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

A. Kulik and A. Pilipenko

UDC 519.21

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 a fixed hyperplane $H$. It is assumed that $X^\varepsilon(0) = x^0 \in H$, the drift $a(x)$ has a Hölder asymptotics as $x$ approaches $H$, and the limit ODE $dX(t) = a(X(t))\,dt$ does not have a unique solution. It is shown that if the drift pushes the solution away from $H$, then the limit process with certain probabilities selects some extreme solutions of the limit ODE. If the drift attracts the solution to $H$, then the limit process satisfies an ODE with certain averaged coefficients. To prove the last result, we formulate an averaging principle, which is quite general and new.

1. Introduction

Consider an ordinary differential equation (ODE)

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

where $a$ is a continuous function of linear growth satisfying a local Lipschitz condition everywhere except the point $u = 0$. Then the solution of (1.1) may be not unique. Indeed, for

$$a(u) = \sqrt{|u|} \text{sgn}(u),$$

the ODE (1.1) has multiple solutions $\pm t^2/4$, $t \geq 0$.

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

$$du_\varepsilon(t) = a(u_\varepsilon(t))\,dt + \varepsilon dW(t),\quad u_\varepsilon(0) = 0,$$

where $W$ is a Wiener process. Equation (1.2) has a unique strong solution due to the Zvonkin–Veretennikov theorem [22]. It is easy to see that a family of distributions of $\{u_\varepsilon\}$ is weakly relatively compact because $a$ has

---

1 Wroclaw University of Science and Technology, Wroclaw, Poland; e-mail: kulik.alex.m@gmail.com.
2 Institute of Mathematics, National Academy of Sciences of Ukraine, Kyiv, Ukraine; National Technical University of Ukraine “I. Sikorsky Kyiv Polytechnic Institute”. Kyiv, Ukraine; e-mail: pilipenko.ay@gmail.com.
3 Corresponding author.

Published in Ukrain’s’kyi Matematychnyi Zhurnal, Vol. 72, No. 9, pp. 1254–1285, September, 2020. Ukrainian DOI: 10.37863/umzh.v72i9.6292. Original article submitted August 27, 2020.
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 can be considered as the 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; for the other generalizations see, e.g., [4–9, 12, 15, 18–21] and the references therein. In the multidimensional case, the investigations are much more complicated than in the one-dimensional case. Thus, at present, we still do not have simple sufficient conditions guaranteeing the existence of the limit \( \lim_{\varepsilon \to 0} u_\varepsilon \) and the property of relative compactness for the distributions of \( u_\varepsilon \). This ensures the existence and uniqueness of a weak solution to (1.3) and the property of relative compactness for this limit. As one of the causes of this situation, we can mention the absence of linear ordering in the multidimensional case. Indeed, in the one-dimensional situation, there are only two ways to exit from the point 0: one way to the right and the other to the left. The probability of going to the left or to the right can be easily found because there are explicit formulas for the probabilities of hitting in one-dimensional diffusions. The equation for the limit process outside 0 must satisfy the original ODE because \( a \) is Lipschitz continuous there.

In the present paper, we consider the multidimensional case in which the Lipschitz condition for \( a \) may be violated in a hyperplane. We now describe the corresponding model.

Consider a stochastic differential equation (SDE)

\[
\begin{align*}
\text{du}_\varepsilon(t) &= a(u_\varepsilon(t))dt + \varepsilon \sigma(u_\varepsilon(t))dW(t), \\
u_\varepsilon(0) &= x^0,
\end{align*}
\]

where \( a : \mathbb{R}^d \to \mathbb{R}^d \) and \( \sigma : \mathbb{R}^d \to \mathbb{R}^{d \times m} \) are measurable functions and \( W \) is an \( m \)-dimensional Wiener process.

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

We 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 where \( a \) is globally Lipschitz continuous in the bottom half space \( \mathbb{R}^d_- := \mathbb{R}^{d-1} \times (-\infty, 0) \) and in the top 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 the present paper, we investigate the case where the drift has a Hölder-type asymptotics in the neighborhood of \( H \). Namely, we 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 in \( \mathbb{R}_-^d \), \( b^\pm(x) \neq 0 \), \( x \in H \);

(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 are based on new ideas compared with the proofs from [19]. To illustrate the difference, we briefly recall the results obtained in [19], where the case \( \gamma = 0 \) was considered, and present a sketch of the expected results for the case where \( \gamma \in (0, 1) \).

Case 1 (Vector Field \( a \) Pushes Away from the Hyperplane). By \( n = (0, \ldots, 0, 1) \) we denote a vector normal to the hyperplane \( H \). Assume that \( \gamma = 0 \) and \( \pm(a^\pm(x), n) > 0 \), \( x \in H \). Then there are two solutions \( u^\pm \)
of the equation
\[ du(t) = a(u(t))dt, \]
which start at \( x^0 \in H \) and exit from \( H \) immediately to the top and bottom half spaces, respectively. In [19], it was proved that if \( \gamma = 0 \), then the limit process \( u_0 \) immediately leaves \( H \) and moves as \( u^\pm \) with probabilities proportional to \( \left| (a^\pm(x^0), n) \right| \). The corresponding proof is similar to the proof in the one-dimensional case. It is based on the use of a comparison principle adapted to the multidimensional situation. The investigations for arbitrary \( \gamma \in (0, 1) \) are similar but the corresponding selection probabilities are different.

**Remark 1.1.** In [19], it was assumed 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 from \( H \) on one side of \( H \) and attracts on the other side (e.g., \( (a^\pm(x), n) > 0 \)), then there is a unique solution to (1.4) that starts at \( x^0 \in H \). This solution immediately exits from \( H \) (to the top half space in the analyzed case) and the limit process \( u_0 \) is equal to this solution of the ODE; see [19]. For \( \gamma \in (0, 1) \), the result is similar. Assume that, e.g., \( b^\pm(x^0) > 0 \). Then there exists a unique solution to Eq. (1.4) that exits \( H \) immediately (it is also possible that the ODE may have other solutions staying in \( H \)). Moreover, this solution exits into the top half space and the limit process \( u_0 \) is equal to this solution. We do not prove this result in the present paper. The proof is similar to [19].

**Case 2 (Vector Field \( a \) Pushes Toward the Hyperplane).** Assume that \( \gamma = 0 \) and \( \pm(a^\pm(x), n) < 0, x \in H \). It can be seen that any limit point of \( \{u_\varepsilon\} \) must stay in \( H \) with probability 1. In [19], it was proved that the limit process \( u_0 \) satisfies an ODE on \( H \) with the following drift:

\[ P_H(p_+(x)a^+(x) + p_-(x)a^-(x)), \]

where \( P_H \) is the orthogonal projection onto \( H \) and the coefficients \( p_\pm(x) \) are equal to

\[ \frac{a^+_d(x)}{a^+_d(x) - a^-_d(x)}. \]

Note that this multidimensional result does not have one-dimensional analogs with zero process as the limit. In the multidimensional case, the first \( d - 1 \) coordinates may change, while \( d \)th coordinate remains equal to zero.

The idea of the proof was to analyze the time spent by \( u_\varepsilon \) in the top and bottom half spaces. It was shown that, since any limit process stays in \( H \) and \( u_\varepsilon \) is close to \( H \) for small \( \varepsilon \), the times spent in the top and bottom half spaces in the 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 presented in [19] is independent of the type of noise. A small noise may be an arbitrary process that (a) ensures the existence of solution and (b) converges to 0 uniformly in probability as \( \varepsilon \to 0 \) (however, the corresponding results were formulated solely for the Brownian noise).

The proof from [19] does not work if \( a_d(x) \to 0 \) as \( x \) approaches to \( H \). The time spent in the top and bottom half spaces may depend on the asymptotics of decay of \( a_d \) in the neighborhood of \( H \). In the present paper, we prove the result for the case where \( a \) satisfies assumptions \( A1 \) and \( A2 \) with \( \gamma \in (0, 1) \), \( b^+(x) < 0 \), and \( b^-(x) > 0 \) for \( x \in H \).

It turns out that if we scale the vertical coordinate \( \varepsilon^{-\delta}u_{d,\varepsilon}(t) \) for a special choice of \( \delta > 0 \), then a pair formed by \( (u_{1,\varepsilon}(t), \ldots, u_{d-1,\varepsilon}(t)) \) and \( \varepsilon^{-\delta}u_{d,\varepsilon}(t) \) can be treated as components of a Markov process in “slow” and “fast” times, respectively. Hence, the description of the limit process for \( \{u_\varepsilon\} \) is closely related to the averaging principle.
for Markov processes. In what follows, we show that the limit process satisfies an ODE on $H$ whose coefficients are obtained as a result of averaging of the functions $a_k^\pm$, $k = 1, \ldots, d-1$, over the stationary distribution of a scaled vertical component, while the other components remain frozen. The idea to use scaling in the small-noise problem was efficiently applied in the one-dimensional case for the drift given by a power-type function and the noise in the form of a Lévy $\alpha$-stable (or even more general) process.

**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], which somehow depends on a Wiener process $W$ (that should formally disappear in the limit equation).

The paper is organized as follows: In Section 2, we formulate the problem and our main results. The proofs for the cases where the drift pushes away from $H$ and toward $H$ are presented in Sections 3 and 4, respectively.

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

2. Main Results

We represent $u_\varepsilon(t)$ as a pair $(X_\varepsilon(t), Y_\varepsilon(t))$, where $Y_\varepsilon$ is the last coordinate of $u_\varepsilon$ and $X_\varepsilon$ consists of the first $d-1$ coordinates. In what follows, we study solely the general problem for the pair $(X_\varepsilon(t), Y_\varepsilon(t))$, which can be easily be reformulated for $u_\varepsilon$. To simplify notation, we assume that $X_\varepsilon$ is a $d$-dimensional process but not a $(d-1)$-dimensional process.

The general statement of the problem is as follows: Let $X_\varepsilon$ and $Y_\varepsilon$ 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:

\[
\begin{align*}
   dX_\varepsilon(t) &= \psi(X_\varepsilon(t), Y_\varepsilon(t)) \, dt + \varepsilon \, b(X_\varepsilon(t), Y_\varepsilon(t)) \, dB(t), \\
   dY_\varepsilon(t) &= \varphi(X_\varepsilon(t), Y_\varepsilon(t)) \, dt + \varepsilon \beta(X_\varepsilon(t), Y_\varepsilon(t)) \, dW(t), \\
   X_\varepsilon(0) &= x^0, \quad Y_\varepsilon(0) = 0,
\end{align*}
\]

(2.1)

where $B$ and $W$ 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

\[ \psi(x, y) = \psi^+(x, y) \mathbb{1}_{y>0} + \psi^-(x, y) \mathbb{1}_{y<0} \] and \[ \varphi(x, y) = \varphi^+(x, y) \mathbb{1}_{y>0} + \varphi^-(x, y) \mathbb{1}_{y<0}, \]

where $\psi^\pm$ and $\varphi^\pm$ are functions bounded and continuous in $x$ and $y$.

We assume that domains of $\psi^\pm$ and $\varphi^\pm$ coincide with the entire spaces $x \in \mathbb{R}^d$ and $y \in \mathbb{R}$, despite the fact that we use only their values in the corresponding half spaces. The functions $\psi$ and $\varphi$ may have jump discontinuities on $H$.

\[ \varphi^+(x, 0) \neq 0 \text{ for any } x \in \mathbb{R}^d. \]

(B2)
(B3) \( \beta(x, y) = \beta^+(x, y)1_{y \geq 0} + \beta^-(x, y)1_{y < 0} \), where \( \beta^\pm \) are functions bounded, continuous, and separated from zero in the entire space \( \mathbb{R}^d \times \mathbb{R} \) and \( b \) is a function 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 of

\[
\begin{align*}
  d\hat{X}_\varepsilon(t) & = \frac{\psi}{\beta^2}(\hat{X}_\varepsilon(t), \hat{Y}_\varepsilon(t))dt + \varepsilon \frac{b}{\beta}(\hat{X}_\varepsilon(t), \hat{Y}_\varepsilon(t))dB(t), \\
  d\hat{Y}_\varepsilon(t) & = \varepsilon dW(t), \\
  \hat{X}_\varepsilon(0) & = x^0, \quad \hat{Y}_\varepsilon(0) = 0.
\end{align*}
\]

Note that all coefficients may be discontinuous in \( H \) but the processes spend zero time there with probability 1. Thus, any redefinition of the coefficients in \( H \) does not affect the equations.

