DOI: 10.37863/umzh.v72i9.6292

UDC 519.21

**A. Kulik** (Wroclaw Univ. Sci. and Technology, Poland),
**A. Pilipenko** (Inst. Math. Nat. Acad. Sci. Ukraine, Kyiv,
and Nat. Techn. Univ. Ukraine "Igor Sikorsky Kyiv Polytechnic Institute")

# ON REGULARIZATION BY A SMALL NOISE OF MULTIDIMENSIONAL ODES WITH NON-LIPSCHITZ COEFFICIENTS*

# ПРО РЕГУЛЯРИЗАЦІЮ МАЛИМ ШУМОМ БАГАТОВИМІРНИХ ЗВИЧАЙНИХ ДИФЕРЕНЦІАЛЬНИХ РІВНЯНЬ З НЕЛІПШИЦЕВИМИ КОЕФІЦІЄНТАМИ

In this paper we solve a selection problem for multidimensional SDE $dX^\varepsilon(t) = a(X^\varepsilon(t))\,dt + \varepsilon\sigma(X^\varepsilon(t))\,dW(t)$, where the drift and diffusion are locally Lipschitz continuous outside of a fixed hyperplane $H$. It is assumed that $X^\varepsilon(0) = x^0 \in H$, the drift $a(x)$ has a Hoelder asymptotics as $x$ approaches $H$, and the limit ODE $dX(t) = a(X(t))\,dt$ does not have a unique solution. We show that if the drift pushes the solution away from $H$, then the limit process with certain probabilities selects some extremal solutions to the limit ODE. If the drift attracts the solution to $H$, then the limit process satisfies an ODE with some averaged coefficients. To prove the last result we formulate an averaging principle, which is quite general and new.

Статтю присвячено знаходженню границі розв'язків багатовимірних стохастичних диференціальних рівнянь $dX^\varepsilon(t) = a(X^\varepsilon(t))\,dt + \varepsilon\sigma(X^\varepsilon(t))\,dW(t)$ при $\varepsilon \to 0$, де коефіцієнти переносу та дифузії є локально ліпшицевими функціями ззовні фіксованої гіперплощини $H$. Припускається, що $X^\varepsilon(0) = x^0 \in H$, коефіцієнт переносу $a(x)$ має гельдерову асимптотику, коли $x$ наближається до $H$, і граничне звичайне диференціальне рівняння $dX(t) = a(X(t))\,dt$ може не мати єдиного розв'язку. Доведено, що якщо перенос відштовхує розв'язок від гіперплощини $H$, то граничний процес із певними ймовірностями вибирає деякі екстремальні розв'язки граничного диференціального рівняння. Якщо перенос притягує розв'язок до $H$, то граничний процес задовольняє звичайне диференціальне рівняння з усередненими коефіцієнтами. Для доведення останнього результату сформульовано новий достатньо загальний принцип усереднення.

**1. Introduction.** Consider an ODE

$$\frac{du(t)}{dt} = a(u(t)),$$
$$u(0) = 0, \tag{1.1}$$

where $a$ is a continuous function of linear growth that satisfies a local Lipschitz condition everywhere except of the point $u = 0$. Then uniqueness of the solution to (1.1) may fail; e.g., for $a(u) = \sqrt{|u|}\,\mathrm{sgn}(u)$, the ODE (1.1) has multiple solutions $\pm t^2/4,\ t \ge 0$.

Consider a perturbation of (1.1) by a small noise:

$$du_\varepsilon(t) = a(u_\varepsilon(t))dt + \varepsilon dW(t),$$
$$u_\varepsilon(0) = 0, \tag{1.2}$$

* The work of A. Kulik was supported by the Polish National Science Center (grant 2019/33/B/ST1/02923). Research of A. Pilipenko was partially supported by Norway-Ukrainian cooperation in mathematical education Eurasia 2016-Long-term CPEA-LT-2016/10139, DFG project "Stochastic dynamics with interfaces" (no. 452119141) and by the Alexander von Humboldt Foundation within the Research Group Linkage Programme "Singular diffusions: analytic and stochastic approaches" between the University of Potsdam and the Institute of Mathematics of the National Academy of Sciences of Ukraine.

where $W$ is a Wiener process. Equation (1.2) has a unique strong solution due to the Zvonkin – Veretennikov theorem [22]. It easy to see that a family of distributions of $\{u_\varepsilon\}$ is weakly relatively compact because $a$ has a linear growth. Moreover, any limit point of $\{u_\varepsilon\}$ as $\varepsilon \to 0$ satisfies equation (1.1) because $a$ is continuous. Hence, if the limit $\lim_{\varepsilon\to 0} u_\varepsilon$ (in distribution) exists, then this limit may be considered as a natural selection of a solution to (1.1).

The corresponding problem was originated in papers by Bafico and Baldi [2, 3], who considered the one-dimensional case; other generalizations see, for example, in [4 – 9, 12, 15, 18 – 21] and references therein. Investigations in multidimensional case are much complicated than in the one-dimensional one. There are still no simple sufficient conditions that ensure existence of a limit $\lim_{\varepsilon\to 0} u_\varepsilon$ and a characterization of this limit. One of the reason for this is the absence of the linear ordering in the multidimensional case. Indeed, in the one-dimensional situation the are only two ways to exit from the point 0: one way to the right and another to the left. The probability of going left or right can be easily obtained since there are explicit formulas for hitting probabilities for one-dimensional diffusions. The equation for the limit process outside of 0 must satisfy the original ODE because $a$ is Lipschitz continuous there.

In this paper we consider the multidimensional case, where the Lipschitz condition for $a$ may fail at a hyperplane. Let us describe the corresponding model. Consider an SDE

$$\begin{gathered} du_\varepsilon(t) = a(u_\varepsilon(t))dt + \varepsilon\sigma(u_\varepsilon(t))dW(t), \\ u_\varepsilon(0) = x^0, \end{gathered} \tag{1.3}$$

where $a\colon \mathbb{R}^d \to \mathbb{R}^d$, $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are measurable functions, $W$ is an $m$-dimensional Wiener process.

Assume that $a$ and $\sigma$ are of linear growth, $\sigma$ is continuous and satisfies the uniform ellipticity condition. This ensures existence and uniqueness of a weak solution to (1.3) and relative compactness for the distributions of $\{u_\varepsilon\}$.

Set $H := \mathbb{R}^{d-1} \times \{0\}$. Suppose that the initial starting point $x^0 \in H$ and that the drift $a$ satisfies the local Lipschitz property in $\mathbb{R}^d \setminus H$.

Note that the definition of $a$ on $H$ is inessential because $u_\varepsilon$ spends zero time in $H$ with probability 1 due to the nondegeneracy of the diffusion coefficient.

The case when $a$ is globally Lipschitz continuous in the lower half-space $\mathbb{R}^d_- := \mathbb{R}^{d-1} \times (-\infty, 0)$ and the upper half-space $\mathbb{R}^d_+ := \mathbb{R}^{d-1} \times (0,\infty)$ was investigated in [19]. The result was formulated in terms of the vertical components of $a^\pm(x^0) := \lim_{x\to x^0, x\in\mathbb{R}^d_\pm} a(x)$. In this paper we investigate the case when the drift has Hölder-type asymptotic in a neighborhood of $H$. Namely, we will assume that

**A1)** *$a_d(x) = |x_d|^\gamma b(x)$, where $\gamma < 1$, $x_d$ is the $d$th coordinate of $x = (x_1, \ldots, x_d)$, and $b$ is a globally Lipschitz continuous function in $\mathbb{R}^d_+$ and $\mathbb{R}^d_-$, $b^\pm(x) \neq 0$, $x \in H$;*

**A2)** *$a_k$, $k = 1, \ldots, d-1$, are globally Lipschitz functions in $\mathbb{R}^d_+$ and $\mathbb{R}^d_-$.*

This case has new features, and the proofs will be based on new ideas compared to the proofs from [19]. To illustrate the difference, let us recall briefly results of [19], where the case $\gamma = 0$ was considered, and sketch the expected results in the case $\gamma \in (0, 1)$.

*Case* 1 (*The vector field $a$ pushes outwards the hyperplane*). Denote by $\mathbf{n} = (0, \ldots, 0, 1)$ the normal vector to the hyperplane $H$. Assume that $\gamma = 0$ and $\pm(a^{\pm}(x), \mathbf{n}) > 0,\ x \in H$. Then there are two solutions $u^{\pm}$ to

$$du(t) = a(u(t))dt \tag{1.4}$$

that start at $x^0 \in H$ and exit from $H$ immediately to the upper and the lower half-spaces, respectively. It was proved in [19] that if $\gamma = 0$, then the limit process $u_0$ immediately leaves $H$ and moves as $u^{\pm}$ with probabilities proportional to $\left|(a^{\pm}(x^0), \mathbf{n})\right|$. The corresponding proof was similar to the one-dimensional situation. It used some comparison principle adapted to the multidimensional situation. Investigations for arbitrary $\gamma \in (0, 1)$ will be similar, but selection probabilities will be different.

***Remark* 1.1.** It was assumed in [19] that the noise is additive, i.e., $\sigma$ is the identity matrix and $m = d$. The case of multiplicative noise is completely analogous.

***Remark* 1.2.** If $\gamma = 0$ and the vector field $a$ pushes away $H$ from one side of $H$ and attracts from another side (for example, $(a^{\pm}(x), \mathbf{n}) > 0$), then there is a unique solution to (1.4) that starts at $x^0 \in H$. This solution exits from $H$ immediately (to the upper half-space in our case) and the limit process $u_0$ equals this solution of the ODE, see [19]. If $\gamma \in (0, 1)$, the result is similar. Assume, for example, that $b^{\pm}(x^0) > 0$. Then there exists a unique solution to (1.4) that exit $H$ immediately (there may be other solutions that stay in $H$). Moreover, this solution exits to the upper half-space and the limit process $u_0$ equals this solution. We do not prove this result in this paper. The proof is similar to [19].

*Case* 2 (*The vector field $a$ pushes towards the hyperplane*). Assume that $\gamma = 0$ and $\pm(a^{\pm}(x), \mathbf{n}) < 0,\ x \in H$. It can be seen that any limit point of $\{u_\varepsilon\}$ must stay at $H$ with probability 1. It was proved in [19] that the limit process $u_0$ satisfies an ODE on $H$ with the drift $P_H(p_+(x)a^+(x) + p_-(x)a^-(x))$, where $P_H$ is the orthogonal projection to $H$ and the coefficients $p_{\pm}(x)$ are equal to $\dfrac{a_d^{\mp}(x)}{a_d^-(x) - a_d^+(x)}$. Note that this multidimensional result has no one-dimensional analogues, where the limit is zero process. In multidimensional case the first $d-1$ coordinates may change while $d$th coordinate stays zero.

The idea of proof was to analyze the time spent by $u_\varepsilon$ in upper and lower half-spaces. It was seen that since any limit process stays at $H$ and $u_\varepsilon$ is close close to $H$ for small $\varepsilon$, then the times spent in upper and lower half-spaces in a neighborhood of $x \in H$ are proportional to the $d$th coordinates $a_d^-(x)$ and $a_d^+(x)$, respectively (they are not zero if $\gamma = 0$). Note that the proof in [19] was independent of the type of a noise. The small noise might be arbitrary process that (a) ensures existence a solution and (b) converges to 0 uniformly in probability as $\varepsilon \to 0$ (however, the corresponding results were formulated for Brownian noise only).

The proof from [19] does not work if $a_d(x) \to 0$ as $x$ approaches to $H$. The time spent in upper and lower half-spaces might depend on the asymptotic of decay of $a_d$ in a neighborhood of $H$. In this paper we prove the result when $a$ satisfies assumptions **A1**, **A2** with $\gamma \in (0, 1)$, $b^+(x) < 0$ and $b^-(x) > 0$ for $x \in H$.

It appears that if we scale the vertical coordinate $\varepsilon^{-\delta} u_{d,\varepsilon}(t)$ for a special choice of $\delta > 0$, then a pair $\big(u_{1,\varepsilon}(t), \ldots, u_{d-1,\varepsilon}(t)\big)$ and $\varepsilon^{-\delta} u_{d,\varepsilon}(t)$ can be considered as components of a Markov process in a "slow" and "fast" time, respectively. Hence the description of the limit process for $\{u_\varepsilon\}$ is closely related to the averaging principle for Markov processes. We will see that the limit process satisfies an ODE on $H$ whose coefficients are an averaging of functions $a_k^{\pm}$, $k = 1, \ldots, d-1$, over a stationary distribution of a scaled vertical component given the other components were frozen. The idea to use some scaling for small-noise problem was effectively used in one-dimensional case if the drift is a power-type function and the noise is a Levy $\alpha$-stable process or even more general.

***Remark* 1.3.** The case $\gamma = 1$ is critical. If $a_d(x) \sim x_d b(x)$, where $b^{\pm}(x) \neq 0$, $x \in H$, then the limit process may be non-Markov and satisfy certain equation [19] that depends somehow on a Wiener process $W$ (that formally should disappear in a limit equation).

The paper is organized as follows. In Section 2 we formulate the problem and the main results. The proofs for the cases when the drift pushes outwards $H$ and towards $H$ are given in Sections 3 and 4, respectively.

In Subsection 2.3 we also formulate an averaging principle, which is quite general and new result. The proof of averaging principle is postponed to Section 5.

**2. Main results.** Let us represent $u_\varepsilon(t)$ as a pair $\big(X_\varepsilon(t), Y_\varepsilon(t)\big)$, where $Y_\varepsilon$ is the last coordinate of $u_\varepsilon$ and $X_\varepsilon$ consists of the first $d-1$ coordinates. Below we study only the general problem for the pair $\big(X_\varepsilon(t), Y_\varepsilon(t)\big)$, which can be easily be reformulated for $u_\varepsilon$. For notational convenience, we assume below that $X_\varepsilon$ is a $d$-dimensional process but not $(d-1)$-dimensional one.

The general setup is the following. Let $X_\varepsilon, Y_\varepsilon$ be stochastic processes with values in $\mathbb{R}^d$ and $\mathbb{R}$, respectively. Assume that the pair $X_\varepsilon$, $Y_\varepsilon$ satisfies the following SDE:

$$\begin{gathered} dX_\varepsilon(t) = \psi\big(X_\varepsilon(t), Y_\varepsilon(t)\big)\, dt + \varepsilon\, b\big(X_\varepsilon(t), Y_\varepsilon(t)\big) dB(t), \\ dY_\varepsilon(t) = \varphi\big(X_\varepsilon(t), Y_\varepsilon(t)\big) Y_\varepsilon^{\gamma}(t)\, dt + \varepsilon\beta\big(X_\varepsilon(t), Y_\varepsilon(t)\big)\, dW(t), \\ X_\varepsilon(0) = x^0, \qquad Y_\varepsilon(0) = 0, \end{gathered} \tag{2.1}$$

where $B$, $W$ are Wiener processes (multidimensional and one-dimensional), that may be dependent.

Denote

$$y^\gamma := |y|^\gamma (\mathbb{1}_{y>0} - \mathbb{1}_{y<0}),$$
$$H := \mathbb{R}^d \times \{0\}.$$

Assume that

**B1**) $\psi(x,y) = \psi^+(x,y)\mathbb{1}_{y\geq 0} + \psi^-(x,y)\mathbb{1}_{y<0}$ and $\varphi(x,y) = \varphi^+(x,y)\mathbb{1}_{y\geq 0} + \varphi^-(x,y)\mathbb{1}_{y<0}$, where functions $\psi^{\pm}$, $\varphi^{\pm}$ are bounded, continuous in $x$, $y$.

We assume that domains of $\psi^{\pm}$, $\varphi^{\pm}$ are the whole space $x \in \mathbb{R}^d$, $y \in \mathbb{R}$, despite we use their values on the corresponding half-spaces only. The functions $\psi$, $\varphi$ may have jump discontinuity on $H$.

**B2**) $\varphi^{\pm}(x,0) \neq 0$ for any $x \in \mathbb{R}^d$.

**B3**) $\beta(x,y) = \beta^+(x,y)\mathbb{1}_{y\geq 0} + \beta^-(x,y)\mathbb{1}_{y<0}$, where $\beta^{\pm}$ are bounded, continuous and separated from zero function in the whole space $\mathbb{R}^d \times \mathbb{R}$; function $b$ is bounded and continuous in $(\mathbb{R}^d \times \mathbb{R}) \setminus H$.

**B4**) $\gamma \in (0,1)$.

Under assumptions **B1** – **B4** there exists a weak solution to (2.1).

