2017-06-01

On dynamical systems perturbed by a null-recurrent motion: the general case

arXiv:1508.05346v2 [math.PR] 15 Sep 2015

On dynamical systems perturbed by a null-recurrent motion: The general case

Zsolt Pajor-Gyulai *[1,3] and Michael Salins †[2,3]

[1]Courant Institute of Mathematical Sciences, New York University, New York, NY
[2]Department of Mathematics and Statistics, Boston University, Boston, MA
[3]Department of Mathematics, University of Maryland, College Park, MD

Abstract

We consider a perturbed ordinary differential equation where the perturbation is only significant when a one-dimensional null recurrent diffusion is close to zero. We investigate the first order correction to the unperturbed system and prove a central limit theorem type result, i.e., that the normalized deviation process converges in law in the space of continuous functions to a limit process which we identify. We show that this limit process has a component which only moves when the limit of the null-recurrent fast motion equals zero. The set of these times forms a zero-measure Cantor set and therefore the limiting process cannot be described by a standard SDE. We characterize this process by its infinitesimal generator (with appropriate boundary conditions) and we also characterize the process as the weak solution of an SDE that depends on the local time of the fast motion process. We also investigate the long time behavior of such a system when the unperturbed motion is trivial. In this case, we show that the long-time limit is constant on a set of full Lebesgue measure with probability 1, but it has nontrivial drift and diffusion components that move only when the fast motion equals zero.

1 Introduction

In systems of multidimensional diffusion processes where both slow and fast timescales are present, the limiting behavior of the slow component is an interesting mathematical problem with applications in physics, biology and other areas. One possible setting is a system of diffusion processes $(\tilde{X}^\varepsilon(t), \tilde{Y}^\varepsilon(t))$ that depends on a parameter ε representing the ratio of the two time scales. More precisely, suppose that $\tilde{X}^\varepsilon$ changes faster and faster in time as $\varepsilon \downarrow 0$ while $\tilde{Y}^\varepsilon$ changes on the same time scale for all values of ε. Then $\tilde{X}^\varepsilon$ and $\tilde{Y}^\varepsilon$ are called the fast and the slow component respectively.

Classical stochastic averaging theory, first invented by Khasminskii ([16]) expanding the ideas of Bogolyubov for ordinary differential equations ([4]), states that if the fast component X^ε has a finite invariant measure, then as $\varepsilon \to 0$, the distribution of the slow component Y^ε approaches a limit where the X dependence in the drift and diffusion coefficients of Y is

*zsolt@cims.nyu.edu
†msalins@bu.edu

averaged out with respect to this invariant measure. This result was later extended and refined by a vast number of authors (see e.g [10],[11],[19],[18],[23],[27] amongst many others).

In this paper, we are going to study a system where the fast component is a null-recurrent process and therefore has no invariant probability measure while the slow motion is a perturbation of an ordinary differential equation. Namely, let W be a k-dimensional Brownian motion and consider the system

$$d\tilde{X}^\varepsilon(t) = \varepsilon^{-1}\varphi(\tilde{X}^\varepsilon(t), \tilde{Y}^\varepsilon(t))dW(t), \qquad \tilde{X}^\varepsilon(0) = x_0/\varepsilon, \tag{1.1}$$

$$d\tilde{Y}^\varepsilon(t) = b(\tilde{X}^\varepsilon(t), \tilde{Y}^\varepsilon(t))dt + \sigma(\tilde{X}^\varepsilon(t), \tilde{Y}^\varepsilon(t))dW(t), \qquad \tilde{Y}^\varepsilon(0) = y_0, \tag{1.2}$$

where all coefficients are Lipschitz continuous, φ is bounded away from zero and infinity, and the process $\tilde{X}^\varepsilon$ ($\tilde{Y}^\varepsilon$) is one (d) dimensional. Intuitively, the fast component $\tilde{X}^\varepsilon(t)$ is of order $\mathcal{O}(\varepsilon^{-1})$ and therefore the limiting dynamics of the system, as $\varepsilon \downarrow 0$, is governed by the behavior of the coefficients for large absolute value of the x-variable.

In [17], it was shown that if b, φ, and σ have certain limits as $|x| \uparrow \infty$ in the Cesaro sense then $(\varepsilon\tilde{X}^\varepsilon(t), \tilde{Y}^\varepsilon(t))$ converges in law to a process where the x-dependence in (1.1)-(1.2) is replaced by coefficients associated to these limits. More precisely, after the change of coordinates $(X^\varepsilon(t), Y^\varepsilon(t)) = (\varepsilon\tilde{X}^\varepsilon(t), \tilde{Y}^\varepsilon(t))$, the system $(X^\varepsilon(t), Y^\varepsilon(t))$ can be written as

$$d\begin{pmatrix} X^\varepsilon(t) \\ Y^\varepsilon(t) \end{pmatrix} = \begin{pmatrix} 0 \\ b(\varepsilon^{-1}X^\varepsilon(t), Y^\varepsilon(t)) \end{pmatrix} dt + \begin{pmatrix} \varphi(\varepsilon^{-1}X^\varepsilon(t), Y^\varepsilon(t)) \\ \sigma(\varepsilon^{-1}X^\varepsilon(t), Y^\varepsilon(t)) \end{pmatrix} dW(t)$$

It might be tempting to assume that the limit will be given by taking the limit of the coefficients for large positive and negative values of x. However, this is not the complete picture as one has to include the effects of null recurrent averaging. To see this, note that the diffusion matrix can be written as

$$A(x,y) = \frac{1}{2}\begin{pmatrix} |\varphi(x,y)|^2 & \varphi\sigma^T(x,y) \\ \sigma\varphi^T(x,y) & \sigma\sigma^T(x,y) \end{pmatrix} = \frac{1}{2}\frac{1}{1/|\varphi(x,y)|^2}\begin{pmatrix} 1 & \frac{\varphi\sigma^T(x,y)}{|\varphi(x,y)|^2} \\ \frac{\sigma\varphi^T(x,y)}{|\varphi|^2(x,y)} & \frac{\sigma\sigma^T(x,y)}{|\varphi^2(x,y)|} \end{pmatrix}$$

Let $p(x,y) = 1/|\varphi(x,y)|^2$ and note that if we freeze the y dynamics, this is proportional to the sigma finite invariant density of the X^ε process. For a function $K(x,y)$, define

$$K_\pm(x,y) = \chi_{\{x>0\}} \cdot \lim_{u\to+\infty}\frac{1}{u}\int_0^u K(v,y)dv + \chi_{\{x<0\}} \cdot \lim_{u\to-\infty}\frac{1}{u}\int_0^u K(v,y)dv,$$

whenever the Cesaro-limits exist. Using this, introduce the averaged diffusion matrix $\bar{A}_{ij}(x,y) = (pA_{ij})_\pm(x,y)/p_\pm(x,y)$ for $i,j = 1,..,1+d$ and the averaged drift $\bar{b}(x,y) = (0, (pb)_\pm/p_\pm)$.

Theorem 1.1 (Khasminskii, Krylov [17]). *Suppose that φ, σ, b are Lipschitz continuous and, for each x, their first and second order derivatives in y are bounded continuous functions of y. Assume as well that for some positive constants $c_1, c_2, c_3 > 0$, we have*

$$c_1 \leq |\varphi(x,y)| \leq c_2, \qquad \sum_{i=1}^d (A^T(x,y))_{ii} + b_i^2(x,y)) \leq c_3(1+|y|^2)$$

for all $(x,y) \in \mathbb{R}^{1+d}$*. Also assume that* p, pA_{ij}, pb_i *and all their first and second derivatives in* y *have well defined Cesaro limits as* $x \to \pm\infty$*. Then the pair* $(\varepsilon \tilde{X}^\varepsilon(t), \tilde{Y}^\varepsilon(t))$ *converges in distribution in* $C\left([0,\infty), \mathbb{R}^{1+d}\right)$ *to the solution of*

$$d\left(X^0(t), Y^0(t)\right) = \bar{b}(X^0(t), Y^0(t))dt + \sqrt{\bar{A}(X^0(t), Y^0(t))}dW(t),$$

whenever weak uniqueness for this equation holds.

In general, however, the Cesáro limits in question can be quite different for large positive and negative values of x. This implies that the coefficients of this limiting system will be discontinuous at $x = 0$ in the generic situation and the well-posedness of the corresponding martingale problem becomes a challenge. Note that the limit being well defined is an assumption in this theorem. For a list of special cases when well-posedness holds, we refer the reader to [17] (see also [2]).

In this paper, we answer the question what happens when the Cesaro limits of σ are zero while the Cesaro limits of b are equal. Namely, we study the following special case of (1.1)-(1.2).

$$d\tilde{X}^\varepsilon(t) = \varepsilon^{-1}\varphi(\tilde{X}^\varepsilon(t), \tilde{Y}^\varepsilon(t))dW(t), \qquad \tilde{X}^\varepsilon(0) = x_0/\varepsilon, \tag{1.3}$$

$$d\tilde{Y}^\varepsilon(t) = [b_1(\tilde{Y}^\varepsilon(t)) + b_2(\tilde{X}^\varepsilon(t), \tilde{Y}^\varepsilon(t))]dt + \sigma(\tilde{X}^\varepsilon(t), \tilde{Y}^\varepsilon(t))dW(t), \qquad \tilde{Y}^\varepsilon(0) = y_0, \tag{1.4}$$

where b_2 and σ satisfy a certain decay condition in x which guarantee that their Cesáro limits for large $|x|$ vanish so their contribution to the dynamics for large values of $|\tilde{X}^\varepsilon|$ is negligible. In this setting, the result of [17] implies that $Y^\varepsilon(t)$ converges uniformly in probability to the solution $y(t)$ of the ordinary differential equation (ODE)

$$\dot{y}(t) = b_1(y(t)), \qquad y(0) = y_0.$$

In fact, it can be shown that this convergence holds in a stronger sense (see Lemma 3.2). Consequently, $Y^\varepsilon(t)$ can be viewed as a perturbation of $y(t)$ and it is of interest to describe the behavior of their deviation. This system can model, for example, the effect of spatial inhomogenity localized around a one co-dimensional surface (e.g. a crack in a solid) where the homogeneous structures away from the interface can possibly be different on different sides of this interface. This is the situation, for example, in groundwater hydrology when one studies diffusion in composite porous materials, see e.g. [20] and the references therein. In our scaling, the dynamics transversal to the interface is much faster than the one parallel to it.

In [22], the authors considered the simpler case when $\tilde{X}^\varepsilon$ is a Brownian motion independent of the noise driving the slow motion. Namely, the stochastic differential equation (SDE)

$$d\tilde{Y}^\varepsilon(t) = [b_1(Y^\varepsilon(t)) + b_2(\varepsilon^{-1}W_1(t), Y^\varepsilon)]dt + \sigma(\varepsilon^{-1}W_1(t), Y^\varepsilon(t))dW_2(t), \qquad Y^\varepsilon(0) = y_0,$$

where W_1 and W_2 are independent standard Brownian motions, was considered and it was shown that the first order correction to the deterministic limit is of order $\sqrt{\varepsilon}$. Moreover, a functional central limit theorem was proved for the normalized difference $\varepsilon^{-1/2}(Y^\varepsilon(t) - y(t))$ with the limit identified as a process that is Gaussian when conditioned on W_1. This case was significantly simpler than (1.3)-(1.4) as the difficulty arising from the discontinuity at $x = 0$ does not appear and there is no feedback from the Y^ε process to X^ε. This allowed for a proof using only elementary stochastic analysis. In the limit, the perturbation acts only when

$W_1(t) = 0$ (i.e., on a Cantor set of times) and the equation for the limiting solution can be expressed in terms of the local time of W_1 at zero.