By using the transformation-of-time arguments (see, e.g., [13]), we get a solution to

\[
\begin{align*}
  d\hat{X}_\varepsilon(t) & = \psi(\hat{X}_\varepsilon(t), \hat{Y}_\varepsilon(t))dt + \varepsilon b(\hat{X}_\varepsilon(t), \hat{Y}_\varepsilon(t))dB(t), \\
  d\hat{Y}_\varepsilon(t) & = \varepsilon \beta(\hat{X}_\varepsilon(t), \hat{Y}_\varepsilon(t))dW(t), \\
  \hat{X}_\varepsilon(0) & = x^0, \quad \hat{Y}_\varepsilon(0) = 0.
\end{align*}
\]

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

**Remark 2.1.** If \( b \) is nondegenerate, then the existence of solution can be proved without transformation of the 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

\[ \text{sgn}(y)\varphi(x, y)y^\gamma > 0, \quad y \neq 0 \]

and the drift pushes away from the hyperplane \( \mathbb{R}^d \times \{0\} \).

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

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

\[
\begin{align*}
  dX(t) & = \psi(X(t), Y(t))dt, \\
  dY(t) & = \varphi(X(t), Y(t))Y^\gamma(t)dt, \\
  X(0) & = x^0, \quad Y(0) = 0,
\end{align*}
\]

such that \( Y^+(t) > 0 \) and \( Y^-(t) < 0 \) for all \( t > 0 \).
Indeed, we set \( \tilde{Y}(t) := Y^{1-\gamma}(t) \). Then

\[
X(t) = x^0 + \int_0^t \psi\left(X(s), \tilde{Y}^{\frac{1}{1-\gamma}}(s)\right) ds,
\]

\[
\tilde{Y}(t) = (1 - \gamma) \int_0^t \varphi\left(X(s), \tilde{Y}^{\frac{1}{1-\gamma}}(s)\right) ds.
\]

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

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

\[
\tilde{Y}^\pm(t) = (1 - \gamma) \int_0^t \varphi^\pm\left(X^\pm(s), (\tilde{Y}^\pm(s))^{\frac{1}{1-\gamma}}\right) ds
\]

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 \( t > 0 \). By the inverse change of variables, we get the desired functions

\[
Y^\pm(t) = \left(\tilde{Y}^\pm(t)\right)^{\frac{1}{1-\gamma}}.
\]

The solution does not explode for a finite time because \( \psi^\pm \) and \( \varphi^\pm \) are bounded by assumption \( B_1 \).

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

\[
p^- \delta_{(X-, Y^-)} + p^+ \delta_{(X+, Y^+)},
\]

where

\[
p^\pm = \frac{\left( \frac{\varphi^\pm(x^0, 0)}{(\beta^\pm(x^0, 0))^2} \right)^{\frac{1}{1+\gamma}}}{\left( \frac{\varphi^- (x^0, 0)}{(\beta^- (x^0, 0))^2} \right)^{\frac{1}{1+\gamma}} + \left( \frac{\varphi^+ (x^0, 0)}{(\beta^+ (x^0, 0))^2} \right)^{\frac{1}{1+\gamma}}}
\]

(2.2)

and \( \delta_{(X^+, Y^+)} \) and \( \delta_{(X^-, Y^-)} \) denote the unit masses 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 \((X^+(t), Y^+(t))\) (resp., \((X^-(t), Y^-(t))\)) with probability 1.
Remark 2.3. If we have the inequalities $\varphi^+(x^0,0) > 0$ and $\varphi^-(x^0,0) < 0$ only at the initial point (and, hence, in a certain neighborhood due to the continuity of coefficients), then the functions $(X^\pm(t), Y^\pm(t))$ are well defined up to the time $\tau_{\mu}^\pm := \inf \{ t > 0 : Y^\pm(t) = 0 \}$ of the first return to $H$. In this case, we have the convergence (in distribution) for the stopped processes:

$$
\left( X_\varepsilon(\cdot \land \tau_{\mu}^+ \land \tau_{\mu}^-), Y_\varepsilon(\cdot \land \tau_{\mu}^+ \land \tau_{\mu}^-) \right) \Rightarrow p_- \delta(X^-(\cdot \land \tau_{\mu}^+ \land \tau_{\mu}^-), Y^-(\cdot \land \tau_{\mu}^+ \land \tau_{\mu}^-)) + p_+ \delta(X^+(\cdot \land \tau_{\mu}^+ \land \tau_{\mu}^-), Y^+(\cdot \land \tau_{\mu}^+ \land \tau_{\mu}^-)).
$$

The proof is basically the same but involves (in addition) the routine localization arguments.

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

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

Theorem 2.2. For any $T > 0$, the following uniform convergence in probability is true:

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

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

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

and

$$
\overline{\psi}(x) = \psi^+(x,0) \frac{\left( \frac{(\beta^+(x,0))^2}{\varphi^+(x,0)} \right)^{\frac{1}{\gamma+1}} \left( \frac{(\beta^-(x,0))^2}{\varphi^-(x,0)} \right)^{\frac{1}{\gamma+1}}}{\varphi^+(x,0)^{\frac{1}{\gamma+1}}} + \psi^-(x,0) \frac{\left( \frac{(\beta^-(x,0))^2}{\varphi^-(x,0)} \right)^{\frac{1}{\gamma+1}} \left( \frac{(\beta^+(x,0))^2}{\varphi^+(x,0)} \right)^{\frac{1}{\gamma+1}}}{\varphi^-(x,0)^{\frac{1}{\gamma+1}}}, \quad (2.3)
$$

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 exit law from the set $\{ x \in \mathbb{R}^d : \psi^+(x,0) < 0 \}$.
where \( \pi(x) \) is the stationary distribution for the SDE

\[
dy(x)(t) = (\varphi^+(x,0)\mathbb{I}_{y(x)(t)>0} + \varphi^-(x,0)\mathbb{I}_{y(x)(t)<0})(y(x)(t))^\gamma dt + \beta(x,0) dW(t).
\]

Hence,

\[
\lim_{\varepsilon \to 0} \psi(x) = \psi^+(x,0)\pi(x)(0,\infty)) + \psi^-(x,0)\pi(x)(-\infty,0)),
\]

i.e., the drift of the limit equation is the averaging of \( \psi^\pm \) over the stationary distribution of SDE with frozen variable \( x \). The corresponding relation between the averaging principle and the procedure of averaging of coefficients in the limit equation for the small noise perturbation problem becomes clear from the proof.

In the next section, we formulate the 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. Assume that, for \( \varepsilon > 0 \), the processes \( X_\varepsilon(t) \) and \( Y_\varepsilon(t) \) take values in \( \mathbb{R}^d \) and \( \mathbb{R}^k \) and have the form

\[
X_\varepsilon(t) = X_\varepsilon(0) + \int_0^t a^\varepsilon(X_\varepsilon(s),Y_\varepsilon(t)) \, ds + \int_0^t \sigma^\varepsilon(X_\varepsilon(s),Y_\varepsilon(s)) \, dB^\varepsilon_s
\]

\[
+ \int_0^t \int_{\mathbb{R}^m} e^\varepsilon(X_\varepsilon(s-),Y_\varepsilon(s-),u) \left[ N^\varepsilon(du,ds) - 1_{|u| \leq \rho} \nu^\varepsilon(du)ds \right] + \xi_\varepsilon(t),
\]

\[
Y_\varepsilon(t) = Y_\varepsilon(0) + \varepsilon^{-1} \int_0^t A^\varepsilon(X_\varepsilon(s),Y_\varepsilon(s)) \, ds + \varepsilon^{-1/2} \int_0^t \Sigma^\varepsilon(X_\varepsilon(s),Y_\varepsilon(s)) \, dW^\varepsilon_s
\]

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

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

The factor \( \varepsilon^{-1} \) in the intensity measure for \( Q^\varepsilon(dz,dt) \) and the factors \( \varepsilon^{-1} \) and \( \varepsilon^{-1/2} \) at the integrals w.r.t. \( ds \) and \( dW^\varepsilon_s \) in the equation for \( Y_\varepsilon \) mean that the evolution of the component \( Y_\varepsilon \) occurs on the “fast” time scale \( \varepsilon^{-1} t \), which is also called the “microscopic” time scale. The component \( X_\varepsilon \) evolves on 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 that it should be asymptotically small in the following sense:

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

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

in probability.

The aim of this section is to get the averaging principle for the “slow” component $X_\varepsilon$. We emphasize that the adopted framework is quite general. In particular,

- the two-scale system (2.4) is *fully coupled* in a sense that the coefficients of the “slow” component depend on the “fast” component, and vice versa;
- it is allowed that the noises for the “slow” and “fast” components can be dependent;
- the coefficients of the “slow” component can be discontinuous.

We now introduce additional assumptions for system (2.4). Note that all assumptions listed below are quite natural and nonrestrictive.

$H_1$ (bounds for the coefficients). There exists a constant $C$ such that

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

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

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

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

$H_2$ (bounds for the Lévy measures). There exist constants $C$ and $p > 0$ such that

$$\int_{\mathbb{R}^m} (|u|^2 \land 1) \nu^\varepsilon(du) \leq C, \quad \int_{\mathbb{R}^l} (|z|^2 1_{|z| \leq 1} + |z|^p 1_{|z| > 1}) \mu^\varepsilon(dz) < \infty.$$

$H_3$ (the coefficients of the fast component are convergent). There exist continuous functions $A(x,y)$, $\Sigma(x,y)$, and $C(x,y,z)$ such that

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

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

To introduce the next condition, we define the *weak convergence* of a family of Lévy measures on $\mathbb{R}^m$ in the following way:

$$\nu^\varepsilon(du) \Rightarrow \nu(du)$$
if, for every continuous function $\varphi$ whose support is compactly embedded into $\mathbb{R}^m \setminus \{0\}$, we have

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

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

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

In addition,

$$\nu\left(\{u: |u| = \rho\}\right) = 0, \quad \mu\left(\{z: |z| = \rho\}\right) = 0. \quad (2.5)$$

Condition (2.5) implies that the cutoff functions $1_{|u| \leq \rho}$ and $1_{|z| \leq \rho}$ used in (2.4) are a.s. continuous w.r.t. the measures $\nu(du)$ and $\mu(dz)$, respectively. Note that there exists an at most countable set of levels $\rho$ such that (2.5) is not true. Hence, one can always choose $\rho$ satisfying this condition. Clearly, changing the cutoff level would affect the drift coefficients respectively.

Further, we assume that the drift of the fast component realizes attraction to the origin.

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

$$A^\varepsilon(x, y) \cdot y \leq -c|y|^\kappa + 1, \quad |y| \geq r. \quad (2.6)$$

In addition, the following balance condition is satisfied:

$$\kappa + p > 1, \quad (2.7)$$

where $p$ is a constant introduced in assumption H$_2$.

Consider a family of “frozen microscopic equations”

$$dy(t) = A(x, y(t)) \, dt + \Sigma(x, y(t-)) \, dW_t$$

$$+ \int_{\mathbb{R}^l} C(x, y(t-), z) \left[ Q(dz, ds) - 1_{|z| \leq 1} \mu(dz)ds \right], \quad (2.8)$$

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.

**F$_0$** (“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 $P^{(x)}_t(y, dy')$ for the corresponding family of transition probabilities.
We also denote by

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

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

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

**F1** ("frozen microscopic dynamics" is locally mixing). There exists \(h > 0\) such that, for any \(R > 0\), one can find \(\rho = \rho_R > 0\) such that, for any \(x, y_1,\) and \(y_2\) with \(|x| \leq R, |y_1| \leq R,\) and \(|y_2| \leq R,\)

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

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

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

Note that assumptions **F1** and **H5** 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 explicit rate; see Proposition 5.1 in what follows.

For the coefficients of the "slow" component, we assume a weaker analog of **H3**, where the convergence and continuity of the limit coefficients may be not true on an exceptional set, which should be, in a certain sense, negligible.

