7]). Via the *stratification method* (for the detailed exposition of this topic see [9], Section 2) this criterium can be extended to any probability space with a measurable group of admissible transformations ($\Leftrightarrow$ *admissible family*), that generates measurable stratification of the probability space (for a details, see Section 2 below). The key point is that the criterium for the convergence in variations of the family of induced measures is local in the following sense: such a convergence holds true, as soon as the initial probability measure is restricted to any set, where the gradient (w.r.t. given admissible family) of the limiting functional is non-degenerate. We impose a condition (condition **N** in Theorem 1.3 below), that implies existence of an admissible family, such that the solution to (0.1) possesses a gradient w.r.t. this family, and this gradient is non-degenerate with positive probability. Under this condition, the (local) criterium for convergence in variation provides the *local Doeblin condition in a small ball* (Lemma 3.1 below). Together with *topological irreducibility* of the process (provided, in our settings, by condition **S** of Theorem 1.3), this gives condition **LD**.

In our construction, we use time-stretching transformations of the Lévy jump noise. Such a choice is not the only possible; for instance, one can use groups of transformations, varying values of the jumps (see [8]), and obtain another version of the non-degeneracy condition **N**. In order to make exposition compact, we do not give exact formulation and proof of the correspondent statement, although the general scheme is totally the same. Let us just outline that the main advantage of choice of the differential structure, made in the present article, is that, for the Lévy jump noise, time-stretching transformations, unlike transformations of the phase variable, are admissible without any regularity claim on the Lévy measure of the noise.

The structure of the paper is the following. In Section 1, we formulate the main statements of the paper. In Section 2, we give necessary background from the stochastic calculus involving time-stretching transformations of the Lévy jump noise, and its applications to the convergence in variation of the distributions of the solutions to SDE's with such noise. In Section 3, we prove the main statements of the paper. Sufficient conditions for the basic conditions **R**,**N**,**S** from Section 2, easy to deal with, are given in Section 4. In Section 5, we give counterexamples showing that, in general, none of the basic conditions can be removed without losing ergodicity of the process.

1. The main results

Let us introduce notation. Everywhere below ν is a Poisson point measure on $\mathbb{R}^+\times\mathbb{R}^d$, Π is its Lévy measure and $\tilde\nu(dt,du)\equiv\nu(dt,du)-dt\Pi(du)$ is the correspondent compensated point measure. We denote by $p(\cdot)$ the point process associated with ν and by $\mathcal{D}$ the domain of $p(\cdot)$. Later on, we will impose such a conditions on the coefficients a,c of the equation (0.1), that this equation, endowed by the initial condition $X(0)=x\in\mathbb{R}^m$, has the unique strong solution $\{X(t)\equiv X(x,t),t\geq 0\}$, that is a process with cádlág trajectories. We denote by P_μ the distribution in $\mathbb{D}(\mathbb{R}^+,\mathbb{R}^m)$ of the solution $X(\cdot)$ to (0.1) with $\mathrm{Law}(X(0))=\mu$, by E_μ the expectation w.r.t. P_μ, and by P^t_μ the distribution of $X(t)$ w.r.t. P_μ. In particular, we denote $\mathsf{P}_x\equiv\mathsf{P}_{\delta_x}, P^t_x\equiv P^t_{\delta_x}, x\in\mathbb{R}^m$.

All the functions used below are supposed to be measurable (jointly measurable w.r.t. (x,u), if necessary). The gradient w.r.t. the variable z is denoted by ∇_z. The unit matrix in $\mathbb{R}^m$ is denoted by $I_{\mathbb{R}^m}$. We use the same notation $\|\cdot\|$ for the Euclidean norms both in $\mathbb{R}^m$ and $\mathbb{R}^d$, and for an appropriate matrix norms. The open ball in $\mathbb{R}^m$ with the center x and radius R is denoted by $B_{\mathbb{R}^m}(x,R)$. The space of probability measures on $\mathbb{R}^m$ is denoted by $\mathcal{P}$, the total variation norm is denoted by $\|\cdot\|_{var}$. The (closed) support of the measure $\mu\in\mathcal{P}$ is denoted by $\mathrm{supp}\,\mu$. For $\mu\in\mathcal{P}$ and non-negative measurable function $\phi:\mathbb{R}^m\to\mathbb{R}$, we denote

$$\phi(\mu)\equiv\int_{\mathbb{R}^m}\phi(x)\mu(dx)\in\mathbb{R}^+\cup\{+\infty\}.$$

The coefficient a is supposed to belong to the class $C^1(\mathbb{R}^m,\mathbb{R}^m)$ and to satisfy the linear growth condition. In our considerations, we will deal with the following two types of SDE's with a jump noise.

A. Moderate non-additive noise. The SDE of the type (0.1) with the jump coefficient c dependent on space variable x. We claim the following standard conditions to hold true:

$$\|c(x,u)-c(y,u)\|\leq K(1+\|u\|)\|x-y\|,\quad \|c(x,u)\|\leq\psi_*(x)\|u\|,\quad u\in\mathbb{R}^d,x,y\in\mathbb{R}^m \tag{1.1}$$

with some constant $K\in\mathbb{R}^+$ and some function $\psi_*:\mathbb{R}^d\to\mathbb{R}^+$ satisfying linear growth condition. We also claim the following specific moment condition:

$$\int_{\mathbb{R}^d}\sup_{\|x\|\leq R}\left[\|c(x,u)\|+\|\nabla_x c(x,u)\|_{\mathbb{R}^{m^2}}\right]\Pi(du)<+\infty,\quad R\in\mathbb{R}^+ \tag{1.2}$$

(the gradient $\nabla_x c(x,u)$ is supposed to exist, and to be continuous w.r.t. x). We interpret this condition in a sense that the jump part of the equation is *moderate*.

B. Arbitrary additive noise. The SDE of the type (0.1) with the jump coefficient c that does not depend on space variable x: $c(x,u)=c(u)$ and $\|c(u)\|\leq K\|u\|$. No moment conditions like (1.2) are imposed on the jump part. In this case, (0.1) is a non-linear analogue (0.2) of the Ornstein-Uhlenbeck equation. Making a change $\nu(\cdot)\mapsto\nu_c(\cdot),\nu_c([0,t]\times A)\equiv\nu([0,t]\times c^{-1}(A))$, one can reduce (0.2) to the same equation with $m=d$ and $c(u)=u$. In order to simplify notation, in a sequel we consider such an equations only.

In the both cases given above, equation (0.1), endowed by the initial condition $X(0)=x$, has the unique strong solution, that is a Feller Markov process with cádlág trajectories, and $\{P_x^t(\cdot),t\in\mathbb{R}^+,x\in\mathbb{R}^m\}$ is its transition probability. In the case **A**, the trajectories of this solution a.s. have bounded variation on every finite interval.

Denote, like in [20],

$$\mathcal{Q}=\{f\in C^2(\mathbb{R}^m,\mathbb{R})|\exists\tilde f \text{ locally bounded such that } \int_{\|u\|>1}f(x+c(x,u))\Pi(du)\leq\tilde f(x),\quad x\in\mathbb{R}^m\},$$

and, for $f\in\mathcal{Q}$, write

$$\mathcal{A}f(x)=\int_{\mathbb{R}^d}\left[f(x+c(x,u))-f(x)-(\nabla f(x),c(x,u))_{\mathbb{R}^m}\cdot\mathbb{1}_{\|u\|\leq1}\right]\Pi(du),\quad x\in\mathbb{R}^m.$$

Let us formulate two general statements concerning convergence rate of P_μ^t to the ergodic distribution of X and estimates for the β-mixing coefficients of X.

Theorem 1.1. *Suppose that the* **LD** *condition holds true together with the following recurrence condition:*

R. *There exist function* $\boldsymbol{\phi}\in\mathcal{Q}$ *and constants* $\boldsymbol{\alpha},\boldsymbol{\gamma}>0$ *such that*

$$\mathcal{A}\boldsymbol{\phi}\leq-\boldsymbol{\alpha}\boldsymbol{\phi}+\boldsymbol{\gamma}\quad\text{and } \boldsymbol{\phi}(x)\to+\infty,\quad\|x\|\to+\infty.$$

Then the process X *possesses unique invariant distribution* $\mu_{inv}\in\mathcal{P}$, *and there exist positive constants* $\mathsf{C}_1,\mathsf{C}_2$ *such that, for every* $\mu\in\mathcal{P}$ *with* $\boldsymbol{\phi}(\mu)<+\infty$,

$$\left\|P_\mu^t-\mu_{inv}\right\|_{var}\leq\mathsf{C}_1[\boldsymbol{\phi}(\mu)+1]\exp\left[-\mathsf{C}_2t\right],\quad t\in\mathbb{R}^+. \tag{1.3}$$

Recall that the β-mixing coefficient for X is defined by

$$\beta_\mu(t)\equiv\sup_{s\in\mathbb{R}^+}\mathsf{E}_\mu\left\|\mathsf{P}_\mu(\cdot|\mathcal{F}_0^s)-\mathsf{P}_\mu(\cdot)\right\|_{var,\mathcal{F}_{t+s}^\infty},\quad t\in\mathbb{R}^+,$$

where $\mathcal{F}_a^b\equiv\sigma(X(s),s\in[a,b]),\mathsf{P}_\mu(\cdot|\mathcal{F}_0^s)$ denotes the conditional distribution of P_μ w.r.t. $\mathcal{F}_0^s$, and

$$\|\varkappa\|_{var,\mathcal{G}}\equiv\sup_{\substack{B_1\cap B_2=\emptyset,B_1\cup B_2=\mathbb{R}^m\\ B_1,B_2\in\mathcal{G}}}[\varkappa(B_1)-\varkappa(B_2)].$$

If $\mu=\mu_{inv}$, then $\beta(\cdot)\equiv\beta_{\mu_{inv}}(\cdot)$ is the mixing coefficient of the stationary version of the process X.

Theorem 1.2. *Suppose conditions of Theorem 1.1 to hold true. Then*

(i) *for every* $\mu\in\mathcal{P}$ *with* $\boldsymbol{\phi}(\mu)<+\infty$

$$\beta_\mu(t)\leq\mathsf{C}_1[\boldsymbol{\phi}(\mu)+1]\exp[-\mathsf{C}_2t]; \tag{1.4}$$

(ii) $\boldsymbol{\phi}(\mu_{inv})<+\infty$, *and thus the mixing coefficient* $\beta(\cdot)$ *allows the exponential estimate (1.4).*

In order to shorten exposition, we do not formulate here typical applications of the estimates of the type (1.4), such as Central Limit Theorem, referring the reader to the literature (see, for instance, Theorem 4 [28]).

The following theorem, that gives sufficient conditions for condition **LD** to hold true, is the main result of the present paper. We need some additional notation.

In the case **A**, put $\tilde{a}(\cdot) = a(\cdot) - \int_{\|u\|\leq 1} c(\cdot, u)\Pi(du)$, and denote

$$\Delta(x, u) = [\tilde{a}(x + c(x, u)) - \tilde{a}(x)] - [\nabla_x c(x, u)]\tilde{a}(x).$$

In the case **B**, denote $\Delta(x, u) = [a(x + u)) - a(x)]$ (note that if in the case **B** condition (1.2) holds true, then these two formulas define the same function). Denote, by $\{\mathcal{E}_s^t, 0 \leq s \leq t\}$, the solution to the linear SDE in $\mathbb{R}^{m\times m}$

$$\mathcal{E}_s^t = I_{\mathbb{R}^m} + \int_s^t \nabla a(X(r))\mathcal{E}_s^r\, dr + \int_s^t \int_{\|u\|\leq 1} \nabla_x c(X(r-))\mathcal{E}_s^{r-}\tilde{\nu}(dr, du) + \int_s^t \int_{\|u\|>1} \nabla_x c(X(r-))\mathcal{E}_s^{r-}\nu(dr, du)$$

($\mathcal{E}_s^t$ is supposed to be right continuous w.r.t. both time variables s, t), and define the random linear space S_t as the span (in $\mathbb{R}^m$) of the set

$$\{\mathcal{E}_\tau^t \Delta(X(\tau-), p(\tau)), \tau \in \mathcal{D} \cap (0, t)\}.$$

Theorem 1.3. *Let the two following conditions to hold true.*

N. *There exist $x_* \in \mathbb{R}^m, t_* > 0$ such that*

$$\mathsf{P}_{x_*}(S_{t_*} = \mathbb{R}^m) > 0.$$

S. *For any $R > 0$ there exists $t = t(R)$ such that*

$$x_* \in \operatorname{supp} P_x^t, \quad \|x\| \leq R.$$

Then condition **LD** *holds true.*

In Section 4 below, we give some sufficient conditions for **R**,**N**,**S** to hold true, formulated in the terms of the coefficients of the equation (0.1) and Lévy measure of the jump noise.

2. Time-stretching transformations and convergence in variation of induced measures

In this section, we give necessary background for the technique involving time-stretching transformations of the Lévy point measure, with its applications to the problem of convergence in variation of the distributions of the solutions to SDE's with jumps. This technique is our tool in the proof of Theorem 1.3. Some of the statements we give without proofs, referring to the recent papers [18],[17].

Denote $H = L_2(\mathbb{R}^+), H_0 = L_\infty(\mathbb{R}^+) \cap L_2(\mathbb{R}^+), Jh(\cdot) = \int_0^\cdot h(s)\, ds, h \in H$. For a fixed $h \in H_0$, we define the family $\{T_h^t, t \in \mathbb{R}\}$ of transformations of the axis $\mathbb{R}^+$ by putting $T_h^t x, x \in \mathbb{R}^+$ equal to the value at the point $s = t$ of the solution to the Cauchy problem

$$z'_{x,h}(s) = Jh(z_{x,h}(s)), \quad s \in \mathbb{R}, \qquad z_{x,h}(0) = x. \tag{2.1}$$

Denote $T_h \equiv T_h^1$, then $T_{sh} \circ T_{th} = T_{(s+t)h}$ ([18]). This means that $\mathcal{T}_h \equiv \{T_{th}, t \in \mathbb{R}\}$ is a one-dimensional group of transformations of the time axis $\mathbb{R}^+$. It follows from the construction that $\frac{d}{dt}|_{t=0} T_{th}x = Jh(x), x \in \mathbb{R}^+$. We call T_h the *time stretching transformation* because, for $h \in C(\mathbb{R}^+) \cap H_0$, it can be informally described in the following way: every infinitesimal segment dx of the time axis should be stretched by $e^{h(x)}$ times, and then all the stretched segments should be glued together, preserving initial order of the segments ([18]).