In the current paper, we prove a similar result for the general system (1.3)-(1.4) using the martingale method. We show that if $\sigma \neq 0$, the pair $(\varepsilon \tilde{X}^\varepsilon(t), \varepsilon^{-1/2}(\tilde{Y}^\varepsilon(t) - y(t)))$ converges to a non-trivial process with singular diffusion on the interface which we describe. On the other hand if $\sigma \equiv 0$, then the pair $(\varepsilon \tilde{X}^\varepsilon(t), \varepsilon^{-1}(\tilde{Y}^\varepsilon(t) - y(t)))$ converges to a nontrivial process with singular drift on the interface. Moreover, we show that in the absence of the unperturbed system ($b_1 \equiv 0$), the interesting limit behavior happens on longer time scales (of order ε^{-2}). The long-time limit has both singular drift and singular diffusion in the sense that there is a proper diffusion process on the interface, with coefficients that are averages with respect to a sigma-finite invariant measure of the fast motion, that is sampled at the random times when the limit of the fast motion is in the origin. Once again, these times form a Cantor set and the limiting slow motion is constant for almost all times.

It is worth pointing out that in some sense, the original result of [17] and the results in this paper parallel the first few steps of the program contained in ([10], Chapter 7) when the fast process is positively recurrent (See Section 7).

The plan of this paper is as follows. In Section 2, we specify the conditions on the coefficients in (1.3)-(1.4), describe the limit processes, and state our convergence results. In Section 3, we derive the main technical tools needed for our analysis. In Section 4, we show that the families of processes under consideration are tight while in Section 5, we prove that all subsequential limits converge to the solution of the associated martingale problem. In Section 6, we prove that these martingale problems are well-posed and characterize their unique solutions as the weak solutions of certain SDE-s. In Section 7, we mention some open problems.

Acknowledgements

The authors are grateful to D. Dolgopyat for introducing them to the problem and to L. Koralov and D. Dolgopyat for their helpful suggestions during invaluable discussions and for reading the manuscript. We also thank P.E. Jabin for a discussion on Section 7. While working on the paper, Z. Pajor-Gyulai was partially supported by the NSF grants number 1309084 and DMS1101635. M. Salins was partially supported by the NSF grant number 1407615.

2 Assumptions and main results

For convenience, we will make the change of coordinates $(X^\varepsilon(t), Y^\varepsilon(t)) = (\varepsilon \tilde{X}^\varepsilon(t), \tilde{Y}^\varepsilon(t))$ and consider the system

$$dX^\varepsilon(t) = \varphi(\varepsilon^{-1} X^\varepsilon(t), Y^\varepsilon(t)) dW(t), \qquad X^\varepsilon(0) = x_0, \quad (2.1)$$

$$dY^\varepsilon(t) = [b_1(Y^\varepsilon(t)) + b_2(\varepsilon^{-1} X^\varepsilon(t), Y^\varepsilon(t))]dt + \sigma(\varepsilon^{-1} X^\varepsilon(t), Y^\varepsilon(t)) dW(t), \quad Y^\varepsilon(0) = y_0, \quad (2.2)$$

where $W(t)$ is a k-dimensional Wiener process, $\varphi(x, y)$ is a k-dimensional vector field, $b_1(y)$ and $b_2(x, y)$ are d-dimensional vector fields, and $\sigma(x, y)$ is a $d \times k$-matrix. In this way, $X^\varepsilon(t)$ is a one-dimensional process and $Y^\varepsilon(t)$ is a d-dimensional process.

We make the following assumptions:

1. φ, b_2, and σ are Lipschitz continuous and $b_1(y)$ is a twice continuously differentiable d-dimensional vector field with bounded derivatives.

2. Decay conditions hold in x:

$$\hat{b}(x) := \sup_{y\in\mathbb{R}} |b_2(x,y)|_{\mathbb{R}^d} \in L^1(\mathbb{R}), \qquad \hat{\sigma}^2(x) = \sup_{y\in\mathbb{R}} \mathrm{Tr}\sigma\sigma^T(x,y) = \sup_{y\in\mathbb{R}} \sum_{i,j=1}^{d} \sigma_{ij}^2(x,y) \in L^1(\mathbb{R}). \tag{2.3}$$

3. X^ε is non-degenerate. More precisely, if $|\varphi(x,y)|^2 = \sum_{i=1}^k \varphi_i^2(x,y)$, we require that there are numbers $0 < c_1 \le c_2 < \infty$ such that

$$c_1 < |\varphi(x,y)|^2 < c_2 \qquad \forall (x,y) \in \mathbb{R}^{1+d}$$

4. $1/|\varphi(x,y)|^2$ has Cesaro-limits in x, i.e. the limits

$$\lim_{x\to\pm\infty} \frac{1}{x} \int_0^x \frac{1}{|\varphi(u,y)|^2} du$$

exist. Define

$$a_\pm(x,y) = \chi_{\{x\ge 0\}} \left(\lim_{u\to\infty} \frac{1}{u} \int_0^u \frac{1}{|\varphi(v,y)|^2} dv \right)^{-1} + \chi_{\{x<0\}} \left(\lim_{u\to-\infty} \frac{1}{u} \int_0^u \frac{1}{|\varphi(v,y)|^2} dv \right)^{-1}. \tag{2.4}$$

Occasionally, we will use the notation $a_+(y)$ to denote the restriction of $a_\pm$ to $\{(x,y) \in \mathbb{R}^{d+1} : x > 0\}$ (similarly we will use $a_-(y)$).

It follows from Assumption 1 that the ODE

$$\frac{dy}{dt} = b_1(y(t)) \qquad y(0) = y_0,$$

which serves as the unperturbed part of the slow motion, has a unique globally defined solution.

Before stating our main result, let us introduce the construction which we use to describe the limit processes. Let $H_+ = \{(x,y) : x > 0\}$ and $H_- = \{(x,y) : x < 0\}$ be two half-spaces in $\mathbb{R}^{1+d}$ and let $\mathcal{L}_\pm$ be second order differential operators acting on functions over $H_\pm$ respectively. Denote the interface by $H_0 = \{(x,y) : x = 0\}$. Also let $\beta_i(y), \alpha_{ij}(y) \in C(\mathbb{R}^d)$, $i,j = 1,...,d$, be such that $(\alpha_{ij}(y))$ is uniformly elliptic and such that the SDE

$$dU(t) = \beta(U(t))dt + \sqrt{\alpha(U(t))}dW(t)$$

has a unique weak solution. Let $p_\pm(y) \ge 0$ with $p_+(y)+p_-(y) = 1$. Let $C_0(\mathbb{R}^{1+d})$ be the Banach space of bounded, continuous functions $f : \mathbb{R}^{1+d} \to \mathbb{R}$ decaying at infinity endowed with the supremum norm.

We construct an operator $\bar{\mathcal{L}}$ on $C_0(\mathbb{R}^{1+d})$ as follows. Let its domain $\mathcal{D}(\bar{\mathcal{L}}) \subseteq C_0(\mathbb{R}^{1+d})$ consist of $f \in C_0(\mathbb{R}^{1+d})$ such that

- The restriction of f to H_+, H_-, and along H_0 are smooth.
- $\partial_{y_i} f(x,y) \in C^1(\mathbb{R}^{1+d})$ for $i = 1,...,d$.
- $\mathcal{L}_+ f(0+,y) = \mathcal{L}_- f(0-,y)$, for all $y \in \mathbb{R}^d$.

- $\partial_x f(x,y)$ has right and left limits as $x \to 0$ and the gluing-boundary condition

$$p_+(y)\partial_x f(0+,y) - p_-(y)\partial_x f(0-,y) + \sum_{i=1}^{d} \beta_i(y)\partial_{y_i} f(0,y) + \frac{1}{2}\sum_{i,j=1}^{d} \alpha_{ij}(y)\partial_{y_i y_j} f(0,y) = 0 \tag{2.5}$$

is satisfied for all $(x,y) \in \mathbb{R}^{1+d}$.

For any $f \in \mathcal{D}(\bar{\mathcal{L}})$, let $\bar{\mathcal{L}}f(x,y) = \mathcal{L}_\pm f(x,y)$ for $x \in H_\pm$ and define $\bar{\mathcal{L}}f(0,y)$ by continuity.

This is the adaptation of a special case of the well-known Wentzell type boundary conditions (see [30], [31] or the survey [29]). In general, it is hard to prove that the closure of such an operator is the generator of a continuous strong Markov process $Z(t) = (X(t), Y(t))$. Nevertheless, we will show that when $p_+(y) \equiv p_-(y) \equiv 1/2$ and $\mathcal{L}_\pm$ has a special form, $\bar{\mathcal{L}}$ does indeed generate a Markov semigroup on $C_0(\mathbb{R}^{1+d})$. In fact, in the cases of interest we construct the corresponding Markov process probabilistically. However, we believe that the closure of such an operator $\bar{\mathcal{L}}$ is the generator of a Markov semigroup in greater generality.

Intuitively, the process $Z(t)$ can be described as follows. Inside the half-spaces H_+ and H_-, it coincides with the diffusion generated by $\mathcal{L}_\pm$ up to the point when it hits H_0. On H_0, $Z(t)$ follows a diffusion process for an infinitesimal amount of time, after which it leaves H_0 in the direction of $H_\pm$ with probability $p_\pm$. The set of times at which $Z(t) \in H_0$ has Lebesgue measure zero and consequently the displacement due to the effect of the boundary is singular in time (the support of the increments is a Cantor set in $\mathbb{R}$). To make this intuition rigorous and obtain a characterization of $Z(t)$, let us recall the definition of the local time of a semimartingale (see Theorem 7.1 in [13]).

Definition 2.1. *The symmetric local time of a semimartingale S with quadratic variation $\langle S\rangle$ is the unique nonnegative random field*

$$L = \{L^S(t,x); (t,x) \in [0,\infty) \times \mathbb{R}\}$$

such that the following hold:

1. *The mapping $(t,x) \to L^S(t,x)$ is measurable and $L^S(t,x)$ is adapted.*

2. *For each $x \in \mathbb{R}$, the mapping $t \to L^S(t,x)$ is non-decreasing and constant on each open interval where $S(t) \neq x$.*

3. *For every Borel measurable $f : \mathbb{R} \to [0,\infty)$, we have*

$$\int_0^t f(S_s)d\langle S\rangle_s = \int_{\mathbb{R}} f(x)L^S(t,x)dx \qquad a.s.$$

4. *$L^S(t,x)$ is a.s. jointly continuous in t and x for $x \neq 0$, while the one sided limits exist at $x = 0$ and*

$$L^S(t,0) = \frac{1}{2}(L^S(t,0+) + L^S(t,0-)) \tag{2.6}$$

Note that instead of (2.6), we could have choosen an arbitrary convex combination. Other common choices include $L^S(t,0) = L^S(t,0+)$ (right local time) or $L^S(t,0) = L^S(t,0-)$ (left local time). For our current purposes, however, the symmetric local time satisfying (2.6) is the most suitable as it makes our formulas more transparent.