**H6** (the coefficients of the slow component are convergent). There exist functions \(a(x, y), \sigma(x, y),\) and \(c(x, y, u)\) and an open set \(B \subset \mathbb{R}^d \times \mathbb{R}^k\) such that, for any compact set \(K \subset B,\)

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

uniformly on \(K\) and, for any \(R > 1,\)

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

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

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

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.
We define the procedure of averaging of the limit drift coefficient for the macroscopic component w.r.t. the family of IPMs for the frozen microscopic component:

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

Further, we consider the limit diffusion matrix and compensated/noncompensated jump kernels for the macroscopic component:

$$b(x) = \sigma(x)\sigma(x)^{\ast}, \quad K_{(\rho)}(x,y,A) = \nu\{u : |u| \leq \rho, c(x,y,u) \in A\}, \quad K^{(\rho)}(x,y,A) = \nu\{u : |u| > \rho, c(x,y,u) \in A\},$$

and introduce the corresponding averaged characteristics as follows:

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

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

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

Finally, we introduce an auxiliary technical assumption.

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

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

Remark 2.5. It is easy to present a sufficient condition for the validity of $A_0$. Namely, it is sufficient to assume that, in addition to $H_0$–$H_6$, $F_0$, and $F_1$, the transition probabilities $P_t^{(x)}(y,dv)$ are continuous in $x$ w.r.t. the total variation convergence for each $y \in \mathbb{R}^k$, $t \geq t_0$. Then, in view of convergence (5.6), the same continuity holds for the family of the IPMs $\pi(x)(dy)$. The indicated continuity combined with $H_1$, $H_2$, $H_4$, and $H_6$, yields the required continuity of the averaged coefficients.

We are now ready to formulate our main statement.

Theorem 2.3. Assume that $H_0$–$H_6$, $F_0$, $F_1$, and $A_0$ are satisfied,

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

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

\[
L\varphi(x) = \nabla \varphi(x) \cdot \pi(x) + \frac{1}{2} \nabla^2 \varphi(x) \cdot \overline{b}(x)
\]

\[
+ \int_{\mathbb{R}^m} (\varphi(x + v) - \varphi(x) - \nabla \varphi(x) \cdot v) \overline{K}(\rho)(x, dv)
\]

\[
+ \int_{\mathbb{R}^m} (\varphi(x + v) - \varphi(x)) \overline{K}(\rho)(x, dv)
\]

\[
= \nabla \varphi(x) \cdot \overline{\pi}(x) + \frac{1}{2} \nabla^2 \varphi(x) \cdot \overline{b}(x)
\]

\[
+ \int_{\mathbb{R}^k} \int_{\mathbb{R}^m} \left( \varphi(x + c(x, y, u)) - \varphi(x) - \nabla \varphi(x) \cdot c(x, y, u) \mathbb{1}_{|u| \leq \rho} \right) \nu(du)\pi(x)(dy), \tag{2.9}
\]

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 completely repeats 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 the boundedness of the functions \( \varphi, \psi, b, \) and \( \beta \).

Therefore, to prove the theorem, it suffices to verify that any subsequence \( \{(X_{\varepsilon_{n_k}}, Y_{\varepsilon_{n_k}})\} \) contains a subsequence \( \{(X_{\varepsilon_{n_{k_k}}}, Y_{\varepsilon_{n_{k_k}}})\} \) that converges to the desired limit. Without loss of generality, we can assume that \( \{(X_\varepsilon, Y_\varepsilon)\} \) is itself weakly convergent.

**Step 2.** Estimation of the time spent by \( Y_\varepsilon \) in a neighborhood of 0.

We use the following general statement:

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

\[
d\eta_\varepsilon(t) = a_\varepsilon(t)\eta_\varepsilon(t)dt + \varepsilon b_\varepsilon(t)dW(t),
\]

\[
\eta_\varepsilon(0) = 0,
\]

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

\[
a_\varepsilon(t) \geq A > 0, \quad 0 < C_1 \leq b_\varepsilon(t) \leq C_2
\]

for all \( \omega, t, \varepsilon \).
Let

\[ \tau_\epsilon(\delta) := \inf \{ t \geq 0 : |\eta_\epsilon(t)| \geq \delta \} . \]

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

\[ \forall \delta > 0 \; \exists \varepsilon_0 > 0 \; \forall \varepsilon \in (0, \varepsilon_0) : \mathbb{E} \tau_\epsilon(\delta) \leq K \delta^{1-\gamma} . \]

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

Without loss of generality we can assume that

\[ \psi^\pm(x, 0) \geq c_1 > 0 \quad \text{and} \quad 0 < c_2 \leq \beta(x) \leq c_3 \quad \text{for all} \quad x \in \mathbb{R}^d, \]

where \( c_{1,2,3} \) are positive constants. This assumption does not restrict generality because the general case can be considered by using localization. Under this additional assumption, Lemma 3.1 applied to

\[ \tau_\epsilon(\delta) := \inf \{ t \geq 0 : |Y_\epsilon(t)| \geq \delta \} , \]

and the Chebyshev inequality yield

\[ \forall \delta > 0 \; \exists \varepsilon_0 > 0 \; \forall \varepsilon \in (0, \varepsilon_0) : \mathbb{P}(\tau_\epsilon(\delta) \geq \delta^{\frac{1+\gamma}{\alpha}}) \leq K \delta^{\frac{1+\gamma}{\alpha}} . \]  

(3.2)

**Remark 3.1.** It follows from the construction of \( Y^\pm \) that inequality (3.2) is also valid for

\[ \tau^\pm(\delta) := \inf \{ t \geq 0 : |Y^\pm(t)| \geq \delta \} . \]

**Step 3.** It is easy to see from (3.2) that, with high probability, the random variable \( \tau_\epsilon(\delta) \) is dominated by \( \delta^{\frac{1+\gamma}{\alpha}} \). Further, it follows from the standard estimates for the moments of SDEs that, for small \( t \), we have

\[ \mathbb{E} \sup_{s \in [0, t]} |X_\epsilon(s) - x^0|^2 \leq Ct , \]

where the constant \( C \) can be selected independently of \( \epsilon \in [0, 1] \).

Hence, we get the following estimates:

\[ \exists C_1 > 0 \; \forall \delta > 0 \; \exists \varepsilon_0 > 0 \; \forall \varepsilon \in (0, \varepsilon_0) : \mathbb{P}\left( \sup_{t \in [0, \tau_\epsilon(\delta)]} |X_\epsilon(t) - x^0| \geq \delta^{\frac{1+\gamma}{\alpha}} \right) \leq C_1 \delta^{\frac{1+\gamma}{\alpha}} , \]  

(3.3)

\[ \mathbb{P}\left( \sup_{t \in [0, \tau_\epsilon(\delta)]} |X^\pm(t) - x^0| + |Y^\pm(t)| \geq 2\delta^{\frac{1+\gamma}{\alpha}} \right) \leq C_1 \delta^{\frac{1+\gamma}{\alpha}} . \]  

(3.4)

We now verify, e.g., (3.3):

\[ \mathbb{P}\left( \sup_{t \in [0, \tau_\epsilon(\delta)]} |X_\epsilon(t) - x^0| \geq \delta^{\frac{1+\gamma}{\alpha}} \right) \leq \mathbb{P}(\tau_\epsilon(\delta) > \delta^{\frac{1+\gamma}{\alpha}}) + \mathbb{P}\left( \sup_{t \in [0, \frac{\delta}{2}]} |X_\epsilon(t) - x^0| \geq \delta^{\frac{1+\gamma}{\alpha}} \right) \]

where
Let \( \varepsilon > 0 \) be such that
\[
\sup_{t \in [0,\tau_\varepsilon(\delta)]} |Y_\varepsilon(t)| = |Y_\varepsilon(\tau_\varepsilon(\delta))| = \delta \quad \text{a.s.}
\]
by the definition of \( \tau_\varepsilon(\delta) \).

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

If we want to highlight that \( y > 0 \) (or \( y < 0 \)), then the corresponding solution is denoted by
\[
(X^{+,x,y}(t), Y^{+,x,y}(t)) \quad \text{or} \quad (X^{-,x,y}(t), Y^{-,x,y}(t)), \quad \text{respectively}.
\]
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 obtain
\[
\sup_{t \in [0,T]} (|X_\varepsilon(t) - X^+(t)| + |Y_\varepsilon(t) - Y^+(t)|)
\]
\[
\leq \sup_{t \in [0,T]} \left( |X_\varepsilon(\tau_\varepsilon(\delta)) + t - X^{+,x,y}(\tau_\varepsilon(\delta)), \bar{Y}_\varepsilon(\tau_\varepsilon(\delta))(t)|
\right.
\]
\[
+ |Y_\varepsilon(\tau_\varepsilon(\delta)) + t - Y^{+,x,y}(\tau_\varepsilon(\delta)), \bar{Y}_\varepsilon(\tau_\varepsilon(\delta))(t)|
\]
\[
+ \sup_{t \in [0,T]} \left( |X^{+,x,y}(\tau_\varepsilon(\delta)), \bar{Y}_\varepsilon(\tau_\varepsilon(\delta))(t) - X^+(\tau_\varepsilon(\delta)) + t|
\right.
\]
\[
+ |Y^{+,x,y}(\tau_\varepsilon(\delta)), \bar{Y}_\varepsilon(\tau_\varepsilon(\delta))(t) - Y^+(\tau_\varepsilon(\delta)) + t|
\]
\[
+ \sup_{t \in [0,T]} \left( |X^+(\tau_\varepsilon(\delta)) + t - X^+(t)| + |Y^+(\tau_\varepsilon(\delta)) + t - Y^+(t)|
\right.
\]
\[
+ \sup_{t \in [0,\tau_\varepsilon(\delta)]} \left( |X_\varepsilon(t) - x^0| + |Y_\varepsilon(t)|
\right)
\]
\[
= I_1 + \ldots + I_4.
\]

We choose a small number \( \delta > 0 \) and, after this, select \( \varepsilon_0 > 0 \) from (3.2). It follows from (3.3), (3.4), and the construction of \((X^+, Y^+)\) in Subsection 2.1 that \( I_2, I_3, \) and \( I_4 \) are small with high probability.
To estimate $I_1$, we need the following statement for integral equations: Let $f(t) = (f_X(t), f_Y(t))$ be a non-random continuous function. Assume that functions $X_{(f)}^{\pm,x,y}$ and $Y_{(f)}^{\pm,x,y}$ satisfy the integral equation

$$X_{(f)}^{\pm,x,y}(t) = x + \int_0^t \psi^\pm(X_{(f)}^{\pm,x,y}(s), Y_{(f)}^{\pm,x,y}(s)) \, ds + f_X(t),$$

$$Y_{(f)}^{\pm,x,y}(t) = \int_0^t \varphi^\pm(X_{(f)}^{\pm,x,y}(s), Y_{(f)}^{\pm,x,y}(s)) \gamma(s) \, ds + f_Y(t), \quad t \in [0, T],$$

$$X_{(f)}^{\pm,x,y}(0) = x, \quad Y_{(f)}^{\pm,x,y}(0) = y.$$

**Remark 3.2.** We do not assume that a pair $X_{(f)}^{\pm,x,y}$, $Y_{(f)}^{\pm,x,y}$ is a unique solution. We also recall that the domains of $\psi^\pm$ and $\varphi^\pm$ coincide with the whole space.

**Lemma 3.2.**

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

\[ \forall y \in \left[\frac{1}{R}, R\right] : \quad \left| X_{(f)}^{+,x,y}(t) - X_{(f)}^{-,x,y}(t) \right| + \left| Y_{(f)}^{+,x,y}(t) - Y_{(f)}^{-,x,y}(t) \right| \leq \delta, \]

\[ \forall y \in [-R, -1/R] : \quad \left| X_{(f)}^{-,x,y}(t) - X_{(f)}^{-,x,y}(t) \right| + \left| Y_{(f)}^{-,x,y}(t) - Y_{(f)}^{-,x,y}(t) \right| \leq \delta. \]

The proof of the lemma is standard. Note that if $\alpha$ is sufficiently small, then

$$Y_{(f)}^{\pm,x,y}(t) \neq 0, \quad t \in [0, T],$$

and the coefficients of the integral equations are locally Lipschitz continuous for $y \neq 0$.