Indeed, it follows from the standard compactness arguments that there exists a weak solution to

$$d\hat{X}_\varepsilon(t) = \frac{\psi}{\beta^2}\big(\hat{X}_\varepsilon(t), \hat{Y}_\varepsilon(t)\big)dt + \varepsilon\,\frac{b}{\beta}\big(\hat{X}_\varepsilon(t), \hat{Y}_\varepsilon(t)\big)dB(t),$$

$$d\hat{Y}_\varepsilon(t) = \varepsilon dW(t),$$

$$\hat{X}_\varepsilon(0) = x^0, \qquad \hat{Y}_\varepsilon(0) = 0.$$

Note that all coefficients may be discontinuous in $H$ but the processes spend zero time there with probability 1. Any redefinition of coefficients in $H$ does not affect the equations.

By using the transformation of time arguments (see, for example, [13]), we get a solution to

$$d\hat{\hat{X}}_\varepsilon(t) = \psi\Big(\hat{\hat{X}}_\varepsilon(t), \hat{\hat{Y}}_\varepsilon(t)\Big)dt + \varepsilon\, b\Big(\hat{\hat{X}}_\varepsilon(t), \hat{\hat{Y}}_\varepsilon(t)\Big)dB(t),$$

$$d\hat{\hat{Y}}_\varepsilon(t) = \varepsilon\beta\Big(\hat{\hat{X}}_\varepsilon(t), \hat{\hat{Y}}_\varepsilon(t)\Big)dW(t),$$

$$\hat{\hat{X}}_\varepsilon(0) = x^0, \qquad \hat{\hat{Y}}_\varepsilon(0) = 0.$$

Finally, Girsanov's theorem yields existence of a weak solution to (2.1).

***Remark* 2.1.** If $b$ is nondegenerate, then existence of a solution can be proved without transformation of time arguments.

***2.1. Repulsion from the hyperplane.*** In this subsection, we assume that $\varphi^\pm(x,0) > 0$ for all $x \in \mathbb{R}^d$. Then $\operatorname{sgn}(y)\varphi(x,y)y^\gamma > 0,\ y \neq 0$ and the drift pushes away from the hyperplane $\mathbb{R}^d \times \{0\}$.

Suppose that assumptions **B1 – B4** hold true and functions $\psi^\pm, \varphi^\pm$ are locally Lipschitz continuous in $(x,y) \in \mathbb{R}^d \times \mathbb{R}$.

Then there are unique solutions $(X^+(t), Y^+(t))$ and $(X^-(t), Y^-(t))$ to the unperturbed system (i.e., $\varepsilon = 0$):

$$dX(t) = \psi\big(X(t), Y(t)\big)\,dt,$$

$$dY(t) = \varphi\big(X(t), Y(t)\big)Y^\gamma(t)\,dt,$$

$$X(0) = x^0, \qquad Y(0) = 0,$$

such that $Y^+(t) > 0$ and $Y^-(t) < 0$ for all $t > 0$.

Indeed, set $\tilde{Y}(t) := Y^{1-\gamma}(t)$. Then

$$X(t) = x^0 + \int_0^t \psi\Big(X(s), \tilde{Y}^{\frac{1}{1-\gamma}}(s)\Big)\,ds,$$

$$\tilde{Y}(t) = (1-\gamma)\int_0^t \varphi\Big(X(s), \tilde{Y}^{\frac{1}{1-\gamma}}(s)\Big)\,ds.$$

Since $\gamma \geq 0$, the functions $(x,\tilde{y}) \to \psi^\pm(x,\tilde{y})$ and $(x,\tilde{y}) \to \varphi^\pm(x,\tilde{y})$ are locally Lipschitz continuous. So, equations

$$X^\pm(t) = x^0 + \int_0^t \psi^\pm\Big(X^\pm(s), (\tilde{Y}^\pm(s))^{\frac{1}{1-\gamma}}\Big)\,ds,$$

$$\tilde{Y}^{\pm}(t) = (1-\gamma)\int\limits_0^t \varphi^{\pm}\Big(X^{\pm}(s), (\tilde{Y}^{\pm}(s))^{\frac{1}{1-\gamma}}\Big)\,ds$$

have unique solutions $(X^{\pm}(t), \tilde{Y}^{\pm}(t))$ and these solutions are such that $\tilde{Y}^{+}(t) > 0$ and $\tilde{Y}^{-}(t) <$ $< 0$ for all $t > 0$. Making the inverse change of variables we get the desired functions $Y^{\pm}(t) =$ $= \big(\tilde{Y}^{\pm}(t)\big)^{\frac{1}{1-\gamma}}$.

The solution does not explode in a finite time because $\psi^{\pm}, \varphi^{\pm}$ are bounded by assumption **B1**.

**Theorem 2.1.** *The distribution of* $(X_\varepsilon, Y_\varepsilon)$ *in* $C([0,T])$ *converges weakly as* $\varepsilon \to 0$ *to the measure*

$$p_{-}\delta_{(X^{-},Y^{-})} + p_{+}\delta_{(X^{+},Y^{+})},$$

*where*

$$p_{\pm} = \frac{\left(\dfrac{\varphi^{\pm}(x^0,0)}{\big(\beta^{\pm}(x^0,0)\big)^2}\right)^{\frac{1}{\gamma+1}}}{\left(\dfrac{\varphi^{-}(x^0,0)}{(\beta^{-}(x^0,0))^2}\right)^{\frac{1}{\gamma+1}} + \left(\dfrac{\varphi^{+}(x^0,0)}{\big(\beta^{+}(x^0,0)\big)^2}\right)^{\frac{1}{\gamma+1}}} \tag{2.2}$$

*and* $\delta_{(X^{+},Y^{+})}$, $\delta_{(X^{-},Y^{-})}$ *means the unit mass that concentrated on the functions* $(X^{+},Y^{+})$ *and* $(X^{-},Y^{-})$, *respectively.*

The proof is given in Section 3.

***Remark* 2.2.** If $\pm\varphi^{\pm}(x,0) > 0$ (or $\pm\varphi^{\pm}(x,0) < 0$) for all $x \in \mathbb{R}^d$, then the limit process is $(X^{+}(t), Y^{+}(t))$ (respectively, $(X^{-}(t), Y^{-}(t))$) with probability 1.

***Remark* 2.3.** If we have inequality $\varphi^{+}(x^0,0) > 0$ and $\varphi^{-}(x^0,0) < 0$ only at the initial point (and hence in some neighborhood by continuity of coefficients), then the functions $(X^{\pm}(t), Y^{\pm}(t))$ are well defined up to the moment $\tau_H^{\pm} := \inf\big\{t > 0\colon Y^{\pm}(t) = 0\big\}$ of the first return to $H$. In this case we have the convergence in distribution for the stopped processes:

$$\big(X_\varepsilon(\cdot\wedge\tau_H^{+}\wedge\tau_H^{-}), Y_\varepsilon(\cdot\wedge\tau_H^{+}\wedge\tau_H^{-})\big) \Rightarrow$$

$$\Rightarrow p_{-}\delta_{(X^{-}(\cdot\wedge\tau_H^{+}\wedge\tau_H^{-}),Y^{-}(\cdot\wedge\tau_H^{+}\wedge\tau_H^{-}))} + p_{+}\delta_{(X^{+}(\cdot\wedge\tau_H^{+}\wedge\tau_H^{-}),Y^{+}(\cdot\wedge\tau_H^{+}\wedge\tau_H^{-}))}.$$

The proof is essentially the same, but it involves routine localization arguments in addition.

***2.2. Attraction to the hyperplane.*** In this subsection, we assume that $\varphi^{\pm}(x,0) < 0$ for all $x \in \mathbb{R}^d$.

Suppose that assumptions **B1** – **B4** hold true and $\psi^{\pm}$ are locally Lipschitz in $x$ for any fixed $y$.

**Theorem 2.2.** *For any* $T > 0$ *we have the uniform convergence in probability*

$$\lim_{\varepsilon\to 0}\sup_{t\in[0,T]}\big\|\big(X_\varepsilon(t), Y_\varepsilon(t)\big) - (X(t),0)\big\| = 0,$$

*where* $X(t)$ *is a solution to the ODE*

$$dX(t) = \bar{\psi}(X(t))dt, \quad X(0) = 0,$$

*and*

$$\overline{\psi}(x) = \psi^+(x,0)\frac{\left(\dfrac{(\beta^+(x,0))^2}{\varphi^+(x,0)}\right)^{\frac{1}{\gamma+1}}}{\left(\dfrac{(\beta^-(x,0))^2}{\varphi^-(x,0)}\right)^{\frac{1}{\gamma+1}} + \left(\dfrac{(\beta^+(x,0))^2}{\varphi^+(x,0)}\right)^{\frac{1}{\gamma+1}}} +$$

$$+\psi^-(x,0)\frac{\left(\dfrac{(\beta^-(x,0))^2}{\varphi^-(x,0)}\right)^{\frac{1}{\gamma+1}}}{\left(\dfrac{(\beta^-(x,0))^2}{\varphi^-(x,0)}\right)^{\frac{1}{\gamma+1}} + \left(\dfrac{(\beta^+(x,0))^2}{\varphi^+(x,0)}\right)^{\frac{1}{\gamma+1}}}. \tag{2.3}$$

The proof is given in Section 4.

***Remark* 2.4.** Note that

$$\frac{\varphi^+(x,0)^{-\frac{1}{\gamma+1}}}{\varphi^+(x,0)^{-\frac{1}{\gamma+1}} + \varphi^-(x,0)^{-\frac{1}{\gamma+1}}} = \pi^{(x)}([0,\infty)),$$

$$\frac{\varphi^-(x,0)^{-\frac{1}{\gamma+1}}}{\varphi^+(x,0)^{-\frac{1}{\gamma+1}} + \varphi^-(x,0)^{-\frac{1}{\gamma+1}}} = \pi^{(x)}((-\infty,0)),$$

where $\pi^{(x)}$ is the stationary distribution for the SDE

$$dy^{(x)}(t) = \big(\varphi^+(x,0)1\!\!1_{y^{(x)}(t)>0} + \varphi^-(x,0)1\!\!1_{y^{(x)}(t)<0}\big)(y^{(x)}(t))^\gamma\, dt + \beta\big(x,0)\, dW(t).$$

Hence,

$$\overline{\psi}(x) = \psi^+(x,0)\pi^{(x)}([0,\infty)) + \psi^-(x,0)\pi^{(x)}((-\infty,0)),$$

i.e., the drift of the limit equation is the averaging of $\psi^\pm$ over the stationary distribution of an SDE with frozen $x$ variable. The corresponding relation between the averaging principle and averaging of coefficients in the limit equation for the small noise perturbation problem will be seen from the proof.

In the next subsection, we formulate an averaging principle, which is applied in the proof of Theorem 2.2. We consider more general SDEs than (2.1) because the idea of the proof is universal. The corresponding result may be interesting by itself.

***2.3. Averaging.*** Let for $\varepsilon > 0$ the processes $X_\varepsilon(t), Y_\varepsilon(t)$ take values in $\mathbb{R}^d, \mathbb{R}^k$ and have the form

$$X_\varepsilon(t) = X_\varepsilon(0) + \int_0^t a^\varepsilon\big(X_\varepsilon(s), Y_\varepsilon(t)\big)\, ds + \int_0^t \sigma^\varepsilon\big(X_\varepsilon(s), Y_\varepsilon(s)\big)\, dB_s^\varepsilon +$$

$$+\int_0^t \int_{\mathbb{R}^m} c^\varepsilon\big(X_\varepsilon(s-), Y_\varepsilon(s-), u\big)\Big[N^\varepsilon(du,ds) - 1_{|u|\le\rho}\nu^\varepsilon(du)ds\Big] + \xi_\varepsilon(t), \tag{2.4}$$

$$Y_\varepsilon(t) = Y_\varepsilon(0) + \varepsilon^{-1}\int_0^t A^\varepsilon\big(X_\varepsilon(s), Y_\varepsilon(s)\big)\, ds + \varepsilon^{-1/2}\int_0^t \Sigma^\varepsilon\big(X_\varepsilon(s), Y_\varepsilon(s)\big)\, dW_s^\varepsilon +$$

$$+\int_0^t\int_{\mathbb{R}^l} C^\varepsilon\big(X_\varepsilon(s-),Y_\varepsilon(s-),z\big)\Big[Q^\varepsilon(dz,ds)-1_{|z|\le\rho}\varepsilon^{-1}\mu^\varepsilon(dz)ds\Big],$$

where $B^\varepsilon_t$, $W^\varepsilon_t$ are Brownian motions and $N^\varepsilon(du,dt)$, $Q^\varepsilon(dz,dt)$ are Poisson point measures on a common filtered probability space $(\Omega^\varepsilon,\mathcal{F}^\varepsilon,\mathbf{P}^\varepsilon)$, and the random measures $N^\varepsilon(du,dt)$, $Q^\varepsilon(dz,dt)$ have the intensity measures $\nu^\varepsilon(du)dt$ and $\varepsilon^{-1}\mu^\varepsilon(dz)dt$, respectively. These random measures are involved into the system in the partially compensated form, which is quite typical for the Lévy-driven SDEs; what is a bit unusual is the choice of the cutoff functions $1_{|u|\le\rho}$, $1_{|z|\le\rho}$ with the number $\rho>0$ to be specified separately. This choice will become clear later, when we describe the limit behavior of the Lévy measures $\nu^\varepsilon(du)$, $\mu^\varepsilon(dz)$ as $\varepsilon\to0$. Note that here and below we do not assume a uniqueness of a solution to prelimit equation (2.4).

The factor $\varepsilon^{-1}$ in the intensity measure for $Q^\varepsilon(dz,dt)$ and the factors $\varepsilon^{-1}$, $\varepsilon^{-1/2}$ at the integrals w.r.t. $ds$ and $dW^\varepsilon_s$ in the equation for $Y_\varepsilon$ mean that the evolution of the component $Y_\varepsilon$ happens at the "fast" time scale $\varepsilon^{-1}t$, which we will also call the "microscopic" time scale. The component $X_\varepsilon$ evolves at the "slow", or "macroscopic" time scale $t$; its evolution involves the deterministic term, two stochastic terms (continuous and partially compensated jump parts), and a residual term $\xi_\varepsilon$, for which we do not impose any structural assumptions, and only require it to be asymptotically small in the following sense:

$\mathbf{H_0}$ (Negligibility of the residual term). The process $\xi_\varepsilon(t)$ is an adapted càdlàg process, and, for any $T>0$,

$$\sup_{t\in[0,T]}|\xi_\varepsilon(t)|\to0,\quad \varepsilon\to0,$$

in probability.

The aim of this subsection is to get the averaging principle for the "slow" component $X_\varepsilon$. Let us stress that the framework we adopt is quite general. In particular,

the two-scale system (2.4) is *fully coupled* in the sense that the coefficients of the "slow" component depend on the "fast" one, and vice versa;

the noises for the "slow" and the "fast" component are allowed to be dependent;

the coefficients of the "slow" component can be discontinuous.

Let us introduce further assumptions on the system (2.4). Note that all the assumptions listed below are quite natural and nonrestrictive.

$\mathbf{H_1}$ (Bounds for the coefficients). There exists a constant $C$ such that

$$|a^\varepsilon(x,y)|\le C,\quad |\sigma^\varepsilon(x,y)|\le C,\quad |\Sigma^\varepsilon(x,y)|\le C,\quad |c^\varepsilon(x,y,u)|\le C|u|,\quad |C^\varepsilon(x,y,z)|\le C|z|$$

for all values of $x$, $y$, $u$, $z$.

In addition, for any $R>0$ there exists a constant $C_R$ such that

$$|A^\varepsilon(x,y)|\le C_R,\qquad x\in\mathbb{R}^d,\quad |y|\le R.$$

$\mathbf{H_2}$ (Bounds for the Lévy measures). There exist constants $C$ and $p>0$ such that

$$\int_{\mathbb{R}^m}(|u^2|\wedge1)\nu^\varepsilon(du)\le C,\qquad \int_{\mathbb{R}^l}(|z^2|1_{|z|\le1}+|z|^p1_{|z|>1})\mu^\varepsilon(dz)<\infty.$$

**H$_3$** (The coefficients of the fast component are convergent). There exist continuous functions $A(x,y),\ \Sigma(x,y),\ C(x,y,z)$ such that

$$A^\varepsilon(x,y)\to A(x,y),\qquad \Sigma^\varepsilon(x,y)\to\Sigma(x,y),\qquad\text{and}\qquad C^\varepsilon(x,y,z)\to C(x,y,z)\quad\text{as}\quad\varepsilon\to 0$$

uniformly on every compact set in $\mathbb{R}^d\times\mathbb{R}^k$, $\mathbb{R}^d\times\mathbb{R}^k$, and $\mathbb{R}^d\times\mathbb{R}^k\times\big(\mathbb{R}^l\setminus\{0\}\big)$, respectively.