Assume that $Z(t)$ is the unique solution of the martingale problem given $\bar{\mathcal{L}}$ for appropriate β, α and $p_\pm$. It is natural to conjecture (see also e.g. [12],[21]) that it is also the unique weak solution of an SDE involving singular terms on the interface $x = 0$. Namely, let

$$\mathcal{L}_\pm = \frac{1}{2}A(x,y)\partial_{xx} + B(x,y)\partial_y \tag{2.7}$$

be the diffusion operator on $H_- \cup H_+$,where ∂_y is the d-dimensional gradient with respect to the y variable and $A(x,y)$, $B(x,y)$ are matrix-valued functions that are uniformly elliptic and sufficiently regular on both halfspaces H_i. Assume $A(x,y)$ has a jump discontinuity at $x = 0$ ($B(x,y)$ may or may not be continuous at $x = 0$). Then $Z(t)$ is a weak-solution to the following SDE

$$dZ_1(t) = \sqrt{A(Z_1(t), Z_2(t))}dW(t) + [p_+(Z_2(t)) - p_-(Z_2(t))]dL^{Z_1}(t), \qquad Z_1(0) = x_0, \tag{2.8}$$

$$dZ_2(t) = B(Z_1(t), Z_2(t))dt + \beta(Z_2(t))dL^{Z_1}(t) + \sqrt{\alpha(Z_2(t))}dV^{Z_1}(t), \qquad Z_2(0) = y_0, \tag{2.9}$$

where $\sqrt{\cdot}$ is the matrix square root, $L^{Z_1}(t) = L^{Z_1}(t,0)$ is the local time of Z_1 at 0, W is a one-dimensional Brownian motion. $V^{Z_1}(t)$ is a continuous martingale whose quadratic variation is $\langle V_i^{Z_1}, V_j^{Z_1}\rangle_t = \delta_{ij}L^{Z_1}(t)$ and the cross-variations are $\langle V_i^{Z_1}, W\rangle_t \equiv 0$ for all $i = 1, ..., d$. It is natural to define the weak solution to such an SDE as a quadruple (Z_1, Z_2, W, V^{Z_1}) such that W is a one-dimensional Brownian motion and V^{Z_1} is a continuous martingale satisfying the above variation relations and that (Z_1, Z_2) solves (2.8)-(2.9) in the integral sense with (W, V^{Z_1}) taken as the driving noise. In this way, Z_1 diffuses like a skew-Brownian motion (see [21]) modulated by $Z_2(t)$, and $Z_2(t) - \int_0^t B(Z_1(s), Z_2(s))ds$ moves only when $Z_1(t) = 0$, because $L^{Z_1}(t)$ (and therefore $V^{Z_1}(t)$) is constant when $Z_1 \neq 0$.

Because of the discontinuous coefficients and the coupled nature of the above equations, the well-posedness of such an SDE is in general unclear. We will show, just as one would expect based on the classical situation, that the weak solutions of (2.8)-(2.9) are in one to one correspondence with the solutions of the martingale problem associated to the operator $\bar{\mathcal{L}}$ as discussed above under fairly general circumstances. However, we do not need this general result for the proofs of our main results and therefore defer the proof to the appendix.

We are ready to state our main result on the asymptotic behavior of the process $(X^\varepsilon, Y^\varepsilon)$. As we mentioned before, $Y^\varepsilon(t)$ converges to $y(t)$. Our first main result concerns the rate of this convergence. Let $C([0,\infty), \mathbb{R}^k)$ be the space of all continuous $\mathbb{R}^k$-valued functions on the halfline endowed with the metric

$$\rho(f,g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\sup_{t\in[0,n]}|f(t) - g(t)|}{1 + \sup_{t\in[0,n]}|f(t) - g(t)|}$$

It is well known that weak convergence of probability measures over this space is equivalent to weak convergence of measures in $C([0,T], \mathbb{R}^k)$ with the supremum norm for all $T > 0$. Let us introduce the normalized deviation process $\zeta^\varepsilon(t) = \varepsilon^{-1/2}(Y^\varepsilon(t) - y(t))$.

Theorem 2.2. *The triple $(X^\varepsilon(t), y(t), \zeta^\varepsilon(t))$ converges in law in $C([0,\infty),\mathbb{R}^{1+2d})$ to a Markov process which is the unique solution of the martingale problem associated to the operator $\bar{\mathcal{L}}$ described above with*

$$\mathcal{L}_\pm = \frac{1}{2}a_\pm(x,y)\frac{\partial^2}{\partial x^2} + b_1(y)\cdot\frac{\partial}{\partial y} + (\partial_y b_1(x,y)\cdot\zeta)\cdot\frac{\partial}{\partial\zeta}$$

where $a_\pm$ is as in (2.4) and

$$p_\pm(y) \equiv 1/2, \qquad \beta(y) = 0, \qquad \alpha(y) = \begin{pmatrix} 0 & 0 \\ 0 & \int_{-\infty}^{\infty}\left(\frac{\sigma\sigma^T}{|\varphi|^2}\right)(x,y)dx \end{pmatrix}.$$

Moreover, there is a one dimensional Brownian motion W and a continuous martingale V^{X^0} such that $\langle V_i^{X^0}, V_j^{X_0}\rangle_t = \delta_{ij}L^{X^0}$ and $\langle V_i, W\rangle = 0$ for each $i = 1,\dots,d$, and such that the limit process of the pair $(X^\varepsilon,\zeta^\varepsilon)$ satisfies the inhomogeneous equations

$$dX^0(t) = \sqrt{a_\pm(X^0(t),y(t))}dW(t), \tag{2.10}$$

$$d\zeta^0(t) = \partial_y b_1(X^0(t),y(t))\cdot\zeta^0(t)dt + \sqrt{\left(\int_{-\infty}^{\infty}\left(\frac{\sigma\sigma^T}{|\varphi|^2}\right)(x,y(t))dx\right)}dV^{X^0}(t), \tag{2.11}$$

with $(X^0(0),\zeta^0(0)) = (x_0, 0)$ where $\sqrt{\cdot}$ in (2.11) denotes the matrix square root, and $\partial_y b_1(y)$ is the derivative tensor of the vector field b_1 at $y\in\mathbb{R}^d$, i.e $(\partial_y b_1(y))_{ij} = \partial(b_1)^i/\partial y^j$.

Remark 2.3. *Note that once the processes X^0 and y are constructed, the solution of (2.11) is explicitly given by the variation of parameters formula*

$$\zeta^0(t) = \int_0^t e^{\int_s^t(\partial_y b_1(X^0(r),y(r))dr}\sqrt{\left(\int_{-\infty}^{\infty}\left(\frac{\sigma\sigma^T}{|\varphi|^2}\right)(x,y(s))dx\right)}dV^{X^0}(s).$$

As is apparent from this theorem, the drift part of the perturbation (b_2) does not contribute to the deviations of order $\sqrt{\varepsilon}$. On the other hand, the following result suggests that b_2 does play a role in fluctuations of order ε.

Theorem 2.4. *Let $\sigma\equiv 0$ and $\bar{\zeta}^\varepsilon(t) = \varepsilon^{-1}(Y^\varepsilon(t) - y(t))$. The triple $(X^\varepsilon(t), y(t), \bar{\zeta}^\varepsilon(t))$ converges in law in $C([0,\infty),\mathbb{R}^{1+2d})$ to the unique solution of the martingale problem associated to the operator $\bar{\mathcal{L}}$ described above with*

$$\mathcal{L}_\pm = \frac{1}{2}a_\pm(x,y)\frac{\partial^2}{\partial x^2} + b_1(y)\cdot\frac{\partial}{\partial y} + (\partial_y b_1(x,y)\cdot\zeta)\cdot\frac{\partial}{\partial\zeta}$$

and

$$p_\pm(y) \equiv 1/2, \qquad \beta(y) = \left(0, \int_{-\infty}^{\infty}\left(\frac{b_2}{|\varphi|^2}\right)(x,y)dx\right), \qquad \alpha(y) \equiv 0.$$

Moreover, there is a Brownian motion W such that the limit process of the pair $(X^\varepsilon,\zeta^\varepsilon)$ solves the inhomogeneous equation

$$dX^0(t) = \sqrt{a_\pm(X^0(s),y(s))}dW(t) \qquad X^0(0) = x_0,$$

$$d\zeta^0(t) = \partial_y b_1(X^0(t),y(t))\cdot\zeta^0(t)dt + \left(\int_{-\infty}^{\infty}\left(\frac{b_2}{|\varphi|^2}\right)(x,y(t))dx\right)dL^{X^0}(t) \qquad \zeta^0(0) = 0,$$

where $\sqrt{\cdot}$ in (2.11) denotes the matrix square root, and $\partial_y b_1(y)$ is the derivative tensor of the vector field b_1 at $y\in\mathbb{R}^d$, i.e $(\partial_y b_1(y))_{ij} = \partial(b_1)^i/\partial y^j$.

Remark 2.5. *In this case, the variation of parameters formula implies*

$$\zeta^0(t) = \int_0^t e^{\int_s^t (\partial_y b_1(X^0(r),y(r))dr} \left(\int_{-\infty}^{\infty} \left(\frac{b_2}{|\varphi|^2} \right) (x, y(s)) dx \right) dL^{X^0}(s).$$

When $b_1(y) \equiv 0$, we have $y(t) \equiv 0$ and it follows from Theorem 2.2 that $\sup_{t \in [0,T]} |Y^\varepsilon(t)|$ converges to zero for every fixed T in probability. On long time scales (of order ε^{-2}), however, the process converges to a limit with both singular drift and singular diffusion on the interface. Let $\bar{X}^\varepsilon(t) = \varepsilon X^\varepsilon(\varepsilon^{-2}t)$ and $\bar{Y}^\varepsilon(t) = Y^\varepsilon(\varepsilon^{-2}t)$. This pair satisfies the SDE

$$d\bar{X}^\varepsilon(t) = \varphi(\varepsilon^{-2}\bar{X}^\varepsilon(t), \bar{Y}^\varepsilon(t)) dW(t), \qquad \bar{X}^\varepsilon(0) = \varepsilon x_0, \tag{2.12}$$

$$d\bar{Y}^\varepsilon(t) = \varepsilon^{-2} b_2(\varepsilon^{-2}\bar{X}^\varepsilon(t), \bar{Y}^\varepsilon(t)) dt + \varepsilon^{-1}\sigma(\varepsilon^{-2}\bar{X}^\varepsilon(t), \bar{Y}^\varepsilon(t)) dW(t), \qquad \bar{Y}^\varepsilon(0) = y_0, \tag{2.13}$$

with a Brownian motion W different from the one in (2.1)-(2.2).

Theorem 2.6. *The pair* $(\bar{X}^\varepsilon, \bar{Y}^\varepsilon)$ *converges in law in* $\mathcal{C}([0,\infty); \mathbb{R}^{1+d})$ *to the unique solution of the martingale problem associated to the operator* $\bar{\mathcal{L}}$ *with*

$$\mathcal{L}_\pm = \frac{1}{2} a_\pm(x,y) \frac{\partial^2}{\partial x^2}$$

and

$$p_\pm(y) \equiv 1/2, \qquad \beta(y) = \int_{-\infty}^{\infty} \left(\frac{b_2}{|\varphi|^2} \right)(x,y) dx, \qquad \alpha(y) = \int_{-\infty}^{\infty} \left(\frac{\sigma\sigma^T}{|\varphi|^2} \right)(x,y) dx.$$

Moreover, the limit process is the unique weak solution of the equation

$$d\bar{X}^0(t) = \sqrt{a_\pm(\bar{X}^0(t), \bar{Y}^0(t))} dW(t), \tag{2.14}$$

$$d\bar{Y}^0(t) = \left(\int_{-\infty}^{\infty} \left(\frac{b_2}{|\varphi|^2} \right)(x, \bar{Y}^0(t)) dx \right) dL^{\bar{X}^0}(t) + \left(\sqrt{\int_{-\infty}^{\infty} \left(\frac{\sigma\sigma^T}{|\varphi|^2} \right)(x, \bar{Y}^0(t)) dx} \right) dV^{\bar{X}^0}(t), \tag{2.15}$$

with $(\bar{X}^0(0), \bar{Y}(0)) = (0, y_0)$ *where* $L^{\bar{X}^0}$, W, *and* $V^{\bar{X}^0}$ *are as in* (2.8) *and* (2.9).