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

$$\left| X_\varepsilon(\tau_\varepsilon(\delta) + t) - X_{+,x}(\tau_\varepsilon(\delta)), Y_\varepsilon(\tau_\varepsilon(\delta)) (t) \right|$$

$$+ \left| Y_\varepsilon(\tau_\varepsilon(\delta) + t) - Y_{+,x}(\tau_\varepsilon(\delta)), Y_\varepsilon(\tau_\varepsilon(\delta)) (t) \right|$$

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

where

$$f(t) := \left\{ \begin{array}{ll} \tau_\varepsilon(\delta) + t & \varepsilon \int_{\tau_\varepsilon(\delta)}^t b(X_\varepsilon(s), Y_\varepsilon(s)) \, dB(s), \varepsilon \int_0^t \beta(X_\varepsilon(s), Y_\varepsilon(s)) \, dW(s) \end{array} \right\}. $$
Since $b$ and $\beta$ are bounded, we get uniform convergence in probability:

$$
\varepsilon \sup_{t \in [0,T]} \left( \int_{\tau_{\varepsilon}(\delta)}^{\tau_{\varepsilon}(\delta)+t} b(X_{\varepsilon}(s), Y_{\varepsilon}(s)) dB(s) + \int_0^t \beta(X_{\varepsilon}(s), Y_{\varepsilon}(s)) dW(s) \right) \xrightarrow{\mathbb{P}} 0, \quad \varepsilon \to 0,
$$

for any $\delta > 0$.

This fact, (3.3), (3.5), and Lemma 3.2 imply the following convergence:

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

for any $\delta > 0$.

**Step 5.** The proof of the theorem follows from Step 4 and the following estimate for the probabilities

$$
\mathbb{P}(Y_{\varepsilon}(\tau_{\varepsilon}(\delta)) = \pm \delta):
$$

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

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

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

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

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

$$
|\varphi^\pm(x, y) - \varphi^\pm(x^0, 0)| \leq \nu,
$$

$$
0 < (\beta^\pm(x^0, 0))^2 - \nu < (\beta^\pm(x, y))^2 < (\beta^\pm(x^0, 0))^2 + \nu
$$

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

We set

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

It follows from (3.3) that

$$
\mathbb{P}(\sigma_{\varepsilon}(\delta^{\frac{1+\alpha}{6}}) < \tau_{\varepsilon}(\delta)) < C\delta^{\frac{1+\gamma}{6}}
$$

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 $Y_{\varepsilon}$ leaves $[-\delta, \delta]$ prior to $X_{\varepsilon}$ exits $[-\delta_1, \delta_1]$. Therefore, without loss of generality, we can assume that (3.6) is satisfied for all $(x, y)$.
We set
\[
s_\varepsilon(y) := \begin{cases} \int_0^y \exp \left\{ \frac{2(\varphi^+(x^0, 0) + \nu) z^{\gamma+1}}{\varepsilon^2(\gamma + 1)(\beta^+(x^0, 0)^2 - \nu)} \right\} \, dz, & y \geq 0, \\ \int_0^y \exp \left\{ \frac{2(\varphi^-(x^0, 0) - \nu)|z|^{\gamma+1}}{\varepsilon^2(\gamma + 1)(\beta^-(x^0, 0)^2 + \nu)} \right\} \, dz, & y \leq 0. \end{cases}
\]

Then
\[
\varphi(x, y)y^{\gamma}s_\varepsilon'(y) + \frac{\varepsilon^2}{2}\beta^2(x, y)s_\varepsilon''(y) \leq 0
\]
for all \(x, y\) (recall that we assume that (3.6) is satisfied for all \((x, y)\)). Hence,
\[
0 \geq \mathbf{E}s_\varepsilon(Y_\varepsilon(\tau_\varepsilon(\delta))) = s_\varepsilon(\delta)\mathbf{P}(Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta) + s_\varepsilon(-\delta)\mathbf{P}(Y_\varepsilon(\tau_\varepsilon(\delta)) = -\delta)
\]
\[
= s_\varepsilon(\delta)\mathbf{P}(Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta) + s_\varepsilon(-\delta)(1 - \mathbf{P}(Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta))
\]
\[
= (s_\varepsilon(\delta) - s_\varepsilon(-\delta))\mathbf{P}(Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta) + s_\varepsilon(-\delta).
\]

Therefore,
\[
\limsup_{\varepsilon \to 0} \mathbf{P}(Y_\varepsilon(\tau_\varepsilon(\delta)) = \delta)
\]
\[
\leq \lim_{\varepsilon \to 0} \frac{-s_\varepsilon(-\delta)}{s_\varepsilon(\delta) - s_\varepsilon(-\delta)}
\]
\[
= \lim_{\varepsilon \to 0} \frac{\int_{-\delta}^0 \exp \left\{ \frac{2(\varphi^-(x^0, 0) - \nu)|z|^{\gamma+1}}{\varepsilon^2(\gamma + 1)((\beta^-(x^0, 0)^2 + \nu)} \right\} \, dz + \int_0^\delta \exp \left\{ \frac{2(\varphi^+(x^0, 0) + \nu)z^{\gamma+1}}{\varepsilon^2(\gamma + 1)((\beta^+(x^0, 0)^2 - \nu)} \right\} \, dz}{\int_{-\delta}^0 \exp \left\{ \frac{2(\varphi^-(x^0, 0) - \nu)|z|^{\gamma+1}}{\varepsilon^2(\gamma + 1)((\beta^-(x^0, 0)^2 + \nu)} \right\} \, dz + \int_0^\delta \exp \left\{ \frac{2(\varphi^+(x^0, 0) + \nu)z^{\gamma+1}}{\varepsilon^2(\gamma + 1)((\beta^+(x^0, 0)^2 - \nu)} \right\} \, dz}
\]
\[
= \frac{\left( \frac{\varphi^+(x^0, 0) + \nu}{(\beta^+(x^0, 0)^2 - \nu) \frac{1}{\gamma+1}} \right)^{\frac{1}{\gamma+1}} + \left( \frac{\varphi^-(x^0, 0) - \nu}{(\beta^-(x^0, 0)^2 + \nu) \frac{1}{\gamma+1}} \right)^{\frac{1}{\gamma+1}}}{\left( \frac{\varphi^+(x^0, 0) + \nu}{(\beta^+(x^0, 0)^2 - \nu) \frac{1}{\gamma+1}} \right)^{\frac{1}{\gamma+1}} + \left( \frac{\varphi^-(x^0, 0) - \nu}{(\beta^-(x^0, 0)^2 + \nu) \frac{1}{\gamma+1}} \right)^{\frac{1}{\gamma+1}}}.\]

Here, we have used the following fact:
\[
\int_0^\delta \exp \left\{ \frac{-A z^{\gamma+1}}{\varepsilon^2} \right\} \, dz = \frac{1}{1 + \gamma} \left( \frac{\varepsilon^2}{A} \right)^{\frac{1}{\gamma+1}} \int_0^A e^{-t \frac{1}{\gamma+1} - 1} \, dt
\]
\[ \sim \frac{1}{1+\gamma} \left( \frac{\varepsilon^2}{A} \right)^{\frac{1}{1+\gamma}} \Gamma\left(\frac{1}{1+\gamma}\right), \quad \varepsilon \to 0, \]

for any \( A > 0 \) and \( \delta > 0 \).

Since \( \nu \) is arbitrary, this completes the proof of Lemma 3.3 and Theorem 2.1.

4. Proof of Theorem 2.2

First, we note that

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

Indeed, by the Itô formula, we get

\[ Y_\varepsilon^2(t) \leq C\varepsilon^2 + 2\varepsilon \int_0^t Y_\varepsilon(s)\beta(X_\varepsilon(s), Y_\varepsilon(s)) \, dW(s), \]

where \( C \) is independent of \( \varepsilon \). Hence, we arrive at the estimate

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

It follows from the Doob inequality that

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

This completes the proof of (4.1).

Let \( \delta > 0 \) be a fixed number. Note that

\[ \varepsilon^{-\delta} Y_\varepsilon(t) = \varepsilon^{\delta(\gamma-1)} \int_0^t \varphi(X_\varepsilon(s), Y_\varepsilon(s)) \left( \varepsilon^{-\delta} Y_\varepsilon(t) \right)^\gamma \, dt \]

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

\[ = \int_0^t \varphi(X_\varepsilon(s), \varepsilon^{\delta} \varepsilon^{-\delta} Y_\varepsilon(s)) \left( \varepsilon^{-\delta} Y_\varepsilon(t) \right)^\gamma \, d\left( \varepsilon^{\delta(\gamma-1)} t \right) \]

\[ + \varepsilon^{1-\delta(\gamma-1)} \frac{\delta(\gamma-1)}{2} \int_0^t \beta(X_\varepsilon(s), \varepsilon^{\delta} \varepsilon^{-\delta} Y_\varepsilon(s)) \, dW_\varepsilon\left( \varepsilon^{\delta(\gamma-1)} s \right), \]
where
\[ W_\varepsilon(t) = \varepsilon^{\frac{\delta(\gamma-1)}{2}} W\left( e^{-\delta(\gamma-1)t} \right) \]
is a Wiener process.

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

\[
\bar{Y}_\varepsilon(t) = \int_0^t \varphi(X_\varepsilon(s), e^{\frac{2}{\gamma+1}} \bar{Y}_\varepsilon(s)) \bar{Y}_\varepsilon(s) d\left( e^{\frac{2(\gamma-1)}{\gamma+1}} s \right) + \int_0^t \beta(X_\varepsilon(s), e^{\frac{2}{\gamma+1}} \bar{Y}_\varepsilon(s)) dW_\varepsilon\left( e^{\frac{2(\gamma-1)}{\gamma+1}} s \right).
\]

We set \( \bar{\varepsilon} = e^{\frac{2(1-\gamma)}{\gamma+1}} \). Therefore,

\[
\begin{align*}
dX_\varepsilon(t) &= a_\varepsilon(X_\varepsilon(t), \bar{Y}_\varepsilon(t)) \, dt + b_\varepsilon(X_\varepsilon(t), \bar{Y}_\varepsilon(t)) \, dB(t), \\
d\bar{Y}_\varepsilon(t) &= \alpha_\varepsilon(X_\varepsilon(t), \bar{Y}_\varepsilon(t)) \, d\bar{\varepsilon}^{-1} t + \beta_\varepsilon(X_\varepsilon(t), \bar{Y}_\varepsilon(t)) \, dW_\varepsilon(\bar{\varepsilon}^{-1} t),
\end{align*}
\]

where

\[
\begin{align*}
a_\varepsilon(x, y) &= \psi(x, e^{\frac{2}{\gamma+1}} y) = \psi(x, e^{\frac{1}{\gamma-1}} y), \\
b_\varepsilon(x, y) &= \varepsilon b(x, e^{\frac{2}{\gamma+1}} y) = \varepsilon^{(\gamma+1)/2(1-\gamma)} b(x, e^{\frac{1}{\gamma-1}} y), \\
\alpha_\varepsilon(x, y) &= \varphi(x, e^{\frac{2}{\gamma+1}} y) y^\gamma = \varphi(x, e^{\frac{1}{\gamma-1}} y) y^\gamma, \\
\beta_\varepsilon(x, y) &= \beta(x, e^{\frac{2}{\gamma+1}} y) = \beta(x, e^{\frac{1}{\gamma-1}} y).
\end{align*}
\]

We see that system (4.2) has the form (2.4). We now apply Theorem 2.3 with \( k = 1 \)

\[
\begin{align*}
a^\varepsilon(x, y) &:= a_\varepsilon(x, y), \\
\sigma^\varepsilon(x, y) &= c^\varepsilon(x, y, u) = C^\varepsilon(x, y, z) = 0, \\
A^\varepsilon(x, y) &:= \alpha_\varepsilon(x, y), \\
\Sigma^\varepsilon(x, y) &:= \beta_\varepsilon(x, y), \\
\xi^\varepsilon(t) &:= \varepsilon^{(\gamma+1)/2(1-\gamma)} \int_0^t b(X_\varepsilon(s), e^{\frac{1}{\gamma-1}} \bar{Y}_\varepsilon(s)) dB(s).
\end{align*}
\]

Conditions \( H_0, H_1, \) and \( H_2 \) are obviously true.
The functions from condition $H_3$ are as follows:

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

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

Note that $A$ and $\Sigma$ are discontinuous for $y = 0$ and formally $H_3$ is not satisfied. However, the only place in Theorem 2.3, where we use the continuity of $A$ and $\Sigma$, is the identification of the limit points for the sequence $\{y_{\varepsilon_n}\}$ in the proof of Proposition 5.2. Since the diffusion coefficients $\beta^\pm$ are functions separated from zero, it can be seen that the processes $\{y_{\varepsilon_n}\}$ spend small amounts of time in small neighborhoods of 0 uniformly in $\{\varepsilon_n\}$. This implies that any limit point of $\{y_{\varepsilon_n}\}$ solves (2.8) and Proposition 5.2 is true. This is all what we need for the application of Theorem 2.3.