To introduce the next condition, let us define the *weak convergence* of a family of Lévy measures on $\mathbb{R}^m$ in the following way: $\nu^\varepsilon(du)\Longrightarrow\nu(du)$ if for every continuous function $\varphi$ with a support compactly embedded into $\mathbb{R}^m\setminus\{0\}$,

$$\int\limits_{\mathbb{R}^m}\varphi(z)\,\nu^\varepsilon(dz)\to\int\limits_{\mathbb{R}^m}\varphi(z)\,\nu(dz),\quad\varepsilon\to 0.$$

**H$_4$** (The Lévy measures of the noises are weakly convergent). There exist Lévy measures $\nu(du)$, $\mu(dz)$ on $\mathbb{R}^m,\mathbb{R}^l$ respectively, such that

$$\nu^\varepsilon(du)\Longrightarrow\nu(du)\qquad\text{and}\qquad\mu^\varepsilon(dz)\Longrightarrow\mu(dz)\quad\text{as}\quad\varepsilon\to 0.$$

In addition,

$$\nu\big(\{u\colon|u|=\rho\}\big)=0,\qquad\mu\big(\{z\colon|z|=\rho\}\big)=0.\tag{2.5}$$

Condition (2.5) yield that the cutoff functions $1_{|u|\le\rho},1_{|z|\le\rho}$ used in (2.4) are a.s. continuous w.r.t. the measures $\nu(du),\mu(dz)$, respectively. Note that there exists at most countable set of levels $\rho$ such that (2.5) fails, hence one can always choose $\rho$ to satisfy this condition. Of course, changing the cutoff level would change the drift coefficients respectively.

Next, assume that the drift of the fast component performs an attraction to origin.

**H$_5$** (The drift condition for the microscopic dynamics). There exist $\kappa>0$ and $c,r>0$ such that

$$A^\varepsilon(x,y)\cdot y\le -c|y|^{\kappa+1},\quad |y|\ge r.\tag{2.6}$$

In addition, the *balance condition* holds:

$$\kappa+p>1,\tag{2.7}$$

where $p$ is introduced in the assumption **H$_2$**.

Consider a family of “frozen microscopic equations”

$$dy(t)=A\big(x,y(t)\big)\,dt+\Sigma\big(x,y(t-)\big)\,dW_t+ \\ +\int\limits_{\mathbb{R}^l}C\big(x,y(t-),z\big)\Big[Q(dz,ds)-1_{|z|\le 1}\mu(dz)ds\Big],\tag{2.8}$$

where $W$ is a Wiener process and $Q(dz,dt)$ is an independent Poisson point measure with the intensity measure $\mu(dz)dt$. For the corresponding “frozen dynamics” we introduce a separate family of assumptions.

**$\mathbf{F_0}$** (The "frozen microscopic dynamics" is well defined and Feller). For any $x$ and any initial value $y(0) = y$, the SDE (2.8) has a unique weak solution, which is a Markov process. Furthermore we denote the corresponding family of Markov processes by $y^{(x)},\ x \in \mathbb{R}^d$, and write $P_t^{(x)}(y, dy')$ for the corresponding family of transition probabilities.

We also denote

$$P_t^{\text{frozen}} f(x,y) = \int_{\mathbb{R}^k} f(x,y') P_t^{(x)}(y,dy'), \quad t \geq 0,$$

the semigroup of operators corresponding to the two-component process $(x, y^{(x)})$ in which the first component is constant and the second one is the Markov process specified above. We assume that this semigroup is Feller.

For this family, we assume the following mixing property, which is actually the local Dobrushin condition, uniform in parameter $x$; see [16] (Section 2).

**$\mathbf{F_1}$** (The "frozen microscopic dynamics" is locally mixing). There exists $h > 0$ such that, for any $R > 0$, there exists $\rho = \rho_R > 0$ such that, for any $x,\ y_1,\ y_2$ with $|x| \leq R,\ |y_1| \leq R,\ |y_2| \leq R$

$$\left\| P_h^{(x)}(y_1, dy') - P_h^{(x)}(y_2, dy') \right\|_{TV} \leq 1 - \rho,$$

where $P_t^{(x)}(y, dy')$ denotes the transition probability of the process $y^{(x)}$, and *the total variation* distance between probability measures is defined as

$$\|\lambda_1 - \lambda_2\|_{TV} = \sup_A (\lambda_1(A) - \lambda_2(A)).$$

We note that assumptions $\mathbf{F_1}$, $\mathbf{H_5}$ ensure that, for each $x \in \mathbb{R}^d$, the laws of $y_t^{(x)}$ converge to the invariant probability measure (IPM) $\pi^{(x)}(dy)$ with an explicitly rate; see Proposition 5.1 below.

For the coefficients of the "slow" component, we assume a weaker analogue of $\mathbf{H_3}$, where the convergence and continuity of the limiting coefficients may fail on an exceptional set, which should be negligible, in a sense.

**$\mathbf{H_6}$** (The coefficients of the slow component are convergent). There exist functions $a(x,y)$, $\sigma(x,y),\ c(x,y,u)$ and an open set $B \subset \mathbb{R}^d \times \mathbb{R}^k$ such that, for any compact set $K \subset B$,

$$a^\varepsilon(x,y) \to a(x,y) \qquad \text{and} \qquad \sigma^\varepsilon(x,y) \to \sigma(x,y) \quad \text{as} \quad \varepsilon \to 0$$

uniformly on $K$, and for any $R > 1$

$$c^\varepsilon(x,y,u) \to c(x,y,u), \quad \varepsilon \to 0,$$

uniformly on $K \times \{u \colon R^{-1} \leq |u| \leq R\}$. The set $\Delta = (\mathbb{R}^d \times \mathbb{R}^k) \setminus B$ satisfies

$$\pi^{(x)}\{y : (x,y) \in \Delta\} = 0 \quad \text{for any} \quad x \in \mathbb{R}^d.$$

In addition, the functions $a(x,y),\ \sigma(x,y)$, and $c(x,y,u)$ are continuous on $B$ and $B \times (\mathbb{R}^m \setminus \{0\})$, respectively.

Define the averaging of the limiting drift coefficient for the macroscopic component w.r.t. the family of IPMs for the frozen microscopic one:

$$\overline{a}(x) = \int\limits_{\mathbb{R}^k} a(x,y)\pi^{(x)}(dy).$$

Next, consider the limiting diffusion matrix and compensated/noncompensated jump kernels for the macroscopic component,

$$b(x,y) = \sigma(x,y)\sigma(x,y)^*,$$

$$K_{(\rho)}(x,y,A) = \nu\big(\{u\colon |u| \le \rho, c(x,y,u) \in A\}\big),$$

$$K^{(\rho)}(x,y,A) = \nu\big(\{u\colon |u| > \rho, c(x,y,u) \in A\}\big),$$

and introduce the corresponding averaged characteristics as

$$\overline{b}(x) = \int\limits_{\mathbb{R}^k} b(x,y)\pi^{(x)}(dy),$$

$$\overline{K}_{(\rho)}(x,dv) = \int\limits_{\mathbb{R}^k} K_{(\rho)}(x,y,dv)\pi^{(x)}(dy),$$

$$\overline{K}^{(\rho)}(x,dv) = \int\limits_{\mathbb{R}^k} K^{(\rho)}(x,y,dv)\pi^{(x)}(dy).$$

Finally, we introduce an auxiliary technical assumption.

$\mathbf{A_0}$ The averaged coefficients $\overline{a}(x)$, $\overline{b}(x)$ are continuous. The averaged Lévy kernels $\overline{K}_{(\rho)}(x,dv)$, $\overline{K}^{(\rho)}(x,dv)$ depend on $x$ continuously, in the sense that

$$\overline{K}_{(\rho)}(x',dv) \Longrightarrow \overline{K}_{(\rho)}(x,dv) \qquad \text{and} \qquad \overline{K}^{(\rho)}(x',dv) \Longrightarrow \overline{K}^{(\rho)}(x,dv) \quad \text{as} \quad x' \to x.$$

***Remark* 2.5.** It is easy to give a sufficient condition for $\mathbf{A_0}$ to holds. Namely, it is enough to assume, in addition to $\mathbf{H_0}-\mathbf{H_6}$, $\mathbf{F_0}$, $\mathbf{F_1}$, that the transition probabilities $P_t^{(x)}(y,dy')$ are continuous in $x$ w.r.t. the total variation convergence for each $y \in \mathbb{R}^k$, $t \ge t_0$. Then, because of the convergence (5.6), the same continuity holds for the family of the IPMs $\pi^{(x)}(dy)$. The latter continuity, combined with $\mathbf{H_1}$, $\mathbf{H_2}$, $\mathbf{H_4}$, and $\mathbf{H_6}$, yields the required continuity of the averaged coefficients.

Now we are ready to formulate our main statement.

**Theorem 2.3.** *Assume $\boldsymbol{H_0}-\boldsymbol{H_6}$, $\boldsymbol{F_0}$, $\boldsymbol{F_1}$, and $\boldsymbol{A_0}$ to hold,*

$$X_\varepsilon(0) \to x^0, \quad \varepsilon \to 0,$$

*in probability and $\{Y_\varepsilon(0)\}$ be bounded in probability.*

*Then the family $\{X_\varepsilon, \varepsilon > 0\}$ is weakly compact in $\mathbb{D}\big([0,\infty),\mathbb{R}^d\big)$, and any of its weak limit point as $\varepsilon \to 0$ is a solution to the martingale problem $(L, C_0^\infty)$ with*

$$L\varphi(x) = \nabla\varphi(x)\cdot\overline{a}(x) + \frac{1}{2}\nabla^2\varphi(x)\cdot\overline{b}(x) +$$

$$+ \int\limits_{\mathbb{R}^m} \big(\varphi(x+v) - \varphi(x) - \nabla\varphi(x)\cdot v\big)\overline{K}_{(\rho)}(x,dv) +$$

$$+\int\limits_{\mathbb{R}^m}\big(\varphi(x+v)-\varphi(x)\big)\overline{K}^{(\rho)}(x,dv)=$$

$$=\nabla\varphi(x)\cdot\overline{a}(x)+\frac{1}{2}\nabla^2\varphi(x)\cdot\overline{b}(x)+$$

$$+\int\limits_{\mathbb{R}^k}\int\limits_{\mathbb{R}^m}\Big(\varphi(x+c(x,y,u))-\varphi(x)-\nabla\varphi(x)\cdot c(x,y,u)\mathbb{1}_{|u|\le\rho}\Big)\nu(du)\pi^{(x)}(dy), \tag{2.9}$$

*where* $\varphi\in C_0^\infty$.

*If the martingale problem* (2.9) *is well posed*, *then* $X_\varepsilon$ *weakly converges as* $\varepsilon\to 0$ *to its unique solution with* $X(0)=x^0$.

**3. Proof of Theorem 2.1.** The proof almost copying the proof of Theorem 3.1 in [19]. Thus we only sketch the main steps of the proof.

***Step* 1.** The sequence $\{(X_\varepsilon,Y_\varepsilon)\}$ is weakly relatively compact. The proof follows from boundedness of functions $\varphi$, $\psi$, $b$, $\beta$.

Therefore, to prove the theorem it suffices to verify that any subsequence $\{(X_{\varepsilon_n},Y_{\varepsilon_n})\}$ contains sub-subsequence $\{(X_{\varepsilon_{n_k}},Y_{\varepsilon_{n_k}})\}$ that converges to the desired limit. Without loss of generality we will assume that $\{(X_\varepsilon,Y_\varepsilon)\}$ is weakly convergent by itself.

***Step* 2.** Estimate for the time spent by $Y_\varepsilon$ in a neighborhood of 0.

We will use the following general statement.

**Lemma 3.1.** *Assume that processes* $\{\eta_\varepsilon(t)\}$ *satisfy the SDE*

$$d\eta_\varepsilon(t)=a_\varepsilon(t)\eta_\varepsilon^\gamma(t)dt+\varepsilon b_\varepsilon(t)dW(t),$$

$$\eta_\varepsilon(0)=0,$$

*where* $|\gamma|<1$, *and* $a_\varepsilon(t)$, $b_\varepsilon(t)$ *are* $\mathcal{F}_t$*-adapted processes such that*

$$a_\varepsilon(t)\ge A>0,\quad 0<C_1\le b_\varepsilon(t)\le C_2$$

*for all* $\omega$, $t$, $\varepsilon$.

*Set*

$$\tau_\varepsilon(\delta):=\inf\big\{t\ge 0\colon|\eta_\varepsilon(t)|\ge\delta\big\}.$$

*Then there is a constant* $K=K(A,C_1,C_2)$ *such that*

$$\forall\delta>0\quad\exists\varepsilon_0>0\quad\forall\varepsilon\in(0,\varepsilon_0)\colon\mathbf{E}\tau_\varepsilon(\delta)\le K\delta^{1-\gamma}.$$

The proof of the lemma is quite standard. We postpone it to the Appendix.

Without loss of generality we will assume that

$$\psi^\pm(x,0)\ge c_1>0\qquad\text{and}\qquad 0<c_2\le\beta(x)\le c_3\quad\text{for all}\quad x\in\mathbb{R}^d, \tag{3.1}$$

where $c_{1,2,3}$ are some positive constants. This assumption does not restrict generality, since the general case can be considered using a localization. Under this additional assumption, Lemma 3.1 applied to

$$\tau_\varepsilon(\delta) := \inf\big\{t \ge 0\colon |Y_\varepsilon(t)| \ge \delta\big\}$$

and the Chebyshev inequality yield

$$\forall\delta > 0 \quad \exists\varepsilon_0 > 0 \quad \forall\varepsilon \in (0, \varepsilon_0)\colon \mathbf{P}\big(\tau_\varepsilon(\delta) \ge \delta^{\frac{1-\gamma}{2}}\big) \le K\delta^{\frac{1-\gamma}{2}}. \tag{3.2}$$

***Remark* 3.1.** It can be seen from the construction of $Y^\pm$ that the inequality (3.2) is valid for $\tau^\pm(\delta) := \inf\big\{t \ge 0\colon |Y^\pm(t)| \ge \delta\big\}$ also.

***Step* 3.** We see from (3.2) that with high probability the random variable $\tau_\varepsilon(\delta)$ is dominated by $\delta^{\frac{1-\gamma}{2}}$. It follows from the standard estimates for moments of SDEs that for small $t$ we have

$$\mathbf{E}\sup_{s\in[0,t]}\big|X_\varepsilon(s) - x^0\big|^2 \le Ct,$$

where constant $C$ can be selected independently of $\varepsilon \in [0, 1]$.

So, we have the following estimates:

$$\exists C_1 > 0 \quad \forall\delta > 0 \quad \exists\varepsilon_0 > 0 \quad \forall\varepsilon \in (0, \varepsilon_0)\colon \mathbf{P}\left(\sup_{t\in[0,\tau_\varepsilon(\delta)]}|X_\varepsilon(t) - x^0| \ge \delta^{\frac{1-\gamma}{6}}\right) \le C_1\delta^{\frac{1-\gamma}{6}}, \tag{3.3}$$

$$\mathbf{P}\left(\sup_{t\in[0,\tau_\varepsilon(\delta)]}|X^\pm(t) - x^0| + |Y^\pm(t)| \ge 2\delta^{\frac{1-\gamma}{6}}\right) \le C_1\delta^{\frac{1-\gamma}{6}}. \tag{3.4}$$

Verify, for example, (3.3):

$$\mathbf{P}\left(\sup_{t\in[0,\tau_\varepsilon(\delta)]}|X_\varepsilon(t) - x^0| \ge \delta^{\frac{1-\gamma}{6}}\right) \le$$

$$\le \mathbf{P}\big(\tau_\varepsilon(\delta) > \delta^{\frac{1-\gamma}{2}}\big) + \mathbf{P}\left(\sup_{t\in\left[0,\delta^{\frac{1-\gamma}{2}}\right]}|X_\varepsilon(t) - x^0| \ge \delta^{\frac{1-\gamma}{6}}\right) \le$$

$$\le \delta^{\frac{1-\gamma}{2}} + \frac{\mathbf{E}\sup_{t\in\left[0,\delta^{\frac{1-\gamma}{2}}\right]}|X_\varepsilon(t) - x^0|^2}{\delta^{\frac{1-\gamma}{3}}} \le$$

$$\le \delta^{\frac{1-\gamma}{2}} + \frac{C\delta^{\frac{1-\gamma}{2}}}{\delta^{\frac{1-\gamma}{3}}} \le C_1\delta^{\frac{1-\gamma}{6}}.$$

Note also that

$$\sup_{t\in[0,\tau_\varepsilon(\delta)]}|Y_\varepsilon(t)| = \big|Y_\varepsilon(\tau_\varepsilon(\delta))\big| = \delta \quad \text{a.s.} \tag{3.5}$$

by the definition of $\tau_\varepsilon(\delta)$.

***Step* 4.** We denote by $\big(X^{x,y}(t), Y^{x,y}(t)\big)$ a solution to the corresponding ODE that starts from $x \in \mathbb{R}^d$, $y \ne 0$. This solution never hits $\mathbb{R}^d \times \{0\}$, recall (3.1). We have correctness of the definition of $\big(X^{x,y}(t), Y^{x,y}(t)\big)$ because in all other points coefficients satisfy the local Lipschitz condition.