Denote $\Pi_{fin} = \{\Gamma \in \mathcal{B}(\mathbb{R}^d), \Pi(\Gamma) < +\infty\}$ and define, for $h \in H_0, \Gamma \in \Pi_{fin}$, a transformation T_h^Γ of the random measure ν by

$$[T_h^\Gamma \nu]([0, t] \times \Delta) = \nu([0, T_h t] \times (\Delta \cap \Gamma)) + \nu([0, t] \times (\Delta \backslash \Gamma)), \quad t \in \mathbb{R}^+, \Delta \in \Pi_{fin}.$$

Further we use the standard terminology from the theory of Poisson point measures without any additional discussion. The term "(locally finite) configuration" for a realization of the point measure is frequently used. We suppose that the basic probability space $(\Omega, \mathcal{F}, P)$ satisfies condition $\mathcal{F} = \sigma(\nu)$, i.e. every random variable is a functional of ν. This means that in fact one can treat Ω as the configuration space over $\mathbb{R}^+ \times (\mathbb{R}^d\backslash\{0\})$ with a respective σ-algebra. The image of a configuration of the point measure ν under T_h^Γ can be described in a following way: every point (τ, x) with $x \notin \Gamma$ remains unchanged; for every point (τ, x) with $x \in \Gamma$, its "moment of the jump" τ is transformed to $T_{-h}\tau$; neither any point of the configuration is eliminated nor any new point is added to the configuration.

For $h \in H_0, \Gamma \in \Pi_{fin}$ transformation T_h^Γ is admissible for ν, i.e. the distributions of the point measures ν and $T_h^\Gamma \nu$ are equivalent ([18]). This imply that (recall that $\mathcal{F} = \sigma(\nu)$) the transformation T_h^Γ generates the corresponding transformation of the random variables, we denote it also by T_h^Γ.

Define $\mathcal{C}$ as the set of functionals $f \in \cap_p L_p(\Omega, P)$ satisfying the following condition: for every $\Gamma \in \Pi_{fin}$, there exists the random element $\nabla_H^\Gamma f \in \cap_p L_p(\Omega, P, H)$ such that, for every $h \in H_0$,

$$(\nabla_H^\Gamma f, h)_H = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}[T_{\varepsilon h}^\Gamma \circ f - f] \tag{2.2}$$

with convergence in every $L_p, p < +\infty$.

Denote

$$(\rho^\Gamma, h) = -\int_0^\infty h(t)\, \tilde{\nu}(dt, \Gamma), \quad h \in H_0, \Gamma \in \Pi_{fin}.$$

Lemma 2.1. *(*[18]*, Lemma 3.2). For every $\Gamma \in \Pi_{fin}$, the pair $(\nabla_H^\Gamma, \mathcal{C})$ satisfies the following conditions:*

1) For every $f_1, \dots, f_n \in \mathcal{C}$ and $F \in C_b^1(\mathbb{R}^n)$,

$$F(f_1, \dots, f_n) \in \mathcal{C} \quad \textit{and} \quad \nabla_H F(f_1, \dots, f_n) = \sum_{k=1}^n F_k'(f_1, \dots, f_n) \nabla_H f_k$$

(chain rule).

2) The map $\rho^\Gamma : h \mapsto (\rho^\Gamma, h)$ is a weak random element in H with weak moments of all orders, and

$$E(\nabla_H^\Gamma f, h)_H = -Ef(\rho^\Gamma, h), \quad h \in H, f \in \mathcal{C}$$

(integration-by-parts formula).

3) There exists a countable set $\mathcal{C}_0 \subset \mathcal{C}$ such that $\sigma(\mathcal{C}_0) = \mathcal{F}$.

The construction described before gives us the family $\mathcal{T} = \{T_h^\Gamma, h \in H_0, \Gamma \in \Pi_{fin}\}$ of the admissible transformations of the probability space $(\Omega, \mathcal{F}, P)$, such that the probability P is *logarithmically differentiable* w.r.t. every T_h^Γ with the correspondent logarithmic derivative equal (ρ^Γ, h). This allows us to introduce a derivative w.r.t. such a family, that is an analogue to the Malliavin derivative on the Wiener space or the Sobolev derivative on the finite-dimensional space. However, the structure of the family $\mathcal{T}$ differs from the structure of the family of the linear shifts: for instance, there exist $h, g \in H_0, \Gamma \in \Pi_{fin}$ such that $T_h^\Gamma \circ T_g^\Gamma \neq T_g^\Gamma \circ T_h^\Gamma$. This feature motivates the following "refinement", introduced in [18], of the construction described before.

Definition 2.2. A family $\mathcal{G} = \{[a_i, b_i) \subset \mathbb{R}^+, h_i \in H_0, \Gamma_i \in \Pi_{fin}, i \in \mathbb{N}\}$ is called *a differential grid* (or simply *a grid*) if

(i) for every $i \neq j$, $\Big([a_i, b_i) \times \Gamma_i\Big) \cap \Big([a_j, b_j) \times \Gamma_j\Big) = \emptyset$;

(ii) for every $i \in \mathbb{N}$, $Jh_i > 0$ inside (a_i, b_i) and $Jh_i = 0$ outside (a_i, b_i).

Any grid $\mathcal{G}$ generates a partition of some part of the phase space $\mathbb{R}^+ \times (\mathbb{R}^d\backslash\{0\})$ of the random measure ν into the cells $\{\mathcal{G}_i = [a_i, b_i) \times \Gamma_i\}$. We call the grid $\mathcal{G}$ *finite*, if $\mathcal{G}_i = \emptyset$ for all indices $i \in \mathbb{N}$ except some finite number of them. Although while studying some other problems (such as smoothness of the transition probability density for X, see [18]) we typically use infinite grids, in our current exposition we can restrict ourselves by a finite grids with the number of non-empty cells equal to m (recall that m is the dimension of the phase space for the equation (0.1)). Thus, everywhere below we simplify notation from [18] and consider the grids $\mathcal{G}$ with index i varying from 1 to m.

Denote $T_s^i=T_{sh_i}^{\Gamma_i}$. For any $i\leq m, s,\tilde s\in\mathbb{R}$, the transformations T_s^i,$T_{\tilde s}^i$ commute because so do the time axis transformations T_{sh_i},$T_{\tilde s h_i}$. Transformation T_s^i does not change points of configuration outside the cell $\mathcal{G}_i$ and keeps the points from this cell in it. Therefore, for every $i,\tilde i\leq m, s,\tilde s\in\mathbb{R}$, transformations T_s^i,$T_{\tilde s}^{\tilde i}$ commute, which implies the following proposition ([18]).

Proposition 2.3. *For a given grid $\mathcal{G}$ and $t=(t_1,\dots,t_m)\in\mathbb{R}^m$, define the transformation*

$$T_t^{\mathcal{G}}=T_{t_1}^1\circ T_{t_2}^2\circ\dots\circ T_{t_m}^m.$$

Then $\mathcal{T}^{\mathcal{G}}=\{T_t^{\mathcal{G}},l\in\mathbb{R}^m\}$ is the group of admissible transformations of Ω which is additive in the sense that $T^{\mathcal{G}}_{t^1+t^2}=T^{\mathcal{G}}_{t^1}\circ T^{\mathcal{G}}_{t^2}, t^{1,2}\in\mathbb{R}^m$.

It can be said that, by fixing the grid $\mathcal{G}$, we choose from the whole family of admissible transformations $\{T_h^\Gamma, h\in H_0,\Gamma\in\Pi_{fin}\}$ the additive sub-family, that is more convenient to deal with. The following lemma describes the differential properties of the solution to (0.1) w.r.t. this family. We denote by $X(x,\cdot)$ the strong solution to (0.1) with $X(0)=x$.

Lemma 2.4. **I**. *In the case* **B**, *for every $t\in\mathbb{R}^+,\Gamma\in\Pi_{fin}$, every component $X_k(t), k=1,\dots,m$ of the vector $X(t)$ belongs to the class $\mathcal{C}$. For every $h\in H_0, x\in\mathbb{R}^m$, the process*

$$Y^{h,\Gamma}(x,t)\equiv((\nabla_H^\Gamma X_1(x,t),h)_H,\dots,(\nabla_H^\Gamma X_m(x,t),h)_H)^\top,\quad x\in\mathbb{R}^m,t\in\mathbb{R}^+$$

satisfies the equation

$$Y^{h,\Gamma}(x,t)=\int_0^t\int_\Gamma\Delta(X(x,s-),u)Jh(s)\,\nu(ds,du)+\int_0^t[\nabla a](X(x,s))Y^{h,\Gamma}(x,s)\,ds,\quad t\geq 0. \tag{2.3}$$

II. *In the case* **A**, *for every $x\in\mathbb{R}^m,t\in\mathbb{R}^+,\Gamma\in\Pi_{fin},h\in H_0$, every component of the vector $X(x,t)$ is a.s. differentiable w.r.t. $\{T_{rh}^\Gamma,r\in\mathbb{R}\}$, i.e., there exist a.s. limits*

$$Y_k^{h,\Gamma}(x,t)=\lim_{\varepsilon\to0}\frac{1}{\varepsilon}[T_{\varepsilon h}^\Gamma X_k(x,t)-X_k(x,t)],\quad k=1,\dots,m. \tag{2.4}$$

The process $Y^{h,\Gamma}(x,t)=(Y_1^{h,\Gamma}(x,t),\dots,Y_m^{h,\Gamma}(x,t))^\top$ satisfies the equation

$$Y^{h,\Gamma}(x,t)=\iint_{[0,t]\times\Gamma}\Delta(X(x,s-),u)Jh(s)\,\nu(ds,du)+$$

$$+\int_0^t[\nabla a](X(x,s))Y^{h,\Gamma}(x,s)\,ds+\iint_{[0,t]\times\mathbb{R}^d}[\nabla_x c](X(x,s-),u)Y^{h,\Gamma}(x,s-)\tilde\nu(ds,du),\quad t\geq0. \tag{2.5}$$

Statement **I** is proved in [18], Theorem 4.1; statement **II** is proved in [16], Lemma 4.1.

Remark. Solutions to equations (2.3), (2.5) can be given explicitly:

$$Y^{h,\Gamma}(t)=\iint_{[0,t]\times\Gamma}Jh(s)\cdot\mathcal{E}_s^t\Delta(X(x,s-),u)\nu(ds,du)=\sum_{\tau\in\mathcal{D},p(\tau)\in\Gamma}Jh(\tau)\cdot\mathcal{E}_\tau^t\Delta(X(x,\tau-),p(\tau)). \tag{2.6}$$

For a given grid $\mathcal{G}=\{[a_i,b_i)\subset\mathbb{R}^+,h_i\in H_0,\Gamma_i\in\Pi_{fin},i\leq m\}$, denote $Y^{\mathcal{G},i}=Y^{h_i,\Gamma_i},\quad i=1,\dots,m$ and consider the matrix-valued process

$$Y^{\mathcal{G}}(x,t)\equiv(Y_k^i(x,t))_{i,k=1}^m,\quad t\in\mathbb{R}^+.$$

The following lemma is the key point in our approach. The statement of the lemma is formulated for the cases **A** and **B** simultaneously.

Lemma 2.5. *Let $x\in\mathbb{R}^m,t>0$ be fixed, denote $\Omega_{x,t}\equiv\{\det Y^{\mathcal{G}}(x,t)\neq0\}$. Then*

$$P|_{\Omega_{x,t}}\circ[X(y,t)]^{-1}\xrightarrow{var}P|_{\Omega_{x,t}}\circ[X(x,t)]^{-1},\quad y\to x. \tag{2.7}$$

The proof is based on the following criterium for convergence in variation of induced measures on a finite-dimensional space, obtained in [1].

Theorem 2.6. *Let $F, F_n : \mathbb{R}^m \to \mathbb{R}^m$ be measurable functions, that have the approximative derivatives $\nabla F_n, \nabla F$ a.s. w.r.t Lebesgue measure λ^m, and $E \in \mathcal{B}(\mathbb{R}^m)$ has finite Lebesgue measure. Suppose that $F_n \to F$ and $\nabla F_n \to \nabla F$ in a sense of convergence in measure λ^m and $\det \nabla F \neq 0$ a.s. on E. Then the following statements are equivalent.*

(i) for every measurable $A \subset E$ $\quad \lambda^m|_A \circ F_n^{-1} \xrightarrow{var} \lambda^m|_A \circ F^{-1}, n \to +\infty$;

(ii) for every measurable $A \subset E$ and every $\delta > 0$ there exists a compact set $K_\delta \subset A$ such that $\lambda^m(A \backslash K_\delta) \le \delta$ and $\lim_{n\to+\infty} \lambda^m(F_n(K_\delta)) = \lambda^m(F(K_\delta))$.

In the situation, described in the preamble of the Theorem 2.6, both (i) and (ii) can fail (see, for instance, Example 1.2 [17]). Thus, in order to provide (i) (that is our goal), we should impose some additional conditions on the sequence $\{F_n\}$, sufficient for (ii) to hold true. The following two sufficient conditions were proved in [1], Corollaries 2.5 and 2.7, and in [17], Theorem 3.1, correspondingly.