Condition $H_4$ is satisfied with $\nu = \mu = 0$.

Without loss of generality, we can assume that

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

(4.3)

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

Consider the equation with frozen coefficients

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

$$+ \left(\beta^+(x, 0)\mathbb{1}_{y^{(x)}(t) \geq 0} + \beta^-(x, 0)\mathbb{1}_{y^{(x)}(t) < 0}\right)dW(t).$$  

(4.4)

The existence and uniqueness of weak solution to the equation with frozen coefficients and the strong Markov property follow from [10]. Hence, condition $F_6$ is true.

To verify condition $F_1$, we modify the argument from [16] (Section 3.3.2). Since 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 $x$, $y$, and $y'$. In order to overcome this minor difficulty, we use the following localization argument: Consider the SDE

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

(4.5)

This SDE has a constant diffusion coefficient and a bounded and Hölder continuous drift coefficient. Hence, the standard analytic theory (see, e.g., [11]) implies that the transition probability density $p^{(x,+)}(t, y, y')$ is continuous in $x$, $y$, and $y'$. Thus, for $y_0 = 1$ and every $t_0 > 0$, we find

$$\sup_{|x| \leq R} \left\| P_{t_0}^{(x,+)}(y, dy') - P_{t_0}^{(x,+)}(y_0, dy') \right\|_{TV}$$

$$= \sup_{|x| \leq R} \int_{|y| \leq R} \left| P_{t_0}^{(x,+)}(y, dy') - p_{t_0}^{(x,+)}(y_0, dy') \right| dy' \to 0, \quad y \to y_0.$$
The coefficients of equations (4.4), (4.5) coincide on \([0, 2]\). Thus, the laws of the solutions to these equations stopped at the time of exit from \([0, 2]\) coincide. Taking sufficiently small \(t_0\), we can guarantee that each of these solutions stay in \([0, 2]\) up to the time \(t_0\) with probability \(\geq \frac{5}{6}\) if the initial value \(y\) stays in \([\frac{1}{2}, \frac{3}{2}]\). By the coupling characterization of the TV distance ("coupling lemma"; see, e.g., [16], Theorem 2.2.2), this yields that, for the indicated \(t_0\),

\[
\sup_{y_1, y_2 \in \left[\frac{1}{2}, \frac{3}{2}\right], |x| \leq R} \left\| P^{(x)}_{t_0}(y_1, dy') - P^{(x, +)}_{t_0}(y_2, dy') \right\|_{TV} \leq \left(1 - \frac{5}{6}\right) + \left(1 - \frac{5}{6}\right) = \frac{1}{3}.
\]

Combining these two estimates, we see that there exist sufficiently small \(t_0 > 0\) and \(r > 0\) such that

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

On the right-hand side, we can actually take any number \(\geq \frac{1}{3} + \frac{1}{3} = \frac{2}{3}\). 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 the 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 (by using this representation) that, for any \(t_1 > 0\), there exists \(\delta > 0\) such that

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

We take \(h = t_0 + t_1\) and, for \(x, y_1, y_2\) with \(|x| \leq R\), \(|y_1| \leq R\), and \(|y_2| \leq R\), consider two processes \(Y^1_t\) and \(Y^2_t\), which start at \(y_1\) and \(y_2\), respectively, solve Eq. (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 by their values at the time \(t_1\) (we can construct this process by using the coupling lemma for the probability kernels; see [16], Theorem 2.2.4). Then

\[
\left\| P^{(x)}_{h}(y_1, dy') - P^{(x)}_{h}(y_2, dy') \right\|_{TV} \leq P(Y^1_{h} \neq Y^2_{h})
\]

\[
= \int_{\mathbb{R}^2} 4 \left\| P^{(x)}_{t_0}(z_1, dy') - P^{(x)}_{t_0}(z_2, dy') \right\|_{TV} P^{(x)}_{h}(y_1, dz_1) P^{(x)}_{h}(y_2, dz_2)
\]

\[
\leq 1 - P^{(x)}_{t_1}(y_1, [1-r, 1+r]) P^{(x)}_{t_1}(y_2, [1-r, 1+r])
\]

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

\[
\leq 1 + \left(-1 + \frac{3}{4}\right) P^{(x)}_{t_1}(y_1, [1-r, 1+r]) P^{(x)}_{t_1}(y_2, [1-r, 1+r]) \leq 1 - \frac{\delta^2}{4}
\]

for any \(|x| \leq R\), \(|y_1| \leq R\), and \(|y_2| \leq R\), which completes the proof of \(F_1\).
The IPM \( \pi^{(x)}(dy) \) is equal to (see [14], Exercise 5.40):

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

where

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

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

Condition \( H_6 \) is satisfied with

\[
a(x, y) = \psi_+(x, 0) 1_{y > 0} + \psi_-(x, 0) 1_{y < 0}, \quad \sigma(x, y) = c(x, y, z) = 0, \quad \text{and} \quad B = \mathbb{R}^d \times \{0\}.
\]

The averaged coefficient

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

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

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

is Lipschitz continuous, \( \overline{b}(x) = 0 \), and

\[
\overline{K}_{(\rho)}(x, dv) = \overline{K}^{(\rho)}(x, dv) = 0.
\]

Hence, condition \( A_0 \) holds and the corresponding martingale problem possesses a unique solution.

Together with (4.1), this completes 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 assumption of negligibility \( H_0 \) and the assumptions of boundedness \( H_1 \) and \( H_2 \). Under the assumptions of the theorem,
for any $C^\infty_0$-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 $\varepsilon \to 0$ solves the martingale problem (2.9), it suffices to show that, for any $C^\infty_0$-function $\varphi$, any $s_1, \ldots, s_q < s < t$, and any continuous and bounded function $\Phi : \mathbb{R}^{d \times q} \to \mathbb{R}$, we have

$$
\mathbb{E}^\varepsilon \Phi(X_\varepsilon(s_1), \ldots, X_\varepsilon(s_q)) \left[ \varphi(X_\varepsilon(t)) - \varphi(X_\varepsilon(s)) - \int_s^t L\varphi(X_\varepsilon(r)) \, dr \right] \to 0, \quad \varepsilon \to 0. \quad (5.1)
$$

We denote the expectation w.r.t. $P^\varepsilon$ by $\mathbb{E}^\varepsilon$. Also denote

$$
\tilde{X}_\varepsilon(t) = X_\varepsilon(t) - \xi_\varepsilon(t)
= X_\varepsilon(0) + \int_0^t a^\varepsilon(X_\varepsilon(s), Y_\varepsilon(t)) \, ds + \int_0^t \sigma^\varepsilon(X_\varepsilon(s), Y_\varepsilon(s)) \, dB^\varepsilon_s
+ \int_0^t \int_{\mathbb{R}^m} c^\varepsilon(X_\varepsilon(s-), Y_\varepsilon(s-), u) \left[ N^\varepsilon(du, ds) - 1_{|u| \leq \rho} \nu^\varepsilon(du) \, ds \right]. \quad (5.2)
$$

Note that $L\varphi$ is a bounded and continuous function. Hence, by $H_0$, relation (5.1) is equivalent to

$$
\mathbb{E}^\varepsilon \Phi(\tilde{X}_\varepsilon(s_1), \ldots, \tilde{X}_\varepsilon(s_q)) \left[ \varphi(\tilde{X}_\varepsilon(t)) - \varphi(\tilde{X}_\varepsilon(s)) - \int_s^t L\varphi(\tilde{X}_\varepsilon(r)) \, dr \right] \to 0, \quad \varepsilon \to 0. \quad (5.3)
$$

Denote

$$
b^\varepsilon(x, y) = \sigma^\varepsilon(x, y)(\sigma^\varepsilon(x, y))^*,
K^\varepsilon_{(\rho)}(x, y, A) = \nu^\varepsilon(\{u : |u| \leq \rho, c^\varepsilon(x, y, u) \in A\}),
K^{(\rho),\varepsilon}(x, y, A) = \nu^\varepsilon(\{u : |u| > \rho, c^\varepsilon(x, y, u) \in A\}),
$$

and

$$
\mathcal{L}^\varepsilon \varphi(x, y) = \nabla \varphi(x) \cdot a^\varepsilon(x, y) + \frac{1}{2} \nabla^2 \varphi(x) \cdot b^\varepsilon(x, y)
+ \int_{\mathbb{R}^m} \left( \varphi(x + v) - \varphi(x) - \nabla \varphi(x) \cdot v \right) K^\varepsilon_{(\rho)}(x, dv)
+ \int_{\mathbb{R}^m} \left( \varphi(x + v) - \varphi(x) \right) K^{(\rho),\varepsilon}(x, y, dv).
$$
Then, by the Itô formula, we get

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

Applying \(H_0\) once again, we conclude that, in order to prove (5.1) and (5.3), it suffices to show that, for any \(s_1, \ldots, s_q < t\),

\[
E^\varepsilon \Phi(X_\varepsilon(s_1), \ldots, X_\varepsilon(s_q)) \left( \mathcal{L}^\varepsilon \varphi(X_\varepsilon(t), Y_\varepsilon(t)) - L \varphi(X_\varepsilon(t)) \right) \to 0, \quad \varepsilon \to 0.
\] (5.5)

Prior to 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 \(H_1-H_5, F_0, F_1\) be satisfied. 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 \(|x| \leq R\) and \(|y| \leq R\),

\[
\left\| P^x_t(y, dy') - \pi^x(dy') \right\|_{TV} \leq Ct^{-\frac{\kappa+p-1}{1-\kappa}}.
\] (5.6)

If \(\kappa \geq 1\), then there exists \(a > 0\) such that, for every \(R > 0\) and any \(x, y\) with \(|x| \leq R\) and \(|y| \leq R\),

\[
\left\| P^x_t(y, dy') - \pi^x(dy') \right\|_{TV} \leq Ce^{-at}
\]

where \(C\) is a constant depending on \(R\).

**Proof.** The required statement was actually obtained, although not in the proposed precise form, in [16] (Section 3). The difference between the current situation and the situation studied in [16] is that the ergodic rates were obtained in the cited work for individual processes (at the same time, in the present work, we have a family indexed by \(x\)) and separately for diffusions and Lévy driven SDEs (here, we have both types of noise simultaneously). This difference is not crucial, and we just give a short sketch of the applied reasoning and refer the reader to [16] for details.