If we wish to highlight that $y > 0$ (or $y < 0$), then the corresponding solution is denoted by $\big(X^{+,x,y}(t), Y^{+,x,y}(t)\big)$ $\big($or $\big(X^{-,x,y}(t), Y^{-,x,y}(t)\big)$, respectively$\big)$.

Let $\omega$ be such that $Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta$, i.e., the process $Y_\varepsilon$ hits $\delta$ earlier than $-\delta$. Then for this $\omega$ we have

$$\sup_{t\in[0,T]} \big(|X_\varepsilon(t) - X^+(t)| + |Y_\varepsilon(t) - Y^+(t)|\big) \le$$

$$\le \sup_{t\in[0,T]} \Big(\big|X_\varepsilon(\tau_\varepsilon(\delta)+t) - X^{+,X_\varepsilon(\tau_\varepsilon(\delta)),Y_\varepsilon(\tau_\varepsilon(\delta))}(t)\big| +$$

$$+\big|Y_\varepsilon(\tau_\varepsilon(\delta)+t) - Y^{+,X_\varepsilon(\tau_\varepsilon(\delta)),Y_\varepsilon(\tau_\varepsilon(\delta))}(t)\big|\Big) +$$

$$+ \sup_{t\in[0,T]} \Big(\big|X^{+,X_\varepsilon(\tau_\varepsilon(\delta)),Y_\varepsilon(\tau_\varepsilon(\delta))}(t) - X^+(\tau_\varepsilon(\delta)+t)\big| +$$

$$+\big|Y^{+,X_\varepsilon(\tau_\varepsilon(\delta)),Y_\varepsilon(\tau_\varepsilon(\delta))}(t) - Y^+(\tau_\varepsilon(\delta)+t)\big|\Big) +$$

$$+ \sup_{t\in[0,T]} \Big(\big|X^+(\tau_\varepsilon(\delta)+t) - X^+(t)\big| + \big|Y^+(\tau_\varepsilon(\delta)+t) - Y^+(t)\big|\Big) +$$

$$+ \sup_{t\in[0,\tau_\varepsilon(\delta)]} \big(|X_\varepsilon(t) - x^0| + |Y_\varepsilon(t)|\big) =$$

$$= I_1 + \ldots + I_4.$$

Select small $\delta > 0$ and after that select $\varepsilon_0 > 0$ from (3.2). It follows from (3.3), (3.4), and construction of $(X^+, Y^+)$ in Subsection 2.1 that $I_2,\ I_3,\ I_4$ are small with high probability.

To estimate $I_1$ we need the following statement on integral equations. Let $f(t) = \big(f_X(t), f_Y(t)\big)$ be a nonrandom continuous function, and functions $X^{\pm,x,y}_{(f)},\ Y^{\pm,x,y}_{(f)}$ satisfy the integral equation

$$X^{\pm,x,y}_{(f)}(t) = x + \int\limits_0^t \psi^\pm\big(X^{\pm,x,y}_{(f)}(s), Y^{\pm,x,y}_{(f)}(s)\big)\,ds + f_X(t),$$

$$Y^{\pm,x,y}_{(f)}(t) = \int\limits_0^t \varphi^\pm\big(X^{\pm,x,y}_{(f)}(s), Y^{\pm,x,y}_{(f)}(s)\big)\big(Y^{\pm,x,y}_{(f)}\big)^\gamma(s)\,ds + f_Y(t), \quad t\in[0,T],$$

$$X^{\pm,x,y}_{(f)}(0) = x, \qquad Y^{\pm,x,y}_{(f)}(0) = y.$$

***Remark* 3.2.** We do not assume that a pair $X^{\pm,x,y}_{(f)},\ Y^{\pm,x,y}_{(f)}$ is a unique solution. Recall also that the domains of $\psi^\pm,\ \varphi^\pm$ is the whole space.

**Lemma 3.2.**

$$\forall\delta>0\quad \forall R\ge 1\quad \exists\alpha>0\quad \forall x\in[-R,R]\quad \forall t\in[0,T]\quad \forall f:\quad \|f\|_\infty<\alpha,$$

$$\forall y\in\left[\frac{1}{R}, R\right]\colon\ \big|X^{+,x,y}_{(f)}(t) - X^{+,x,y}(t)\big| + \big|Y^{+,x,y}_{(f)}(t) - Y^{+,x,y}(t)\big| \le \delta,$$

$$\forall y \in \left[-R, -\frac{1}{R}\right]\colon \left|X_{(f)}^{-,x,y}(t) - X^{-,x,y}(t)\right| + \left|Y_{(f)}^{-,x,y}(t) - Y^{-,x,y}(t)\right| \le \delta.$$

The proof of the lemma is standard. Notice that if $\alpha$ is small enough, then $Y_{(f)}^{\pm,x,y}(t) \neq 0$, $t \in [0,T]$, and coefficients of the integral equations are locally Lipschitz continuous if $y \neq 0$.

Let $\omega$ be such that $Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta$. Then

$$\left|X_\varepsilon(\tau_\varepsilon(\delta)+t) - X^{+,X_\varepsilon(\tau_\varepsilon(\delta)),Y_\varepsilon(\tau_\varepsilon(\delta))}(t)\right| + \left|Y_\varepsilon(\tau_\varepsilon(\delta)+t) - Y^{+,X_\varepsilon(\tau_\varepsilon(\delta)),Y_\varepsilon(\tau_\varepsilon(\delta))}(t)\right| =$$
$$= \left(\left|X_{(f)}^{+,x,\delta}(t) - X^{+,x,\delta}(t)\right| + \left|Y_{(f)}^{+,x,\delta}(t) - Y^{+,x,\delta}(t)\right|\right)\Big|_{x=X_\varepsilon(\tau_\varepsilon(\delta))},$$

where

$$f(t) := \left(\varepsilon \int\limits_{\tau_\varepsilon(\delta)}^{\tau_\varepsilon(\delta)+t} b\big(X_\varepsilon(s), Y_\varepsilon(s)\big)dB(s),\ \varepsilon\int\limits_0^t \beta\big(X_\varepsilon(s), Y_\varepsilon(s)\big)dW(s)\right).$$

Since $b$ and $\beta$ are bounded we have the uniform convergence in probability:

$$\varepsilon \sup_{t\in[0,T]}\left(\left|\int\limits_{\tau_\varepsilon(\delta)}^{\tau_\varepsilon(\delta)+t} b\big(X_\varepsilon(s), Y_\varepsilon(s)\big)dB(s)\right| + \left|\int\limits_0^t \beta\big(X_\varepsilon(s), Y_\varepsilon(s)\big)dW(s)\right|\right) \xrightarrow{\mathbf{P}} 0, \quad \varepsilon \to 0,$$

for any $\delta > 0$.

This, (3.3), (3.5), and Lemma 3.2 give us convergence

$$\sup_{t\in[0,T]}\Big(\left|X_\varepsilon(\tau_\varepsilon(\delta)+t) - X^{X_\varepsilon(\tau_\varepsilon(\delta)),Y_\varepsilon(\tau_\varepsilon(\delta))}(t)\right| +$$
$$+\left|Y_\varepsilon(\tau_\varepsilon(\delta)+t) - Y^{X_\varepsilon(\tau_\varepsilon(\delta)),Y_\varepsilon(\tau_\varepsilon(\delta))}(t)\right|\Big) \xrightarrow{\mathbf{P}} 0, \quad \varepsilon \to 0,$$

for any $\delta > 0$.

***Step* 5.** The proof of the theorem follows from Step 4 and the next estimate of probabilities $\mathbf{P}\big(Y_\varepsilon(\tau_\varepsilon(\delta)) = \pm\delta\big)$.

**Lemma 3.3.** *For all* $\mu > 0,\ \delta_0 > 0$ *exists to* $\delta \in (0,\delta_0)$ *such that*

$$p^+ - \mu \le \liminf_{\varepsilon\to 0} \mathbf{P}\big(Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta\big) \le \limsup_{\varepsilon\to 0} \mathbf{P}\big(Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta\big) \le p^+ + \mu,$$
$$p^- - \mu \le \liminf_{\varepsilon\to 0} \mathbf{P}\big(Y_\varepsilon(\tau_\varepsilon(\delta)) = -\delta\big) \le \limsup_{\varepsilon\to 0} \mathbf{P}\big(Y_\varepsilon(\tau_\varepsilon(\delta)) = -\delta\big) \le p^- + \mu,$$

*where* $p^\pm$ *are defined in* (2.2).

***Proof.*** Let $\nu > 0$ be arbitrary. Select $\delta_1 > 0$ such that

$$\left|\varphi^\pm(x,y) - \varphi^\pm(x^0,0)\right| \le \nu, \qquad 0 < \big(\beta^\pm(x^0,0)\big)^2 - \nu < \big(\beta^\pm(x,y)\big)^2 < \big(\beta^\pm(x^0,0)\big)^2 + \nu \quad (3.6)$$

as $|x - x^0| < \delta_1,\ |y| \in [0,\delta_1]$.

Set $\sigma_\varepsilon(\delta) := \inf\big\{t \ge 0\colon |X_\varepsilon(t) - x^0| \ge \delta\big\}$.

It follows from (3.3) that $\mathbf{P}\Big(\sigma_\varepsilon\big(\delta^{\frac{1-\gamma}{6}}\big) < \tau_\varepsilon(\delta)\Big) < C\delta^{\frac{1-\gamma}{6}}$ for small $\varepsilon$. Hence, if $\delta^{\frac{1-\gamma}{6}} < \delta_1$, then with probability greater than $1 - C\delta^{\frac{1-\gamma}{6}}$ the process $Y_\varepsilon$ exits $[-\delta,\delta]$ before $X_\varepsilon$ exits $[-\delta_1,\delta_1]$. Therefore, without loss of generality we will assume that (3.6) is satisfied for all $(x,y)$.

Set

$$s_\varepsilon(y) := \begin{cases} \displaystyle\int_0^y \exp\Big\{-\frac{2(\varphi^+(x^0,0)+\nu)z^{\gamma+1}}{\varepsilon^2(\gamma+1)(\beta^+(x^0,0)^2-\nu)}\Big\}dz, & y \ge 0,\\[2ex] \displaystyle\int_0^y \exp\Big\{-\frac{2(\varphi^-(x^0,0)-\nu)|z|^{\gamma+1}}{\varepsilon^2(\gamma+1)(\beta^-(x^0,0)^2+\nu)}\Big\}dz, & y \le 0.\end{cases}$$

Then

$$\varphi(x,y)y^\gamma s_\varepsilon'(y) + \frac{\varepsilon^2}{2}\beta^2(x,y)s_\varepsilon''(y) \le 0$$

for all $x,\ y$ (recall that we assume that (3.6) is satisfied for all $(x,y)$).

So,

$$0 \ge \mathbf{E}s_\varepsilon\big(Y_\varepsilon(\tau_\varepsilon(\delta))\big) = s_\varepsilon(\delta)\mathbf{P}\big(Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta\big) + s_\varepsilon(-\delta)\mathbf{P}\big(Y_\varepsilon(\tau_\varepsilon(\delta)) = -\delta\big) =$$

$$= s_\varepsilon(\delta)\mathbf{P}\big(Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta\big) + s_\varepsilon(-\delta)\big(1 - \mathbf{P}\big(Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta\big)\big) =$$

$$= \big(s_\varepsilon(\delta) - s_\varepsilon(-\delta)\big)\mathbf{P}\big(Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta\big) + s_\varepsilon(-\delta).$$

Therefore,

$$\limsup_{\varepsilon\to 0}\mathbf{P}\big(Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta\big) \le \lim_{\varepsilon\to 0}\frac{-s_\varepsilon(-\delta)}{s_\varepsilon(\delta) - s_\varepsilon(-\delta)} =$$

$$= \lim_{\varepsilon\to 0}\frac{\displaystyle\int_{-\delta}^0 \exp\Big\{-\frac{2(\varphi^-(x^0,0)-\nu)|z|^{\gamma+1}}{\varepsilon^2(\gamma+1)((\beta^-(x^0,0))^2+\nu)}\Big\}dz}{\displaystyle\int_{-\delta}^0 \exp\Big\{-\frac{2(\varphi^-(x^0,0)-\nu)|z|^{\gamma+1}}{\varepsilon^2(\gamma+1)((\beta^-(x^0,0))^2+\nu)}\Big\}dz + \int_0^\delta \exp\Big\{-\frac{2(\varphi^+(x^0,0)+\nu)z^{\gamma+1}}{\varepsilon^2(\gamma+1)\big(\big(\beta^+(x^0,0)\big)^2-\nu\big)}\Big\}dz} =$$

$$= \frac{\left(\dfrac{\varphi^+(x^0,0)+\nu}{\big(\beta^+(x^0,0)\big)^2-\nu}\right)^{\frac{1}{\gamma+1}}}{\left(\dfrac{\varphi^-(x^0,0)-\nu}{(\beta^-(x^0,0))^2+\nu}\right)^{\frac{1}{\gamma+1}} + \left(\dfrac{\varphi^+(x^0,0)+\nu}{\big(\beta^+(x^0,0)\big)^2-\nu}\right)^{\frac{1}{\gamma+1}}}.$$

Here we used the following:

$$\int\limits_0^\delta \exp\Big\{-\frac{Az^{\gamma+1}}{\varepsilon^2}\Big\}dz = \frac{1}{1+\gamma}\Big(\frac{\varepsilon^2}{A}\Big)^{\frac{1}{1+\gamma}}\int\limits_0^{\frac{A\delta^{\gamma+1}}{\varepsilon^2}} e^{-t}t^{\frac{1}{1+\gamma}-1}dt \sim$$

$$\sim \frac{1}{1+\gamma}\left(\frac{\varepsilon^2}{A}\right)^{\frac{1}{1+\gamma}}\Gamma\left(\frac{1}{1+\gamma}\right), \quad \varepsilon\to 0,$$

for any $A > 0,\ \delta > 0.$

Since $\nu$ was arbitrary, this completes the proof of Lemma 3.3 and Theorem 2.1.

**4. Proof of Theorem 2.2.** At the beginning notice that

$$Y_\varepsilon \Rightarrow 0, \quad \varepsilon\to 0. \tag{4.1}$$

Indeed, by Itô's formula we have

$$Y_\varepsilon^2(t) \le C\varepsilon^2 + 2\varepsilon\int_0^t Y_\varepsilon(s)\beta\big(X_\varepsilon(s), Y_\varepsilon(s)\big)\,dW(s),$$

where $C$ is independent of $\varepsilon$. Hence we get an estimate

$$\sup_{t\in[0,T]} \mathbf{E}Y_\varepsilon^2(t) \le C\varepsilon^2 \quad \forall\varepsilon>0.$$

It follows from the Doob inequality that

$$\sup_{t\in[0,T]}\left|2\varepsilon\int_0^t Y_\varepsilon(s)\beta\big(X_\varepsilon(s), Y_\varepsilon(s)\big)\,dW(s)\right| \xrightarrow{\mathbf{P}} 0, \quad \varepsilon\to 0.$$

This completes the proof of (4.1).

Let $\delta > 0$ be a fixed number. Notice that

$$\varepsilon^{-\delta}Y_\varepsilon(t) = \varepsilon^{\delta(\gamma-1)}\int_0^t \varphi\big(X_\varepsilon(s), Y_\varepsilon(s)\big)\big(\varepsilon^{-\delta}Y_\varepsilon(t)\big)^\gamma\,dt +$$

$$+\varepsilon^{1-\delta-\frac{\delta(\gamma-1)}{2}}\varepsilon^{\frac{\delta(\gamma-1)}{2}}\int_0^t \beta\big(X_\varepsilon(s), Y_\varepsilon(s)\big)\,dW(s) =$$

$$=\int_0^t \varphi\big(X_\varepsilon(s), \varepsilon^\delta\varepsilon^{-\delta}Y_\varepsilon(s)\big)\big(\varepsilon^{-\delta}Y_\varepsilon(t)\big)^\gamma\,d\big(\varepsilon^{\delta(\gamma-1)}t\big)+$$

$$+\varepsilon^{1-\frac{\delta(\gamma+1)}{2}}\int_0^t \beta\big(X_\varepsilon(s), \varepsilon^\delta\varepsilon^{-\delta}Y_\varepsilon(s)\big)\,dW_\varepsilon\big(\varepsilon^{\delta(\gamma-1)}s\big),$$

where $W_\varepsilon(t) = \varepsilon^{\frac{\delta(\gamma-1)}{2}}W\big(\varepsilon^{-\delta(\gamma-1)}t\big)$ is a Wiener process.