Proposition 2.7. I. *Let $F, F_n \in W^1_{p,loc}(\mathbb{R}^m, \mathbb{R}^m)$ with $p \ge m$ ($W^1_{p,loc}$ denotes the local Sobolev space), and $F_n \to F, n \to \infty$ w.r.t. Sobolev norm $\|\cdot\|_{W^1_p(\mathbb{R}^m,\mathbb{R}^m)}$ on every ball. Then*

$$\lambda^m|_A \circ F_n^{-1} \xrightarrow{var} \lambda^m|_A \circ F^{-1}, n \to +\infty \text{ for every measurable } A \subset \{\det \nabla F \neq 0\}. \tag{2.8}$$

II. *Let in the situation, described in the preamble of the Theorem 2.6, the sequence $\{F_n\}$ be **uniformly approximatively Lipschitz**. This, by definition, means that for every $\delta > 0, R < +\infty$ there exist a compact set $K_{\delta,R}$ and a constant $L_{\delta,R} < +\infty$ such that $\lambda^m(B_{\mathbb{R}^m}(0,R)\backslash K_\delta) < \delta$ and every function $F_n|_{K_\delta}$ is a Lipschitz function with the Lipschitz constant $L_{\delta,R}$. Then (2.8) holds true.*

Proof of the Lemma 2.5. Take the sequence $y_n \to x, n \to +\infty$, and denote $f = X(x,t), f_n = X(y_n,t)$. It is proved in [18] (proof of Theorem 3.1), that the group $\mathcal{T}^{\mathcal{G}} \equiv \{T^{\mathcal{G}}_s, s \in \mathbb{R}^m\}$ generates a measurable parametrization of $(\Omega, \mathcal{F}, P)$, i.e. there exists a measurable map $\Phi : \Omega \to \mathbb{R}^m \times \tilde{\Omega}$ such that $\tilde{\Omega}$ is a Borel measurable space and the image of every orbit of the group $\mathcal{T}^{\mathcal{G}}$ under Φ has the form $L \times \{\varpi\}$, where $\varpi \in \tilde{\Omega}$ and L is a linear subspace of $\mathbb{R}^m$. The linear subspace L differs from $\mathbb{R}^m$ exactly in the case, when the orbit $T^{\mathcal{G}}\{\omega\}$ is built for such an ω, that, for some $i = 1, \dots, m$, $\nu((a_i, b_i) \times \Gamma_i) = 0$ (i.e., some T^i does not change ω). For every ω of such a type $\det Y^{\mathcal{G}}(x,t) = 0$, and thus we need to investigate the laws of f, f_n, restricted to $\Omega_{\mathcal{G}} \equiv \{\nu((a_i, b_i) \times \Gamma_i) > 0, i = 1, \dots, m\}$, only.

The measure $P|_{\Omega_{\mathcal{G}}}$ can be decomposed into a regular family of conditional distributions such that every conditional distribution is supported by an orbit of the group $\mathcal{T}^{\mathcal{G}}$ (see, for instance, [23]). Therefore we can write

$$P(A) = \int_{\tilde{\Omega}} P_\varpi([A]_\varpi)\pi(\varpi), \quad A \in \mathcal{F} \cap \Omega_{\mathcal{G}}, \tag{2.9}$$

where $[A]_\varpi = \{s \in \mathbb{R}^m | (s,\varpi) \in A\}$, π is the image of $P|_{\Omega_{\mathcal{G}}}$ under the natural projection $\Omega \to \tilde{\Omega}$ and $P_\cdot(\cdot)$ is a *probabilistic kernel*, i.e. $P_\cdot(A)$ is a measurable function for every $A \in \mathcal{B}(\mathbb{R}^m)$ and $P_\varpi(\cdot)$ is a probability measure on $\mathbb{R}^m$ for every $\varpi \in \tilde{\Omega}$. Any functional g on $\Omega_{\mathcal{G}}$ now can be considered as a functional on $\mathbb{R}^m \times \tilde{\Omega}$, $g = \{g(s,\varpi), s \in \mathbb{R}^m, \varpi \in \tilde{\Omega}\}$. Below, we denote $[g]_\varpi(\cdot) \equiv g(\cdot,\varpi) : \mathbb{R}^m \to \mathbb{R}^m$, $\varpi \in \tilde{\Omega}$. One can write, for $A \subset \Omega_{\mathcal{G}}$, that

$$(P|_A \circ g^{-1})(\cdot) = \int_{\tilde{\Omega}} [P_\varpi|_{A_\varpi} \circ [g]^{-1}_\varpi](\cdot)\pi(d\varpi).$$

Therefore, in order to prove the statement of Lemma 2.5, it is sufficient to prove that

$$P_\varpi|_{[\Omega_{x,t}]_\varpi} \circ [f_n]_\varpi \xrightarrow{var} P_\varpi|_{[\Omega_{x,t}]_\varpi} \circ [f]_\varpi \quad \text{for } \pi\text{-almost all } \varpi \in \tilde{P}. \tag{2.10}$$

Denote $\rho_i = (\rho^{\Gamma_i}, h_i), i = 1, \dots, m$, and let $\rho^{\mathcal{G}}$ be $\mathbb{R}^m$-valued function such that $(\rho^{\mathcal{G}}, t) = \sum_{i=1}^m t_i\rho_i, t \in \mathbb{R}^m$. Then one can show that, for π-almost all $\varpi \in \tilde{\Omega}$, the measure P_ϖ possesses the logarithmic derivative equal to $[\rho^{\mathcal{G}}]_\varpi$ (we do not give the detailed exposition here, since this fact is quite analogous to the one for logarithmically differentiable measures on linear spaces, see [5]). One can deduce from the explicit formula for $\rho^{\mathcal{G}}$ that there exists $c > 0$ such that $E\exp[(\rho^{\mathcal{G}}, t)] < +\infty, \|t\| \leq c$, and, therefore, that for π-almost all $\varpi \in \tilde{\Omega}$

$$\int_{\mathbb{R}^m} \exp[([\rho^{\mathcal{G}}]_\varpi(s), t)] P_\varpi(ds) < +\infty, \quad \|t\| \leq c. \tag{2.11}$$

Due to Proposition 4.3.1 [4], for every ϖ such that (2.11) holds true, the measure P_ϖ has the form $P_\varpi(dx) = p_\varpi(x)\lambda^m(dx)$, where the function p_ϖ is continuous and positive. Therefore, in order to prove (2.10), it is enough to prove that, for π-almost all such ϖ and every $R > 0$,

$$\lambda^m|_{B_{\mathbb{R}^m}(0,R)\cap[\Omega_{x,t}]_\varpi} \circ [f_n]_\varpi \xrightarrow{var} \lambda^m|_{B_{\mathbb{R}^m}(0,R)\cap[\Omega_{x,t}]_\varpi} \circ [f]_\varpi \quad \text{for } \pi\text{-almost all } \varpi \in \tilde{P}. \tag{2.12}$$

In order to prove (2.12), let us introduce auxiliary notions and give their relations with the notions of Sobolev and approximative derivatives.

Definition 2.8. For a given function $F : \mathbb{R}^m \to \mathbb{R}^m$ and σ-finite measure $\varkappa$ on $\mathcal{B}(\mathbb{R}^m)$ we say that F is direction-wise $\varkappa$-a.s. differentiable, if there exists function $\nabla F : \mathbb{R}^m \to \mathbb{R}^{m\times m}$ such that, for every $t \in \mathbb{R}^m$, $\frac{1}{\varepsilon}[F(\cdot + t\varepsilon) - F(\cdot)] \to (\nabla F(\cdot), t), \varepsilon \to 0$ $\varkappa$-a.s. We say that g is direction-wise differentiable in the $L_{p,loc}(\varkappa)$ sense, if $F \in L_{p,loc}(\varkappa)$ and, for every $t \in \mathbb{R}^m$, $\frac{1}{\varepsilon}[F(\cdot + t\varepsilon) - F(\cdot)] \to (\nabla F(\cdot), t), \varepsilon \to 0$ in $L_{p,loc}(\varkappa)$.

Proposition 2.9. *1. Let $\varkappa(dx) = p(x)\lambda^m(dx)$ with $p(x) \geq C_R > 0, \|x\| \leq R$ for any $R > 0$. Then every function F, that is direction-wise $\varkappa$-a.s. differentiable, is also direction-wise λ^m-a.s. differentiable, and every function F, that is direction-wise differentiable in $L_{p,loc}(\varkappa)$ sense, is also direction-wise differentiable in $L_{p,loc}(\lambda^m)$ sense. The function ∇F from the definition of $\varkappa$-differentiability (either in a.s. or $L_{p,loc}$ sense) λ^m-a.s. coincides with the one from the definition of λ^m-differentiability.*

2. If F is direction-wise $L_{p,loc}(\lambda^m)$-differentiable, then $F \in W_{p,loc}(\mathbb{R}^m, \mathbb{R}^m)$ and ∇F coincides with its Sobolev derivative.

3. If F is direction-wise λ^m-a.s. differentiable, then F has approximative derivative at λ^m-almost all points $x \in \mathbb{R}^m$, and ∇F coincides with its approximative derivative.

Proof. Statements 1 and 2 immediately follow from the definition. Statement 3 follow from Theorem 3.1.4 [11] and the trivial fact, that the usual differentiability at some point w.r.t. given direction implies approximative differentiability at the same point w.r.t. this direction.

Now, we can finish the proof of Lemma 2.5. Denote $f_{n,\varpi} = [f_n]_\varpi, f_\varpi = [f]_\varpi$. By the construction, $[T_r^{\mathcal{G}} f_n]_\varpi(s) = [f_n]_\varpi(s + r), \quad s, r \in \mathbb{R}^m, \varpi \in \tilde{\Omega}$. This, together with (2.9) and Lemma 2.4, provides that there exists $\tilde{\Omega}_0 \subset \tilde{\Omega}$ with $\pi(\tilde{\Omega}\backslash\tilde{\Omega}_0) = 0$ such that, for every $\varpi \in \tilde{\Omega}_0$, (2.11) holds true, and the functions $f_{n,\varpi}, f_\varpi$ are either direction-wise P_ϖ-differentiable (in the case **A**), or direction-wise differentiable in the $L_{p,loc}(P_\varpi)$ sense (in the case **B**), and $\nabla f_{n,\varpi} = [Y^{\mathcal{G}}(y_n, t)]_\varpi, \nabla f_\varpi = [Y^{\mathcal{G}}(x, t)]_\varpi$. This, in particular, means that $[\Omega_{x,t}]_\varpi = \{\det \nabla f_\varpi \neq 0\}$ for $\varpi \in \tilde{\Omega}_0$.

Now, we can apply the standard theorem on L_p-continuity of the solution to an SDE w.r.t. initial condition (see Theorem 4, Chapter 4.2 [12]), and obtain that $f_n \to f, [Y^{\mathcal{G}}(y_n, t)] \to [Y^{\mathcal{G}}(x, t)], n \to \infty$, in L_p sense. This implies that, for π-almost all $\varpi \in \tilde{\Omega}_0$, $f_{n,\varpi} \to f_\varpi, \nabla f_{n,\varpi} \to \nabla f_\varpi, n \to \infty$ in $L_p(P_\varpi)$ sense, and, therefore, in $L_{p,loc}(\lambda^m)$ sense. In the case **B**, convergence (2.12) (and thus the statement of the lemma) follows straightforwardly from the statement **I** of Proposition 2.7. In the case **A**, convergence (2.12) follows from the statement **II** of the same Proposition, and Lemma 3.3 [17], that provides that, for π-almost all $\varpi \in \tilde{\Omega}$, the sequence $\{\nabla f_{n,\varpi}\}$ is uniformly approximatively Lipschitz on $\mathbb{R}^m$. The lemma is proved.

3. Proofs of the main results

3.1. **Proof of Theorem 1.3.** We prove Theorem 1.3 in two steps. First, we use Lemma 2.5 and show that, under condition **N**, the local Doeblin condition holds true inside some small ball.

Lemma 3.1. *Under condition* **N**, *there exists $\varepsilon_* > 0$, such that*

$$\inf_{x,y\in B(x_*,\varepsilon_*)}\int_{\mathbb{R}^m}\left[P_x^{t_*}\wedge P_y^{t_*}\right](dz)>0. \tag{3.1}$$

Proof. Suppose that the grid $\mathcal{G}$ is such that, in the notation of Lemma 2.5,

$$P(\Omega_{x_*,t_*})>0, \tag{3.2}$$

and denote $P_{x_*,x}(dz)=\mathsf{P}_x(\Omega_{x_*,t_*},X(x,t_*)\in dz)$. One can see that $P_{x_*,x}(dz)=p_{x_*,x}(z)P_x^{t_*}(dz)$ with $p_{x_*,x}\le 1$, and therefore

$$\int_{\mathbb{R}^m}\left[P_{x_*,x}\wedge P_{x_*,y}\right](dz)\le\int_{\mathbb{R}^m}\left[P_x^{t_*}\wedge P_y^{t_*}\right](dz),\quad x,y\in\mathbb{R}^m.$$

Thus, in order to prove (3.1), it is enough to prove that

$$\inf_{x,y\in B(x_*,\varepsilon_*)}\int_{\mathbb{R}^m}\left[P_{x_*,x}\wedge P_{x_*,y}\right](dz)>0.$$

The latter inequality follows from the condition (3.2), Lemma 2.5 and relation

$$\int_{\mathbb{R}^m}\left[P_{x_*,x}\wedge P_{x_*,y}\right](dz)=P(\Omega_{x_*,t_*})-\frac{1}{2}\|P_{x_*,x}(\cdot)-P_{x_*,y}(\cdot)\|_{var}\to P(\Omega_{x_*,t_*}),\quad x,y\to x_*.$$

Thus, the only thing left to show is that, under condition **N**, the grid $\mathcal{G}$ can be chosen in such a way that (3.2) holds true. Denote, by $\mathcal{J}_m$, the family of all rational partitions of $(0,t_*)$ of the length $2m$; any $J\in\mathcal{J}_m$ is the set of the type

$$J=\{a_1,b_1,a_2,b_2,\dots,a_m,b_m\},\quad\text{with}\quad 0<a_1<b_1<\dots<a_m<b_m<t_*,\quad a_i,b_i\in\mathbb{Q},\quad i=1,\dots,m.$$

Denote, for the set J of such a type and $r\in\mathbb{N}$,

$$\Omega_{J,r}=\Bigg\{\forall\, i=1,\dots,m\quad\exists!\,\tau_i\in(a_i,b_i)\cap\mathcal{D},\ \|p(\tau_i)\|\ge\frac{1}{r},\ i=1,\dots,m$$
$$\text{and span}\,\{\mathcal{E}^t_{\tau_i}\Delta(X(x_*,\tau_i-),p(\tau_i)),i=1,\dots,m\}=\mathbb{R}^m\Bigg\}.$$

Then, elementary considerations show that

$$\Omega_{x_*,t_*}=\bigcup_{J\in\mathcal{J}_m,r\in\mathbb{N}}\Omega_J.$$

Therefore, under condition **N**, there exist $J^*=\{a_1^*,b_1^*,\dots,a_m^*,b_m^*\}$ and $r^*\in\mathbb{N}$ such that $P(\Omega_{J^*,r^*})>0$. Let $h_i^*\in H_0,i=1,\dots,m$ be arbitrary functions such that, for any i, $Jh_i^*>0$ inside (a_i^*,b_i^*) and $Jh_i^*=0$ outside (a_i^*,b_i^*). Consider the grid $\mathcal{G}^*$ with

$$\Gamma_i=\begin{cases}\{u|\|u\|\ge\frac{1}{r}\}, & i\le m\\ \emptyset, & i>m\end{cases},\quad a_i=a_i^*,b_i=b_i^*,h_i=h_i^*,\ i\le m,\quad a_i,b_i,h_i\text{ are arbitrary for }i>m.$$

Then formula (2.6) shows that, on the set Ω_{J^*,r^*},

$$Y^{\mathcal{G}^*,i}(x_*,t_*)\equiv Y^{h_i,\Gamma_i}(x_*,t_*)=Jh_i(\tau_i)\mathcal{E}_{\tau_i}\Delta(X(x_*,\tau_i-),p(\tau_i)),\quad i=1,\dots,m.$$

Since $Jh_i(\tau_i)\ne 0$ by the construction, the vectors $\{Y^{\mathcal{G}^*,i}(x_*,t_*)\}$ are linearly independent iff so are the vectors $\{\mathcal{E}_{\tau_i}\Delta(X(x_*,\tau_i-),p(\tau_i))\}$. Thus,

$$P(\Omega_{x_*,t_*})=P(\det Y(x_*,t_*)\ne 0)\ge P(\Omega_{J^*,r^*})>0,$$

that gives (3.2). The lemma is proved.