As one can see from (2.15), the singular drift and diffusion on the interface are both significant on these longer time scales. Also, a careful look at the explicit formula for $\beta(y)$ and $\alpha(y)$ reveals that they are essentially averages with respect to a σ-finite invariant measure of the fixed y process $dX^y(t) = \varphi(X^y(t), y) dW(t)$ (with density $c/|\varphi(\cdot, y)|^2$ for any $c > 0$). In this sense, Theorem 2.6 is an averaging result where the singular term on the interface follows the averaged dynamics of the slow component with respect to the invariant measure of the fast motion. Due to the null recurrence of the fast dynamics, however, the natural timescale for this averaged process is not proportional to t (as it is due to the ergodic behavior of fast-slow systems in the positive recurrent case). Instead, it is proportional to the amount of time that the fast component asymptotically spends in the essential support of b and σ, or in other words to the local time at $x = 0$. Since our fast motion is roughly Brownian, this time scale is of order $\sqrt{t}$.

Note that even though the theorems above are stated for a flat interface, they are applicable to situations when the interface is a more general hypersurface. We also remark that we chose the initial conditions x_0, y_0 somewhat arbitrarily. The modifications one needs to make to include the case when x_0, y_0 can depend on ε differently are straightforward.

3 Auxiliary lemmas

First we derive an estimate on certain integral functionals of the fast process (2.1)

Lemma 3.1. *For any $p \geq 1$, there exists $C_p > 0$ such that for all $\psi \in L^1(\mathbb{R})$, all stopping times $0 < \tau_1 < \tau_2$ with $\mathrm{E}\tau_2^{p/2} < \infty$, and all $\varepsilon > 0$,*

$$\mathrm{E}\left|\frac{1}{\varepsilon}\int_{\tau_1}^{\tau_2} \psi(\varepsilon^{-1}X^\varepsilon(r))dr\right|^p \leq C_p|\psi|^p_{L^1(\mathbb{R})}\mathrm{E}|\tau_2 - \tau_1|^{\frac{p}{2}}. \tag{3.1}$$

Proof. If $f \in C^2(\mathbb{R})$, then by Ito formula,

$$f(\varepsilon^{-1}X^\varepsilon(\tau_2)) - f(\varepsilon^{-1}X^\varepsilon(\tau_1)) = \tag{3.2}$$
$$= \frac{1}{\varepsilon}\int_{\tau_1}^{\tau_2} f'(\varepsilon^{-1}X^\varepsilon(r))\varphi(\varepsilon^{-1}X^\varepsilon(r), Y^\varepsilon(r))dW(r) + \frac{1}{2\varepsilon^2}\int_{\tau_1}^{\tau_2} f''(\varepsilon^{-1}X^\varepsilon(r))|\varphi(\varepsilon^{-1}X^\varepsilon(r), Y^\varepsilon(r))|^2 dr$$

almost surely. In fact, by the Meyer-Tanaka formula, the above expression is valid even if f'' is not continuous, but is only L^1. We want to construct f such that $f''(x) = |\psi(x)|$. Since $\psi \in L^1(\mathbb{R})$, the function

$$\Psi(x) = \int_{-\infty}^{x} |\psi(y)|dy$$

is well defined, continuous and bounded ($0 \leq \Psi(x) \leq |\psi|_{L^1(\mathbb{R})}$). Then we can define

$$f(x) = \int_0^x \Psi(y)dy.$$

This function is Lipschitz continuous because

$$|f(x_2) - f(x_1)| \leq \left|\int_{x_1}^{x_2} \Psi(y)dy\right| \leq |\psi|_{L^1(\mathbb{R})}|x_2 - x_1|. \tag{3.3}$$

Now, because $c_1 \leq |\varphi(x,y)|^2$,

$$\frac{1}{\varepsilon}\int_{\tau_1}^{\tau_2} |\psi(\varepsilon^{-1}X^\varepsilon(r))|dr \leq c\varepsilon\frac{1}{2\varepsilon^2}\int_{\tau_1}^{\tau_2} |\psi(\varepsilon^{-1}X^\varepsilon(r))||\varphi(\varepsilon^{-1}X^\varepsilon(r), Y^\varepsilon(r))|^2 dr,$$

where here and in what follows, c is some (possibly different from line to line) positive constant independent of ε. Plugging this into (3.2),

$$\frac{1}{\varepsilon}\int_{\tau_1}^{\tau_2} |\psi(\varepsilon^{-1}X^\varepsilon(r))|dr \leq$$
$$\leq c\left(\varepsilon(f(\varepsilon^{-1}X^\varepsilon(\tau_2)) - f(\varepsilon^{-1}X^\varepsilon(\tau_1))) - \int_{\tau_1}^{\tau_2} \Psi(\varepsilon^{-1}X^\varepsilon(r))\varphi(\varepsilon^{-1}X^\varepsilon(r), Y^\varepsilon(r))dW(r)\right)$$

By the elementary inequality $|a+b|^p \leq 2^{p-1}(|a|^p + |b|^p)$,

$$\mathrm{E}\left|\frac{1}{\varepsilon}\int_{\tau_1}^{\tau_2} \psi(\varepsilon^{-1}X^\varepsilon(r))dr\right|^p \leq c\Bigg(\mathrm{E}\varepsilon^p\left|f(\varepsilon^{-1}X^\varepsilon(\tau_2)) - f(\varepsilon^{-1}X^\varepsilon(\tau_1))\right|^p$$
$$+\mathrm{E}\left|\int_{\tau_1}^{\tau_2} \Psi(\varepsilon^{-1}X^\varepsilon(r))\varphi(\varepsilon^{-1}X^\varepsilon(r), Y^\varepsilon(r))dW(r)\right|^p\Bigg).$$

By (3.3),

$$\varepsilon|f(\varepsilon^{-1}X^\varepsilon(\tau_2)) - f(\varepsilon^{-1}X^\varepsilon(\tau_1))| \leq$$
$$\leq \varepsilon|\psi|_{L^1(\mathbb{R})}|\varepsilon^{-1}X^\varepsilon(\tau_2) - \varepsilon^{-1}X^\varepsilon(\tau_1)| = |\psi|_{L^1(\mathbb{R})}\left|\int_{\tau_1}^{\tau_2}\varphi(\varepsilon^{-1}X^\varepsilon(r), Y^\varepsilon(r))dW(r)\right|,$$

and therefore,

$$\mathrm{E}\left|\frac{1}{\varepsilon}\int_{\tau_1}^{\tau_2}\psi(\varepsilon^{-1}X^\varepsilon(r))dr\right|^p \leq c|\psi|^p_{L^1(\mathbb{R})}\mathrm{E}\left|\int_{\tau_1}^{\tau_2}\varphi(\varepsilon^{-1}X^\varepsilon(r), Y^\varepsilon(r))dW(r)\right|^p +$$
$$+ c\mathrm{E}\left|\int_{\tau_1}^{\tau_2}\Psi(\varepsilon^{-1}X^\varepsilon(r))\varphi(\varepsilon^{-1}X^\varepsilon(r), Y^\varepsilon(r))dW(r)\right|^p$$

Each of these expectations can be bounded by the BDG inequality, and we see that

$$\mathrm{E}\left|\frac{1}{\varepsilon}\int_{\tau_1}^{\tau_2}\psi(\varepsilon^{-1}X^\varepsilon(r))dr\right|^p \leq c|\psi|^p_{L^1(\mathbb{R})}\left|\mathrm{E}\int_{\tau_1}^{\tau_2}|\varphi(\varepsilon^{-1}X^\varepsilon(r), Y^\varepsilon(r))|^2dr\right|^{p/2} +$$
$$+ c\left|\mathrm{E}\int_{\tau_1}^{\tau_2}\Psi^2(\varepsilon^{-1}X^\varepsilon(r))|\varphi(\varepsilon^{-1}X^\varepsilon(r), Y^\varepsilon(r))|^2dr\right|^{p/2}.$$

All of these integrands are bounded because $|\varphi| \leq c_2$ and $\Psi(x) \leq |\psi|_{L^1(\mathbb{R})}$. Therefore, we can conclude that

$$\mathrm{E}\left|\frac{1}{\varepsilon}\int_{\tau_1}^{\tau_2}\psi(\varepsilon^{-1}X^\varepsilon(r))dr\right|^p \leq C_p|\psi|^p_{L^1(\mathbb{R})}\mathrm{E}|\tau_2 - \tau_1|^{\frac{p}{2}}.$$

as required. □

Next we show that the slow motion (2.2) converges to the unperturbed system in a strong sense.

Lemma 3.2. *For every $p \geq 1$, there exists a constant C_p such that*

$$\sup_{t\in[0,T]}\mathrm{E}|Y^\varepsilon(t) - y(t)|^p < C_p\max\{\varepsilon^pT^{p/4}, \varepsilon^{p/2}T^{p/2}\}e^{C_pT^{p-1}}.$$

Proof. Note that

$$Y^\varepsilon(t) - y(t) = \int_0^t(b_1(Y^\varepsilon(s)) - b_1(y(s))ds + \int_0^t b_2\left(\varepsilon^{-1}X^\varepsilon(s), Y^\varepsilon(s)\right)ds +$$
$$\int_0^t\sigma\left(\varepsilon^{-1}X^\varepsilon(s), Y^\varepsilon(s)\right)dW(s) = I_1^\varepsilon(t) + I_2^\varepsilon(t) + I_3^\varepsilon(t).$$

By the Lipschitz continuity of b_1 and Jensen's inequality,

$$\mathrm{E}|I_1^\varepsilon(t)|^p \leq T^{p-1}\mathrm{Lip}(b_1)^p\int_0^t\mathrm{E}|Y^\varepsilon(s) - y(s)|^p.$$

On the other hand, recall that $\hat{b}(x) = \sup_{y\in\mathbb{R}^d}|b_2(x,y)|$ and note

$$\sup_{t\in[0,T]}\mathrm{E}|I_2^\varepsilon(t)|^p \leq \mathrm{E}\left(\int_0^T\hat{b}(\varepsilon^{-1}X^\varepsilon(s))ds\right)^p \leq C_p\varepsilon^pT^{p/2},$$

where in the last inequality we used Lemma 3.1 with $\tau_1 = 0$ and $\tau_2 = T$. Finally, it is easy to see that the scalar quadratic variation of I_3 is

$$\langle I_3 \rangle_t = \int_0^t \mathrm{Tr}\sigma\sigma^T(\varepsilon^{-1}X^\varepsilon(s), Y^\varepsilon(s))ds \le \int_0^t \hat{\sigma}^2(\varepsilon^{-1}X^\varepsilon(s))ds,$$

(where we recall $\hat{\sigma}^2(x) = \sup_{y\in\mathbb{R}^d} \mathrm{Tr}\sigma\sigma^T(x,y)$) and therefore by the Burkholder-Davis-Gundy inequality and Lemma 3.1 we have

$$\sup_{t\in[0,T]} \mathrm{E}|I_3^\varepsilon(t)|^p \le C_p \mathrm{E}\left(\int_0^T \hat{\sigma}^2(\varepsilon^{-1}X^\varepsilon(s))ds\right)^{p/2} < C_p \varepsilon^{p/2} T^{p/4}. \tag{3.4}$$

The result now follows from Gronwall's lemma. $\square$

Next, we recall Lemma 3.5 from [17], which explains why the Cesaro limits of $\frac{1}{|\varphi|^2}$ describe the limiting behavior of X^ε.