If $1-\dfrac{\delta(\gamma+1)}{2} = 0$, i.e., $\delta = \dfrac{2}{\gamma+1}$, then the process $\tilde{Y}_\varepsilon(t) := \varepsilon^{-\delta}Y_\varepsilon(t) = \varepsilon^{\frac{-2}{\gamma+1}}Y_\varepsilon(t)$ satisfies the SDE

$$\tilde{Y}_\varepsilon(t) = \int_0^t \varphi\big(X_\varepsilon(s), \varepsilon^{\frac{2}{\gamma+1}}\tilde{Y}_\varepsilon(s)\big)\tilde{Y}_\varepsilon^\gamma(s)\, d\Big(\varepsilon^{\frac{2(\gamma-1)}{\gamma+1}} s\Big) + \int_0^t \beta\big(X_\varepsilon(s), \varepsilon^{\frac{2}{\gamma+1}}\tilde{Y}_\varepsilon(s)\big)\, dW_\varepsilon\Big(\varepsilon^{\frac{2(\gamma-1)}{\gamma+1}} s\Big).$$

Set $\tilde{\varepsilon} = \varepsilon^{\frac{2(1-\gamma)}{\gamma+1}}$. Therefore,

$$\begin{aligned} dX_\varepsilon(t) &= a_{\tilde{\varepsilon}}\big(X_\varepsilon(t), \tilde{Y}_\varepsilon(t)\big)\, dt + b_{\tilde{\varepsilon}}\big(X_\varepsilon(t), \tilde{Y}_\varepsilon(t)\big)\, dB(t), \\ d\tilde{Y}_\varepsilon(t) &= \alpha_{\tilde{\varepsilon}}(X_\varepsilon(t), \tilde{Y}_\varepsilon(t))\, d\tilde{\varepsilon}^{-1}t + \beta_{\tilde{\varepsilon}}\big(X_\varepsilon(t), \tilde{Y}_\varepsilon(t)\big)\, dW_{\tilde{\varepsilon}}\big(\tilde{\varepsilon}^{-1}t\big), \end{aligned} \tag{4.2}$$

where

$$a_{\tilde{\varepsilon}}(x,y) = \psi\big(x, \varepsilon^{\frac{2}{\gamma+1}} y\big) = \psi\big(x, \tilde{\varepsilon}^{\frac{1}{1-\gamma}} y\big),$$

$$b_{\tilde{\varepsilon}}(x,y) = \varepsilon b\big(x, \varepsilon^{\frac{2}{\gamma+1}} y\big) = \tilde{\varepsilon}^{(\gamma+1)/2(1-\gamma)} b\big(x, \tilde{\varepsilon}^{\frac{1}{1-\gamma}} y\big),$$

$$\alpha_{\tilde{\varepsilon}}(x,y) = \varphi\big(x, \varepsilon^{\frac{2}{\gamma+1}} y\big) y^\gamma = \varphi\big(x, \tilde{\varepsilon}^{\frac{1}{1-\gamma}} y\big) y^\gamma,$$

$$\beta_{\tilde{\varepsilon}}(x,y) = \beta\big(x, \varepsilon^{\frac{2}{\gamma+1}} y\big) = \beta\big(x, \tilde{\varepsilon}^{\frac{1}{1-\gamma}} y\big).$$

We see that the system (4.2) has the form (2.4). Let us apply Theorem 2.3, where $k = 1$,

$$a^\varepsilon(x,y) := a_{\tilde{\varepsilon}}(x,y), \qquad \sigma^\varepsilon(x,y) = c^\varepsilon(x,y,u) = C^\varepsilon(x,y,z) = 0,$$

$$A^\varepsilon(x,y) := \alpha_{\tilde{\varepsilon}}(x,y), \qquad \Sigma^\varepsilon(x,y) := \beta_{\tilde{\varepsilon}}(x,y),$$

$$\xi_\varepsilon(t) := \tilde{\varepsilon}^{(\gamma+1)/2(1-\gamma)} \int_0^t b(X_\varepsilon(s), \tilde{\varepsilon}^{\frac{1}{1-\gamma}} \tilde{Y}_{\tilde{\varepsilon}}(s)) dB(s).$$

Conditions $\mathbf{H_0}$, $\mathbf{H_1}$, and $\mathbf{H_2}$ are obviously true.

Functions from condition $\mathbf{H}_3$ are

$$A(x,y) = \big(\varphi^+(x,0)\mathbb{1}_{y>0} + \varphi^-(x,0)\mathbb{1}_{y<0}\big) y^\gamma,$$

$$\Sigma(x,y) = \beta^+(x,0)\mathbb{1}_{y\geq 0} + \beta^-(x,0)\mathbb{1}_{y<0}, \quad C(x,y,z) = 0.$$

Note that $A$ and $\Sigma$ are discontinuous at $y = 0$ and formally $\mathbf{H}_3$ is not satisfied. However, the only place in Theorem 2.3, where we used the continuity of $A$ and $\Sigma$, is the identification of limit points for the sequence $\{y_{\varepsilon_n}\}$ in the proof of Proposition 5.2. Since diffusion coefficients $\beta^\pm$ are separated from zero functions, it can be seen that processes $\{y_{\varepsilon_n}\}$ spend small time in small neighborhoods of 0 uniformly on $\{\varepsilon_n\}$. This yields that any limit point of $\{y_{\varepsilon_n}\}$ solves (2.8) and Proposition 5.2 holds true. This is all we need for Theorem 2.3 application.

Condition $\mathbf{H}_4$ is satisfied with $\nu = \mu = 0$.

Without loss of generality we will assume that

$$\varphi^\pm(x,0) \leq c < 0 \quad \text{for all} \quad x \in \mathbb{R}^d, \tag{4.3}$$

where $c$ is a constant. The general case can be considered using a localization. Hence, condition $\mathbf{H_5}$ is satisfied with $\kappa = \gamma$.

Consider equation with frozen coefficients

$$dy^{(x)}(t) = \big(\varphi^+(x,0)1\!\!1_{y^{(x)}(t)>0} + \varphi^-(x,0)1\!\!1_{y^{(x)}(t)<0}\big)(y^{(x)}(t))^\gamma\, dt+$$
$$+\Big(\beta^+(x,0)1\!\!1_{y^{(x)}(t)\geq 0} + \beta^-(x,0)1\!\!1_{y^{(x)}(t)<0}\Big)dW(t). \tag{4.4}$$

Existence and uniqueness of a weak solution to equation with frozen coefficients, and the strong Markov property follows from [10]. Hence, condition $\mathbf{F_0}$ holds true.

To verify condition $\mathbf{F_1}$, we modify the argument from [16] (Section 3.3.2). Because the diffusion coefficient in (4.4) is discontinuous, we do not have a good reference to state that the transition probability density $p^{(x)}(t,y,y')$ is continuous in $x,\ y,\ y'$. In order to overcome this minor difficulty we use the following localization argument. Consider the SDE

$$dy^{(x,+)}(t) = \varphi^+(x,0)\big(|y^{(x,+)}(t)|\wedge 2\big)^\gamma \operatorname{sgn}\big(y^{(x,+)}(t)\big)\, dt + \beta^+(x,0)\, dW(t). \tag{4.5}$$

This is an SDE with a constant diffusion coefficient and bounded and Hölder continuous drift coefficient, hence the standard analytic theory (see, e.g., [11]) yields that its transition probability density $p^{(x,+)}(t,y,y')$ is continuous in $x,\ y,\ y'$. Then for $y_0 = 1$ and every $t_0 > 0$ it holds that

$$\sup_{|x|\leq R}\big\|P^{(x,+)}_{t_0}(y,dy') - P^{(x,+)}_{t_0}(y_0,dy')\big\|_{TV} =$$
$$= \sup_{|x|\leq R}\int \big|p^{(x,+)}_{t_0}(y,y') - p^{(x,+)}_{t_0}(y_0,y')\big|\, dy' \to 0, \quad y\to y_0.$$

The coefficients of the equations (4.4), (4.5) coincide on $[0,2]$, and thus the laws of the solutions to these equations, stopped at the moment of exit from $[0,2]$, coincide. Taking $t_0$ small enough, we can guarantee that each of these solutions stay in $[0,2]$ up to the time $t_0$ with probability $\geq \dfrac{5}{6}$ if the initial value $y$ stays in $\left[\dfrac{1}{2},\dfrac{3}{2}\right]$. By the coupling characterization of the TV distance (the "Coupling lemma", e.g., [16], Theorem 2.2.2), this yields that, for such $t_0$,

$$\sup_{y_1,y_2\in[\frac12,\frac32],|x|\leq R}\Big\|P^{(x)}_{t_0}(y_1,dy') - P^{(x,+)}_{t_0}(y_2,dy')\Big\|_{TV} \leq \left(1-\frac{5}{6}\right)+\left(1-\frac{5}{6}\right) = \frac{1}{3}.$$

Combining these two estimates we see that there exist $t_0 > 0$ and $r > 0$ small enough, so that

$$\sup_{y_1,y_2\in[1-r,1+r],|x|\leq R}\Big\|P^{(x)}_{t_0}(y_1,dy') - P^{(x)}_{t_0}(y_2,dy')\Big\|_{TV} < \frac{3}{4};$$

in the right-hand side we could actually take any number $> \dfrac{1}{3}+\dfrac{1}{3}=\dfrac{2}{3}$. This proves the local Dobrushin condition in a small ball centered at $y_0 = 1$. To extend this condition to a large ball $|y|\leq R$, we use another standard argument, based on the support theorem. Namely, $y^{(x)}$ can be represented as an image of a Brownian motion under the time change and the change of measure (see [13]). Since the Wiener measure in $C_0(0,\infty)$ has a full topological support, it is easy to show using this representation that, for any $t_1 > 0$, there exists $\delta > 0$ such that

$$P^{(x)}_{t_1}\big(y,[1-r,1+r]\big)\geq\delta, \qquad |x|\leq R, \quad |y|\leq R.$$

Take $h = t_0 + t_1$ and for $x,\ y_1,\ y_2$ with $|x| \le R,\ |y_1| \le R,\ |y_2| \le R$ consider two processes $Y_t^1$, $Y_t^2$ which start at $y_1,\ y_2$, respectively, solve (4.4) independently up to the time $t_1$, and then provide the maximal coupling probability on the time interval $[t_1, t_1 + t_0]$, conditioned on their values at the time $t_1$ (we can construct such a process using the Coupling lemma for probability kernels [16], Theorem 2.2.4). Then

$$\left\| P_h^{(x)}(y_1, dy') - P_h^{(x)}(y_2, dy') \right\|_{TV} \le \mathbf{P}\big(Y_h^1 \ne Y_h^2\big) =$$

$$= \int_{\mathbf{R}^2} 4 \left\| P_{t_0}^{(x)}(z_1, dy') - P_{t_0}^{(x)}(z_2, dy') \right\|_{TV} P_h^{(x)}(y_1, dz_1) P_h^{(x)}(y_2, dz_2) \le$$

$$\le 1 - P_{t_1}^{(x)}\big(y_1, [1-r, 1+r]\big) P_{t_1}^{(x)}\big(y_2, [1-r, 1+r]\big) +$$

$$+ \int_{[1-r,1+r]^2} \left\| P_{t_0}^{(x)}(z_1, dy') - P_{t_0}^{(x)}(z_2, dy') \right\|_{TV} P_h^{(x)}(y_1, dz_1) P_h^{(x)}(y_2, dz_2) \le$$

$$\le 1 + \left(-1 + \frac{3}{4}\right) P_{t_1}^{(x)}\big(y_1, [1-r, 1+r]\big) P_{t_1}^{(x)}\big(y_2, [1-r, 1+r]\big) \le$$

$$\le 1 - \frac{\delta^2}{4}$$

for any $|x| \le R,\ |y_1| \le R,\ |y_2| \le R$, which completes the proof of $\mathbf{F_1}$.

The IPM $\pi^{(x)}(dy)$ equals (see [14], Exercise 5.40):

$$\pi^{(x)}(dy) = c(x) \left( \exp\left\{ \frac{\varphi^+(x,0)}{(\beta^+(x,0))^2} \frac{y^{\gamma+1}}{(\gamma+1)} \right\} \mathbb{1}_{y \ge 0} + \exp\left\{ \frac{\varphi^-(x,0)}{(\beta^-(x,0))^2} \frac{y^{\gamma+1}}{(\gamma+1)} \right\} \mathbb{1}_{y<0} \right) dy,$$

where

$$c(x)^{-1} = \int_{\mathbb{R}} \left( \exp\left\{ \frac{\varphi^+(x,0)}{(\beta^+(x,0))^2} \frac{y^{\gamma+1}}{(\gamma+1)} \right\} \mathbb{1}_{y \ge 0} + \exp\left\{ \frac{\varphi^-(x,0)}{(\beta^-(x,0))^2} \frac{y^{\gamma+1}}{(\gamma+1)} \right\} \mathbb{1}_{y<0} \right) dy =$$

$$= \frac{\Gamma(\frac{1}{\gamma+1})}{(\gamma+1)} \left( \left( \frac{(\gamma+1)(\beta^+(x,0))^2}{\varphi^+(x,0)} \right)^{\frac{1}{\gamma+1}} + \left( \frac{(\gamma+1)(\beta^-(x,0))^2}{\varphi^-(x,0)} \right)^{\frac{1}{\gamma+1}} \right).$$

Condition $\mathbf{H_6}$ is satisfied with $a(x,y) = \psi_+(x,0)\mathbb{1}_{y>0} + \psi_-(x,0)\mathbb{1}_{y<0},\ \sigma(x,y) = c(x,y,z) = 0$, and $B = \mathbb{R}^d \times \{0\}$.

The averaged coefficient

$$\overline{a}(x) = \psi^+(x,0)\pi^{(x)}([0,\infty)) + \psi^-(x,0)\pi^{(x)}((-\infty,0)) =$$

$$= \psi^+(x,0) \frac{\left( \frac{(\beta^+(x,0))^2}{\varphi^+(x,0)} \right)^{\frac{1}{\gamma+1}}}{\left( \frac{(\beta^-(x,0))^2}{\varphi^-(x,0)} \right)^{\frac{1}{\gamma+1}} + \left( \frac{(\beta^+(x,0))^2}{\varphi^+(x,0)} \right)^{\frac{1}{\gamma+1}}} +$$

$$+\psi^-(x,0)\frac{\left(\dfrac{(\beta^-(x,0))^2}{\varphi^-(x,0)}\right)^{\frac{1}{\gamma+1}}}{\left(\dfrac{(\beta^-(x,0))^2}{\varphi^-(x,0)}\right)^{\frac{1}{\gamma+1}}+\left(\dfrac{(\beta^+(x,0))^2}{\varphi^+(x,0)}\right)^{\frac{1}{\gamma+1}}}$$

is Lipschitz continuous, $\bar{b}(x)=0,\ \overline{K}_{(\rho)}(x,dv)=\overline{K}^{(\rho)}(x,dv)=0$. So, condition $\mathbf{A_0}$ holds true and the corresponding martingale problem has a unique solution.

This with (4.1) concludes the proof.