The last step in the proof of Theorem 1.3 is to combine statement of the previous lemma with the condition **S** and show, that the local Doeblin condition holds true in any bounded region of $\mathbb{R}^m$.

Lemma 3.2. *Under conditions* **N** *and* **S**, *condition* **LD** *holds true with*

$$T(R)=t(R)+t_*,\quad R>0.$$

Proof. The process X is a Feller one; this follows, for instance, from Theorem 4, Chapter 4.2 [12]. Therefore, the function $x\mapsto P_x^t(O)$ is lower semicontinuous for any open set O and any $t>0$. This, together with condition **S**, provides that, for any $R>0$,

$$\delta(R)\equiv\inf_{\|x\|\leq R}P_x^{t(R)}(B(x_*,\varepsilon_*))>0.$$

Denote

$$\gamma_*=\inf_{x,y\in B(x_*,\varepsilon_*)}\int_{\mathbb{R}^m}\left[P_x^{t_*}\wedge P_y^{t_*}\right](dz)=\inf_{x,y\in B(x_*,\varepsilon_*)}\left[1-\frac{1}{2}\left\|P_x^{t_*}-P_y^{t_*}\right\|_{var}\right]>0,$$

then, for any $x,y\in B(x_*,\varepsilon_*)$ and any $A\in\mathcal{B}(\mathbb{R}^m)$.

$$P_x^{t_*}(A)+P_y^{t_*}(\mathbb{R}^m\backslash A)\leq 2-2\gamma_*.$$

Take some $R>0$ and denote $T=T(R)=t(R)+t_*$. Take two independent processes X^1,X^2, satisfying equations of the type (0.1) with the independent point measures ν^1,ν^2 and starting from the points x,y. Then, for $x,y\in B(0,R)$ and any given $A\in\mathcal{B}(\mathbb{R}^m)$, we can write

$$P_x^T(A)+P_y^T(\mathbb{R}^m\backslash A)=E\Big[P^{t_*}_{X^1(t(R))}(A)+P^{t_*}_{X^2(t(R))}(\mathbb{R}^m\backslash A)\Big]\leq$$

$$\leq E\Big[2\mathbf{1}_{\{X^1(t(R))\notin B(t_*,\varepsilon_*)\}\cup\{X^2(t(R))\notin B(t_*,\varepsilon_*)\}}+(2-2\gamma_*)\mathbf{1}_{\{X^1(t(R)),\{X^2(t(R))\in B(t_*,\varepsilon_*)\}}\Big]\leq 2-2\gamma_*\delta^2(R).$$

Therefore,

$$\inf_{\|x\|,\|y\|\leq R}\int_{\mathbb{R}^m}[P_x^T\wedge P_y^T](dz)=1-\frac{1}{2}\sup_{\|x\|,\|y\|\leq R}\left\|P_x^T-P_y^T\right\|_{var}\geq\gamma_*\delta^2(R)>0.$$

This completes the proof of Lemma 3.2 and Theorem 1.3.

3.2. **Proofs of Theorems 1.1,1.2.** Statements, close to those of of Theorems 1.1,1.2, are well known in different settings, and there exists several well developed ways to prove such kind of a statements. For instance, statement of Theorem 1.1 can be derived straightforwardly from Theorems 5.1, 6.1 [22], since condition **LD** provides that, for any time-discretized process $X^\Delta\equiv\{X(k\Delta),k\in\mathbb{Z}_+\}$ (so called Δ-*skeleton chain*), any compact set is a petite set. However, it is difficult to obtain on this way an explicit expressions (or estimates) for the constants $\mathsf{C}_1,\mathsf{C}_2$, involved in the principal estimates (1.3),(1.4). Therefore, we use another way to prove (1.3),(1.4), based on the coupling technique. In general, we follow the scheme of the proof, proposed for diffusion processes in [27],[28], but our construction of the coupling slightly differs from the one used there. This allows us to exclude from the construction auxiliary conditions, such as Harnack inequality used in [27] or condition (T) used in [28], that are unnatural and restrictive in the context of SDE's with a jump noise.

Let us start with the construction of the coupling used in the proof. Since various coupling constructions are used widely in the literature, we restrict our exposition by the sketch of the construction only, and omit technical details. First, let us give two basic "bricks" of our construction. Everywhere below we call "coupling" any $\mathbb{R}^m\times\mathbb{R}^m$-valued process $Y=(Y^1,Y^2)$ such that the laws of Y^1,Y^2 coincide with $\mathsf{P}_{\mu_1},\mathsf{P}_{\mu_2}$ with some given $\mu_1,\mu_2\in\mathcal{P}$.

1. Simple coupling. We call $Y=(Y^1,Y^2)$ a simple coupling with a starting point $y=(y_1,y_2)\in\mathbb{R}^{2m}$, if it is a coupling with $\mu_{1,2}=\delta_{y_{1,2}}$, and the processes Y^1,Y^2 are

(a) independent, if $y_1\neq y_2$;

(b) equal one to another, if $y_1=y_2$.

In order to show that such process exists one should simply consider two equations of the type (0.1) with random point measures ν_1,ν_2 that are either independent in the case $y_1\neq y_2$, or equal one to another in the case $y_1=y_2$.

2. Gluing coupling. We call $Y=(Y^1,Y^2)$ a gluing coupling with a starting point $y=(y_1,y_2)\in\mathbb{R}^{2m}$ and terminal time $T>0$, if it is a coupling with $\mu_{1,2}=\delta_{y_{1,2}}$, and

(a) $Y^1=Y^2$, if $y_1=y_2$;

(b) $P(Y^1(T)=Y^2(T))=\int_{\mathbb{R}^m}[P^T_{y_1}\wedge P^T_{y_2}](dz)$, if $y_1\neq y_2$.

One can show that such process exists in a following way (one need to consider the case $y_1 \neq y_2$ only). First, due to the standard *Coupling lemma* (also called *Dobrushin lemma*, see [10]), there exists a probability measure $\varkappa$ on $\mathbb{R}^m \times \mathbb{R}^m$ such that

$$\varkappa(\{z = (z_1, z_2)|z_1 = z_2\}) = \int_{\mathbb{R}^m} [P^T_{y_1} \wedge P^T_{y_2}](dz).$$

Then, this measure is considered as a distribution of Y at the moment T, and the distribution of the whole trajectory of Y is defined by this measure and the family of conditional distributions $\{P(Y \in \cdot|Y(T) = z), z \in \mathbb{R}^m \times \mathbb{R}^m\}$ (such a construction is correct since $\mathbb{D}(\mathbb{R}^+, \mathbb{R}^m \times \mathbb{R}^m)$ is a Borel measurable space). Any conditional distribution $P(Y \in \cdot|Y(T) = z), z = (z_1, z_2) \in \mathbb{R}^m \times \mathbb{R}^m$ can be constructed, for instance, as the product of the measures

$$P(X \in \cdot|X(0) = y_1, X(T) = z_1), \quad P(X \in \cdot|X(0) = y_2, X(T) = z_2).$$

Both simple and gluing coupling can be constructed simultaneously on one probability space $(\Omega_*, \mathcal{F}_*, P)$ for all $y = (y_1, y_2) \in \mathbb{R}^m \times \mathbb{R}^m$ in a way, that is jointly measurable in probability variably ω and space variable y. For the simple coupling this follows from the standard theorem on a measurable modification (note that, by the construction, this coupling is continuous in probability w.r.t. y on the sets $\{y_1 \neq y_2\}$ and $\{y_1 = y_2\}$). For the gluing coupling one can verify this using the *lemma about three random variables* ([29]). Further we denote both these couplings with a starting point $y = (y_1, y_2)$ by Y^{y_1,y_2}.

Now, we can describe our construction. We fix $T, R > 0$, that will be defined later. Construct the probability space $(\Omega, \mathcal{F}, P)$ as an infinite product

$$\Omega = \Omega_0 \times \mathop{\times}_{k=1}^{\infty} \Omega_k, \quad \mathcal{F} = \mathcal{F}_0 \otimes \bigotimes_{k=1}^{\infty} \mathcal{F}_k, \quad P = P_0 \times \mathop{\times}_{k=1}^{\infty} P_k,$$

where, for $k \geq 1$, $(\Omega_k, \mathcal{F}_k, P_k) = (\Omega_*, \mathcal{F}_*, P_*)$. Given $\mu_1, \mu_2 \in \mathcal{P}$, construct on $(\Omega_0, \mathcal{F}_0, P_0)$ two independent $\mathbb{R}^m$-valued elements $Z^{1,2}$ with $\mathrm{Law}(Z^{1,2}) = \mu_{1,2}$. Next, consider the simple coupling $\{Y_1^{y_1,y_2}, (y_1, y_2) \in \mathbb{R}^m \times \mathbb{R}^m\}$ that is defined on $(\Omega_1, \mathcal{F}_1, P_1)$, and consider the process $Z_1(t) = Y^{y_1,y_2}(t)\Big|_{y_1=Z^1, y^2=Z^2}, t \geq 0$. Denote

$$Q_1 = \inf\{t|\|Z_1^1(t)\| \leq R, \|Z_1^2(t)\| \leq R\}.$$

Consider the gluing coupling $\{Y_2^{y_1,y_2}, (y_1, y_2) \in \mathbb{R}^m \times \mathbb{R}^m\}$ that is defined on $(\Omega_2, \mathcal{F}_2, P_2)$, and consider the process $Z_2(t) = Y^{y_1,y_2}(t - Q_1)\Big|_{y_1=Z_1^1(Q_1), y^2=Z_1^2(Q_1)}, t \geq Q_1$. Denote $Q_2 = Q_1 + T$. Repeat this construction iteratively: take the next "independent copy" of the simple coupling, substitute the terminal value $Z_2(Q_2)$ as the starting point in it, and wait till the random moment Q_3 when both its coordinates appear inside the ball $\{\|x\| \leq R\}$. Then take the next "independent copy" of the gluing coupling, substitute the terminal value $Z_3(Q_3)$ as the starting point in it, wait till the moment $Q_4 = Q_3 + T$, and so on. Define the process $\{Y(t) = (Y^1(t), Y^2(t)), t \in \mathbb{R}^+\}$, by $Y(t) = Z_k(t), t \in [Q_{k-1}, Q_k)$; below we call this process "switching coupling". It has the following properties by the construction:

(i) $\mathrm{Law}(Y^i(\cdot)) = \mathsf{P}_{\mu_i}, i = 1, 2$;

(ii) for any $k \in \mathbb{N}$, $Y^1(t) = Y^2(t), t \geq Q_k$ as soon as $Y^1(Q_k) = Y^2(Q_k)$;

(iii) $P\Big(Y^1(Q_{2k}) = Y^2(Q_{2k})\Big|Y^1(Q_{2k-1}) \neq Y^2(Q_{2k-1})\Big) \geq \inf_{\|x\|,\|y\| \leq R} \int_{\mathbb{R}^m} [P_x^T \wedge P_y^T](dz), k \in \mathbb{N}$.

Denote $k_* = \min\{k|Y^1(Q_k) = Y^2(Q_k)\}$ and put $Q_* = Q_{k_*}$; Q_* is the "gluing moment" for the coordinates of the switching coupling Y. Let us give some estimates that, together with the property (iii) and condition **LD**, allow one to control the tail probabilities for Q_*. Everywhere below, we suppose that $\boldsymbol{\phi} \geq 0$. This does not restrict generality, since one can replace $\boldsymbol{\phi}$ by $\boldsymbol{\phi} + C$ with a properly chosen constant C.

Fix some $c \in (0, 1)$ and take $R > 0$ such that $\boldsymbol{\phi}(x) > \max\left[\frac{\boldsymbol{\gamma}}{c\boldsymbol{\alpha}}, 1\right]$ for $\|x\| > R$. Denote $L \equiv \inf\{t|\|X(t)\| \leq R\}$.

Lemma 3.3. *For any $\mu \in \mathcal{P}$ with $\boldsymbol{\phi}(\mu) < +\infty$,*

(*a*)
$$\boldsymbol{\phi}(P_\mu^t) \leq \boldsymbol{\phi}(\mu)\exp[-\boldsymbol{\alpha}t] + \frac{\boldsymbol{\gamma}}{\boldsymbol{\alpha}};$$

(b) $$\mathsf{P}_\mu(L>t)\leq \boldsymbol{\phi}(\mu)\exp[-(1-c)\boldsymbol{\alpha}t].$$

Proof. Inequality (a) is a standard corollary of the Dynkin formula and condition $\mathcal{A}\boldsymbol{\phi}\leq -\boldsymbol{\alpha}\boldsymbol{\phi}+\boldsymbol{\gamma}$ (see, for instance, beginning of the proof of Theorem 6.1 [22]). Consider, together with the process X, the process $\tilde X(\cdot)=X(\cdot\wedge L)$ (i.e., the process X, stopped at the first moment of its visit to the ball $\{\|x\|\leq R\}$). By the construction, its (extended) generator $\tilde{\mathcal{A}}$ satisfies, for $|x|>R$, the condition

$$\tilde{\mathcal{A}}\boldsymbol{\phi}(x)=\mathcal{A}\boldsymbol{\phi}(x)\leq -\boldsymbol{\alpha}\boldsymbol{\phi}(x)+\boldsymbol{\gamma}\leq -(1-c)\boldsymbol{\alpha}\boldsymbol{\phi}(x).$$

Then, writing down the relation, analogous to (a), for the process $\tilde X$, we obtain that

$$\mathsf{E}_\mu(\boldsymbol{\phi}(\tilde X(t))1\!\!1_{\|\tilde X(t)\|>R}\leq \int_{\|x\|>R}\boldsymbol{\phi}(x)\mu(dx)\cdot\exp[-(1-c)\boldsymbol{\alpha}t]\leq \boldsymbol{\phi}(\mu)\exp[-(1-c)\boldsymbol{\alpha}t].$$

Since, by the construction, $\boldsymbol{\phi}(x)\geq 1$ for $\|x\|>R$, this implies (b). The lemma is proved.