Lemma 3.3. *For any $f \in C(\mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}^d)$ and any $T < \infty$, we have*

$$\sup_{t\in[0,T]} \left| \int_0^t f(s, X^\varepsilon(s), Y^\varepsilon(s))(|\varphi|^2(\varepsilon^{-1}X^\varepsilon(s), Y^\varepsilon(s)) - a_\pm(X^\varepsilon(s), Y^\varepsilon(s)))ds \right| \to 0, \tag{3.5}$$

as $\varepsilon \downarrow 0$ in probability where $a_\pm$ is defined in (2.4) *Moreover, (3.5) holds for every family of stopping times $\{\tau^\varepsilon\}_{\varepsilon\le\varepsilon_0}$ in place of T such that $\{X^\varepsilon(\tau^\varepsilon)\}_{\varepsilon\le\varepsilon_0}$ is tight.*

Before proving Lemma 3.3, we establish the following fact about Cesaro limits.

Lemma 3.4. *Assume that $g : \mathbb{R} \to \mathbb{R}$ is bounded and has the Cesaro limit*

$$\bar{g} := \lim_{x\to+\infty} \frac{1}{x} \int_0^x g(\xi)d\xi.$$

Then for any $f \in L^1_{\mathrm{loc}}(\mathbb{R})$, and $R > 0$,

$$\lim_{\varepsilon\to0} \sup_{x\in[0,R]} \left| \bar{g} \int_0^x f(\xi)d\xi - \int_0^x f(\xi)g(\varepsilon^{-1}\xi)d\xi \right| = 0. \tag{3.6}$$

Proof. First, we consider the case where $f \equiv 1$ (see [17]). In this case, for $x \in \mathbb{R}$,

$$\int_0^x (\bar{g} - g(\varepsilon^{-1}\xi))d\xi = x\left(\bar{g} - \frac{\varepsilon}{x}\int_0^{x/\varepsilon} g(\xi)d\xi\right) =: x\alpha\left(\frac{x}{\varepsilon}\right). \tag{3.7}$$

Because $\bar{g}$ is the Cesaro limit of g, for any fixed x, the above expression converges to zero as $\varepsilon \to 0$. This convergence is uniform in $|x| \le R$ because for $|x| \le R$,

$$\left|x\alpha\left(\frac{x}{\varepsilon}\right)\right| = \left|\int_0^x (\bar{g} - g(\varepsilon^{-1}\xi))d\xi\right| \le \min\left\{2x|g|_{L^\infty}, R\left|\bar{g} - \frac{\varepsilon}{x}\int_0^{x/\varepsilon} g(\xi)d\xi\right|\right\}. \tag{3.8}$$

Next, we consider the case where $f \in C^1$. In this case, by integrating by parts,

$$\int_0^x f(\xi)(\bar{g} - g(\varepsilon^{-1}\xi))d\xi = f(x)x\alpha\left(\frac{x}{\varepsilon}\right) - \int_0^x f'(\xi)\xi\alpha\left(\frac{\xi}{\varepsilon}\right)d\xi. \tag{3.9}$$

Therefore,

$$\sup_{0\le x\le R}\left|\int_0^x f(\xi)(\bar{g} - g(\varepsilon^{-1}\xi))d\xi\right| \le \sup_{0\le x\le R}\left|x\alpha\left(\frac{x}{\varepsilon}\right)\right|\left(|f|_{C^1[0,R]}\right)(1+R) \tag{3.10}$$

and it follows from (3.8) that the convergence is uniform.

Finally, for a general $f \in L^1_{\text{loc}}(\mathbb{R})$ and $R > 0$, we can approximate f in $L^1([0,R])$ by a sequence $\{f_n\} \subset C^1([0,R])$. Then for $n \in \mathbb{N}$,

$$\sup_{0\le x\le R}\left|\int_0^x f(\xi)(\bar{g} - g(\varepsilon^{-1}\xi))d\xi\right| \le 2|g|_{L^\infty([-R,R])}|f-f_n|_{L^1([-R,R])} + \sup_{0\le x\le R}\left|\int_0^x f_n(\xi)(\bar{g} - g(\varepsilon^{-1}\xi))d\xi\right|.$$

By first choosing n large and then ε small we can make the above expression arbitrarily small. □

Proof of Lemma 3.3. Very similar to the proof of Lemma 3.5 in [17], we only include a sketch of the proof. For $\varepsilon > 0$, define the auxiliary function $u^\varepsilon(t,x,y)$ to solve

$$\begin{cases} |\varphi(\varepsilon^{-1}x,y)|^2 u^\varepsilon_{xx}(t,x,y) = f(t,x,y)(|\varphi(\varepsilon^{-1}x,y)|^2 - a_\pm(x,y)). \\ u^\varepsilon(t,0,y) = u^\varepsilon_x(t,0,y) = 0, \quad \lim_{\varepsilon\to 0}\sup_{|x|\le R}|u^\varepsilon(t,x,y)| = 0. \end{cases} \tag{3.11}$$

By dividing by $|\varphi|^2$ and integrating, we see that for $x > 0$ (the situation for $x < 0$ is similar),

$$u^\varepsilon_x(t,x,y) = \int_0^x f(t,\xi,y)d\xi - a_+(y)\int_0^x \frac{f(t,\xi,y)}{|\varphi(\varepsilon^{-1}\xi,y)|^2}d\xi.$$

By Lemma 3.4, this converges to zero as $\varepsilon \to 0$ uniformly in $|x| \le R$ if and only if $a_+(y)$ is the reciprocal of the Cesaro limit of $\frac{1}{|\varphi|^2}$. By integrating in x once more, we show that u^ε also converges to zero as $\varepsilon \to 0$ uniformly in $|x| \le R$. The proof concludes by applying Ito formula to $u_\varepsilon(t, X^\varepsilon(t), Y^\varepsilon(t))$, identifying the integral in (3.5) as the Ito-correction term, a standard localization argument and estimating the resulting other terms. The last statement of the lemma follows similarly using that $X^\varepsilon(\tau^\varepsilon)$ is in a large enough compact set for all $\varepsilon < \varepsilon_0$ with high probability. □

Finally, we state the version of the Ito-Tanaka-Meyer formula that is relevant to our situation (a special case of [24]).

Lemma 3.5. *Let (Z_1, Z_2, W, V^{Z_1}) be a weak solution of (2.8)-(2.9) with $p_+ = p_- = 1/2$ and assume that $f \in C(\mathbb{R}^{1+d})$ has continuous first and second order derivatives in y and for fixed*

$y \in \mathbb{R}^d$, $f(.,y) \in C^1(\mathbb{R}_+ \cup \mathbb{R}_-)$. *Then*

$$f(Z_1(t), Z_2(t)) = f(Z_1(0), Z_2(0)) + \int_0^t \partial_x^{sym} f(Z_1(s), Z_2(s))\sqrt{A(Z_1(s), Z_2(s))}dW(s)$$
$$+ \sum_{i=1}^d \int_0^t \partial_{y_i} f(Z_1(s), Z_2(s))\sqrt{\alpha(Z_2(s))}dV^{Z_1}(s)$$
$$+ \int_0^t \Big[p_+(Y(s))f_x(0+, Z_2(t)) - p_-(Y(s))f_x(0-, Z_2(t))$$
$$+ \sum_{i=1}^d \partial_{y_i} f(Z_1(s), Z_2(s))\beta_i(Z_2) + \frac{1}{2}\sum_{i,j=1}^d \partial_{y_i y_j} f(Z_1(t), Z_2(t))\alpha_{ij}(Z_2(t))\Big] dL^{Z_1}(t)$$
$$+ \frac{1}{2}\int_0^t \partial_{xx} f(Z_1(s), Z_2(s))A(Z_1(s), Z_2(s))ds + \int_0^t \sum_{i=1}^d \partial_{y_i} f(Z_1(s), Z_2(s))B(Z_1(s), Z_2(s))ds,$$

where used the symmetrized x-derivative

$$\partial_x^{sym} f(x,y) = \frac{\partial_x f(x+,y) + \partial_x f(x-,y)}{2}.$$

Note that the symmetrized derivative was necessary by our choice to use the symmetric local time.

4 Tightness

In this section we are going to show that the laws of the families in Theorem 2.2, Theorem 2.4, and Theorem 2.6 are tight in the space $C([0,\infty); \mathbb{R}^{1+d})$. To prove tightness, we will show that there exists $c > 0$, $p > 0$, $q > 1$ such that for all $T > 0$ and $0 \le s \le t \le T$, and $\varepsilon > 0$,

$$\mathrm{E}|X^\varepsilon(t) - X^\varepsilon(s)|^p \le c|t-s|^q,$$

$$\mathrm{E}|\zeta^\varepsilon(t) - \zeta^\varepsilon(s)|^p \le c|t-s|^q$$

and in the long-time case

$$\mathrm{E}|\bar{X}^\varepsilon(t) - \bar{X}^\varepsilon(s)|^p \le c|t-s|^q,$$
$$\mathrm{E}|\bar{Y}^\varepsilon(t) - \bar{Y}^\varepsilon(s)|^p \le c|t-s|^q.$$

All of these results are consequences of Lemma 3.1.

Theorem 4.1. *Let $(X^\varepsilon, \zeta^\varepsilon)$ satisfy the conditions of Theorem 2.2 or Theorem 2.4. Then the family of laws of the pairs $\{(X^\varepsilon, \zeta^\varepsilon)\}_{\varepsilon>0}$ are tight. Let $(\bar{X}^\varepsilon, \bar{Y}^\varepsilon)$ satisfy the conditions of Theorem 2.6, then the family of laws of the pairs $\{(\bar{X}^\varepsilon, \bar{Y}^\varepsilon)\}_{\varepsilon>0}$ are tight.*

Proof. We prove tightness in the context of Theorem 2.2, as the tightness for the other systems can be proved analogously. In this case,

$$X^\varepsilon(t) - X^\varepsilon(s) = \int_s^t \varphi(\varepsilon^{-1}X^\varepsilon(r), Y^\varepsilon(r))dW(r).$$

By the BDG inequality and the fact that $|\varphi|^2 < c_2$,

$$\mathrm{E}|X^\varepsilon(t) - X^\varepsilon(s)|^p \le c_2|t-s|^{p/2}.$$

As for ζ^ε, because $Y^\varepsilon(t) = y(t) + \sqrt{\varepsilon}\zeta^\varepsilon(t)$,

$$\zeta^\varepsilon(t) - \zeta^\varepsilon(s) = \int_s^t \left(\frac{b_1(y(r) + \sqrt{\varepsilon}\zeta^\varepsilon(r)) - b_1(y(r))}{\sqrt{\varepsilon}}\right) dr + \varepsilon^{-1/2}\int_s^t b_2(\varepsilon^{-1}X^\varepsilon(r), y(r) + \sqrt{\varepsilon}\zeta^\varepsilon(r))dr$$
$$+ \varepsilon^{-1/2}\int_s^t \sigma(\varepsilon^{-1}X^\varepsilon(r), y(r) + \sqrt{\varepsilon}\zeta^\varepsilon(r))dW(r) = I_1^\varepsilon(s,t) + I_2^\varepsilon(s,t) + I_3^\varepsilon(s,t).$$

We estimate $I_2^\varepsilon(s,t)$ and $I_3^\varepsilon(s,t)$ by using Lemma 3.1. By (2.3) and Lemma 3.1

$$\mathrm{E}|I_2^\varepsilon(s,t)|^p \le \mathrm{E}\left|\varepsilon^{-1/2}\int_0^t \hat{b}(\varepsilon^{-1}X^\varepsilon(r))dr\right|^p \le c\varepsilon^{1/2}|t-s|^{p/2}, \tag{4.1}$$

and similarly by the BDG inequality,

$$\mathrm{E}|I_3^\varepsilon(s,t)|^p \le c\mathrm{E}\left(\varepsilon^{-1}\int_s^t \hat{\sigma}^2(\varepsilon^{-1}X^\varepsilon(r))dr\right)^{p/2} \le c|t-s|^{p/4}. \tag{4.2}$$

Finally, if $p > 1$, Lemma 3.2 implies

$$\mathrm{E}|I_1^\varepsilon(s,t)|^p \le c_{p,T}\|b_1\|^p_{C^1(\mathbb{R}^{d+1})}|t-s|^p. \tag{4.3}$$

We choose $p > 4$, $q = p/4$ and tightness follows from (4.1), (4.2), and (4.3) and the Kolmogorov criterion above. $\square$

As a result of tightness, for every subsequence $\varepsilon_n \to 0$ there exists a ε_{n_k} such that $(X^{\varepsilon_{n_k}}, \zeta^{\varepsilon_{n_k}})$ converges in distribution. In the next sections we characterize the limit and show that it is the same process for all such subsequences. Consequently, the entire sequence converges in law to that unique limit.