**5. Proof of Theorem 2.3.** The weak compactness of the family $\{X_\varepsilon,\varepsilon>0\}$ in $\mathbb{D}([0,\infty),\mathbb{R}^d)$ follows, in a standard way, from the negligibility assumption $\mathbf{H_0}$ and the boundedness assumptions $\mathbf{H}_1,\ \mathbf{H}_2$. Under the assumptions of the theorem, for any $C_0^\infty$-function $\varphi$ the function $L\varphi$ is continuous and bounded. Hence, in order to prove that any weak limit point of the family $\{X_\varepsilon,\varepsilon>0\}$ as $\varepsilon\to0$ solves the martingale problem (2.9), it is enough to show that, for any $C_0^\infty$-function $\varphi$, any $s_1,\ldots,s_q<s<t$, and any continuous and bounded function $\Phi\colon\mathbb{R}^{d\times q}\to\mathbb{R}$

$$\mathbf{E}^\varepsilon\Phi\big(X_\varepsilon(s_1),\ldots,X_\varepsilon(s_q)\big)\left[\varphi\big(X_\varepsilon(t)\big)-\varphi(X_\varepsilon(s))-\int_s^t L\varphi(X_\varepsilon(r))\,dr\right]\to0,\quad\varepsilon\to0,\tag{5.1}$$

we denote by $\mathbf{E}^\varepsilon$ the expectation w.r.t. $\mathbf{P}^\varepsilon$. Denote

$$\widetilde{X}_\varepsilon(t)=X_\varepsilon(t)-\xi_\varepsilon(t)=$$

$$=X_\varepsilon(0)+\int_0^t a^\varepsilon\big(X_\varepsilon(s),Y_\varepsilon(t)\big)\,ds+\int_0^t\sigma^\varepsilon\big(X_\varepsilon(s),Y_\varepsilon(s)\big)\,dB_s^\varepsilon+$$

$$+\int_0^t\int_{\mathbb{R}^m}c^\varepsilon\big(X_\varepsilon(s-),Y_\varepsilon(s-),u\big)\Big[N^\varepsilon(du,ds)-1_{|u|\le\rho}\nu^\varepsilon(du)ds\Big].\tag{5.2}$$

Observe that $L\varphi$ is a bounded and continuous function. So, by $\mathbf{H_0}$ relation (5.1) is equivalent to

$$\mathbf{E}^\varepsilon\Phi\big(\widetilde{X}_\varepsilon(s_1),\ldots,\widetilde{X}_\varepsilon(s_q)\big)\left[\varphi\big(\widetilde{X}_\varepsilon(t)\big)-\varphi(\widetilde{X}_\varepsilon(s))-\int_s^t L\varphi(\widetilde{X}_\varepsilon(r))\,dr\right]\to0,\quad\varepsilon\to0.\tag{5.3}$$

Denote

$$b^\varepsilon(x,y)=\sigma^\varepsilon(x,y)\big(\sigma^\varepsilon(x,y)\big)^*,\qquad K_{(\rho)}^\varepsilon(x,y,A)=\nu^\varepsilon\big(\{u\colon|u|\le\rho,c^\varepsilon(x,y,u)\in A\}\big),$$

$$K^{(\rho),\varepsilon}(x,y,A)=\nu^\varepsilon\big(\{u\colon|u|>\rho,c^\varepsilon(x,y,u)\in A\}\big),$$

and

$$\mathcal{L}^\varepsilon\varphi(x,y)=\nabla\varphi(x)\cdot a^\varepsilon(x,y)+\frac12\nabla^2\varphi(x)\cdot b^\varepsilon(x,y)+$$

$$+\int_{\mathbb{R}^m}\Big(\varphi(x+v)-\varphi(x)-\nabla\varphi(x)\cdot v\Big)K_{(\rho)}^\varepsilon(x,dv)+$$

$$+\int_{\mathbb{R}^m}\Big(\varphi(x+v)-\varphi(x)\Big)K^{(\rho),\varepsilon}(x,y,dv).$$

Then, by the Itô formula, we have

$$\varphi\big(\widetilde{X}_\varepsilon(t)\big)-\varphi(\widetilde{X}_\varepsilon(s))=\int_s^t\mathcal{L}^\varepsilon\varphi(X_\varepsilon(r),Y_\varepsilon(r))\,dr+(\text{martingale part}).\tag{5.4}$$

Applying $\mathbf{H_0}$ once again, we get that, to prove (5.1) and (5.3), it is enough to prove, for any $s_1,\ldots,s_q<t$,

$$\mathbf{E}^\varepsilon\Phi\big(X_\varepsilon(s_1),\ldots,X_\varepsilon(s_q)\big)\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big)-L\varphi\big(X_\varepsilon(t)\big)\Big)\to 0,\quad \varepsilon\to 0.\tag{5.5}$$

Before proving (5.5), we formulate and prove two auxiliary statements.

***5.1. Auxiliaries, I: uniform ergodic rate for the frozen microscopic dynamics.***

**Proposition 5.1.** *Let conditions $\boldsymbol{H_1}$–$\boldsymbol{H_5}$, $\boldsymbol{F_0}$, $\boldsymbol{F_1}$ hold. If $\kappa\in(0,1)$ and $p>0$ are from these conditions, then for every $R>0$ there exists $C$ such that for any $x$, $y$ with $|x|\le R$, $|y|\le R$*

$$\big\|P_t^{(x)}(y,dy')-\pi^{(x)}(dy')\big\|_{TV}\le Ct^{-\frac{p+\kappa-1}{1-\kappa}}.\tag{5.6}$$

*If $\kappa\ge 1$, then there exists $a>0$ such that, for every $R>0$ and any $x,y$ with $|x|\le R$, $|y|\le R$,*

$$\big\|P_t^{(x)}(y,dy')-\pi^{(x)}(dy')\big\|_{TV}\le Ce^{-at}$$

*with a constant $C$ depending on $R$.*

***Proof.*** The required statement is actually obtained, though not in this precise form, in [16] (Section 3). The difference between the current situation and the one studied in [16] is that the ergodic rates were obtained there for individual processes (while here we have a family indexed by $x$) and separately for diffusions and Lévy driven SDEs (while here we have both types of the noise involved simultaneously). This difference is not crucial, and we just give a short outline of the argument, referring to [16] for details.

The convergence conditions $\mathbf{H_3}$, $\mathbf{H_4}$ yield that the bounds from the conditions $\mathbf{H_1}$, $\mathbf{H_2}$ and the drift condition $\mathbf{H_5}$ remain true for the limiting coefficients $A(x,y)$, $\Sigma(x,y)$, $C(x,y,z)$ and Lévy measure $\mu(du)$. Then we have the following: if $V\in C^2$ is a function such that $V(y)\ge 1$ and $V(y)=|y|^p$, $|y|\ge 2$, then, for any $x\in\mathbb{R}^d$, the semimartingale decomposition holds

$$V(y^{(x)}(t))=V\big(y^{(x)}(0)\big)+\int_0^t\mathcal{A}V\big(x,y^{(x)}(s)\big)\,ds+(\text{martingale part}),\tag{5.7}$$

where the function $\mathcal{A}V(x,y)$ satisfies

$$\mathcal{A}V(x,y)\le\begin{cases}C_V-a_VV(y)^{\frac{p+\kappa-1}{p}}, & \kappa\in(0,1),\\ C_V-a_VV(y), & \kappa\ge 1,\end{cases}\tag{5.8}$$

with some constants $C_V$, $a_V>0$. For the proof of this statement, see [17] (Proposition 2.5).

Given (5.7), (5.8) we can proceed analogously to [16] (Sections 3.3, 3.4). Namely, for $\kappa \in (0,1)$ we use [16] (Theorem 3.2.3) and [16] (Example 3.2.6) to show that

$$\mathbf{E}_y \widetilde{V}\big(y^{(x)}(h)\big) - \widetilde{V}(y) \le \widetilde{C}_V - \widetilde{c}_V \widetilde{V}(y)^{\frac{p+\kappa-1}{p}}, \tag{5.9}$$

where $h$ is the same as in the assumption $\mathbf{F_1}$, $\widetilde{C}_V,\ \widetilde{c}_V > 0$ are some new constants, and $\widetilde{V}$ is a new function which is equivalent to $V$ in the sense that, for some positive constants $c_1,\ c_2$

$$c_1 V \le \widetilde{V} \le c_2 V.$$

Following the proof of [16] (Theorem 3.2.3) and calculations of [16] (Example 3.2.6) line by line, we easily see that, because the constants $C_V,\ c_V$ in (5.8) do not depend on $x$, the constants $\widetilde{C}_V,\ \widetilde{c}_V$, $c_1, c_2$ and the function $\widetilde{V}$ can be chosen uniformly for $x \in \mathbb{R}^d$.

Inequality (5.9) is actually the *Lyapunov condition* for the *skeleton chain* $y_k^{(x,h)} = y^{(x)}(kh)$, $k \ge 0$, for the process $y^{(x)}$, see [16] (Section 2.8). Combined with the local Dobrushin condition assumed in $\mathbf{F_1}$, we get by [16] (Corollary 2.8.10) the inequality

$$\Big\| P_{kh}^{(x)}(y, dy') - \pi^{(x)}(dy') \Big\|_{TV} \le C(1+k)^{-\frac{p+\kappa-1}{1-\kappa}} \widetilde{V}(y), \quad \kappa \in (0,1),$$

where we have used the identity

$$\frac{p+\kappa-1}{p}\left(1 - \frac{p+\kappa-1}{p}\right)^{-1} = \frac{p+\kappa-1}{1-\kappa}.$$

Since Lyapunov condition and the local Dobrushin condition are uniform in $x$, the constant $C$ here can be chosen uniformly for $x \in \mathbb{R}^d$; one can easily check this following line by line the proofs of [16] (Corollary 2.8.10) and the theorems it is based on: [16] (Theorems 2.7.5 and 2.8.6). Since the total variation distance $\big\|P_t^{(x)}(y, dy') - \pi^{(x)}(dy')\big\|_{TV}$ is nonincreasing in $t$ and $V(y)$ is locally bounded, this completes the proof of the required statement in the case $\kappa \in (0,1)$.

For $\kappa \ge 1$, we can argue in a completely analogous way, using [16] (Corollary 2.8.3).

***5.2. Auxiliaries, II: weak convergence of the microscopic dynamics to the frozen one.*** Consider the following *microscopic* analogue of (2.4). Assume that $(x_\varepsilon, y_\varepsilon)$ is a solution (maybe nonunique) to the equations

$$\begin{gathered} x_\varepsilon(t) = x_\varepsilon(0) + \varepsilon \int_0^t a^\varepsilon\big(x_\varepsilon(s), y_\varepsilon(t)\big)\, ds + \varepsilon^{1/2} \int_0^t \sigma^\varepsilon\big(x_\varepsilon(s), y_\varepsilon(s)\big)\, db_s^\varepsilon + \\ + \int_0^t \int_{\mathbb{R}^m} C^\varepsilon\big(x_\varepsilon(s-), y_\varepsilon(s-), u\big) \Big[n^\varepsilon(du, ds) - 1_{|u|\le\rho} \varepsilon \nu^\varepsilon(du) ds\Big] + \zeta^\varepsilon(t), \\ y_\varepsilon(t) = y_\varepsilon(0) + \int_0^t A^\varepsilon\big(x_\varepsilon(s), y_\varepsilon(s)\big)\, ds + \int_0^t \Sigma^\varepsilon\big(x_\varepsilon(s), y_\varepsilon(s)\big)\, dw_s^\varepsilon + \\ + \int_0^t \int_{\mathbb{R}^l} c^\varepsilon\big(x_\varepsilon(s-), y_\varepsilon(s-), z\big) \Big[q^\varepsilon(dz, ds) - 1_{|z|\le\rho} \mu^\varepsilon(dz) ds\Big], \end{gathered} \tag{5.10}$$

where $b_t^\varepsilon,\ w_t^\varepsilon$ are Brownian motions and $n^\varepsilon(du,dt),\ q^\varepsilon(dz,dt)$ are Poisson point measures on a common filtered probability space $\big(\widetilde\Omega^\varepsilon,\widetilde{\mathcal F}^\varepsilon,\widetilde{\mathbf P}^\varepsilon\big)$, and the random measures $n^\varepsilon(du,dt),\ q^\varepsilon(dz,dt)$ have the intensity measures $\varepsilon\nu^\varepsilon(du)dt$ and $\mu^\varepsilon(dz)dt$, respectively, $\zeta^\varepsilon(t)$ is an adapted càdlàg process.

System (5.10) naturally appears, e.g., if we consider the original system (2.4) at the "microscopic time scale" $\varepsilon t$ with an initial time shift by $t_0$:

$$x_\varepsilon(t)=X_\varepsilon(t_0+\varepsilon t),\qquad y_\varepsilon(t)=Y_\varepsilon(t_0+\varepsilon t),\qquad\text{and}\qquad \zeta^\varepsilon(t)=\xi_\varepsilon(t_0+\varepsilon t)-\xi_\varepsilon(t_0).\quad(5.11)$$

For a fixed pair of functions $(\rho(\varepsilon),\varrho(\varepsilon))$ such that $\rho(\varepsilon)\to0$ and $\varrho(\varepsilon)\to0$ as $\varepsilon\to0$, and constants $R>0,\ T>0$ denote by $\mathcal K(\rho,\varrho,R,T)$ the class of all families $\big\{(x_\varepsilon,y_\varepsilon),\varepsilon>0\big\}$ which satisfy (5.10) on *some* probability space with nonrandom initial values $x_\varepsilon(0),\ y_\varepsilon(0),\ |x_\varepsilon(0)|\le R$, $|y_\varepsilon(0)|\le R$ and

$$\widetilde{\mathbf P}^\varepsilon\left(\sup_{s\le T}|\zeta_\varepsilon(s)|>\rho(\varepsilon)\right)\le\varrho(\varepsilon).$$

**Proposition 5.2.** *Let conditions* $\boldsymbol H_1$–$\boldsymbol H_5$, $\boldsymbol F_0$ *hold. Then, for any* $0<t<T$ *and any bounded continuous function* $f:\mathbb R^d\times\mathbb R^k\to\mathbb R$ *and* $R>0$,

$$\sup_{\{(x_\varepsilon,y_\varepsilon)\}\in\mathcal K(\rho,\varrho,R,T)}\left|\widetilde{\mathbf E}^\varepsilon f(x_\varepsilon(t),y_\varepsilon(t))-P_t^{\text{frozen}}f(x,y)\Big|_{x=x_\varepsilon(0),y=y_\varepsilon(0)}\right|\to0,\quad\varepsilon\to0,\qquad(5.12)$$

*where*

$$P_t^{\text{frozen}}f(x,y)=\int\limits_{\mathbb R^k}f(y')P_t^{(x)}(y,dy'),\quad t\ge0.$$

***Proof.*** Assuming the contrary, we will have that there exists a sequence $x_{\varepsilon_n}(\cdot),\ y_{\varepsilon_n}(\cdot)$ of solutions to (5.10) with $|x_{\varepsilon_n}(0)|\le R,\ |y_{\varepsilon_n}(0)|\le R$ such that

$$\left(\widetilde{\mathbf E}^{\varepsilon_n}f(x_{\varepsilon_n}(t),y_{\varepsilon_n}(t))-P_t^{(x)}f(x,y)\Big|_{x=x_{\varepsilon_n}(0),y=y_{\varepsilon_n}(0)}\right)\not\to0,\quad n\to\infty.\qquad(5.13)$$

Without loss of generality, after passing to a subsequence, we can assume that $x_{\varepsilon_n}(0)\to x_*$ and $y_{\varepsilon_n}(0)\to y_*$ as $n\to\infty$. Then it is easy to show that, for any $c>0$,

$$\lim_{n\to\infty}\widetilde{\mathbf P}^{\varepsilon_n}\left(\sup_{s\in[0,T]}|x_{\varepsilon_n}(s)-x_*|>c\right)=0.\qquad(5.14)$$

Next, denote by $\mathbf P^*$ the law in $\mathbf D\big([0,T],\mathbb R^k\big)$ of $y^{(x_*)}$ with $y^{(x_*)}(0)=y_*$. Since the $\mathbf P^*$-probability for $y(\cdot)$ to have a jump at the point $t$ is 0, the function $F(y(\cdot))=f(y(t))$ is a.s. continuous on $\mathbf D\big([0,T],\mathbb R^k\big)$. Thus, in order to prove that (5.13) fails, it is enough to show that the laws of $y_{\varepsilon_n}$, $n\ge1$, weakly converge in $\mathbf D\big([0,T],\mathbb R^k\big)$ to $\mathbf P^*$. Such a statement is quite standard, and we just outline its proof here.

By (5.14), the continuity assumption $\mathbf H_3$, and convergence of the noise $\mathbf H_4$ it is easy to prove that any weak limit point to $\{y_{\varepsilon_n}\}$ solves (2.8). By the weak uniqueness assumption $\mathbf F_0$, this yields that any weak limit point to $\{y_{\varepsilon_n}\}$ has the law $\mathbf P^*$.

That is, to prove the required weak convergence it is enough to prove that $\{y_{\varepsilon_n}\}$ is weakly compact in $\mathbf D\big([0,T],\mathbb R^k\big)$.

To prove the weak compactness, we use $L_2$-moment bounds for the increments of the process $y_{\varepsilon_n}$ combined with a truncation of the large jumps. Namely, by $\mathbf{H_2}$ for any fixed $\delta > 0$ there exists $Q_\delta$ such that

$$\widetilde{\mathbf{P}}^\varepsilon\Big(N\big([0,T]\times\{|z|>Q_\delta\}\big)>0\Big)<\delta, \quad \varepsilon>0.$$

Thus it is enough to prove weak compactness for every "truncated" family $\{y_{\varepsilon_n,Q}\}$, $Q>0$, where $y_{\varepsilon,Q}$ satisfies an analogue of (5.10) with the integral for $q^\varepsilon$ taken over $\{|z|\le Q\}$ instead of $\mathbb{R}^l$. For such a "truncated" family, applying [17] (Proposition 2.5) we get

$$\big|y_{\varepsilon_n,Q}(s)\big|^2=\big|y_{\varepsilon_n,Q}(0)\big|^2+\int\limits_0^s H(r)\,dr+(\text{martingale part}), \tag{5.15}$$

where $H$ is bounded. Combining this with the maximal martingale inequality, we get that

$$\widetilde{\mathbf{E}}^{\varepsilon_n}\sup_{s\in[0,T]}\big|y_{\varepsilon_n,Q}(s)\big|^2$$

is bounded. Since the coefficient $A^\varepsilon(x,y)$ is bounded locally in $y$, the above bound and the (uniform) bounds for $C^\varepsilon$, $\mu^\varepsilon$ from $\mathbf{H_1}$, $\mathbf{H_2}$ yield the required weak compactness of $\{y_{\varepsilon_n}\}$. Summarizing all the above, we have that $\{y_{\varepsilon_n}\}$ weakly converges to $\mathbf{P}^*$. Combined with (5.14), this contradicts to (5.13) and proves the required statement.