Corollary 3.4. *There exists an invariant measure μ_* for X, such that $\boldsymbol{\phi}(\mu_*)\leq \dfrac{\boldsymbol{\gamma}}{\boldsymbol{\alpha}}$.*

Proof. Take some $\mu\in\mathcal{P}$ with $\boldsymbol{\phi}(\mu)<+\infty$ and consider the family of measures $\{\mu^t,t\in\mathbb{R}^+\}$,

$$\mu^t\equiv\frac{1}{t}\int_0^t P_\mu^s\,ds$$

(the so called *Khasminskii's averages*). It follows from (a) that $\limsup_t\boldsymbol{\phi}(\mu^t)\leq\dfrac{\boldsymbol{\gamma}}{\boldsymbol{\alpha}}$, and this, together with the Fatou lemma, provides that

(i) the family $\{\mu^t,t\in\mathbb{R}^+\}$ possesses some weak partial limit μ_* as $t\to+\infty$;

(ii) $\boldsymbol{\phi}(\mu_*)\leq\dfrac{\boldsymbol{\gamma}}{\boldsymbol{\alpha}}$.

Moreover, the weak partial limit μ_* is an invariant measure for X (the proof of this fact is simple and standard, so we omit the detailed exposition here). This completes the proof.

Estimates of the Lemma 3.3 can be extended from the process X to the coupling $Y=(Y^1,Y^2)$. The only delicate point here is that Y does not have to be a Markov process, so we need some accuracy in writing down the analogues of the estimates (a),(b). Denote

$$\boldsymbol{\psi}(y)\equiv\boldsymbol{\phi}(y_1)+\boldsymbol{\phi}(y_2)\text{ and }\|y\|_\infty\equiv\max[\|y_1\|,\|y_2\|],\quad y=(y_1,y_2)\in\mathbb{R}^m\times\mathbb{R}^m,$$

then $\boldsymbol{\psi}(y)\to+\infty$, $\|y\|_\infty\to+\infty$. Take $\tilde R>0$ such that $\boldsymbol{\psi}(y)>\max\left[\frac{2\boldsymbol{\gamma}}{c\boldsymbol{\alpha}},1\right]$ for $\|y\|_\infty>\tilde R$. Denote $\tilde L\equiv\inf\{t|\|Y(t)\|_\infty\leq\tilde R\}$.

Lemma 3.5. *(a) Let $\mu_1,\mu_2\in\mathcal{P}$ with $\boldsymbol{\phi}(\mu_1),\boldsymbol{\phi}(\mu_2)<+\infty$, and Y be an arbitrary coupling with* Law $(Y^{1,2})=\mu_{1,2}$. *Then*

$$E\boldsymbol{\psi}(Y(t))\leq[\boldsymbol{\phi}(\mu_1)+\boldsymbol{\phi}(\mu_2)]\exp[-\boldsymbol{\alpha}t]+\frac{2\boldsymbol{\gamma}}{\boldsymbol{\alpha}}.$$

(b) Let $\mu_1,\mu_2\in\mathcal{P}$ with $\boldsymbol{\phi}(\mu_1),\boldsymbol{\phi}(\mu_2)<+\infty$, and Y be the simple coupling with Law $(Y^{1,2})=\mu_{1,2}$. *Then*

$$P(\tilde L>t)\leq 2[\boldsymbol{\phi}(\mu_1)+\boldsymbol{\phi}(\mu_2)]\exp[-(1-c)\boldsymbol{\alpha}t].$$

Proof. The first statement follows immediately from Lemma 3.3 and equality $E\boldsymbol{\psi}(Y(t))=\mathsf{E}_{\mu_1}\boldsymbol{\phi}(X(t))+\mathsf{E}_{\mu_2}\boldsymbol{\phi}(X(t))$. In order to prove (b), one should consider separately the cases $Y^1(0)=Y^2(0)$ and $Y^1(0)\neq Y^2(0)$. In the first case, both coordinates Y^1,Y^2 are the same and move like the process X, i.e., the statement (b) of the Lemma 3.3 implies that

$$P(\tilde L>t,Y^1(0)=Y^2(0))\leq[\boldsymbol{\phi}(\mu_1)+\boldsymbol{\phi}(\mu_2)]\exp[-(1-c)\boldsymbol{\alpha}t].$$

In the second case, the the joint dynamics of the coordinates Y^1,Y^2 is described by the Markov process, and the generator $\hat{\mathcal{A}}$ of this process satisfies the relation

$$\hat{\mathcal{A}}\boldsymbol{\psi}(y)=\mathcal{A}\boldsymbol{\phi}(y_1)+\mathcal{A}\boldsymbol{\phi}(y_2)\leq-\boldsymbol{\alpha}\boldsymbol{\psi}(y)+2\boldsymbol{\gamma},\quad y=(y_1,y_2)\in\mathbb{R}^{m\times m}.$$

Applying the same arguments with those used in the proof of Lemma 3.3, we obtain that

$$P(\tilde{L} > t, Y^1(0) \neq Y^2(0)) \leq [\boldsymbol{\phi}(\mu_1) + \boldsymbol{\phi}(\mu_2)] \exp[-(1-c)\boldsymbol{\alpha} t].$$

This provides the needed estimate. The lemma is proved.

Using the Hölder inequality, we obtain that

$$P(Q_* > t) = \sum_{k=0}^{\infty} P\Big(Q_* > t, t \in (Q_{2k}, Q_{2k+2}]\Big) \leq \sum_{k=0}^{\infty} \Big[P(Y^1(Q_{2k}) \neq Y^1(Q_{2k}))\Big]^{\frac{1}{2}} \Big[P(Q_{2k+2} \geq t)\Big]^{\frac{1}{2}}. \tag{3.3}$$

Denote

$$\delta(T, R) = \inf_{\|x\|, \|y\| \leq R} \int_{\mathbb{R}^m} [P_x^T \wedge P_y^T](dz).$$

By the construction, the event $\{Y^1(Q_{2k}) = Y^1(Q_{2k}\}$ does not depend on the values of the process Y up to the moment Q_{2k-2}, and its probability is not less than $(1 - \delta(T, R))$. Thus, we have an estimate

$$P(Y^1(Q_{2k}) \neq Y^1(Q_{2k})) \leq (1 - \delta(T, R))^k. \tag{3.4}$$

Now let, in the construction of the switching coupling Y above, R to be taken equal to $\tilde{R}$ given prior to Lemma 3.5, and $T = T(\tilde{R})$ (see notation in condition **LD**). Then $\delta(T, R) > 0$, and (3.4) gives an exponential (w.r.t. k) estimate for $P(Y^1(Q_{2k}) \neq Y^1(Q_{2k}))$. Next, $P(Q_{2k+2} \geq t)$ can be estimated in the following way. Denote $\Delta_k = Q_{k+1} - Q_k$, $\mathcal{G}^k = \mathcal{F}_{Q_k}, k \geq 0$, where $\{\mathcal{F}_t, t \in \mathbb{R}^+\}$ is the filtration generated by Y. Then, for every $k \geq 0$, $\Delta_{2k-1} = T$ and, from the construction of the switching coupling Y and statement (b) of Lemma 3.5, we have that

$$P(\Delta_{2k} > t | \mathcal{F}_{2k}) \leq 2[\boldsymbol{\phi}(Y^1(Q_{2k})) + \boldsymbol{\phi}(Y^2(Q_{2k}))] \exp[-(1-c)\boldsymbol{\alpha} t], \quad k \geq 0. \tag{3.5}$$

From the statement (a) of Lemma 3.5, applied to $t = T$ and $\mu_{1,2} = \mathrm{Law}(Y^{1,2}(Q^{2k-1}))$, we obtain that

$$E[\boldsymbol{\phi}(Y^1(Q_{2k})) + \boldsymbol{\phi}(Y^2(Q_{2k})) | \mathcal{F}_{2k-1}] \leq \frac{2\boldsymbol{\gamma}}{\boldsymbol{\alpha}} + 2 \sup_{\|x\| \leq R} \boldsymbol{\phi}(R), \quad k \geq 1,$$

here we used that $\|T^{1,2}(Q_{2k-1})\| \leq \tilde{R}$. This, together with (3.5), gives that

$$P(\Delta_{2k} > t | \mathcal{F}_{2k-1}) \leq \Big[\frac{4\boldsymbol{\gamma}}{\boldsymbol{\alpha}} + 4 \sup_{\|x\| \leq R} \boldsymbol{\phi}(R))\Big] \exp[-(1-c)\boldsymbol{\alpha} t], \quad k \geq 1,$$

and, consequently,

$$E\Big(\exp[\frac{1-c}{2}\Delta_{2k}] | \mathcal{F}_{2k-1}\Big) \leq \Big[\frac{4\boldsymbol{\gamma}}{\boldsymbol{\alpha}} + 4 \sup_{\|x\| \leq R} \boldsymbol{\phi}(R))\Big] \int_0^{\infty} \frac{1-c}{2} e^{-\frac{1-c}{2}z} dz = \Big[\frac{4\boldsymbol{\gamma}}{\boldsymbol{\alpha}} + 4 \sup_{\|x\| \leq R} \boldsymbol{\phi}(R))\Big], \quad k \geq 1.$$

Analogously, we have that

$$E(\exp[\frac{1-c}{2}\Delta_0]) \leq 2[\boldsymbol{\phi}(\mu_1) + \boldsymbol{\phi}(\mu_2)].$$

At last, $\Delta_{2k+1} = T, k \geq 0$. This and two previous estimates provide that

$$\begin{aligned} E \exp[\frac{1-c}{2} Q_{2k+2}] &= E \prod_{j=0}^{2k+1} \exp[\frac{1-c}{2}\Delta_j] \leq 2[\boldsymbol{\phi}(\mu_1) + \boldsymbol{\phi}(\mu_2)] \cdot \exp[(k+1)\frac{1-c}{2}T] \cdot \Big[\frac{4\boldsymbol{\gamma}}{\boldsymbol{\alpha}} + 4 \sup_{\|x\| \leq R} \boldsymbol{\phi}(R))\Big]^k = \\ &= 2\exp[\frac{1-c}{2}T][\boldsymbol{\phi}(\mu_1) + \boldsymbol{\phi}(\mu_2)] \cdot \exp[kD], \quad k \geq 0, \quad \text{with} \quad D = \frac{1-c}{2}T + \ln\Big[\frac{4\boldsymbol{\gamma}}{\boldsymbol{\alpha}} + 4 \sup_{\|x\| \leq R} \boldsymbol{\phi}(R))\Big]. \end{aligned} \tag{3.6}$$

Take $p = \max\Big[1, -\frac{2D}{\ln(1-\delta(T,R))}\Big]$, then, by Hölder inequality,

$$E \exp[\frac{1-c}{2p} Q_{2k+2}] \leq \max\Big\{2\exp[\frac{1-c}{2}T][\boldsymbol{\phi}(\mu_1) + \boldsymbol{\phi}(\mu_2)], 1\Big\} \cdot (1 - \delta(T, R))^{-\frac{k}{2}},$$

and, by Chebyshev inequality,

$$P(Q_{2k+2} \geq t) \leq \exp[-\frac{1-c}{2p}t] \cdot \max\Big\{2\exp[\frac{1-c}{2}T][\boldsymbol{\phi}(\mu_1) + \boldsymbol{\phi}(\mu_2)], 1\Big\} \cdot (1 - \delta(T, R))^{-\frac{k}{2}}.$$

This estimate, together with (3.3),(3.4), gives that

$$P(Q^*>t)\leq \max\left\{2\exp[\frac{1-c}{2}T][\phi(\mu_1)+\phi(\mu_2)],1\right\}\cdot\left[1-(1-\delta(T,R))^{\frac14}\right]^{-1}\cdot\exp[-\frac{1-c}{4p}t],\quad t\in\mathbb{R}^+.$$

Now, we can use standard arguments (see [27]) and complete the proofs of Theorems 1.1, 1.2. In order to prove (1.3), consider the switching coupling with $\mu_1=\mu$ and $\mu_2=\mu_*$ given by Corollary 3.4. Then

$$\left\|P^t_\mu-\mu_*\right\|_{var}\leq P(Q^*>t)\leq \tilde{\mathsf{C}}_1[\phi(\mu)+1]\exp[-\mathsf{C}_2 t],\quad t\in\mathbb{R}^+$$

with $\tilde{\mathsf{C}}_1=\max\left[\frac{\gamma}{\alpha},1\right]\cdot 2\exp[\frac{1-c}{2}T]\cdot\left[1-(1-\delta(T,R))^{\frac12}\right]^{-1}$, $\mathsf{C}_2=\frac{1-c}{4p}$. Analogously, for a given $s,t\in\mathbb{R}^+$ consider the switching coupling with $\mu_1=\mu$ and $\mu_2=P^s_\mu$. Then, by statement (a) of Lemma 3.3, $\phi(\mu_2)\leq\phi(\mu)+\frac{\gamma}{\alpha}$, and

$$\mathsf{E}_\mu\left\|\mathsf{P}_\mu(\cdot|\mathcal{F}^s_0)-\mathsf{P}_\mu(\cdot)\right\|_{var,\mathcal{F}^\infty_{t+s}}\leq P(Q^*>t)\leq \mathsf{C}_1[\phi(\mu)+1]\exp[-\mathsf{C}_2 t],\quad t\in\mathbb{R}^+$$

with $\mathsf{C}_1=2\tilde{\mathsf{C}}_1$. These two estimates imply (1.3),(1.4) with $\mu_{inv}=\mu_*$. Theorems 1.1,1.2 are proved.