5 Convergence to the martingale problem

5.1 Freidlin-Wentzell type result

We are going to show that the processes in Theorems 2.2, 2.4, and 2.6 each converge to a solution of a martingale problem. In all three cases, the main tool in proving convergence is a Freidlin-Wentzell type result ([9], see also [12]). Assume we have a family of $(1+k)$-dimensional processes Z^ε. Depending on the situation, Z^ε will be either $Z^\varepsilon = (X^\varepsilon, y, \zeta^\varepsilon)$ $(k = 2d)$ or $Z^\varepsilon = (\bar{X}^\varepsilon, \bar{Y}^\varepsilon)$ $(k = d)$. We are going to use the notation $\mathrm{E}_z(\cdot)$ to the expectation when Z^ε has initial point $z = (x, y, \zeta)$ or $z = (x, y)$ respectively. Recall that $H_\pm = \{z : \pm x > 0\}$

Theorem 5.1. *Let $\bar{\mathcal{L}}_i$ be second order differential operators on H_i, $i = \{+, -\}$ and let $D_i \subseteq C_b^\infty(H_i)$ be a set of test functions that have bounded derivatives of all order. Let $\tau^\varepsilon = \inf\{t \ge 0 : |X^\varepsilon(t)| \le l(\varepsilon)\}$ for some $l(\varepsilon) \downarrow 0$ as $\varepsilon \downarrow 0$. Assume that there is a $\lambda_0 \ge 0$ such that for each $\lambda > \lambda_0$, the following assumptions hold.*

1. *We have for every $f \in D_i$*

$$\mathrm{E}_z \left[e^{-\lambda \tau^\varepsilon} f(Z^\varepsilon(\tau^\varepsilon)) - f(z) + \int_0^{\tau^\varepsilon} e^{-\lambda t} \left(\lambda f(Z^\varepsilon(t)) - \bar{\mathcal{L}}_i f(Z^\varepsilon(t)) \right) dt \right] = \mathcal{O}(k(\varepsilon)) \qquad (5.1)$$

for some $k(\varepsilon) \downarrow 0$ as $\varepsilon \downarrow 0$ uniformly for $z \in H_i$.

2. *For $i \in \{+,-\}$, there is a $u_{i,\lambda} \in D_i$ such that*

$$\bar{\mathcal{L}}_i u_{i,\lambda} = \lambda u_{i,\lambda} \qquad \forall z \in H_i, |x| < 1$$

and such that $u_{\pm,\lambda}(0,y) = 1$ and $\pm \partial_x u_{\pm,\lambda}(0,y) < c < 0$.

3. *There is a $\delta(\varepsilon) \downarrow 0$ as $\varepsilon \downarrow 0$ such that $\delta(\varepsilon)/l(\varepsilon) \to \infty$, $\delta(\varepsilon)/k(\varepsilon) \to \infty$ and*

$$\sup_{z \in \mathbb{R}^{1+d}} \mathrm{E}_z \int_0^\infty e^{-\lambda t} \chi_{(-\delta(\varepsilon), \delta(\varepsilon))}(X^\varepsilon(t)) dt \to 0 \qquad (5.2)$$

as $\varepsilon \downarrow 0$.

4. *Define*

$$\theta^\delta = \inf\{t \geq 0 : |X^\varepsilon(t)| \geq \delta(\varepsilon)\}. \qquad (5.3)$$

There are $p_+(y)$, $p_-(y) > 0$ such that $p_+ + p_- = 1$ and

$$\mathrm{P}_z(Z^\varepsilon(\theta^\delta) \in H_i) \to p_i(y)$$

as $\varepsilon \downarrow 0$ uniformly in $z \in \mathbb{R}^{1+d}$, $|x| < l(\varepsilon)$.

5. *There are $\beta_i(y), \alpha_{ij}(y)$ for $i,j = 1, ..., d$ such that*

$$\frac{1}{\delta} \mathrm{E}_z(Y_i^\varepsilon(\theta^\delta) - y_i) \to \beta_i(y) \qquad \frac{1}{\delta} \mathrm{E}(Y_i^\varepsilon(\theta^\delta) - y_i)(Y_j^\varepsilon(\theta^\delta) - y_j) \to \alpha_{ij}(y),$$

as $\varepsilon \downarrow 0$, uniformly for $|x| < l(\varepsilon)$. Moreover, we have $\sup_{|x|<l(\varepsilon)} \mathrm{E}|Z^\varepsilon(\theta^\delta) - z|^3 = o(\delta)$.

Then for every $f \in C_0(\mathbb{R}^{1+d})$ such that $f|_{H_i} \in D_i$ and the boundary-gluing conditions (2.5) *hold, we have*

$$\text{ess sup} \left| \mathrm{E}_z \left[\int_{t_0}^\infty e^{-\lambda t} \left[\lambda f(Z^\varepsilon(t)) - \bar{\mathcal{L}} f(Z^\varepsilon(t)) \right] dt - e^{-\lambda t_0} f(Z^\varepsilon(t_0)) | \mathcal{F}_{t_0} \right] \right| \to 0,$$

as $\varepsilon \downarrow 0$ for every $t_0 \geq 0$, $\lambda > \lambda_0$, where $\mathcal{F}_{t_0} = \sigma(Z^\varepsilon(s); s \in [0, t_0])$. In particular, for all such functions, every weak limit of Z^ε as $\varepsilon \downarrow 0$ satisfies the martingale problem for $\bar{\mathcal{L}}$ defined by $\bar{\mathcal{L}} f(z) = \bar{\mathcal{L}}_i f(z)$ for $z \in H_i$.

Proof. By the Markov property of $(Z^\varepsilon, \mathrm{P}_z)$, it is sufficient to show that

$$\Delta(\varepsilon, z) = \mathrm{E}_z \int_0^\infty e^{-\lambda t} \left[\lambda f(Z^\varepsilon(t)) - \bar{\mathcal{L}} f(Z^\varepsilon(t)) \right] dt - f(z) \to 0$$

as $\varepsilon \downarrow 0$ uniformly with respect to $z \in \mathbb{R}^{1+d}$.

Introduce the sequence of stopping times $\sigma_0 = 0$, $\phi_n = \inf\{t > \sigma_n : |X^\varepsilon(t)| = l(\varepsilon)\}$ and $\sigma_{n+1} = \inf\{t > \phi_n : |X^\varepsilon(t)| > \delta(\varepsilon)\}$. Using this sequence, we can decompose the above expression to terms corresponding to downcrossings (on $[\sigma_n, \phi_n]$) and upcrossings (on $[\phi_n, \sigma_{n+1}]$) respectively. More precisely,

$$\Delta(\varepsilon, z) = \sum_{n=0}^{\infty} \mathrm{E}_z \left[e^{-\lambda\phi_n} f(Z^\varepsilon(\phi_n)) - e^{-\lambda\sigma_n} f(Z^\varepsilon(\sigma_n)) + \int_{\sigma_n}^{\phi_n} e^{-\lambda t} \left[\lambda f(Z^\varepsilon(t)) - \bar{\mathcal{L}} f(Z^\varepsilon(t))\right] dt \right] +$$
$$+ \sum_{n=0}^{\infty} \mathrm{E}_z \left[e^{-\lambda\sigma_{n+1}} f(Z^\varepsilon(\sigma_{n+1})) - e^{-\lambda\phi_n} f(Z^\varepsilon(\phi_n)) + \int_{\phi_n}^{\sigma_{n+1}} e^{-\lambda t} \left[\lambda f(Z^\varepsilon(t)) - \bar{\mathcal{L}} f(Z^\varepsilon(t))\right] dt \right]$$

Note that by the strong Markov property, we can bound the downcrossing terms by

$$\sum_{n=0}^{\infty} \mathrm{E}_z e^{-\lambda\sigma_n} \mathrm{E}_{z=Z^\varepsilon(\sigma_n)} \left[e^{-\lambda\tau^\varepsilon} f(Z^\varepsilon(\tau^\varepsilon)) - f(z) + \int_0^{\tau^\varepsilon} e^{-\lambda t} \left[\lambda f(Z^\varepsilon(t)) - \bar{\mathcal{L}}_\pm f(Z^\varepsilon(t))\right] dt \right] \leq$$
$$\leq \mathcal{O}(k(\varepsilon)) \sum_{n=0}^{\infty} \mathrm{E} e^{-\lambda\sigma_n} \tag{5.4}$$

where we used (5.1). To bound the sum on the right hand side, first note that

$$\sum_{n=0}^{\infty} \mathrm{E}_z e^{-\lambda\sigma_n} \leq 1 + \sum_{n=1}^{\infty} \mathrm{E}_z e^{-\lambda\sigma_{n-1}} \mathrm{E}_{Z^\varepsilon(\sigma_{n-1})} e^{-\lambda\tau^\varepsilon} \tag{5.5}$$

Let $|x| = \delta(\varepsilon)$ and apply (5.1) to $u_{i,\lambda}$ given by Condition 2. to get

$$\mathrm{E}_z e^{-\lambda\tau^\varepsilon} u_{i,\lambda}(Z^\varepsilon(\tau^\varepsilon)) - u_{i,\lambda}(z) = \mathcal{O}(k(\varepsilon)).$$

This implies that

$$\mathrm{E}_z e^{-\lambda\tau^\varepsilon} = \mathrm{E}_z e^{-\lambda\tau^\varepsilon}(1 - u_{i,\lambda}(Z^\varepsilon(\tau^\varepsilon))) + u_{i,\lambda}(z) + \mathcal{O}(k(\varepsilon)) \tag{5.6}$$

Since $|X^\varepsilon(\tau^\varepsilon)| = l(\varepsilon)$, we have $|1 - u_{i,\lambda}(Z^\varepsilon(\tau^\varepsilon)| = \mathcal{O}(l(\varepsilon))$ whereas the condition on the derivative of $u_{i,\lambda}$ normal to the interface implies $u_{i,\lambda}(z) \leq 1 - C\delta(\varepsilon)$ for some $C > 0$ since $|z| = \delta(\varepsilon)$. Plugging these back to (5.6) gives

$$\sup_{|z|=\delta(\varepsilon)} \mathrm{E}_z e^{-\lambda\tau^\varepsilon} \leq 1 - C'\delta(\varepsilon)$$

for small enough ε for some $C' > 0$. This, together with (5.5), implies that $\sum_{n=0}^{\infty} \mathrm{E}_z e^{-\lambda\sigma_n} = \mathcal{O}(1/\delta(\varepsilon))$ which in turn implies that the bound in (5.4) converges to zero.