***5.3. End of the proof of Theorem* 2.3.** In this subsection we complete the proof of (5.5). This will conclude proof of the theorem. Denote

$$\mathcal{L}\varphi(x,y)=\nabla\varphi(x)\cdot a(x,y)+\frac{1}{2}\nabla^2\varphi(x)\cdot b(x,y)+$$

$$+\int\limits_{\mathbb{R}^m}\Big(\varphi(x+v)-\varphi(x)-\nabla\varphi(x)\cdot v\Big)K_{(\rho)}(x,y,dv)+$$

$$+\int\limits_{\mathbb{R}^m}\Big(\varphi(x+v)-\varphi(x)\Big)K^{(\rho)}(x,y,dv)=$$

$$=\nabla\varphi(x)\cdot a(x,y)+\frac{1}{2}\nabla^2\varphi(x)\cdot b(x,y)+$$

$$+\int\limits_{\mathbb{R}^m}\Big(\varphi(x+c(x,y,u))-\varphi(x)-\nabla\varphi(x)\cdot c(x,y,u)1\!\!1_{|u|\le\rho}\Big)\nu(du). \tag{5.16}$$

Next, since the set $B$ from the condition $\mathbf{H_6}$ is open, there exists a sequence of continuous functions $\chi_j(x,y)$, $j\ge 1$, such that

(i) $0\le\chi_j(x,y)\le 1$, $j\ge 1$;

(ii) each $\chi_j$ has a support compactly embedded to $B$;

(iii) for each $x$, $y$, $\chi_j(x,y)\nearrow\chi_\infty(x,y)=1_B(x,y)$, $j\to\infty$.

Recall the notation $\overline{f}(x)=\int_{\mathbb{R}^k}f(x,y)\pi^{(x)}(dy)$.

The following lemma collects several simple statements used in the proof.

**Lemma 5.1.** *The following properties hold*:

(a) *there exists* $C > 0$ *such that* $|\mathcal{L}^\varepsilon\varphi(x,y)| \le C,\ \varepsilon > 0$, *for all* $x,\ y$;

(b) $\mathcal{L}^\varepsilon\varphi \to \mathcal{L}\varphi,\ \varepsilon \to 0$, *uniformly on each compactum* $K \subset B$;

(c) *there exists* $C > 0$ *such that* $|\mathcal{L}\varphi(x,y)| \le C$ *for all* $(x,y) \in B$;

(d) $\overline{\chi_j}(x) \to 1,\ j \to \infty$, *uniformly on* $\{|x| \le R\}$ *for each* $R > 0$;

(e) $\overline{\chi_j\mathcal{L}\varphi}(x) \to L\varphi(x),\ j \to \infty$, *uniformly on* $\{|x| \le R\}$ *for each* $R > 0$, *where* $L$ *is defined in* (2.9);

(f) *for any* $T > 0$,

$$\mathbf{E}^\varepsilon\big|\widetilde{X}_\varepsilon(t) - \widetilde{X}_\varepsilon(s)\big|^2 \wedge 1 \le C|t-s|, \quad s,t \in [0,T],$$

*where* $\widetilde{X}$ *is defined in* (5.2), *and*

$$\sup_{t\in[0,T],\varepsilon>0} \mathbf{P}^\varepsilon\big(|X_\varepsilon(t)| > R\big) \to 0, \quad R \to \infty;$$

(g) *for any* $T > 0$,

$$\sup_{t\in[0,T],\varepsilon>0} \mathbf{P}^\varepsilon\big(|Y_\varepsilon(t)| > R\big) \to 0, \quad R \to \infty.$$

***Proof.*** Statement (a) follows directly from the assumptions $\mathbf{H_1}$, $\mathbf{H_2}$. Statement (b) can be derived, in a standard way, using the convergence assumptions $\mathbf{H_4}$, $\mathbf{H_6}$ and the bounds from the assumptions $\mathbf{H_1}$, $\mathbf{H_2}$. Statement (c) follows from (a) and (b).

To prove statement (d), we first mention that each function $\overline{\chi_j}$ is continuous by the assumption $\mathbf{F_0}$. These functions converge monotonously, at each $x \in \mathbb{R}^d$, to the function

$$\overline{\chi_\infty}(x) = \int\limits_{\mathbb{R}^k} 1_B(x,y)\pi^{(x)}(dy) \equiv 1,$$

where the last identity holds by the assumption $\mathbf{H_6}$. Then the required uniform convergence follow by the Dini theorem.

To prove statement (e), we first use statements (c) and (d) to get

$$\big|\overline{\chi_j\mathcal{L}\varphi}(x) - \overline{\mathcal{L}\varphi}(x)\big| = \big|\overline{\chi_j\mathcal{L}\varphi}(x) - \overline{\chi_\infty\mathcal{L}\varphi}(x)\big| \le C(1 - \overline{\chi_j}(x)) \to 0, \quad j \to \infty,$$

uniformly for $x$ with $|x| \le R$. Then the required statement follows by the identity

$$\overline{\mathcal{L}\varphi}(x) = \int\limits_{\mathbb{R}^k} \mathcal{L}\varphi(x,y)\pi^{(x)}(dy) =$$

$$= \int\limits_{\mathbb{R}^k} \Big(\nabla\varphi(x)\cdot a(x,y) + \frac{1}{2}\nabla^2\varphi(x)\cdot b(x,y)\Big)\pi^{(x)}(dy) +$$

$$+ \int\limits_{\mathbb{R}^k}\int\limits_{\mathbb{R}^m} \Big(\varphi(x+v) - \varphi(x) - \nabla\varphi(x)\cdot v\Big)K_{(\rho)}(x,dv)\pi^{(x)}(dy) +$$

$$+\int\limits_{\mathbb{R}^k}\int\limits_{\mathbb{R}^m}\Big(\varphi(x+v)-\varphi(x)\Big)K^{(\rho)}(x,y,dv)\pi^{(x)}(dy)=$$

$$=\nabla\varphi(x)\cdot\overline{a}(x)+\frac{1}{2}\nabla^2\varphi(x)\cdot\overline{b}(x)+\int\limits_{\mathbb{R}^m}\Big(\varphi(x+v)-\varphi(x)-\nabla\varphi(x)\cdot v\Big)\overline{K}_{(\rho)}(x,dv)+$$

$$+\int\limits_{\mathbb{R}^m}\Big(\varphi(x+v)-\varphi(x)\Big)\overline{K}^{(\rho)}(x,dv)=$$

$$=L\varphi(x).$$

Statement (f) can be obtained using the same “truncation of large jumps” argument as in the proof of Proposition 5.2 and the bounds from the assumptions $\mathbf{H_1}$, $\mathbf{H}_2$; we omit the details.

To prove statement (g), we treat $Y_\varepsilon(t)$ as the value of the process $y_\varepsilon$ from (5.10) taken at the (large) time instant $\tau=\varepsilon^{-1}t$ with $\zeta_\varepsilon(\tau)=\xi_\varepsilon(\varepsilon\tau)$, i.e., $y_\varepsilon(\tau)=Y_\varepsilon(\varepsilon\tau)$. Without loss of generality we can assume that the constant $\kappa$ in the assumption $\mathbf{H_5}$ satisfies $\kappa\le 1$. Then by [17] (Theorem 2.8), for every $p_Y<p+\kappa-1$,

$$\sup_{\tau\ge 0,\varepsilon>0}\mathbf{E}^\varepsilon\big|y^\varepsilon(\tau)\big|^{p_Y}<\infty,$$

here we have used that the initial values $y_\varepsilon(0)=Y_\varepsilon(0)$ are bounded. This immediately yields (g).

Lemma 5.1 is proved.

Now we are ready to prove (5.5). Fix $N>0$, and write denote by $\mathbf{P}^\varepsilon_{t-\varepsilon N},\mathbf{E}^\varepsilon_{t-\varepsilon N}$ the conditional probability and conditional expectation w.r.t. $\mathcal{F}^\varepsilon_{t-\varepsilon N}$. For $\varepsilon$ small enough, we have $s_q<t-\varepsilon N$ and thus

$$\mathbf{E}^\varepsilon\Phi\big(X_\varepsilon(s_1),\ldots,X_\varepsilon(s_q)\big)\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big)-L\varphi\big(X_\varepsilon(t)\big)\Big)=$$

$$=\mathbf{E}^\varepsilon\Phi\big(X_\varepsilon(s_1),\ldots,X_\varepsilon(s_q)\big)\mathbf{E}^\varepsilon_{t-\varepsilon N}\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big)-L\varphi\big(X_\varepsilon(t)\big)\Big).$$

By the assumption $\mathbf{H}_0$, there exist functions $\rho(\varepsilon)\to 0$, $\varrho(\varepsilon)\to 0$ such that

$$\mathbf{P}^\varepsilon\left(\sup_{s\in[0,T]}|\xi_\varepsilon(s)|>\rho(\varepsilon)\right)\le\varrho(\varepsilon),$$

where $T>t$ is a fixed number. For a given $R>0$, consider the $\mathcal{F}^\varepsilon_{t-\varepsilon N}$-measurable set

$$\Omega^\varepsilon_{t,N,R}=\left\{\omega:\mathbf{P}^\varepsilon_{t-\varepsilon N}\left(\sup_{s\in[0,T]}|\xi_\varepsilon(s)|>\rho(\varepsilon)\right)\le R\varrho(\varepsilon)\right\},$$

then by the Markov inequality

$$\mathbf{P}^\varepsilon(\Omega^\varepsilon\setminus\Omega^\varepsilon_{t,N,R})\le\frac{1}{R\varrho(\varepsilon)}\mathbf{E}^\varepsilon\left[\mathbf{P}^\varepsilon_{t-\varepsilon N}\left(\sup_{s\in[0,T]}|\xi_\varepsilon(s)|>\rho(\varepsilon)\right)\right]=$$

$$=\frac{1}{R\varrho(\varepsilon)}\mathbf{P}^\varepsilon\left(\sup_{s\in[0,T]}|\xi_\varepsilon(s)|>\rho(\varepsilon)\right)\le$$

$$\le \frac{\varrho(\varepsilon)}{R\varrho(\varepsilon)} = \frac{1}{R}.$$

We have seen in the proof of Lemma 5.1 that $L\varphi = \overline{\chi_\infty \mathcal{L}\varphi}$, thus by statement (c) of this lemma the function $L\varphi$ is bounded. The functions $\Phi$, $\mathcal{L}^\varepsilon$ are bounded, as well,

$$\Big|\mathbf{E}^\varepsilon \Phi\big(X_\varepsilon(s_1),\ldots,X_\varepsilon(s_q)\big)\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big) - L\varphi\big(X_\varepsilon(t)\big)\Big)\Big| \le$$

$$\le C\mathbf{P}^\varepsilon(|X_\varepsilon(t-\varepsilon N)|>R) + C\mathbf{P}^\varepsilon(|Y_\varepsilon(t-\varepsilon N)|>R) + \frac{C}{R} +$$

$$+ C\mathbf{E}^\varepsilon 1_{\widetilde{\Omega}^\varepsilon_{t,N,R}}\left|\mathbf{E}^\varepsilon_{t-\varepsilon N}\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big) - L\varphi\big(X_\varepsilon(t)\big)\Big)\right|, \tag{5.17}$$

where we denote

$$\widetilde{\Omega}^\varepsilon_{t,N,R} = \Omega^\varepsilon_{t,N,R} \cap \big\{|X_\varepsilon(t-\varepsilon N)| \le R, |Y_\varepsilon(t-\varepsilon N)| \le R\big\}.$$

Fix $j \ge 1$, decompose

$$\mathbf{E}^\varepsilon_{t-\varepsilon N}\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big) - L\varphi\big(X_\varepsilon(t)\big)\Big) =$$

$$= \mathbf{E}^\varepsilon_{t-\varepsilon N}\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big) - \mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big)\chi_j\big(X_\varepsilon(t),Y_\varepsilon(t)\big)\Big) +$$

$$+ \mathbf{E}^\varepsilon_{t-\varepsilon N}\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big) - \mathcal{L}\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big)\Big)\chi_j\big(X_\varepsilon(t),Y_\varepsilon(t)\big) +$$

$$+ \Big(\mathbf{E}^\varepsilon_{t-\varepsilon N}\mathcal{L}\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big)\,\chi_j\big(X_\varepsilon(t),Y_\varepsilon(t)\big) - P_N^{\text{frozen}}(\chi_j\mathcal{L}\varphi)\big(X_\varepsilon(t-\varepsilon N),Y_\varepsilon(t-\varepsilon N)\big)\Big) +$$

$$+ \Big(P_N^{\text{frozen}}(\chi_j\mathcal{L}\varphi)\big(X_\varepsilon(t-\varepsilon N),Y_\varepsilon(t-\varepsilon N)\big) - \overline{\chi_j\mathcal{L}\varphi}(X_\varepsilon(t-\varepsilon N))\Big) +$$

$$+ \Big(\overline{\chi_j\mathcal{L}\varphi}(X_\varepsilon(t-\varepsilon N)) - L\varphi(X_\varepsilon(t-\varepsilon N))\Big) +$$

$$+ \Big(L\varphi(X_\varepsilon(t-\varepsilon N)) - L\varphi\big(X_\varepsilon(t)\big)\Big). \tag{5.18}$$

Let us estimate each term in the decomposition (5.18). For the first term, we simply write using Lemma 5.1(a)

$$\Big|\mathbf{E}^\varepsilon_{t-\varepsilon N}\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big) - \mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big)\chi_j\big(X_\varepsilon(t),Y_\varepsilon(t)\big)\Big)\Big| \le$$

$$\le C\mathbf{E}^\varepsilon_{t-\varepsilon N}\Big(1 - \chi_j\big(X_\varepsilon(t),Y_\varepsilon(t)\big)\Big). \tag{5.19}$$

For the second term, we recall that the support of $\chi_j$ is compactly embedded to $B$, thus, by Lemma 5.1(b),

$$\Big|\mathbf{E}^\varepsilon_{t-\varepsilon N}\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big) - \mathcal{L}\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big)\Big)\chi_j\big(X_\varepsilon(t),Y_\varepsilon(t)\big)\Big| \le$$

$$\le \sup_{x,y}\big|\mathcal{L}^\varepsilon\varphi(x,y) - \mathcal{L}\varphi(x,y))\big|\chi_j(x,y) \to 0, \quad \varepsilon \to 0. \tag{5.20}$$

To estimate the third term in (5.18), observe first that the function $\chi_j\mathcal{L}\varphi$ is continuous, which folows from $\mathbf{H_4}$, $\mathbf{H_6}$ similarly to Lemma 5.1(b). Next, define the pair $x_\varepsilon,\ y_\varepsilon$ by (5.11) with $t_0 = t-\varepsilon N$ and take $\widetilde{\mathbf{P}}^\varepsilon = \mathbf{P}^\varepsilon_{t-\varepsilon N,\omega}$, the regular version of the conditional probability. Then, for a.a. $\omega\in\widetilde{\Omega}^\varepsilon_{t,N,R}$, the pair $x_\varepsilon,\ y_\varepsilon$ w.r.t. the probability $\mathbf{P}^\varepsilon_{t-\varepsilon N,\omega}$ belongs to the class $\mathcal{K}(\rho, 2R\varrho, R, 2N)$ in the notation introduced before Proposition 5.2. Applying this proposition, we get

$$\mathbf{E}^\varepsilon 1_{\widetilde{\Omega}^\varepsilon_{t,N,R}}\Big|\mathbf{E}^\varepsilon_{t-\varepsilon N}\mathcal{L}\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big)\,\chi_j\big(X_\varepsilon(t),Y_\varepsilon(t)\big)-$$

$$-P^{\text{frozen}}_N(\chi_j\mathcal{L}\varphi)\big(X_\varepsilon(t-\varepsilon N),Y_\varepsilon(t-\varepsilon N)\big)\Big|\to 0,\quad \varepsilon\to 0. \tag{5.21}$$

To estimate the fourth term, we use Proposition 5.1; without loss of generality we assume that $\kappa<1$. Since the function $\chi_j\mathcal{L}\varphi$ is bounded, Proposition 5.1 yields

$$\mathbf{E}^\varepsilon 1_{\widetilde{\Omega}^\varepsilon_{t,N,R}}\Big|P^{\text{frozen}}_N(\chi_j\mathcal{L}\varphi)\big(X_\varepsilon(t-\varepsilon N),Y_\varepsilon(t-\varepsilon N)\big)-\overline{\chi_j\mathcal{L}\varphi}(X_\varepsilon(t-\varepsilon N))\Big|\le CN^{-\frac{p+\kappa-1}{1-\kappa}}. \tag{5.22}$$