4. Sufficient conditions.

In this section, we give some sufficient conditions for **R**,**N**,**S** to hold true.

4.1. **Condition R.** There exists a wide range of conditions, that are sufficient for the recurrence condition **R**, see [20], Section 2.3. Here, we give only one condition of such a type, that is an analogue of Lemma 2.4 [20], but with a condition (11) of this Lemma replaced by an essentially weaker one (condition 3 below).

Proposition 4.1. *Suppose that the following conditions hold true.*
1. There exist $R,\alpha>0$ such that

$$(a(x),x)_{\mathbb{R}^m}\leq-\alpha\|x\|^2,\quad \|x\|\geq R.$$

2. There exists $q\in(0,+\infty)$ such that

$$\int_{\|u\|>1}\|u\|^q\Pi(du)<+\infty.$$

3. The function c can be decomposed into a sum $c=c_1+c_2$ with c_1,c_2 such that
3a. for some function ψ, $\frac{\psi(x)}{\|x\|}\to 0$ as $\|x\|\to\infty$,

$$\|c_1(x,u)\|\leq\psi(x)\|u\|,\quad u\in\mathbb{R}^d,x\in\mathbb{R}^m;$$

3b. $$\|x+c_2(x,u)\|\leq\|x\|,\quad x\in\mathbb{R}^m,\|u\|>1,\quad c_2(\cdot,u)\equiv 0,\quad \|u\|\leq 1.$$

*Then condition **R** holds true.*

Remark. In the case **B**, condition 3 holds true automatically with $c_1=c,c_2=0$.

Proof. Consider $\phi\in C^2(\mathbb{R}^m)$ such that $\phi(x)=\|x\|^q,\|x\|\geq R$. Without losing generality we can suppose that the constant R is chosen in such a way that $\delta_r\equiv\sup_{\|x\|\geq r}\frac{\psi(x)}{\|x\|}\leq\frac12, r\geq R$. Then, for $\|x\|\geq 2R$,

$$\mathcal{A}\phi(x)=q(a(x),x)_{\mathbb{R}^m}\|x\|^{q-2}+\int_{\mathbb{R}^m}\left[\|x+c(x,u)\|^q-\|x\|^q-q\mathbf{1}_{\|u\|\leq1}(c(x,u),x)_{\mathbb{R}^m}\|x\|^{q-2}\right]\Pi(du)=$$

$$=q(a(x),x)_{\mathbb{R}^m}\|x\|^{q-2}+\int_{\|u\|\leq1}\left[\|x+c_1(x,u)\|^q-\|x\|^q-q\mathbf{1}_{\|u\|\leq1}(c_1(x,u),x)_{\mathbb{R}^m}\|x\|^{q-2}\right]\Pi(du)+$$

$$+\int_{\|u\|>1}\left[\|x+c(x,u)\|^q-\|x\|^q\right]\Pi(du).$$

Due to condition 3b,

$$\int_{\|u\|>1}\left[\|x+c(x,u)\|^q-\|x\|^q\right]\Pi(du)\leq\int_{\|u\|>1}\left[\|x+c(x,u)\|^q-\|x+c_2(x,u)\|^q\right]\Pi(du)=$$

$$= \int_{\|u\|>1} \Big[\|x(u)+c_1(x,u)\|^q - \|x(u)\|^q\Big]\Pi(du),$$

where $x(u)=x+c_2(x,u)$. If O is an open subset of $\mathbb{R}^m$, $\{x+sc, s\in[0,1]\}\subset O$ and $\phi\in C^2(O)$, then

$$|\phi(x+c)-\phi(x)-(\nabla\phi(x),c)_{\mathbb{R}^m}| \le \|c\|^2 \sup_{s\in[0,1]} \|[\nabla^2\phi](x+sc)\| \tag{4.1}$$

(the Taylor's formula). Then, for $r\ge 2R$,

$$\int_{\|u\|\le 1} \Big[\|x+c_1(x,u)\|^q - \|x\|^q - q(c_1(x,u),x)_{\mathbb{R}^m}\|x\|^{q-2}\Big]\Pi(du) \le \mathrm{const}\cdot[\|x\|(1-\delta_r)]^{q-2}[\|x\|\delta_r]^2, \quad \|x\|\ge r,$$

here we applied (4.1) with $\phi(x)=\|x\|^q$, $O=\{\|x\|>r(1-\delta_r)\}$ and $c=c_1(x,u)$.

If $q\ge 1$, then, applying the inequality $|\phi(x+c)-\phi(x)|\le\|c\|\sup_{s\in[0,1]}\|[\nabla\phi](x+sc)\|$, with the same ϕ, $x=x(u)$ and $c=c_1(x,u)$, we obtain analogously

$$\left|\int_{\|u\|>1}\Big[\|x(u)+c_1(x,u)\|^q - \|x(u)\|^q\Big]\Pi(du)\right| \le \mathrm{const}\cdot[\|x\|(1-\delta_r)]^{q-1}[\|x\|\delta_r], \quad \|x\|\ge r,$$

here we used that $\|x(u)\|\le\|x\|$. If $q<1$, then we apply inequality $a^q+b^q\ge(a+b)^q, a,b\in\mathbb{R}^+$ and write

$$\left|\int_{\|u\|>1}\Big[\|x(u)+c_1(x,u)\|^q - \|x(u)\|^q\Big]\Pi(du)\right| \le \int_{\|u\|>1}\|c_1(x,u)\|^q\Pi(du)\le \mathrm{const}\cdot[\psi(x)]^q.$$

Thus,

$$\mathcal{A}\boldsymbol{\phi}(x)\le -q\alpha\|x\|^q + C\|x\|^q\Big\{(1-\delta_r)^{q-2}[\delta_r]^2+(1-\delta_r)^{q-1}\delta_r+\delta_r^{2q}\Big\} =$$
$$= \boldsymbol{\phi}(x)\left[-q\alpha + C\Big\{(1-\delta_r)^{q-2}[\delta_r]^2+(1-\delta_r)^{q-1}\delta_r+\delta_r^{2q}\Big\}\right], \quad \|x\|\ge r$$

with some constant C. Since $\delta_r\to 0, r\to+\infty$, this gives **R**. The proposition is proved.

4.2. **Condition N.** In [16], [18] the conditions were given, sufficient for the set $\{S_t=\mathbb{R}^m\}$ to have probability one. Below we give a more mild version of these conditions, sufficint for this set to have non-zero probability. Denote, for $x\in\mathbb{R}^m$, $\Theta_x=\{u|I_{\mathbb{R}^m}+\nabla_x c(x,u)$ is invertible$\}$, and put

$$\hat\Delta(x,u)=[I_{\mathbb{R}^m}+\nabla_x c(x,u)]^{-1}\Delta(x,u), \quad u\in\Theta_x.$$

Denote, by $S^m\equiv\{v\in\mathbb{R}^m|\|v\|_{\mathbb{R}^m}=1\}$, the unit sphere in $\mathbb{R}^m$.

Proposition 4.2. *Suppose that there exists $x_*\in\mathbb{R}^m$ such that*

$$\forall\varepsilon>0, v\in S^m \quad \Pi\Big(u\in\Theta_{x_*}\Big|(\hat\Delta(x_*,u),v)_{\mathbb{R}^m}\neq 0, \|c(x_*,u)\|<\varepsilon\Big)>0. \tag{4.2}$$

Then condition **N** *holds true with this x_* and arbitrary $t_*>0$.*

Proof. We need to prove that, on some Ω_0 with $\mathsf{P}_{x_*}(\Omega_0)>0$,

$$\{\mathcal{E}_\tau^t\Delta(X(\tau-),p(\tau)),\tau\in\mathcal{D}\cap(0,t)\}=\mathbb{R}^m. \tag{4.3}$$

Below, the set Ω_0 will be constructed explicitly, and in particular, on the set Ω_0, the matrix $\mathcal{E}_0^t$ will be non-degenerate (and therefore, for any variables θ,τ, $0\le\theta\le\tau\le t$, the matrix $\mathcal{E}_\theta^\tau$ also will be non-degenerate). Then, on this set,

$$\{\mathcal{E}_\tau^t\Delta(X(\tau-),p(\tau)),\tau\in\mathcal{D}\}=\mathcal{E}_0^t\{\mathcal{E}_0^{\tau-}\hat\Delta(X(\tau-),p(\tau)),\tau\in\mathcal{D}\cap(0,t)\},$$

and (4.3) is equivalent to

$$\{[\mathcal{E}_0^{\tau-}]^{-1}\hat\Delta(X(\tau-),p(\tau)),\tau\in\mathcal{D}\cap(0,t)\}=\mathbb{R}^m. \tag{4.4}$$

For $n\ge1$, consider the set $\mathcal{D}^n=\{\tau\in\mathcal{D},\|p(\tau)\|\ge\frac1n\}$. This set is a.s. locally finite, and therefore can be enumerated increasingly, $\mathcal{D}^n=\{\tau_1^n,\tau_2^n,\dots\}$. Denote, for $k\le m$, $S_k^n=\{[\mathcal{E}_0^{\tau_j^n-}]^{-1}\hat\Delta(X(\tau_j^n-),p(\tau_j^n)),j\le k\}$.

By the construction, S^n_k is a linear span of a finite family of vectors. Let us consider the k-th vector from this family,

$$[\mathcal{E}_0^{\tau_k^n-}]^{-1}\hat{\Delta}(X(\tau_k^n-),p(\tau_k^n)).$$

One can construct the measurable map $V:(\mathbb{R}^m)^{k-1}\to S^m$ such that

$$\forall j=1,\dots,k-1\quad V(x_1,\dots,x_{k-1})\perp x_j,\quad x_1,\dots,x_{k-1}\in\mathbb{R}^m.$$

We will write $v^n_{k-1}\equiv\Big([\mathcal{E}_0^{\tau_j^n-}]^{-1}\Big)^*\cdot V(\{[\mathcal{E}_0^{\tau_j^n-}]^{-1}\hat{\Delta}(X(\tau_j^n-),p(\tau_j^n)),j<k\})$, where $(M)^*$ denotes the adjoint matrix for M. The random vector v^n_{k-1} is well defined on the set $\{\mathcal{E}^{\tau_k^n-}$ is invertible$\}\in\mathcal{F}_{\tau_{k^n}-}$, and is $\mathcal{F}_{\tau_{k^n}-}$-measurable.

The value $p(\tau_k^n)$ is independent of $\mathcal{F}_{\tau_{k^n}-}$, and its distribution is equal to $\frac{1}{\lambda_n}\Pi(\cdot\cap\{\|u\|\geq 1\})$, where $\lambda_n=\Pi(\|u\|\geq 1)$. Therefore, on the set $\{\mathcal{E}^{\tau_k^n-}$ is invertible$\}$,

$$P([\mathcal{E}_0^{\tau_k^n-}]^{-1}\hat{\Delta}(X(\tau_k^n-),p(\tau_k^n))\text{ is well defined and }\notin S^n_{k-1}|\mathcal{F}_{\tau_{k^n}-})=$$

(4.5)
$$=\frac{1}{\lambda_n}\Pi\Big(u\in\Theta_x\Big|\,(\hat{\Delta}(x,u),v)_{\mathbb{R}^m}\neq 0\Big)\Big|_{x=X(\tau_k^n-),v=v^n_{k-1}}\geq\frac{1}{\lambda_n}\inf_{v\in S^m}\Pi\Big(u\in\Theta_x\Big|\,(\hat{\Delta}(x,u),v)_{\mathbb{R}^m}\neq 0\Big)\Big|_{x=X(\tau_k^n-)}.$$

For a given $\varepsilon>0$, consider, for $n\geq 1$, the maps

$$f_n:S^m\ni v\mapsto\Pi\Big(u\in\Theta_{x_*}\Big|\,(\hat{\Delta}(x_*,u),v)_{\mathbb{R}^m}\neq 0,\|c(x_*,u)\|<\varepsilon,\|u\|>\frac{1}{n}\Big).$$

Since functions $c,\nabla_x c,\hat{\Delta}$ are continuous w.r.t. x on their domain, every f_n is lower semicontinuous. For every $v\in S^m$, $f_n(v)$ monotonously tends to a positive limit as $n\uparrow\infty$. Therefore, due to Dini theorem, there exists $n=n(\varepsilon)\in\mathbb{N}$ such that

$$\inf_{v\in S^m}\Pi\Big(u\in\Theta_{x_*}\Big|\,(\hat{\Delta}(x_*,u),v)_{\mathbb{R}^m}\neq 0,\|c(x_*,u)\|<\varepsilon,\|u\|>\frac{1}{n}\Big)>0.$$

Analogously, the function

$$\mathbb{R}^m\ni x\to\inf_{v\in S^m}\Pi\Big(u\in\Theta_x\Big|\,(\hat{\Delta}(x,u),v)_{\mathbb{R}^m}\neq 0,\|c(x,u)\|<\varepsilon,\|u\|>\frac{1}{n(\varepsilon)}\Big)$$

is lower semicontinuous, and thus there exists $\delta=\delta(\varepsilon)>0$ such that

$$\inf_{v\in S^m,x\in B(x_*,\delta)}\Pi\Big(u\in\Theta_x\Big|\,(\hat{\Delta}(x,u),v)_{\mathbb{R}^m}\neq 0,\|c(x,u)\|<\varepsilon,\|u\|>\frac{1}{n(\varepsilon)}\Big)>0.\tag{4.6}$$

Define iteratively $\varepsilon_1,\dots,\varepsilon_m$ in the following way. Take $\varepsilon_m>0$ arbitrary, and put

$$\varepsilon_{k-1}=\min\Big[\frac{1}{6}\delta(\varepsilon_k),\frac{1}{2}\varepsilon_k\Big],\quad k=2,\dots,m.$$

By the construction, $\delta(\varepsilon_k)>3\sum_{l<k}\varepsilon_l, k=2,\dots,m$. Put $n=\max_{l\leq k}n(\varepsilon_l)$. Let us show that, for these $n\in\mathbb{N}$, $\varepsilon_1,\dots,\varepsilon_m>0$, and properly chosen $r\in(0,\frac{t_*}{m})$, the set

$$\Omega_0\equiv[\cap_{k=1}^m\Omega_k^r]\cap\Omega^r$$

has non-zero probability, where

$$\Omega^r=\Big\{\mathcal{E}_r^{t_*}\text{ is invertible }\Big\},$$

$$\Omega_k^r=\Big\{\mathcal{D}^n\cap((k-1)r,kr]=\{\tau_k^r\},\ \mathcal{E}_{(k-1)r}^{kr}\text{ is invertible},\ \|X(\tau_k^n-)-X((k-1)r)\|\leq\varepsilon_{k-1},\|X(kr)-X(\tau_k^n)\|\leq\varepsilon_k,$$
$$\|X(\tau_k^n)-X(\tau_k^n-)\|\leq\varepsilon_k\text{ and }[\mathcal{E}_0^{\tau_k^n-}]^{-1}\hat{\Delta}(X(\tau_k^n-),p(\tau_k^n)\notin\text{span}\,\{[\mathcal{E}_0^{\tau_j^n-}]^{-1}\hat{\Delta}(X(\tau_j^n-),p(\tau_j^n)),j<k\}\Big\}.$$

It is easy to verify that $P[\Omega^r|\mathcal{F}_{mr}]>0$ a.s. (the proof is omitted), so we need to verify that $P(\cap_{k=1}^m\Omega_k^r)>0$.