The convergence conditions \(H_3, H_4\) imply that the bounds from the conditions \(H_1, H_2\) and the drift condition \(H_5\) remain true for the limit coefficients \(A(x, y), \Sigma(x, y), C(x, y, z)\) and the Lévy measure \(\mu(du)\). Thus, we get the following assertion: If \(V \in C^2\) is a function such that

\[
V(y) \geq 1 \quad \text{and} \quad V(y) = |y|^p, \quad |y| \geq 2,
\]

then, for any \(x \in \mathbb{R}^d\), the following semimartingale decomposition is true:

\[
V(y^{(x)}(t)) = V(y^{(x)}(0)) + \int_0^t AV(x, y^{(x)}(s)) \, ds + \text{(martingale part),} \tag{5.7}
\]
where the function $\mathcal{A}V(x,y)$ satisfies

\[
\mathcal{A}V(x,y) \leq \begin{cases} 
C_V - a_V V(y) \frac{p+\kappa-1}{p}, & \kappa \in (0,1), \\
C_V - a_V V(y), & \kappa \geq 1,
\end{cases}
\]  

(5.8)

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

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

\[
E_y \tilde{V}(y^{(x)}(h)) - \tilde{V}(y) \leq \tilde{C}_V - \tilde{c}_V V(y) \frac{p+\kappa-1}{p},
\]  

(5.9)

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

$$c_1 V \leq \tilde{V} \leq c_2 V.$$  

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

Inequality (5.9) is actually the Lyapunov condition for the skeleton chain

$$y^{(x,h)}_k = y^{(x)}(kh), \quad k \geq 0,$$

for the process $y^{(x)}$; see [16] (Section 2.8). In combination with the local Dobrushin condition assumed in $F_1$, by virtue of [16] (Corollary 2.8.10), we get the inequality

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

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 the Lyapunov condition and the local Dobrushin condition are uniform in $x$, the constant $C$ can be chosen, in this case, uniformly for $x \in \mathbb{R}^d$. One can easily check this following the proofs of [16] (Corollary 2.8.10) and the theorems on which this corollary is based (Theorems 2.7.5 and 2.8.6 in [16]). Since the total variation distance

$$\left\| P_t(x,y,dy') - \pi(x)(dy') \right\|_{TV}$$

is nonincreasing in $t$ and $V(y)$ is locally bounded, this completes the proof of the required assertion in the case $\kappa \in (0,1)$.

For $\kappa \geq 1$, we can argue in a completely similar way by using [16] (Corollary 2.8.3).
5.2. Auxiliaries II: Weak Convergence of the Microscopic Dynamics to the Frozen Dynamics. Consider the following microscopic analog of (2.4). Assume that \((x_\varepsilon, y_\varepsilon)\) is a solution (maybe not unique) of the following equations:

\[
x_\varepsilon(t) = x_\varepsilon(0) + \varepsilon \int_0^t a^\varepsilon(x_\varepsilon(s), y_\varepsilon(t)) \, ds + \varepsilon^{1/2} \int_0^t \sigma^\varepsilon(x_\varepsilon(s), y_\varepsilon(s)) \, dB_s^\varepsilon \\
+ \int_0^t \int_{\mathbb{R}^m} C^\varepsilon(x_\varepsilon(s-), y_\varepsilon(s-), u) \left[ n^\varepsilon(du, ds) - 1_{|u| \leq \rho} \varepsilon \nu^\varepsilon(du) \, ds \right] + \zeta^\varepsilon(t),
\]

\[
y_\varepsilon(t) = y_\varepsilon(0) + \int_0^t A^\varepsilon(x_\varepsilon(s), y_\varepsilon(s)) \, ds + \int_0^t \Sigma^\varepsilon(x_\varepsilon(s), y_\varepsilon(s)) \, dw_s^\varepsilon \\
+ \int_0^t \int_{\mathbb{R}^l} c^\varepsilon(x_\varepsilon(s-), y_\varepsilon(s-), z) \left[ q^\varepsilon(dz, ds) - 1_{|z| \leq \rho} \mu^\varepsilon(dz) \, ds \right],
\]

(5.10)

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

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

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

(5.11)

For a fixed pair of functions \((\rho(\varepsilon), q(\varepsilon))\) such that \(\rho(\varepsilon) \to 0\) and \(q(\varepsilon) \to 0\) as \(\varepsilon \to 0\) and constants \(R > 0\) and \(T > 0\), we denote by \(\mathcal{K}(\rho, q, R, T)\) the class of all families \(\{(x_\varepsilon, y_\varepsilon), \varepsilon > 0\}\) satisfying (5.10) in a certain probability space with nonrandom initial values \(x_\varepsilon(0)\) and \(y_\varepsilon(0)\) such that \(|x_\varepsilon(0)| \leq R\), \(|y_\varepsilon(0)| \leq R\), and

\[
\bar{P}^\varepsilon \left( \sup_{s \leq T} |\zeta_\varepsilon(s)| > \rho(\varepsilon) \right) \leq q(\varepsilon).
\]

Proposition 5.2. Let conditions \(H_1 - H_5\) and \(F_0\) be satisfied. Then, for any \(0 < t < T\), any bounded continuous function \(f : \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}\), and \(R > 0\),

\[
\sup_{\{(x_\varepsilon, y_\varepsilon) \in \mathcal{K}(\rho, q, R, T)\}} \left| \bar{E}^\varepsilon f(x_\varepsilon(t), y_\varepsilon(t)) - P_t^{\text{frozen}} f(x, y) \right|_{x=x_\varepsilon(0), y=y_\varepsilon(0)} \to 0, \quad \varepsilon \to 0,
\]

(5.12)

where

\[
P_t^{\text{frozen}} f(x, y) = \int_{\mathbb{R}^k} f(y') P_t^{(x)}(y', dy'), \quad t \geq 0.
\]
Proof. Assume the contrary, i.e., that there exists a sequence \( x_{\varepsilon_n}(\cdot) \), \( y_{\varepsilon_n}(\cdot) \) of solutions to (5.10) with 
\[
| x_{\varepsilon_n}(0) | \leq R \quad \text{and} \quad | y_{\varepsilon_n}(0) | \leq R \quad \text{such that}
\]
\[
\left( \mathbb{E}^{\varepsilon_n} f(x_{\varepsilon_n}(t), y_{\varepsilon_n}(t)) - P^x(t) f(x, y) \right)_{| x=x_{\varepsilon_n}(0), y=y_{\varepsilon_n}(0) |} \not\to 0, \quad n \to \infty. \tag{5.13}
\]

Without loss of generality, after passing to a subsequence, we can assume that
\[
x_{\varepsilon_n}(0) \to x_* \quad \text{and} \quad y_{\varepsilon_n}(0) \to y_* \quad \text{as} \quad n \to \infty.
\]

Then it is easy to show that, for any \( c > 0 \),
\[
\lim_{n \to \infty} \mathbb{P}^{\varepsilon_n} \left( \sup_{s \in [0,T]} | x_{\varepsilon_n}(s) - x_* | > c \right) = 0. \tag{5.14}
\]

Further, we denote by \( P^* \) the law in \( D([0, T], \mathbb{R}^k) \) of \( y^{(x_*)} \) with \( y^{(x_*)}(0) = y_* \). Since the \( P^* \)-probability for \( y(\cdot) \) to have a jump at the point \( t \) is equal to 0, the function \( F(y(\cdot)) = f(y(t)) \) is a.s. continuous on \( D([0, T], \mathbb{R}^k) \). Thus, in order to prove that (5.13) is not true, it is sufficient to show that the laws of \( y_{\varepsilon_n}, n \geq 1 \), weakly converge in \( D([0, T], \mathbb{R}^k) \) to \( P^* \). This statement is quite standard, and, therefore, we only outline its proof in what follows.

By (5.14), the continuity assumption \( H_3 \), and the convergence of noise \( H_4 \), one can easily prove that any weak limit point for \( \{ y_{\varepsilon_n} \} \) solves (2.8). By the weak uniqueness assumption \( F_0 \), this implies that any weak limit point for \( \{ y_{\varepsilon_n} \} \) has the law \( P^* \).

Thus, in order to prove the required weak convergence it is sufficient to show that \( \{ y_{\varepsilon_n} \} \) is weakly compact in \( D([0, T], \mathbb{R}^k) \).

To prove the property of weak compactness, we use \( L_2 \)-moment bounds for the increments of the process \( y_{\varepsilon_n} \) combined with the truncation of large jumps. Namely, by \( H_2 \), for any fixed \( \delta > 0 \), there exists \( Q_{\delta} \) such that
\[
\mathbb{P}^{\varepsilon} \left( N([0, T] \times \{ |z| > Q_{\delta} \}) > 0 \right) < \delta, \quad \varepsilon > 0.
\]

Thus, it is sufficient to prove weak compactness for every “truncated” family \( \{ y_{\varepsilon_n, Q} \} \), \( Q > 0 \), where \( y_{\varepsilon_n, Q} \) satisfies an analog of (5.10) with the integral for \( q^s \) taken over \( \{ |z| \leq Q \} \) instead of \( \mathbb{R}^l \). For this “truncated” family, by applying [17] (Proposition 2.5), we get
\[
| y_{\varepsilon_n, Q}(s) |^2 = | y_{\varepsilon_n, Q}(0) |^2 + \int_0^s H(r) \, dr + \text{(martingale part)}, \tag{5.15}
\]

where \( H \) is bounded. Combining this fact with the maximal martingale inequality, we conclude that
\[
\mathbb{E}^{\varepsilon_n} \sup_{s \in [0,T]} | y_{\varepsilon_n, Q}(s) |^2
\]
is bounded. Since the coefficient \( A^\varepsilon(x, y) \) is bounded locally in \( y \), the above bound and the (uniform) bounds for \( C^\varepsilon \) and \( \mu^\varepsilon \) from \( H_1 \) and \( H_2 \) yield the required weak compactness of \( \{ y_{\varepsilon_n} \} \). Summarizing all arguments presented above, we conclude that \( \{ y_{\varepsilon_n} \} \) weakly converges to \( P^* \). In combination with (5.14), this contradicts (5.13) and proves the required statement.
5.3. End of the Proof of Theorem 2.3. In this section, we complete the proof of relation (5.5) and, hence, the proof of the theorem. Denote

\[ L \phi(x, y) = \nabla \phi(x) \cdot a(x, y) + \frac{1}{2} \nabla^2 \phi(x) \cdot b(x, y) \]

\[ + \int_{\mathbb{R}^m} \left( \varphi(x + v) - \varphi(x) - \nabla \varphi(x) \cdot v \right) K(\rho)(x, y, dv) \]

\[ + \int_{\mathbb{R}^m} \left( \varphi(x + v) - \varphi(x) \right) K(\rho)(x, y, dv) \]

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

\[ + \int_{\mathbb{R}^m} \left( \varphi(x + c(x, y, u)) - \varphi(x) - \nabla \varphi(x) \cdot c(x, y, u) \mathbb{1}_{|u| \leq \rho} \right) \nu(du). \]  

(5.16)

Further, since the set \( B \) from condition \( H_6 \) is open, there exists a sequence of continuous functions \( \chi_j(x, y) \), \( j \geq 1 \), such that

(i) \( 0 \leq \chi_j(x, y) \leq 1, j \geq 1 \);

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

(iii) for each \( x \) and \( y \), we have \( \chi_j(x, y) \rightharpoonup \chi_\infty(x, y) = 1_B(x, y) \) as \( j \to \infty \).

Recall the notation

\[ \overline{f}(x) = \int_{\mathbb{R}^k} f(x, y) \pi(x)(dy). \]

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

Lemma 5.1. The following properties hold:

(a) there exists \( C > 0 \) such that \( |L^\varepsilon \phi(x, y)| \leq C, \varepsilon > 0, \) for all \( x, y \);

(b) \( L^\varepsilon \phi \to L \phi, \varepsilon \to 0 \), uniformly on each compact set \( K \subset B \);

(c) there exists \( C > 0 \) such that \( |L \phi(x, y)| \leq C \) for all \((x, y) \in B \);

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

(e) \( \chi_j L \phi(x) \to L \phi(x), j \to \infty \), uniformly on \( \{|x| \leq R\} \) for each \( R > 0 \), where \( L \) is defined in (2.9);
(f) for any $T > 0$,
\[
\mathbb{E}^\varepsilon \left| \tilde{X}_\varepsilon(t) - \tilde{X}_\varepsilon(s) \right|^2 \leq 1 \leq C|t - s|, \quad s, t \in [0, T],
\]
where $\tilde{X}$ is defined in (5.2), and
\[
\sup_{t \in [0, T], \varepsilon > 0} \mathbb{P}^\varepsilon (|X_\varepsilon(t)| > R) \to 0, \quad R \to \infty;
\]

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

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

To prove statement (d), we first mention that each function $\chi_j$ is continuous by assumption $F_0$. These functions monotonically converge, for each $x \in \mathbb{R}^d$, to the function
\[
\chi_\infty(x) = \int_{\mathbb{R}^k} 1_{B}(x, y)\pi(x)(dy) = 1,
\]
where the last identity holds by assumption $H_6$. Then the required uniform convergence follows from the Dini theorem.