For the fifth term, we have

$$\mathbf{E}^\varepsilon 1_{\widetilde{\Omega}^\varepsilon_{t,N,R}}\Big|\big(\overline{\chi_j\mathcal{L}\varphi}(X_\varepsilon(t-\varepsilon N))-L\varphi(X_\varepsilon(t-\varepsilon N))\big)\Big|\le\sup_{|x|\le R}\big|\overline{\chi_j\mathcal{L}\varphi}(x)-L\varphi(x)\big|. \tag{5.23}$$

For the sixth term, we get

$$\mathbf{E}^\varepsilon 1_{\widetilde{\Omega}^\varepsilon_{t,N,R}}\Big|L\varphi(X_\varepsilon(t-\varepsilon N))-L\varphi\big(X_\varepsilon(t)\big)\Big|\to 0,\quad \varepsilon\to 0, \tag{5.24}$$

by Lemma 5.1(f) and uniform continuity of $L\varphi$ on compacts. Summarizing the estimates (5.17) and (5.19) – (5.24), we get

$$\limsup_{\varepsilon\to 0}\Big|\mathbf{E}^\varepsilon\Phi\big(X_\varepsilon(s_1),\ldots,X_\varepsilon(s_q)\big)\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big)-L\varphi\big(X_\varepsilon(t)\big)\Big)\Big|\le$$

$$\le C\sup_{s\le t,\varepsilon>0}\mathbf{P}^\varepsilon(|X_\varepsilon(s)|>R)+C\sup_{s\le t,\varepsilon>0}\mathbf{P}^\varepsilon\big(|Y_\varepsilon(s)|>R\big)+\frac{C}{R}+$$

$$+CN^{-\frac{p+\kappa-1}{1-\kappa}}+\sup_{|x|\le R}\big|\overline{\mathcal{L}\varphi\chi_j}(x)-L\varphi(x)\big|+$$

$$+C\limsup_{\varepsilon\to 0}\mathbf{E}^\varepsilon 1_{\widetilde{\Omega}^\varepsilon_{t,N,R}}E^\varepsilon_{t-\varepsilon N}\Big(1-\chi_j\big(X_\varepsilon(t),Y_\varepsilon(t)\big)\Big). \tag{5.25}$$

Similarly to (5.21) – (5.23), we have

$$\limsup_{\varepsilon\to 0}\mathbf{E}^\varepsilon 1_{\widetilde{\Omega}^\varepsilon_{t,N,R}}E^\varepsilon_{t-\varepsilon N}\Big(1-\chi_j\big(X_\varepsilon(t),Y_\varepsilon(t)\big)\Big)\le CN^{-\frac{p+\kappa-1}{1-\kappa}}+\sup_{|x|\le R}(1-\overline{\chi_j}(x)),$$

thus,

$$\limsup_{\varepsilon\to 0}\Big|\mathbf{E}^\varepsilon\Phi\big(X_\varepsilon(s_1),\ldots,X_\varepsilon(s_q)\big)\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big)-L\varphi\big(X_\varepsilon(t)\big)\Big)\Big|\le$$

$$\le C\sup_{s\le t,\varepsilon>0}\mathbf{P}^\varepsilon(|X_\varepsilon(s)|>R)+C\sup_{s\le t,\varepsilon>0}\mathbf{P}^\varepsilon\big(|Y_\varepsilon(s)|>R\big)+\frac{C}{R}+$$

$$+CN^{-\frac{p+\kappa-1}{1-\kappa}} + \sup_{|x|\le R} \left|\overline{\mathcal{L}\varphi\chi_j}(x) - L\varphi(x)\right| + C \sup_{|x|\le R} (1 - \overline{\chi_j}(x)). \tag{5.26}$$

The constants $R,\ N,\ j$ in the above inequality are arbitrary. Taking first $j\to\infty, N\to\infty$ for a fixed $R$, we get, by Lemma 5.1(d), (e), that

$$\limsup_{\varepsilon\to 0} \left|\mathbf{E}^\varepsilon \Phi\big(X_\varepsilon(s_1),\ldots,X_\varepsilon(s_q)\big)\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big) - L\varphi\big(X_\varepsilon(t)\big)\Big)\right| \le$$

$$\le C \sup_{s\le t,\varepsilon>0} \mathbf{P}^\varepsilon(|X_\varepsilon(s)|>R) + C \sup_{s\le t,\varepsilon>0} \mathbf{P}^\varepsilon\big(|Y_\varepsilon(s)|>R\big) + \frac{C}{R}. \tag{5.27}$$

Then by Lemma 5.1(f), (g) we can pass to the limit $R\to\infty$ and finally get

$$\limsup_{\varepsilon\to 0} \left|\mathbf{E}^\varepsilon \Phi\big(X_\varepsilon(s_1),\ldots,X_\varepsilon(s_q)\big)\Big(\mathcal{L}^\varepsilon\varphi\big(X_\varepsilon(t),Y_\varepsilon(t)\big) - L\varphi\big(X_\varepsilon(t)\big)\Big)\right| = 0.$$

This proves (5.5) and completes the entire proof.

**6. Appendix.** ***Proof of Lemma* 3.1.** Set

$$\zeta_\varepsilon(t) = \inf\left\{ s\ge 0\colon \int_0^s b_\varepsilon^2(z)dz \ge t\right\}.$$

Making the change of time $\tilde\eta_\varepsilon(t) := \eta_\varepsilon(\zeta_\varepsilon(t))$, we see that $\tilde\eta_\varepsilon(t)$ satisfies assumptions of this lemma with another constant $\tilde A>0$ and a new Wiener process $\tilde W(t) = W(\zeta_\varepsilon(t))$ but with $\tilde b_\varepsilon(t)\equiv 1$. Since $(C_2)^{-2}t \le \zeta_\varepsilon(t) \le (C_1)^{-2}t$ without loss of generality we will assume that $b_\varepsilon(t)\equiv 1$.

Set $L_\varepsilon := Ax^\gamma \dfrac{d}{dx} + \dfrac{\varepsilon^2}{2}\dfrac{d^2}{dx^2}$. Denote

$$v_\varepsilon(x) := \int_0^{|x|} \exp\left\{\frac{-2Ay^{\gamma+1}}{(\gamma+1)\varepsilon^2}\right\}\left(\int_0^y \frac{2}{\varepsilon^2}\exp\left\{\frac{2Az^{\gamma+1}}{(\gamma+1)\varepsilon^2}\right\}dz\right)dy.$$

We have $L_\varepsilon v_\varepsilon(x)\ge 1,\ \operatorname{sgn}(x)v_\varepsilon'(x)\ge 0,$ and $v_\varepsilon(0)=0.$

Then, by Ito's formula, we have

$$\mathbf{E}v_\varepsilon(\eta_\varepsilon(\tau_\varepsilon(\delta)\wedge n)) = \mathbf{E}\int_0^{\tau_\varepsilon(\delta)\wedge n} \big(a_\varepsilon(s)\eta_\varepsilon^\gamma(s)v_\varepsilon'(\eta_\varepsilon(s)) + \frac{\varepsilon^2}{2}v_\varepsilon''(\eta_\varepsilon(s))\big)ds \ge$$

$$\ge \mathbf{E}\int_0^{\tau_\varepsilon(\delta)\wedge n} \big(A\eta_\varepsilon^\gamma(s)v_\varepsilon'(\eta_\varepsilon(s)) + \frac{\varepsilon^2}{2}v_\varepsilon''(\eta_\varepsilon(s))\big)ds = \mathbf{E}\int_0^{\tau_\varepsilon(\delta)\wedge n} L_\varepsilon v_\varepsilon(\eta_\varepsilon(s))ds \ge$$

$$\ge \mathbf{E}\int_0^{\tau_\varepsilon(\delta)\wedge n} 1ds = \mathbf{E}\tau_\varepsilon(\delta)\wedge n.$$

Passing $n\to\infty$ and applying the Fatou lemma we get a.s. finiteness of $\tau_\varepsilon(\delta)$. Since $v_\varepsilon(\eta_\varepsilon(\tau_\varepsilon(\delta))) = = v_\varepsilon(\delta) = v_\varepsilon(-\delta)$, we get the estimate

$$\mathbf{E}\tau_\varepsilon(\delta) \le v_\varepsilon(\delta).$$

Let $x > 0$ be arbitrary. Changing the variables $s := \dfrac{z^{\gamma+1}}{\varepsilon^2}$ and $t := \dfrac{y^{\gamma+1}}{\varepsilon^2}$ we obtain

$$v_\varepsilon(x) = \frac{2\varepsilon^{\frac{2}{\gamma+1}}}{(\gamma+1)\varepsilon^2} \int\limits_0^{\frac{|x|^{\gamma+1}}{\varepsilon^2}} \exp\left\{ -\frac{2At}{\gamma+1}\right\} \left( \int\limits_0^{t^{\frac{1}{\gamma+1}}\varepsilon^{\frac{2}{\gamma+1}}} \exp\left\{ \frac{2Az^{\gamma+1}}{(\gamma+1)\varepsilon^2}\right\} dz \right) t^{\frac{-\gamma}{\gamma+1}} dt =$$

$$= \frac{2\varepsilon^{\frac{4}{\gamma+1}}}{(\gamma+1)^2\varepsilon^2} \int\limits_0^{\frac{|x|^{\gamma+1}}{\varepsilon^2}} \exp\left\{ -\frac{2At}{\gamma+1}\right\} \left( \int\limits_0^t \exp\left\{ \frac{2As}{\gamma+1}\right\} s^{\frac{-\gamma}{\gamma+1}} ds \right) t^{\frac{-\gamma}{\gamma+1}} dt =$$

$$= \frac{2}{(\gamma+1)^2} \varepsilon^{\frac{2(1-\gamma)}{\gamma+1}} \int\limits_0^{\frac{|x|^{\gamma+1}}{\varepsilon^2}} \exp\left\{ -\frac{2At}{\gamma+1}\right\} \left( \int\limits_0^t \exp\left\{ \frac{2As}{\gamma+1}\right\} s^{\frac{-\gamma}{\gamma+1}} ds \right) t^{\frac{-\gamma}{\gamma+1}} dt. \tag{6.1}$$

It follows from L'Hôpital's rule that for any $\alpha > 0$ and $\beta > -1$:

$$\int\limits_0^t e^{\alpha s} s^\beta ds \sim \alpha^{-1} e^{\alpha t} t^\beta, \quad t \to +\infty.$$

So,

$$\int\limits_0^t \exp\left\{ \frac{2As}{\gamma+1}\right\} s^{\frac{-\gamma}{\gamma+1}} ds \sim \frac{\gamma+1}{2A} \exp\left\{ \frac{2At}{\gamma+1}\right\} t^{\frac{-\gamma}{\gamma+1}}, \quad t \to +\infty.$$

Applying this and L'Hôpital's rule, we get

$$\int\limits_0^u \exp\left\{ -\frac{-2At}{\gamma+1}\right\} \left( \int\limits_0^t \exp\left\{ \frac{2As}{\gamma+1}\right\} s^{\frac{-\gamma}{\gamma+1}} ds \right) t^{\frac{-\gamma}{\gamma+1}} dt \sim$$

$$\sim \frac{\gamma+1}{2A} \int\limits_0^u \exp\left\{ -\frac{-2At}{\gamma+1}\right\} \left( \exp\left\{ \frac{2At}{\gamma+1}\right\} t^{\frac{-\gamma}{\gamma+1}} \right) t^{\frac{-\gamma}{\gamma+1}} dt =$$

$$= \frac{\gamma+1}{2A} \int\limits_0^u t^{\frac{-2\gamma}{\gamma+1}} dt = \frac{(\gamma+1)^2}{2A(1-\gamma)} u^{\frac{1-\gamma}{\gamma+1}}, \quad u \to +\infty.$$

Therefore, we get from (6.1) the following equivalence for any fixed $x \neq 0$ as $\varepsilon \to 0$ :

$$v_\varepsilon(x) \underset{\varepsilon\to 0}{\sim} K\varepsilon^{\frac{2(1-\gamma)}{\gamma+1}} \left( \frac{|x|^{\gamma+1}}{\varepsilon^2} \right)^{\frac{-2\gamma}{\gamma+1}+1} =$$

$$= K_2 \varepsilon^{\frac{2(1-\gamma)}{\gamma+1}} \left( \frac{|x|^{\gamma+1}}{\varepsilon^2} \right)^{\frac{1-\gamma}{\gamma+1}} =$$

$$= K_2 \varepsilon^{\frac{2(1-\gamma)}{\gamma+1}} |x|^{1-\gamma} \varepsilon^{\frac{2(\gamma-1)}{\gamma+1}} = K_2 |x|^{1-\gamma},$$

where $K$ is a constant independent of $\delta$.

This yields that for any fixed $\delta \geq 0$:

$$\limsup_{\varepsilon\to 0} \mathbf{E}\tau_\varepsilon(\delta) \leq \limsup_{\varepsilon\to 0} \mathbf{E}v_\varepsilon(\delta) = K_2 \delta^{1-\gamma}.$$

Lemma 3.1 is proved.

## References


1. F. Attanasio, F. Flandoli, *Zero-noise solutions of linear transport equations without uniqueness: an example*, C. R. Acad. Sci. Paris, Ser. I, **347**, 753 – 756 (2009).
2. R. Bafico, *On the convergence of the weak solutions of stochastic differential equations when the noise intensity goes to zero*, Boll. Unione Mat. Ital., **5**, 308 – 324 (1980).
3. R. Bafico, P. Baldi, *Small random perturbations of Peano phenomena*, Stochastics, **6**, No. 2, 279 – 292 (1982).
4. R. Buckdahn, Y. Ouknine, M. Quincampoix, *On limiting values of stochastic differential equations with small noise intensity tending to zero*, Bull. Sci. Math., **133**, 229 – 237 (2009).
5. V. S. Borkar, K. Suresh Kumar, *A new Markov selection procedure for degenerate diffusions*, J. Theor. Probab., **23**, No. 3, 729 – 747 (2010).
6. F. Delarue, F. Flandoli, *The transition point in the zero noise limit for a 1D Peano example*, Discrete Contin. Dyn. Syst., **34**, No. 10, 4071 – 4083 (2014).
7. F. Delarue, F. Flandoli, D. Vincenzi, *Noise prevents collapse of Vlasov – Poisson point charges*, Commun. Pure and Appl. Math., **67**, No. 10, 1700 – 1736 (2014).
8. F. Delarue, M. Maurelli, *Zero noise limit for multidimensional SDEs driven by a pointy gradient* (2019), arXiv preprint arXiv:1909.08702.
9. N. Dirr, S. Luckhaus, M. Novaga, *A stochastic selection principle in case of fattening for curvature flow*, Calc. Var. Part. Different. Equat., **13**, No. 4, 405 – 425 (2001).
10. H. J. Engelbert, W. Schmidt, *Strong Markov continuous local martingales and solutions of one-dimensional stochastic differential equations (Part III)*, Math. Nachr., **151**, No. 1, 149 – 197 (1991).
11. A. Friedman, *Partial differential equations of parabolic type*, Prentice-Hall, New York (1964).
12. S. Herrmann, *Phénomène de Peano et grandes déviations*, C. R. Acad. Sci. Paris, Sér. I, **332**, No. 11, 1019 – 1024 (2001).
13. N. Ikeda, S. Watanabe, *Stochastic differential equations and diffusion processes*, North-Holland Math. Library, **24**, North-Holland Co. Publ., Amsterdam etc. (1981).
14. I. Karatzas, S. E. Shreve, *Brownian motion and stochastic calculus*, Springer, New York (1988).
15. I. G. Krykun, S. Ya. Makhno, *The Peano phenomenon for Ito equations*, J. Math. Sci., **192**, No. 4, 441 – 458 (2013).
16. A. Kulik, *Ergodic behavior of Markov processes*, de Gruyter, Berlin, Boston (2017).
17. A. Kulik, I. Pavlyukevich, *Moment bounds for dissipative semimartingales with heavy jumps*, http://arxiv.org/abs/2004.12449.
18. I. Pavlyukevich, A. Pilipenko, *Generalized selection problem with Lévy noise* (2020), arXiv preprint arXiv: 2004.05421.
19. A. Pilipenko, F. N. Proske, *On a selection problem for small noise perturbation in the multidimensional case*, Stoch. and Dyn., **18**, No. 6 (2018), 23 p., doi 10.1142/S0219493718500454.
20. A. Pilipenko, F. N. Proske, *On perturbations of an ODE with non-Lipschitz coefficients by a small self-similar noise*, Statistics & Probab. Lett., **132**, 62 – 73 (2018), https://doi.org/10.1016/j.spl.2017.09.005.
21. D. Trevisan, *Zero noise limits using local times*, Electron. Commun. Probab., **18**, No. 31 (2013), 7 p.
22. A. Veretennikov, *On strong solutions and explicit formulas for solutions of stochastic integral equations*, Sb. Math., **39**, No. 3, 387 – 403 (1981).


Received 27.08.20