To bound the upcrossing terms, first note that the sum of the integrals $\int_{\phi_n}^{\sigma_{n+1}}$ can be bounded by

$$||(\lambda - \bar{\mathcal{L}})f||_\infty \mathrm{E}_z \int_0^\infty e^{-\lambda t} \chi_{(-\delta(\varepsilon),\delta(\varepsilon))}(Z^\varepsilon(t))dt \to 0$$

as $\varepsilon \downarrow 0$ uniformly in z by (5.2), where we used that f has bounded derivatives of all order. The remaining terms can be written as

$$\sum_{n=0}^{\infty} \mathrm{E}_z(e^{-\lambda\sigma_{n+1}} - e^{\lambda\phi_n}) f(Z^\varepsilon(\sigma_{n+1})) + \sum_{n=0}^{\infty} e^{-\lambda\phi_n}(f(Z^\varepsilon(\sigma_{n+1})) - f(Z^\varepsilon(\phi_n))). \tag{5.7}$$

where the first term can immediately be bounded by

$$||f||_\infty \sup_{z\in\mathbb{R}^{1+d}} \mathrm{E}_z \int_0^\infty \lambda e^{-\lambda t} \chi_{(-\delta(\varepsilon),\delta(\varepsilon))}(Z^\varepsilon(t))dt \to 0.$$

For the second term in (5.7), we can use the strong Markov property to write

$$\sum_{n=0}^\infty \mathrm{E}_z e^{-\lambda\phi_n} \mathrm{E}_{z'=Z^\varepsilon(\phi_n)} \left(f(Z^\varepsilon(\theta^\delta)) - f(z) \right). \tag{5.8}$$

Using $|x'| = l(\varepsilon)$ and Taylor's formula, we get

$$\begin{aligned}
&\mathrm{E}_{z'} \left(f(Z^\varepsilon(\theta^\delta)) - f(z) \right) = E_{z'}[f(Z^\varepsilon(\theta^\delta)) - f(0,y')] + \mathcal{O}(l(\varepsilon)) = \\
&\quad = \delta \mathrm{P}_z(Z^\varepsilon(\theta^\delta) \in H_+)\partial_x f(0+,y') - \delta \mathrm{P}_z(Z^\varepsilon(\theta^\delta) \in H_-)\partial_x f(0-,y') + \\
&\quad + \delta \sum_{i=1}^d \partial_i f(0,y') \frac{1}{\delta} \mathrm{E}_{z'}(Y_i^\varepsilon(\theta^\delta) - y_i') + \delta \sum_{i,j=1}^d \partial_i\partial_j f(0,y') \frac{1}{\delta} E_{z'}(Y_i^\varepsilon(\theta^\delta) - y_i')(Y_j^\varepsilon(\theta^\delta) - y_j') + \\
&\quad + o(\delta),
\end{aligned}$$

where we used the last statement of condition 5. Now using condition 4., the rest of condition 5. and the gluing conditions, we get that this entire expression is $o(\delta)$ uniformly in $|x'| = l(\varepsilon)$. Noting that

$$\sum_{n=0}^\infty \mathrm{E}_z e^{-\lambda\phi_n} \le \sum_{n=0}^\infty e^{-\lambda\sigma_n} = \mathcal{O}(1/\delta),$$

(5.8) converges to zero and the proof is finished. That every weak limit of Z^ε converges to the solution of the martingale problem of $\bar{\mathcal{L}}$ follows exactly as in [9] after noticing that if the Laplace transform of a function is identically zero for sufficiently large λ then the function itself is identically zero. $\square$

We are going to need some estimates on the exit time θ^δ.

Lemma 5.2.

$$\mathrm{E}_z\theta^\delta \le C\delta^2, \qquad \mathrm{E}_z(\theta^\delta)^2 \le C\delta^4.$$

Moreover, there is a $\delta_0 > 0$ such that $\delta \in (0, \delta_0]$ and $|x| < \delta$ implies

$$\mathrm{E}_z e^{u\theta^\delta} \le \frac{1}{\cos(Cu\delta)}.$$

Proof. Let W be a Brownian motion such that $X^\varepsilon(\cdot)$ and $x + W(\int_0^\cdot |\varphi|^2(s)ds)$ have the same distribution and therefore so do θ_W^δ (the hitting time of δ by W) and $\int_0^{\theta^\delta} |\varphi(s)|^2 ds$. The results follow from this, $|\varphi|^2 \in [c_1, c_2]$ and well-known Brownian formulas. $\square$

5.2 Martingale Convergence for the long-time asymptotics

Now we show that the system of equations in Theorem 2.6 satisfies the conditions of Theorem 5.1 with $D_i = \mathcal{C}_b^\infty(H_i)$. As $b_1 \equiv 0$, we denote $b \equiv b_2$ in this case. Let $Z^\varepsilon = (\bar{X}^\varepsilon, \bar{Y}^\varepsilon)$ solve

$$\begin{cases} d\bar{X}^\varepsilon(t) = \varphi(\varepsilon^{-2}\bar{X}^\varepsilon(t), \bar{Y}^\varepsilon(t))dW(t), \\ d\bar{Y}^\varepsilon(t) = \varepsilon^{-2}b(\varepsilon^{-2}\bar{X}^\varepsilon(t), \bar{Y}^\varepsilon(t))dt + \varepsilon^{-1}\sigma(\varepsilon^{-2}\bar{X}^\varepsilon(t), \bar{Y}^\varepsilon(t))dW(t). \end{cases} \tag{5.9}$$

Lemma 5.3. *Let $f \in D_i$ $i = -, +$. If we choose $l(\varepsilon)$ so that $l(\varepsilon) \to 0$ and $\varepsilon^{-2}l(\varepsilon) \to +\infty$, then (5.1) is satisfied for $Z^\varepsilon = (\bar{X}^\varepsilon, \bar{Y}^\varepsilon)$ and $\bar{\mathcal{L}}f(x, y) = \frac{1}{2}a_\pm(x, y)f_{xx}(x, y)$.*

Proof. By Ito's formula and a standard localization argument,

$$K(\varepsilon, z) = \mathrm{E}_z\left[e^{-\lambda\tau^\varepsilon}f(Z^\varepsilon(\tau^\varepsilon)) - f(z) - \int_0^{\tau^\varepsilon} e^{-\lambda t}\left(\lambda f(Z^\varepsilon(t)) - \frac{1}{2}a_\pm^2(Z^\varepsilon(t))f_{xx}(Z^\varepsilon(t))\right)dt\right]$$

$$= \varepsilon^{-2}\mathrm{E}_z\int_0^{\tau^\varepsilon} e^{-\lambda t}\left(\sum_{i=1}^d b_i(\varepsilon^{-2}\bar{X}^\varepsilon(t), \bar{Y}^\varepsilon(t))f_{y_i}(Z^\varepsilon(t))\right)dt$$
$$+\frac{1}{2}\varepsilon^{-2}\mathrm{E}_z\int_0^{\tau^\varepsilon} e^{-\lambda t}\left(\sum_{i,j=1}^d (\sigma\sigma^T)_{ij}(\varepsilon^{-2}\bar{X}^\varepsilon(t), \bar{Y}^\varepsilon(t))f_{y_iy_j}(Z^\varepsilon(t))\right)dt$$
$$+\varepsilon^{-1}\mathrm{E}_z\int_0^{\tau_\varepsilon} e^{-\lambda t}\left(\sum_{i=1}^d(\sigma\varphi^T)_i(\varepsilon^{-2}\bar{X}^\varepsilon(t), \bar{Y}^\varepsilon(t))f_{xy_i}(Z^\varepsilon(t))\right)dt$$
$$+\frac{1}{2}\mathrm{E}_z\int_0^{\tau^\varepsilon} e^{-\lambda t}\left(|\varphi|^2(\varepsilon^{-2}\bar{X}^\varepsilon(t), \bar{Y}^\varepsilon(t)) - a_\pm(Z^\varepsilon(t))\right)f_{xx}(Z^\varepsilon(t))dt.$$
$$:= I_1^\varepsilon + I_2^\varepsilon + I_3^\varepsilon + I_4^\varepsilon.$$

We estimate each term separately. Note that

$$I_1^\varepsilon \leq c\varepsilon^{-2}\mathrm{E}_z\int_0^{\tau^\varepsilon} e^{-\lambda t}\hat{b}(\varepsilon^{-2}\bar{X}^\varepsilon(t))\|f\|_{C^1}dt,$$

then because $\tau^\varepsilon = \inf\{t > 0 : |\bar{X}^\varepsilon| \leq l(\varepsilon)\}$, we have

$$\begin{aligned} I_1^\varepsilon &\leq c\varepsilon^{-2}\|f\|_{C^1}\mathrm{E}_z\int_0^\infty e^{-\lambda t}\hat{b}(\varepsilon^{-2}\bar{X}^\varepsilon(t))\mathbb{1}_{\{|\bar{X}^\varepsilon(t)|>l(\varepsilon)\}}dt \\ &\leq c\varepsilon^{-2}\|f\|_{C^1}\sum_{k=0}^\infty e^{-\lambda k}\mathrm{E}_z\int_k^{k+1}\hat{b}(\varepsilon^{-2}\bar{X}^\varepsilon(t))\mathbb{1}_{\{|\bar{X}^\varepsilon(t)|>l(\varepsilon)\}}dt. \end{aligned}$$

By Lemma 3.1, there exists a constant c, independent of k and ε such that

$$\varepsilon^{-2}\mathrm{E}_z\int_k^{k+1}\hat{b}(\varepsilon^{-2}\bar{X}^\varepsilon(t))\mathbb{1}_{\{|\bar{X}^\varepsilon(t)|>l(\varepsilon)\}}dt \leq c\int\limits_{\{|x|>\varepsilon^{-2}l(\varepsilon)\}}\hat{b}(x)dx.$$

Because $\hat{b} \in L^1(\mathbb{R})$ and $\varepsilon^{-2}l(\varepsilon) \to +\infty$, it follows that $I_1^\varepsilon \to 0$ uniformly in $z \in H_i$. The analysis of I_2^ε is analogous to that of I_1^ε and we can conclude that $I_2^\varepsilon \to 0$ unifomly in $z \in H_i$. To estimate I_3^ε, we recall that $\|\varphi\|_C^2 \leq c_2$, and therefore a little manipulation shows that

$$|I_3^\varepsilon| \leq \varepsilon^{-1}c\|f\|_{C^2}\mathrm{E}_z\int_0^{\tau^\varepsilon} e^{-\lambda t}\hat{\sigma}(\varepsilon^{-2}\bar{X}^\varepsilon(t))dt.$$

By Hölder's inequality for the integral $\mathrm{E}_z \int_0^{\tau^\varepsilon} e^{-\lambda t} \cdot dt$,

$$
\begin{aligned}
|I_3^\varepsilon| &\le \varepsilon^{-1} c \|f\|_{C^2} \left(\mathrm{E}_z \int_0^{\tau^\varepsilon} e^{-\lambda t} dt \right)^{1/2} \left(\mathrm{E}_z \int_0^{\tau^\varepsilon} e^{-\lambda t} \hat{\sigma}^2 (\varepsilon^{-2} \bar{X}^\varepsilon(t)) dt \right)^{1/2} \\
&\le c \|f\|_{C^2} \left(\varepsilon^{-2} \mathrm{E}_z \int_0^{\tau^\varepsilon} e^{-\lambda t} \hat{\sigma}^2 (\varepsilon^{-2} \bar{X}^\varepsilon(t)) dt \right)^{1/2}.
\end{aligned}
$$

Therefore, $I_3^\varepsilon \to 0$ unifomly in $z \in H_i$ by the same arguments used for I_1^ε. We show that $I_4^\varepsilon \to 0$ by applying Lemma 3.3. □

Lemma 5.4. *For each $\lambda > 0$, and $i \in \{+, -\}$, there is a $u_{i,\lambda} \in D_i$ such that*

$$\frac{1}{2} a_i(z) \partial_{xx} u_{i,\lambda}(z) = \lambda u_{i,\lambda}(z) \qquad \forall z \in H_i, |x| < 1$$

and such that $u_{\pm,\lambda}(0, y) = 1$ and $\pm \partial_x u_{\pm,\lambda}(0, y) < c < 0$.