The process X can be described in the following way: at the moments $\tau_k^n, k\geq 1$, it has the jumps of the value $c(X(\tau_k^n-), p(\tau_k^n))$, and on every interval of the type $(\tau_{k-1}^n, \tau_n^k), k\geq 1$, it moves due to SDE

$$dX^n(t) = a^n(X^n(t))\,dt + \int_{\|u\|<\frac{1}{n}} c(X^n(t-),u)\tilde\nu(dt,du), \tag{4.7}$$

where $a^n(x)=a(x)+\int_{\|u\|\in[\frac{1}{n},1]}c(x,u)\Pi(du), \tau_0^n=0$. Denote, by $X^n(x,\cdot)$, the solution to (4.7) with $X^n(0)=x$ and, by $\mathcal{E}^{n,x,\cdot}_{\cdot}$, the correspondent stochastic exponent (both $X^n(x,\cdot)$ and $\mathcal{E}^{n,x}$ are independent of the point process $p|_{\mathcal{D}^n}$). Then, for a given $n\in\mathbb{N}, \varepsilon_1,\dots,\varepsilon_m>0$, one can choose $r\in(0,\frac{t_*}{m})$ small enough for

$$D\equiv \inf_{x\in B(x_*,3\sum_{l=1}^m\varepsilon_m)} P\Big(\forall s\leq r\,\|X^n(x,s)-x\|<\min_l\varepsilon_l, \mathcal{E}_0^{n,x,s} \text{ is invertible}\Big)>0.$$

Now, let us estimate $P(\Omega_k^r|\cap_{l<k}^m\Omega_l^r)$. The set $\cap_{l<k}^m\Omega_l^r$ belongs to $\mathcal{F}_{(k-1)r}$, and, on this set,

$$\|X((k-1)r)-x_*\|\leq 3\sum_{l<k-1}\varepsilon_l+2\varepsilon_{k-1}.$$

Now we can take subsequently conditional expectations first w.r.t. $\mathcal{F}_{\tau_k^n-}\vee\sigma(p(\tau_k^n))$, then w.r.t. $\mathcal{F}_{\tau_k^n-}$ (on this step, we use (4.5)), and, at last, w.r.t. $\mathcal{F}_{(k-1)r}$, and write that, on this set,

$$P\Big(\Omega_k^n\cap\{\|X((\tau_k^n-)-x_*\|\leq 3\sum_{l<k}\varepsilon_l, \|X((\tau_k^n)-X((\tau_k^n-)\|\leq\varepsilon_k, \|X((kr)-X((\tau_k^n)\|\leq\varepsilon_k\}|\mathcal{F}_{(k-1)r}\Big)\geq$$

$$\geq D\cdot(r\lambda_n e^{-r\lambda_n})\cdot\frac{\gamma}{\lambda_n}\cdot D = r\gamma D^2 e^{-r\lambda_n},$$

where

$$\gamma=\min_{l\leq k}\inf_{x\in B(x_*,3\sum_{l<k})}\Pi\Big(u\in\Theta_x\Big|\,(\hat\Delta(x,u),v)_{\mathbb{R}^m}\neq 0, \|c(x,u)\|<\varepsilon_k, \|u\|>\frac{1}{n}\Big)>0$$

by the construction. Therefore,

$$P(\cap_{k=1}^m\Omega_k^r)>0,$$

that gives the needed statement. The proposition is proved.

Let us also give sufficient condition for $\mathbf{N}$, in which conditions on the coefficients a, c and Lévy measure of the noise are separated.

For $w\in S^d$ (recall that S^d denotes the unit sphere in $\mathbb{R}^d$), $\varrho\in(0,1)$, denote by $V_+(w,\varrho)\equiv\{y\in\mathbb{R}^d|(y,w)_{\mathbb{R}^d}\geq\varrho\|y\|_{\mathbb{R}^d}\}$ the one-sided cone with the axis $\langle w\rangle\equiv\{tw, t\in\mathbb{R}\}$, and by $V(w,\varrho)\equiv\{y\in\mathbb{R}^d||(y,w)_{\mathbb{R}^d}|\geq\varrho\|y\|_{\mathbb{R}^d}\}$ the two-sided cone with the same axis.

Proposition 4.3. *Suppose that the following two conditions hold true.*
1. For every $w\in S^d$, there exists $\varrho\in(0,1)$, such that, for every $\delta>0$,

$$\Pi(V(w,\varrho)\cap\{u|\|u\|\leq\delta\})>0.$$

2. For some point x_*, there exists its neighborhood O_{x_*} such that
2a. $c(x,u)=\chi(x)u+\delta(x,u)$, $x\in O_{x_*}$, and

$$\|\delta(x_*,u)\|+\|\nabla_x\delta(x_*,u)\|=o(\|u\|),\quad \|u\|\to 0;$$

2b. the functions $\chi(\cdot)$ and $\tilde a(\cdot)$ belong to $C^1(O_{x_*},\mathbb{R}^{m\times d})$ and $C^1(O_{x_*},\mathbb{R}^m)$ correspondingly, and satisfy the following joint non-degeneracy condition:

$$\text{rank}\Big[\nabla\tilde a(x_*)\chi(x_*)-\nabla\chi(x_*)\tilde a(x_*)\Big]=m.$$

Then condition $\mathbf{N}$ holds true with this x_* and arbitrary $t_*>0$.

Remark. Condition 2b. is formulated in the case $\mathbf{A}$. In the case $\mathbf{B}$, it should be replaced by the condition $\det\nabla a(x_*)\neq 0$, and, in this case, condition 2a. trivially holds true with $\chi(x)\equiv I_{\mathbb{R}^m}$.

Proof. We use Proposition 4.2. Denote $\Big[\nabla\tilde a(x_*)\chi(x_*)-\nabla\chi(x_*)\tilde a(x_*)\Big]=A$. It follows from the condition 2 and explicit formula for $\hat\Delta$, that

$$\hat\Delta(x_*,u)=Au+o(\|u\|),\quad \|u\|\to 0. \tag{4.8}$$

Let $v\in S^m,\varepsilon>0$ be fixed. Consider the linear subspace $L_v=\{u\in\mathbb{R}^d|Au\perp v\}=A^*<v>^\perp$ (A^* is the adjoint matrix for A). This subspace is proper, due to condition $\operatorname{rank}A=m$. Take $w\in S^d$ such that $w\perp L_v$, then, for any $\varrho\in(0,1)$, $V(w,\varrho)\cap L_v=\emptyset$, and, therefore, there exists $c=c(v,\varrho)>0$ such that

$$|(Au,v)_{\mathbb{R}^m}|\geq c\|u\|,\quad u\in V(w,\varrho).$$

This, together with (4.8), provides that

$$|(\hat\Delta(x_*,u),v)_{\mathbb{R}^m}|\geq c\|u\|+o(\|u\|),\quad u\in V(w,\varrho),\|u\|\to 0. \tag{4.9}$$

Take ϱ from the condition 1 of the Proposition, and $\delta_*=\varepsilon\cdot[\psi_*(x_*)]^{-1}$ (ψ_* is given in the condition (1.1)). Then, for every $\delta\in(0,\delta_*)$, the measure Π of the set $V(w,\varrho)\cap\{\|u\|\leq\delta\}$ is positive, and, on this set, $\|c(x_*,u)\|<\varepsilon$. On the other hand (4.9) implies that, for δ small enough,

$$(\hat\Delta(x_*,u),v)_{\mathbb{R}^m}\neq 0 \text{ on the set } V(w,\varrho)\cap\{\|u\|\leq\delta\}.$$

The proposition is proved.

4.3. **Condition S.** One possible way to provide that condition **S** holds true is to use general support theorems for the distribution of the of solution to SDE with a jump noise. For instance, Theorem I [25] provides, in the case **A**, the following result.

Proposition 4.4. *Consider U, the set of sequences $\{(t_n,u_n),n\geq 1\}$, where $\{t_n\}\subset\mathbb{R}^+$ is a strictly increasing sequence with $\lim t_n=+\infty$, and $\{u_n\}\subset\operatorname{supp}\Pi$ is arbitrary. Suppose that, for any given $R,T\in\mathbb{R}^+$, for every x with $\|x\|\leq R$ and $\varepsilon>0$ there exists a sequence $\{(t_n,u_n)\}\in U$ such that the solution to the equation*

$$Z(t)=x+\int_0^t\tilde a(Z(s))\,ds+\sum_{t_n\leq t}c(Z(t_n-),u_n),\quad t\in\mathbb{R}^+,$$

satisfies the condition $\|Z(T)-x_*\|<\varepsilon$.

*Then condition **S** holds true.*

Another possibility is to give some straightforward conditions, that seem to be more suitable in a certain concrete cases. Let us formulate, without a detailed proof, one condition of such a type. Note that, unlike the previous Proposition, the next one does not require moment restriction on the Lévy measure of the noise, and is formulated for the both cases **A** and **B** simultaneously. Denote, for any $x\in\mathbb{R}^m$, $\Pi_x(\cdot)=\Pi(u\in\mathbb{R}^d|c(x,u)\in\cdot)$.

Proposition 4.5. *Suppose that, for every $x\in\mathbb{R}^m,v\in S^m$, there exists $\varrho\in(0,1)$ such that, for any $\delta>0$,*

$$\Pi_x(V_+(v,\varrho)\cap\{\|y\|\leq\delta\})>0. \tag{4.10}$$

*Then $y\in\operatorname{supp}P_x^t$ for every $x,y\in\mathbb{R}^m$, $t>0$, and therefore condition **S** holds true.*

Sketch of the proof. Take $v=\frac{y-x}{\|y-x\|}$, $\varrho\in(0,1)$ from the condition (4.10) for the given x and v, and $\delta_*=\frac12\|y-x\|$. Then there exist $\delta_1,\delta_2,\gamma>0$ such that $0<\delta_1<\delta_2<\delta_*$,

$$\Pi_x(V_+(v,\varrho)\cap\{\|y\|\in[\delta_1,\delta_2]\})>0,$$

and

$$\|(x+c)-y\|\leq\|x-y\|-\gamma,\quad c\in V_+(v,\varrho)\cap\{\|y\|\in[\delta_1,\delta_2]\}.$$

Then arguments, analogous to those given in the proof of Proposition 4.2 allows one to conclude that, for any two points $x\neq y$, there exist $\gamma>0$ and $t>0$ such that, for any $s\in(0,t)$,

$$\mathsf{P}_x(\|X(s)-y\|<\|x-y\|-\frac{\gamma}{2})>0. \tag{4.11}$$

Let $\varepsilon \in (0, \|x-y\|)$ be given, then, since the process X is Feller, one can conclude from (4.11) that there exist $t_\varepsilon > 0$, $\gamma_\varepsilon > 0$ such that, for any $t \le t_\varepsilon$,

$$p_t \equiv \inf_{z:\|z-y\|\in[\varepsilon,\|x-y\|]} \mathsf{P}_z(\|X(s)-y\| < \|z-y\| - \gamma_\varepsilon) > 0.$$

This implies that, for any $x \ne y$ and $\varepsilon > 0$, for any $t \le t_\varepsilon$,

$$\mathsf{P}_x(X(t) \in B(y,\varepsilon)) \ge [p_{\frac{t}{N}}]^N > 0, \quad \text{where } N = \Big[\frac{\|x-y\|}{\gamma_\varepsilon}\Big] + 1.$$

Via the Markov property of the process X, this implies the statement of the Proposition.

4.4. **One-dimensional case. Proof of Proposition 0.1.** In the case $m = 1$, the sufficient conditions given in the previous subsections can be made more precise. For instance, the following version of Proposition 4.2 holds true.

Proposition 4.6. *Let $m = 1$ and suppose that there exists $x_* \in \mathbb{R}$ such that*

$$\Pi\Big(u \in \Theta_{x_*}\Big|\hat{\Delta}(x_*, u) \ne 0\Big) > 0. \tag{4.12}$$

*Then condition **N** holds true for any $t_* > 0$.*

We omit the proof, since it is totally analogous to the one of Proposition 4.2, except one point, that causes the difference between conditions (4.2) and (4.12). For $m = 1$, we have to apply estimate (4.5) only once, for the jump moment τ_1^n. This means that we do not have to control the position of the process after the jump at this moment, and thus, when $m = 1$, the limitation involving ε can be removed from (4.2).

Now, let us prove Proposition 0.1, formulated in the Introduction. Condition **R** is provided by Proposition 4.1. Let us proceed with the conditions **N** and **S**. Since $\Pi(\mathbb{R}\backslash\{0\}) > 0$, either $\Pi((-\infty, 0))$ or $\Pi((0,\infty))$ is non-zero. Let, for instance, $\Pi((0,\infty)) > 0$. Take R large enough for $\sup_{x>R} \frac{a(x)}{x} < 0$, then $y \in \operatorname{supp} P_x^t$ for any $y > R$ and any $x \in \mathbb{R}, t > 0$. This follows from Theorem I [25] in the case $\int_{\mathbb{R}} |u|\Pi(du) < +\infty$, and from Theorem 3 [26] in the case $\int_{\mathbb{R}} |u|\Pi(du) = +\infty$. This provides that **S** holds true with arbitrary $t > 0$ and $x_* > R$. In order to provide **N** for some $x_* > R$, let us use Proposition 4.6. In the case of additive noise, $\hat{\Delta}(x, u) = a(x+u) - a(x)$. Therefore, if $\Pi(\mathbb{R}\backslash\{0\}) > 0$ and (4.12) fails for every $x_* > R$, then there exists a sequence $\{x_n\}$ with $|x_n| \to +\infty$ such that $a(x_{n+1}) = a(x_n)$. This, however, contradicts the condition $\lim \sup_{|x|\to+\infty} \frac{a(x)}{x} < 0$. The proposition is proved.