To prove statement (e), we first use statements (c) and (d) to get
\[
\left| \chi_j \mathcal{L}\varphi(x) - \mathcal{L}\varphi(x) \right| = \left| \chi_j \mathcal{L}\varphi(x) - \chi_\infty \mathcal{L}\varphi(x) \right| \leq C(1 - \chi_j(x)) \to 0, \quad j \to \infty,
\]
uniformly for $x$ with $|x| \leq R$. Then the required statement follows from the identity
\[
\mathcal{L}\varphi(x) = \int_{\mathbb{R}^k} \mathcal{L}\varphi(x, y)\pi(x)(dy)
\]
\[
= \int_{\mathbb{R}^k} \left( \nabla \varphi(x) \cdot a(x, y) + \frac{1}{2} \nabla^2 \varphi(x) \cdot b(x, y) \right)\pi(x)(dy)
\]
\[
+ \int_{\mathbb{R}^k} \int_{\mathbb{R}^m} \left( \varphi(x + v) - \varphi(x) - \nabla \varphi(x) \cdot v \right) K(\rho)(x, dv)\pi(x)(dy)
\]
\[
+ \int_{\mathbb{R}^k} \int_{\mathbb{R}^m} \left( \varphi(x + v) - \varphi(x) \right) K(\rho)(x, y, dv)\pi(x)(dy)
\]
Therefore, by assumption $H_1$ and the bounds from assumptions $H_1$ and $H_2$. 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 for the (large) time $\tau = \varepsilon^{-1} t$ 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 $H_5$ satisfies $\kappa \leq 1$. Thus, by [17] (Theorem 2.8), for every $p_Y < p + \kappa - 1$, we get

$$
\sup_{\tau \geq 0, \varepsilon > 0} \mathbb{E}^\varepsilon |y_\varepsilon(\tau)|^{p_Y} < \infty,
$$

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

Lemma 5.1 is proved.

We are now ready to prove (5.5). We fix $N > 0$, and denote by $P^\varepsilon_{t-\varepsilon N}$ and $E^\varepsilon_{t-\varepsilon N}$ the conditional probability and conditional expectation w.r.t. $\mathcal{F}^\varepsilon_{t-\varepsilon N}$, respectively. For sufficiently small $\varepsilon$, we have $s_q < t - \varepsilon N$ and, therefore,

$$
\mathbb{E}^\varepsilon \Phi(X_\varepsilon(s_1), \ldots, X_\varepsilon(s_q)) \left( \mathcal{L}^\varepsilon \varphi(X_\varepsilon(t), Y_\varepsilon(t)) - L \varphi(X_\varepsilon(t)) \right)
$$

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

By assumption $H_0$, there exist functions $\rho(\varepsilon) \to 0$ and $\varrho(\varepsilon) \to 0$ such that

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

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

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

Thus, by the Markov inequality,

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

$$
= \frac{1}{R \varrho(\varepsilon)} P^\varepsilon \left( \sup_{s \in [0, T]} |\zeta_\varepsilon(s)| > \rho(\varepsilon) \right) \leq \frac{\varrho(\varepsilon)}{R \varrho(\varepsilon)} = \frac{1}{R}.
$$
In the proof of Lemma 5.1, it has been shown that $L\varphi = \frac{1}{\varepsilon}L\varphi$. Thus, by statement (c) of this lemma, the function $L\varphi$ is bounded. The functions $\Phi, \mathcal{L}^\varepsilon$ are also bounded:

$$
\left| E^\varepsilon \Phi(X_\varepsilon(s_1), \ldots, X_\varepsilon(s_q)) \left( \mathcal{L}^\varepsilon \varphi(X_\varepsilon(t), Y_\varepsilon(t)) - L\varphi(X_\varepsilon(t)) \right) \right|
\leq C P^\varepsilon(|X_\varepsilon(t - \varepsilon N)| > R) + C P^\varepsilon(|Y_\varepsilon(t - \varepsilon N)| > R) + \frac{C}{R}
+ C E^\varepsilon 1_{\Omega_{t,N,R}^\varepsilon} \left| E^\varepsilon_{t-\varepsilon N} \left( \mathcal{L}^\varepsilon \varphi(X_\varepsilon(t), Y_\varepsilon(t)) - L\varphi(X_\varepsilon(t)) \right) \right|,
$$

(5.17)

where we have denoted

$$
\Omega_{t,N,R}^\varepsilon = \Omega_{t,N,R} \cap \{|X_\varepsilon(t - \varepsilon N)| \leq R, |Y_\varepsilon(t - \varepsilon N)| \leq R\}.
$$

We fix $j \geq 1$ and decompose

$$
E^\varepsilon_{t-\varepsilon N} \left( \mathcal{L}^\varepsilon \varphi(X_\varepsilon(t), Y_\varepsilon(t)) - L\varphi(X_\varepsilon(t)) \right)
= E^\varepsilon_{t-\varepsilon N} \left( \mathcal{L}^\varepsilon \varphi(X_\varepsilon(t), Y_\varepsilon(t)) - \mathcal{L}^\varepsilon \varphi(X_\varepsilon(t), Y_\varepsilon(t)) \chi_j(X_\varepsilon(t), Y_\varepsilon(t)) \right)
+ \left( E^\varepsilon_{t-\varepsilon N} \mathcal{L} \varphi(X_\varepsilon(t), Y_\varepsilon(t)) \chi_j(X_\varepsilon(t), Y_\varepsilon(t)) \right)
+ \left( \mathcal{L}^\varepsilon \varphi(X_\varepsilon(t), Y_\varepsilon(t)) - \mathcal{L} \varphi(X_\varepsilon(t), Y_\varepsilon(t)) - P^\varepsilon_{\text{frozen}}(\chi_j \mathcal{L} \varphi)(X_\varepsilon(t - \varepsilon N), Y_\varepsilon(t - \varepsilon N)) \right)
+ \left( P^\varepsilon_{\text{frozen}}(\chi_j \mathcal{L} \varphi)(X_\varepsilon(t - \varepsilon N), Y_\varepsilon(t - \varepsilon N)) - \mathcal{L} \varphi(X_\varepsilon(t - \varepsilon N)) \right)
+ \left( \mathcal{L} \varphi(X_\varepsilon(t - \varepsilon N)) - L\varphi(X_\varepsilon(t)) \right).
$$

(5.18)

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

$$
\left| E^\varepsilon_{t-\varepsilon N} \left( \mathcal{L}^\varepsilon \varphi(X_\varepsilon(t), Y_\varepsilon(t)) - \mathcal{L}^\varepsilon \varphi(X_\varepsilon(t), Y_\varepsilon(t)) \chi_j(X_\varepsilon(t), Y_\varepsilon(t)) \right) \right|
\leq C E^\varepsilon_{t-\varepsilon N} \left( 1 - \chi_j(X_\varepsilon(t), Y_\varepsilon(t)) \right).
$$

(5.19)

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

$$
\left| E^\varepsilon_{t-\varepsilon N} \left( \mathcal{L}^\varepsilon \varphi(X_\varepsilon(t), Y_\varepsilon(t)) - \mathcal{L} \varphi(X_\varepsilon(t), Y_\varepsilon(t)) \right) \chi_j(X_\varepsilon(t), Y_\varepsilon(t)) \right|
\leq \sup_{x,y} \left| \mathcal{L}^\varepsilon \varphi(x, y) - \mathcal{L} \varphi(x, y) \right| \chi_j(x, y) \to 0, \quad \varepsilon \to 0.
$$

(5.20)
To estimate the third term in (5.18), we first observe that the function \( \chi_j \mathcal{L} \varphi \) is continuous, which follows from \( \mathbf{H}_4 \) and \( \mathbf{H}_6 \) by analogy with Lemma 5.1(b). Further, we define a pair \( x_\varepsilon, y_\varepsilon \) by (5.11) with \( t_0 = t - \varepsilon N \) and take

\[
\widetilde{P}^\varepsilon = P^\varepsilon_{t - \varepsilon N, \omega};
\]

the regular version of conditional probability. Thus, for a.a. \( \omega \in \tilde{\Omega}^\varepsilon_{t, N, R} \), the pair \( x_\varepsilon, y_\varepsilon \) belongs to the class \( \mathcal{K}(\rho, 2R, \rho, 2N) \) w.r.t. probability \( P^\varepsilon_{t - \varepsilon N, \omega} \) in the notation introduced before Proposition 5.2. Applying this proposition, we get

\[
\mathbb{E}^\varepsilon \mathbb{E}_{\tilde{\Omega}^\varepsilon_{t, N, R}} \bigg| P^\text{frozen}_{t - \varepsilon N} (\chi_j \mathcal{L} \varphi) (x_\varepsilon(t), y_\varepsilon(t)) + \mathcal{L} \varphi (X_\varepsilon(t)) \bigg| \to 0, \quad \varepsilon \to 0.
\]

(5.21)

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

\[
\mathbb{E}^\varepsilon \mathbb{E}_{\tilde{\Omega}^\varepsilon_{t, N, R}} \bigg| P^\text{frozen}_{t - \varepsilon N} (\chi_j \mathcal{L} \varphi) (x_\varepsilon(t), y_\varepsilon(t)) - \chi_j \mathcal{L} \varphi (X_\varepsilon(t)) \bigg| \leq CN^{-\frac{p+\kappa-1}{1-\kappa}}.
\]

(5.22)

For the fifth term, we obtain

\[
\mathbb{E}^\varepsilon \mathbb{E}_{\tilde{\Omega}^\varepsilon_{t, N, R}} \bigg| \mathcal{L} \varphi (X_\varepsilon(t)) - L \varphi (X_\varepsilon(t)) \bigg| \to 0, \quad \varepsilon \to 0,
\]

(5.23)

For the sixth term, we get

\[
\mathbb{E}^\varepsilon \mathbb{E}_{\tilde{\Omega}^\varepsilon_{t, N, R}} \bigg| L \varphi (X_\varepsilon(t)) - \mathcal{L} \varphi (X_\varepsilon(t)) \bigg| \to 0, \quad \varepsilon \to 0,
\]

(5.24)

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

\[
\lim_{\varepsilon \to 0} \sup \mathbb{E}^\varepsilon \Phi (X_\varepsilon(s_1), \ldots, X_\varepsilon(s_q)) \left( \mathcal{L} \varphi (X_\varepsilon(t), Y_\varepsilon(t)) - L \varphi (X_\varepsilon(t)) \right) \leq C \mathbb{P}^\varepsilon (|X_\varepsilon(s)| > R) + C \mathbb{P}^\varepsilon (|Y_\varepsilon(s)| > R) + \frac{C}{R} + CN^{-\frac{p+\kappa-1}{1-\kappa}} + \sup_{|x| \leq R} |\mathcal{L} \varphi \chi_j(x) - L \varphi (x)| + C \limsup_{\varepsilon \to 0} \mathbb{E}^\varepsilon \mathbb{E}_{\tilde{\Omega}^\varepsilon_{t, N, R}} E^\varepsilon_{t - \varepsilon N} \left( 1 - \chi_j (X_\varepsilon(t), Y_\varepsilon(t)) \right).
\]

(5.25)

By analogy with (5.21)–(5.23), we find

\[
\lim_{\varepsilon \to 0} \mathbb{E}^\varepsilon \mathbb{E}_{\tilde{\Omega}^\varepsilon_{t, N, R}} E^\varepsilon_{t - \varepsilon N} \left( 1 - \chi_j (X_\varepsilon(t), Y_\varepsilon(t)) \right) \leq CN^{-\frac{p+\kappa-1}{1-\kappa}} + \sup_{|x| \leq R} (1 - \chi_j(x)).
\]
Thus,

\[
\limsup_{\varepsilon \to 0} \left| E^\varepsilon \Phi(X_\varepsilon(s_1), \ldots, X_\varepsilon(s_q)) \left( \mathcal{L}^\varepsilon \varphi(X_\varepsilon(t), Y_\varepsilon(t)) - L_\varepsilon \varphi(X_\varepsilon(t)) \right) \right|
\]

\[
\leq C \sup_{s \leq t, \varepsilon > 0} P^\varepsilon(|X_\varepsilon(s)| > R) + C \sup_{s \leq t, \varepsilon > 0} P^\varepsilon(|Y_\varepsilon(s)| > R) + \frac{C}{R}
\]

\[
+ C N^{-\frac{q+\gamma-1}{1-\gamma}} + \sup_{|x| \leq R} \left| L_\varphi \chi_j(x) - L \varphi(x) \right| + C \sup_{|x| \leq R} (1 - \chi_j(x)). \tag{5.26}
\]

The constants \( R, N, \) and \( j \) in this inequality are arbitrary. Letting first \( j \to \infty \) and \( N \to \infty \) for fixed \( R \), we conclude, by Lemma 5.1(d), (e), that

\[
\limsup_{\varepsilon \to 0} \left| E^\varepsilon \Phi(X_\varepsilon(s_1), \ldots, X_\varepsilon(s_q)) \left( \mathcal{L}^\varepsilon \varphi(X_\varepsilon(t), Y_\varepsilon(t)) - L_\varepsilon \varphi(X_\varepsilon(t)) \right) \right|
\]

\[
\leq C \sup_{s \leq t, \varepsilon > 0} P^\varepsilon(|X_\varepsilon(s)| > R) + C \sup_{s \leq t, \varepsilon > 0} P^\varepsilon(|Y_\varepsilon(s)| > R) + \frac{C}{R}. \tag{5.27}
\]

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

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

This proves (5.5) and completes the entire proof.