5. Counterexamples

We have seen that three basic conditions **R**,**N**,**S** imply exponential estimates (1.3), (1.4). In this section we give counterexamples that show that, as soon as any of these conditions is removed, the solution to (0.1) may fail to be ergodic (i.e., to possess a unique invariant distribution $\mu_{inv} \in \mathcal{P}$).

The cases, when conditions **R** or **S** are missed, are quite standard and simple, thus we just outline the corresponding examples.

Example 5.1. Let $m = d = 1$, $c(x,u) = u, \Pi = 2\delta_1 + \delta_{-1}$ and $a(x) \in C^1(\mathbb{R})$ is such that $a(x) = -c, |x| \ge 1$, with $c \in (0,1)$. Both conditions **N** and **S** hold true here (this can be provided by the arguments from subsection 4.4), but **R** fails. The law of large numbers provides that, for every $x > 1$,

$$\mathsf{P}_x(\lim_{t\to+\infty} X(t) = +\infty, \inf_{t\in\mathbb{R}^+} X(t) > 1) > 0.$$

This implies that X does not have any invariant probability measure.

Example 5.2. Let $m = d = 1$, $a(x) \in C^1(\mathbb{R})$ be such that $a(x) = -x, |x| \ge 2$ and $a(x) = 0, |x| \le 1$. Let also $\Pi = \delta_1$ and $c(\cdot, 1) \in C^1(\mathbb{R})$ be bounded and such that $c(x,1) = \operatorname{sign} x, |x| \ge 2$ and $x \cdot c(x,1) \ge 0, x \in \mathbb{R}$. Then condition **R** holds true, and **N** holds true for any x_* with $|x_*| > 2$ (Proposition 4.6). Condition **S** fails: starting from any set $A_+ = [1, +\infty)$ or $A_-(-\infty, -1]$, the process X remains in this set with the probability 1. Therefore, there exist at least two different invariant measures for X, supported by these sets.

The last example is more non-trivial, and is concerned with the case where **R**,**S** hold true while **N** does not.

Example 5.3. Let us start with an auxiliary construction. Consider the unit circle $C \equiv \frac{1}{2\pi}S^2$ on the plane $\mathbb{R}^2$, and define the discrete time Markov process Z on C by its transition probability

$$Q(z,\cdot) = (1-3p)\delta_{3z}(\cdot) + p\left[\delta_{\frac{z}{3}}(\cdot) + \delta_{\frac{z+1}{3}}(\cdot) + \delta_{\frac{z+2}{3}}(\cdot)\right], \quad z \in C,$$

where $p \in (0, \frac{1}{6})$ is given, and every arithmetic operation on C is defined as the same operation on $[0,1) \cong C$ modulo 1. If $Z_0 = z$ is any point from $[0,1) \cong C$, then there exists a non-zero probabilities for Z_n to be equal to each point of the type

$$3^{-n}y + \sum_{k=1}^{n} a_k 3^{-k}, \quad a_k \in \{0,1,2\}, k = 1, \dots, n,$$

and therefore, the set $\bigcup_{n\in\mathbb{N}} \operatorname{supp} P(Z_n \in \cdot | Z_0 = z)$ is dense in C, i.e. the process Z is topologically irreducible. Now let us show that Z possesses at least two different invariant measures (in fact, the set of invariant measures here is much larger).

Consider, together with Z, the sequence T_n defined by

$$T_0 = 0, \quad T_{n+1} = \begin{cases} (T_n - 1) \vee 0, & Z_{n+1} = 3Z_n, \\ T_n + 1, & \text{otherwise}, \end{cases} \quad n \geq 0.$$

Then $\{T_n\}$ is a birth-and-death Markov chain with probabilities of birth equal to $b_k \equiv b = 3p$ and probabilities of death equal to $d_k \equiv d = 1 - 3p$. We have that $d > b$ since $p < \frac{1}{6}$, and therefore this chain is ergodic, that means that for any given $\varepsilon > 0$ there exists $L_\varepsilon \in \mathbb{N}$ such that

$$\sup_{n\geq 0} P(T_n \geq L_\varepsilon) < \varepsilon. \tag{5.1}$$

Take $Z_0 = 0$ and consider some weak limit point μ_*^0 for the sequence of Khasminskii's averages

$$\frac{1}{N}\sum_{n\leq N} P(Z_n \in \cdot | Z_0 = 0)$$

(see the proof of Corollary 3.4). By the construction, any digit in the 3-adic representation for Z_n, with the number of the digit greater then T_n, is equal to 0. This means that $\mu_*^0(A_\varepsilon^0) \geq 1 - \varepsilon$, where

$$A_\varepsilon^0 = \{z \in [0,1) |\ \text{all 3-adic digits for } z, \text{ with the number of the digit} \geq L_\varepsilon, \text{ are equal to } 0\}$$

(the inequality holds true since the set A_ε^0 is closed). Therefore, $\mu_*^0(A^0) = 1$, where

$$A^0 = \{z \in [0,1) |\ \text{all 3-adic digits for } z, \text{ except some finite number of the digits, are equal to } 0\}.$$

Analogously, if $Z_0 = \frac{1}{2}$, and $\mu_*^{\frac{1}{2}}$ is any weak limit point for the sequence of Khasminskii's averages

$$\frac{1}{N}\sum_{n\leq N} P(Z_n \in \cdot | Z_0 = \frac{1}{2}),$$

then $\mu_*^{\frac{1}{2}}(A^1) = 1$, where

$$A^1 = \{x \in [0,1) |\ \text{all 3-adic digits for } x, \text{ except some finite number of the digits, are equal to } 1\},$$

and $A^0 \cap A^1 = \emptyset$. This means that μ_*^0 and $\mu_*^{\frac{1}{2}}$ are mutually singular invariant measures for Z.

Now, let us proceed with the construction of the process. Put $m = 2, d = 2$ and $c(x,u) = c_1(x,u_1) + c_2(x,u_2), x \in \mathbb{R}^2, u = (u_1,u_2) \in \mathbb{R}^2$. Let $\Pi = \Pi_1 \times \Pi_2$ with $\Pi_1 = \delta_1, \Pi_2 = (1-3p)\delta_1 + p(\delta_2 + \delta_3 + \delta_4)$. Let the part c_1 to give the radial part of the jump noise:

$$c_1(x,1) = \frac{b(x)}{\|x\|} \cdot x, \quad x \in \mathbb{R}^2,$$

where $b\in C^1(\mathbb{R}^2,\mathbb{R})$ is such that $b(x)=0$ for $\|x\|\le 1$, $b(x)>0$ for $\|x\|\in(1,2)$ and $b(x)=1$ for $\|x\|\ge 2$. For $\|x\|\ge 1$, let the part c_2 to define the "rotational" part of the noise: if x is written in the polar coordinates as (r,θ), then $x+c_2(x,i), i=1,\dots,4$ has the following polar representation:

$$\begin{cases}(r,3\theta), & i=1\\ (r,\frac{\theta+2\pi(i-2)}{3}), & i=2,3,4\end{cases}.$$

For $\|x\|\ge 1$, let the functions $c_2(\cdot,i), i=1,\dots,4$ be defined in an arbitrary way, such that $c_2(\cdot,i)\in C^1(\mathbb{R}^2,\mathbb{R}^2)$, $\|x+c(x,i)\|\ge 1, x\in\mathbb{R}^2, i=1,\dots,4$. The drift coefficient let be equal to $a(x)=-b(x)\cdot x, x\in\mathbb{R}^m$.

By the construction, condition **R** holds true (Proposition 4.1) and condition **S** holds true for every x_* with $\|x^*\|\ge 1$ and every $t>0$ (Theorem I [25]). Let us show that, however, there exist two different invariant measures for X. If $X(0)=x$ is such that $\|x\|\ge 1$, then the processes $R(\cdot)=\|X(\cdot)\|$ and $\Theta(\cdot)=\frac{X(\cdot)}{\|X(\cdot)\|}$ are independent (w.r.t. P_x) Markov processes. The first process possesses at least one invariant measure $\varkappa$, supported by $[1,+\infty)$ (see Corollary 3.4). The second one is the pure jump Markov process with the total intensity of the jump equal, at every point, to $(1-3p)+p+p+p=1$. Its embedded Markov chain coincides, up to the scaling parameter 2π, with the chain Z considered before. Therefore, this process possesses at least two different invariant measures χ_1,χ_2 on S^2. Thus, the process $X(\cdot)$ possesses at least two different invariant measures $\mu_1=\varkappa\times\chi_1, \mu_2=\varkappa\times\chi_2$, supported by $[1,+\infty)\times S^2=\{x|\|x\|\ge 1\}$.

This example shows that the topological irreducibility condition **S**, together with the recurrence condition **R**, is not strong enough to produce ergodicity of the solution to SDE with a jump noise. In order to produce ergodicity, some kind of "smoothing" condition, like non-degeneracy condition **N** in our settings, is needed additionally.

References

[1] *D.E.Alexandrova, V.I.Bogachev, A.Yu.Pilipenko.* On convergence in variation of the induced measures, Mat. Sbornik, 190, 1999, N 9, p.3-20. (in Russian).

[2] *K.Bichteler, J.-B.Gravereaux, J.Jacod.* Malliavin calculus for processes with jumps, New York, Gordon and Breach, 1987.

[3] *J.M.Bismut.* Calcul des variations stochastiques et processus de sauts, Zeit. fur Wahr. 63, 1983, p. 147-235.

[4] *V.I.Bogachev.* Differentiable measures and the Malliavin calculus, J. of Math. Sci, 87, 1997, N 5, p.3577-3731.

[5] *V.I.Bogachev, O.G.Smolyanov* Analytical properties of infinitedimensional distributions, Uspekhi mat. nauk, 45, 1990, N 3, p.3-83 (in Russian).

[6] *D.B.H.Cline, H.H.Pu.* Verifying irreducibility and continuity of a nonlinear time series, Statist. Probab. Lett. 40, 1998, p.139148.

[7] *Yu.A.Davydov* On a convergence in variation of images of one-dimensional measures, Problems of theory of probability distributions. XII. Notes of sci. seminars POMI, vol. 194, Sankt-Peterburg, POMI, 1992, p.42-58.

[8] *Yu.A.Davydov, M.A.Lifshits.* Stratification method in some probability problems, Prob. theory, Math. stat., Theor. cyber., Totals of science and technics, vol. 22, Moscow, VINITI, 1984, p.61-157.

[9] *Yu.A.Davydov, M.A.Lifshits, N.V.Smorodina.* Local Properties of Distributions of Stochastic Functionals, Moscow, Nauka, 1995 (in Russian).

[10] *R.L.Dobrushin.* Prescribing a system of random variables by conditional distributions, Theory of Prob. and Appl. 15, 1970, p.453-486.

[11] *G.Federer.* Geometric Measure Theory, Nauka, Moscow, 1987. (in Russian, translated from G. Federer, Geometric Measure Theory. – New York, Springer, 1969).

[12] *I.I.Gikhman, A.V.Skorokhod.* Stochastic Differential Equations and Their Applications, Kiev, Naukova Dumka, 1982 (in Russian).

[13] *Y.Ishikawa, H.Kunita.* Existence of density for canonical differential equations with jumps, Preprint Nanzan, 2004.

[14] *S.A.Klokov, A.Yu.Veretennikov.* Sub-exponential mixing rate for a class of Markov chains, Math. Comm. 9, 2004, p.9-26

[15] *T.Komatsu, A.Takeuchi.* On the smoothness of PDF of solutions to SDE of jump type, Int. J. Differ. Equ. Appl., 2, 2001, N2, p.141-197.

[16] *A.M.Kulik.* Malliavin Calculus for Lévy Processes With Arbitrary Lévy Measures, Prob. Theor. Math. Stat., 72, 2005, p.67-83.

[17] *A.M.Kulik.* On a convergence in variation for distributions of solutions of SDE's with jumps, Random Operators and Stoch. Equations, 13, 2005, N 3, p.297-312.

[18] *A.M.Kulik.* Stochastic calculus of variations for general Lévy processes and its applications to jump-type SDE's with non-degenerated drift, Preprint, 2006, arxiv.org:math.PR/0606427v2.

[19] *R.Léandre.* Regularites de processus de sauts degeneres (II). – Ann. Inst. Henri Poincare Prob. Stat., 24, 1988, p.209-236.

[20] *H.Masuda.* Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps, Stoch. Proc. and Appl., 117, 2007, p.3556.

[21] *S.P.Meyn, R.L.Tweedie.* Markov Chains and Stochastic Stability, Springer-Verlag London, Ltd., London, 1993.

[22] *S.P. Meyn, R.L. Tweedie.* Stability of Markovian processes. III. FosterLyapunov criteria for continuous-time processes, Adv. Appl. Probab. 25, 1993, p.518548.

[23] *K.R.Partasarathy.* Introduction to Probability and Measure, Springer-Verlag, New York, 1978.

[24] *J.Picard.* On the existence of smooth densities for jump processes, Probab. Theory Rel. Fields, 105, 1996, p.481-511.

[25] *T.Simon.* Support theorem for jump processes, Stoch. Proc. and Appl., 89, 2000, p.1-30.

[26] *T.Simon.* Support dune équation dItô avec sauts en dimension 1, Séminaire de Probabilités 36, 2002, p.314-330.

[27] *A.Yu.Veretennikov* On estimates of mixing rate for stochastic equations, Theory of Prob. and Appl. 32, 1987, p.299-308 (in Russian).

[28] *A.Yu.Veretennikov* On polynomial mixing and rate of convergence for stochastic differential and difference equations, Theory of Prob. and Appl. 44, 1999, p.312-327 (in Russian).

[29] *A.Yu.Veretennikov* Coupling method for Markov chains under integral Doeblin type condition, Theory of Stochastic Processes, 8(24), 2002, N 3-4, p.383-391.

Institute of Mathematics, Ukrainian National Academy of Sciences, 3, Tereshchenkivska Str., Kyiv 01601, Ukraine
E-mail address: kulik@imath.kiev.ua