6. Appendix

**Proof of Lemma 3.1.** We set

\[
\zeta_\varepsilon(t) = \inf \left\{ s \geq 0 : \int_0^s b_\varepsilon^2(z)dz \geq t \right\}.
\]

Making the change of time \( \tilde{\eta}_\varepsilon(t) := \eta_\varepsilon(\zeta_\varepsilon(t)) \), we see that \( \tilde{\eta}_\varepsilon(t) \) satisfies the assumptions of Lemma 3.1 with another constant \( A > 0 \) and a new Wiener process \( \tilde{W}(t) = \tilde{W}(\zeta_\varepsilon(t)) \) but with \( \tilde{b}_\varepsilon(t) \equiv 1 \). Since

\[(C_2)^{-2} t \leq \zeta_\varepsilon(t) \leq (C_1)^{-2} t,\]

without loss of generality, we can assume that \( b_\varepsilon(t) \equiv 1 \).

We set

\[
L_\varepsilon := Ax^\gamma \frac{d}{dx} + \frac{\varepsilon^2}{2} \frac{d^2}{dx^2}.
\]

Denote

\[
v_\varepsilon(x) := \int_0^{|x|} \exp \left\{ \frac{-2Ay^{\gamma+1}}{(\gamma + 1)\varepsilon^2} \right\} \left( \int_0^{|y|} \frac{2A|z|^{\gamma+1}}{(\gamma + 1)\varepsilon^2} \right) dy.
\]
Further, we have
\[ L_\varepsilon v_\varepsilon(x) \geq 1, \quad \text{sgn}(x)v'_\varepsilon(x) \geq 0, \quad \text{and} \quad v_\varepsilon(0) = 0. \]

Thus, by Ito’s formula, we get
\[
\mathbb{E}v_\varepsilon(\eta_\varepsilon(\delta) \land n)) = \mathbb{E} \int_0^{\tau_\varepsilon(\delta) \land n} \left( a_\varepsilon(s)\eta_\varepsilon'(s)v_\varepsilon'(\eta_\varepsilon(s)) + \frac{\varepsilon^2}{2} v''_\varepsilon(\eta_\varepsilon(s)) \right) ds
\]
\[
\geq \mathbb{E} \int_0^{\tau_\varepsilon(\delta) \land n} \left( A\eta_\varepsilon'(s) + \frac{\varepsilon^2}{2} v''_\varepsilon(\eta_\varepsilon(s)) \right) ds = \mathbb{E} \int_0^{\tau_\varepsilon(\delta) \land n} L_\varepsilon v_\varepsilon(\eta_\varepsilon(s)) ds
\]
\[
\geq \mathbb{E} \int_0^{\tau_\varepsilon(\delta) \land n} 1 ds = \mathbb{E}\tau_\varepsilon(\delta) \land n.
\]

Passing to the limit as \( n \to \infty \) and applying the Fatou lemma, we establish the a.s. finiteness of \( \tau_\varepsilon(\delta) \). Since
\[ v_\varepsilon(\eta_\varepsilon(\delta))) = v_\varepsilon(\delta) = v_\varepsilon(-\delta), \]
we get the following estimate:
\[ \mathbb{E}\tau_\varepsilon(\delta) \leq v_\varepsilon(\delta). \]

Let \( x > 0 \) be arbitrary. Changing the variables
\[ s := \frac{\varepsilon^{\gamma+1}}{\varepsilon^2} \quad \text{and} \quad t := \frac{y^{\gamma+1}}{\varepsilon^2}, \]
we obtain
\[
v_\varepsilon(x) = \frac{2\varepsilon^{\gamma+1}}{(\gamma + 1)^2 \varepsilon^2} \int_0^{\frac{|x|^{\gamma+1}}{\varepsilon^2}} \exp \left\{ - \frac{2At}{\gamma + 1} \right\} \left( \int_0^t \exp \left\{ \frac{2Az^{\gamma+1}}{(\gamma + 1)\varepsilon^2} \right\} dz \right) \frac{\varepsilon^{-\gamma}}{t^{\gamma+1}} dt
\]
\[
= \frac{2\varepsilon^{\gamma+1}}{(\gamma + 1)^2 \varepsilon^2} \int_0^{\frac{|x|^{\gamma+1}}{\varepsilon^2}} \exp \left\{ - \frac{2At}{\gamma + 1} \right\} \left( \int_0^t \exp \left\{ \frac{2As}{\gamma + 1} \right\} s^{\frac{-\gamma}{\gamma+1}} ds \right) \frac{\varepsilon^{-\gamma}}{t^{\gamma+1}} dt
\]
\[
= \frac{2}{(\gamma + 1)^2} \int_0^{\frac{|x|^{\gamma+1}}{\varepsilon^2}} \exp \left\{ - \frac{2At}{\gamma + 1} \right\} \left( \int_0^t \exp \left\{ \frac{2As}{\gamma + 1} \right\} s^{\frac{-\gamma}{\gamma+1}} ds \right) \frac{\varepsilon^{-\gamma}}{t^{\gamma+1}} dt. \quad (6.1)
\]
It follows from the L’Hôpital rule that, for any $\alpha > 0$ and $\beta > -1$,

$$\int_{0}^{t} e^{\alpha s} s^\beta ds \sim \alpha^{-1} e^{\alpha t} t^\beta, \quad t \to +\infty.$$ 

Hence,

$$\int_{0}^{t} \exp \left\{ \frac{2As}{\gamma + 1} \right\} s^{\frac{\gamma + 1}{2\gamma + 1}} ds \sim \frac{\gamma + 1}{2A} \exp \left\{ \frac{2At}{\gamma + 1} \right\} t^{\frac{\gamma + 1}{2\gamma + 1}}, \quad t \to +\infty.$$ 

Applying this relation and the L’Hôpital rule, we find

$$\int_{0}^{u} \exp \left\{ -\frac{2At}{\gamma + 1} \right\} \left( \int_{0}^{t} \exp \left\{ \frac{2As}{\gamma + 1} \right\} s^{\frac{\gamma}{\gamma + 1}} ds \right) \frac{t^{\frac{\gamma + 1}{\gamma + 1}}}{t^{\frac{\gamma + 1}{\gamma + 1}}} dt \sim \frac{\gamma + 1}{2A} \int_{0}^{u} \exp \left\{ -\frac{2At}{\gamma + 1} \right\} \left( \exp \left\{ \frac{2At}{\gamma + 1} \right\} t^{\frac{\gamma + 1}{\gamma + 1}} \right) \frac{t^{\frac{\gamma + 1}{\gamma + 1}}}{t^{\frac{\gamma + 1}{\gamma + 1}}} dt = \frac{\gamma + 1}{2A} \int_{0}^{u} t^{\frac{2(n+1)}{\gamma + 1}} dt = \frac{(\gamma + 1)^2}{2A(1-\gamma)} \frac{1}{u^{\gamma + 1}}, \quad u \to +\infty.$$ 

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

$$v_{\varepsilon}(x) \sim_{\varepsilon \to 0} K_{2} \frac{2(1-\gamma)}{\varepsilon^{2}(\gamma + 1)} |x|^{\frac{1-\gamma}{2}} \varepsilon^{\frac{2\gamma + 1}{\gamma + 1}} = K_{2} \frac{2(1-\gamma)}{\varepsilon^{2}(\gamma + 1)} |x|^{\gamma + 1} \frac{1-\gamma}{\gamma + 1} = K_{2} |x|^{\gamma - 1},$$

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

This means that, for any fixed $\delta \geq 0$,

$$\limsup_{\varepsilon \to 0} \mathbf{E}e_{\varepsilon}(\delta) \leq \limsup_{\varepsilon \to 0} \mathbf{E}v_{\varepsilon}(\delta) = K_{2}\delta^{1-\gamma}.$$ 

Lemma 3.1 is proved.

The work of A. Kulik was supported by the Polish National Science Center (Grant 2019/33/B/ST1/02923). The research of A. Pilipenko was partially supported by the Norway–Ukrainian Cooperation in Mathematical
ON REGULARIZATION BY A SMALL NOISE OF MULTIDIMENSIONAL ODEs WITH NON-LIPSCHITZ COEFFICIENTS

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 Program “Singular Diffusions: Analytic and Stochastic Approaches” between the University of Potsdam and the Institute of Mathematics of the National Academy of Sciences of Ukraine.

REFERENCES

1. F. Attanasio and F. Flandoli, “Zero-noise solutions of linear transport equations without uniqueness: an example,” C. R. Math. Acad. Sci. Paris, 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 and P. Baldi, “Small random perturbations of Peano phenomena,” Stochastics, 6, No. 2, 279–292 (1982).

4. R. Buckdahn, Y. Ouknine, and 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 and K. Suresh Kumar, “A new Markov selection procedure for degenerate diffusions,” J. Theoret. Probab., 23, No. 3, 729–747 (2010).

6. F. Delarue and 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, and D. Vincenzi, “Noise prevents collapse of Vlasov–Poisson point charges,” Comm. Pure Appl. Math., 67, No. 10, 1700–1736 (2014).

8. F. Delarue and M. Maurelli, Zero Noise Limit for Multidimensional SDEs Driven by a Pointy Gradient (2019), arXiv:1909.08702

9. N. Dirr, S. Luckhaus, and M. Novaga, “A stochastic selection principle in case of fattening for curvature flow,” Calc. Var. Partial Different. Equat., 13, No. 4, 405–425 (2001).

10. H. J. Engelbert and 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, Inc., Englewood Cliffs, N.J. (1964).

12. S. Herrmann, “Phénomène de Peano et grandes déviations,” C. R. Acad. Sci. Paris, Sér. I Math., 332, No. 11, 1019–1024 (2001).

13. N. Ikeda and S. Watanabe, “Stochastic differential equations and diffusion processes,” North-Holland Mathematical Library, 24, North-Holland Publ., Amsterdam, etc. (1981).

14. I. Karatzas and S. E. Shreve, “Brownian motion and stochastic calculus,” Graduate Texts in Mathematics, 113, Springer-Verlag, New York (1988).

15. I. G. Krykun and S. Ya. Makhno, “The Peano phenomenon for Itô 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 and I. Pavlyukevich, Moment Bounds for Dissipative Semimartingales with Heavy Jumps, http://arxiv.org/abs/ 2004.12449.

18. I. Pavlyukevich and A. Pilipenko, Generalized Selection Problem with Lévy Noise (2020), arXiv: 2004.05421

19. A. Pilipenko and F. N. Proske, “On a selection problem for small noise perturbation in the multidimensional case,” Stochast. Dyn., 18, No. 6, 1850045 (2018); DOI: 10.1142/S0219493718500454

20. A. Pilipenko and F. N. Proske, “On perturbations of an ODE with non-Lipschitz coefficients by a small self-similar noise,” Statist. 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. Comm. Probab., 18, No. 31, 7 pp. (2013).

22. A. Veretennikov, “On strong solutions and explicit formulas for solutions of stochastic integral equations,” Sh. Math., 39, No. 3, 387–403 (1981).