Proof. Set

$$u_{\pm,\lambda}(x, y) = e^{\mp \sqrt{\frac{2\lambda}{a_\pm(y)}} x}.$$

This function satisfies all the requirements. □

Lemma 5.5. *Let $\delta(\varepsilon) \to 0$. Then (5.2) holds for Z^ε.*

Proof. First we observe that

$$\mathrm{E}_z \int_0^\infty e^{-\lambda t} \chi_{(-\delta(\varepsilon), \delta(\varepsilon))} (\bar{X}^\varepsilon(t)) dt \le \sum_{k=0}^\infty e^{-\lambda k} \mathrm{E}_z \int_k^{k+1} \chi_{(-\delta(\varepsilon), \delta(\varepsilon))} (\bar{X}^\varepsilon(t)) dt.$$

By Lemma 3.1, there exists $c > 0$ independent of ε and k such that

$$\mathrm{E}_z \int_k^{k+1} \chi_{(-\delta(\varepsilon), \delta(\varepsilon))} (\bar{X}^\varepsilon(t)) dt \le c |\chi_{(-\delta(\varepsilon), \delta(\varepsilon))}|_{L^1} \le c \delta(\varepsilon).$$

Therefore, for any $z \in \mathbb{R}^{1+d}$,

$$\mathrm{E}_z \int_0^\infty e^{-\lambda t} \chi_{(-\delta(\varepsilon), \delta(\varepsilon))} (\bar{X}^\varepsilon(t)) dt \le \frac{c}{1 - e^{-\lambda}} \delta(\varepsilon) \to 0$$

as $\varepsilon \downarrow 0$. □

Lemma 5.6. *For the system Z^ε, $p_+ = p_- = \frac{1}{2}$ in the sense that*

$$\lim_{\varepsilon \to 0} \mathrm{P}(Z^\varepsilon(\theta^\delta) \in H_+) = \lim_{\varepsilon \to 0} \mathrm{P}(Z^\varepsilon(\theta^\delta) \in H_-) = \frac{1}{2} \tag{5.10}$$

uniformly in $z \in \mathbb{R}^{1+d}$, $|x| < l(\varepsilon)$ as long as $l(\varepsilon)/\delta(\varepsilon) \to 0$.

Proof. Let

$$\mathrm{P}_z(\bar{X}^\varepsilon(\theta^\delta) \in H_+) = f^\varepsilon(x, y)$$

Since the exit only depends on the x component, and the effect of φ is a time change that does not effect the exit probabilities, it is easy to see that the corresponding Brownian formula holds, i.e.

$$f^\varepsilon(x) = \frac{1}{2\delta(\varepsilon)}(x + \delta(\varepsilon))$$

and if $l(\varepsilon) \ll \delta(\varepsilon)$,

$$\lim_{\varepsilon\to 0} \sup_{|x|\le l(\varepsilon)} \left| \mathrm{P}_x(\bar{X}^\varepsilon(\theta^\delta) \in H_+) - \frac{1}{2} \right| \to 0,$$

as $\varepsilon \downarrow 0$. The part for H_- follows by complementation. $\square$

Lemma 5.7. *For $i, j = 1, ..., n$, let*

$$\beta_i(y) = \int_{-\infty}^{\infty} \frac{b_i}{|\varphi|^2}(x, y)dx, \qquad \alpha_{ij}(y) = \int_{-\infty}^{\infty} \frac{(\sigma\sigma^T)_{ij}}{|\varphi|^2}(x, y)dx.$$

Then if $\delta(\varepsilon), l(\varepsilon) \to 0$ such that $\delta(\varepsilon)/l(\varepsilon) \to \infty$, and $\theta^\delta = \inf\{t > 0 : |X^\varepsilon(t)| > \delta\}$,

$$\frac{1}{\delta}\mathrm{E}_z(\bar{Y}_i^\varepsilon(\theta^\delta) - y_i) \to \beta_i(y), \qquad \frac{1}{\delta}\mathrm{E}_z(\bar{Y}_i^\varepsilon(\theta^\delta) - y_i)(\bar{Y}_j^\varepsilon(\theta^\delta) - y_j) \to \alpha_{ij}(y)$$

as $\varepsilon \to 0$ for $|x| < l(\varepsilon)$. Moreover,

$$\sup_{|x|\le l(\varepsilon)} \mathrm{E}|\bar{Y}_i^\varepsilon(\theta^\delta) - y|^3 = o(\delta).$$

Proof. Observe that

$$\frac{1}{\delta}\mathrm{E}_z(\bar{Y}_i^\varepsilon(\theta^\delta) - y_i) = \frac{1}{\delta\varepsilon^2}\mathrm{E}_z \int_0^{\theta^\delta} b_i(\varepsilon^{-2}\bar{X}^\varepsilon(s), \bar{Y}^\varepsilon(s))ds. \tag{5.11}$$

By Lemma 3.1 and the Lipschitz continuity of b, and the fact that

$$|b_i(\varepsilon^{-2}\bar{X}^\varepsilon(t), \bar{Y}^\varepsilon(t)) - b_i(\varepsilon^{-2}\bar{X}(t), y)| \le 2\hat{b}(\varepsilon^{-2}\bar{X}(t))\chi_{\{|\varepsilon^{-2}\bar{X}^\varepsilon(t)|\ge K\}} + c|\bar{Y}^\varepsilon(t) - y|\chi_{\{|\bar{X}^\varepsilon|\le\varepsilon^2 K\}},$$

we have

$$\begin{aligned} I(\delta, \varepsilon) &= \frac{1}{\delta\varepsilon^2} \sup_{|x|\le l(\varepsilon)} \left| \mathrm{E}_z \int_0^{\theta^\delta} b_i(\varepsilon^{-2}\bar{X}^\varepsilon(s), \bar{Y}^\varepsilon(s))ds - \mathrm{E}_z \int_0^{\theta^\delta} b_i(\varepsilon^{-2}\bar{X}^\varepsilon(s), y)ds \right| \\ &\le \frac{2}{\delta}\sqrt{\mathrm{E}\theta^\delta}||\mathbb{1}_{\{x\ge K\}}\hat{b}||_{L^1} + \frac{c}{\delta\varepsilon^2}\mathrm{E}_z \int_0^{\theta^\delta} \mathbb{1}_{\{|\bar{X}^\varepsilon(s)|<K\varepsilon^2\}}|\bar{Y}^\varepsilon(s) - y|ds. \end{aligned} \tag{5.12}$$

Using $\mathrm{E}_z\theta^\delta \le C\delta^2$ for all $|x| \le l(\varepsilon)$, we see that the first term can be made less than $\eta/2$ for any $\eta > 0$ if K is chosen sufficiently large. Fix such a K and note that by Cauchy-Schwarz inequality, the second term in (5.12) can be bounded from above by

$$\frac{1}{\delta\varepsilon^2}\mathrm{E}_z\left[\sup_{t\in[0,\theta^\delta]}|\bar{Y}^\varepsilon(t)-y|\int_0^{\theta^\delta}\mathbb{1}_{\{|\bar{X}^\varepsilon(s)|<K\varepsilon^2\}}ds\right] \le \tag{5.13}$$

$$\le \frac{1}{\delta}\left(\mathrm{E}_z\sup_{t\in[0,\theta^\delta]}|\bar{Y}^\varepsilon(t)-y|^2\right)^{1/2}\left[\mathrm{E}_z\left(\frac{1}{\varepsilon^2}\int_0^{\theta^\delta}\mathbb{1}_{\{|\bar{X}^\varepsilon(s)|<K\varepsilon^2\}}ds\right)^2\right]^{1/2}$$

where the second term is easily seen to be less than $c(\mathrm{E}\theta^\delta)^{1/2} \le c\delta$ by Lemma 3.1. On the other hand, $(a+b)^2 \le 2(a^2+b^2)$ and the BDG inequality imply that

$$\mathrm{E}_z\left(\sup_{t\in[0,\theta^\delta]}|\bar{Y}^\varepsilon(t)-y|^2\right) \le \tag{5.14}$$

$$\le 2\mathrm{E}\left(\frac{1}{\varepsilon^2}\int_0^{\theta}\hat{b}(\varepsilon^{-2}\bar{X}^\varepsilon(s))ds\right)^2 + 2c\mathrm{E}\left(\frac{1}{\varepsilon^2}\int_0^{\theta^\delta}\hat{\sigma}(\varepsilon^{-1}\bar{X}^\varepsilon(s))ds\right) \le c\mathrm{E}_z\theta^\delta \le c\delta^2,$$

where Lemma 3.1 and Lemma 5.2 were used in the last two inequalities. By (5.13), and (5.14), the second term in (5.12) is less than $c\delta < \eta/2$ and consequently $I(\delta,\varepsilon) < \eta$ for sufficiently small δ. Since η was arbitrary, (5.11) implies that

$$\frac{1}{\delta}\mathrm{E}_z(\bar{Y}_i^\varepsilon(\theta^\delta)-y_i) = \frac{1}{\delta\varepsilon^2}\mathrm{E}_z\int_0^{\theta^\delta} b_i(\varepsilon^{-2}\bar{X}^\varepsilon(s),y)ds + o(1).$$

By the properties of the local time, the above expression is equal to

$$\frac{1}{\varepsilon^2\delta}\mathrm{E}_z\int_{-\infty}^{\infty}\frac{b_i}{|\varphi|^2}(\varepsilon^{-2}u,y)L^{\bar{X}^\varepsilon}(u,\theta^\delta)du = \frac{1}{\delta}\mathrm{E}_z\int_{-\infty}^{\infty}\frac{b_i}{|\varphi|^2}(u,y)L^{\bar{X}^\varepsilon}(\varepsilon^2u,\theta^\delta)du$$

where $L^{\bar{X}^\varepsilon}(u,t)$ is the local time of $\bar{X}^\varepsilon$ at u. We calculate by the Tanaka formula that for fixed x,

$$\frac{1}{\delta}\mathrm{E}_zL^{\bar{X}^\varepsilon}(\varepsilon^2u,\theta^\delta) = \frac{1}{\delta}\mathrm{E}\left(|\bar{X}^\varepsilon(\theta^\delta)-\varepsilon^2u| - |x-\varepsilon^2u|\right)$$

Because $\bar{X}^\varepsilon(\theta^\delta) = \pm\delta$, the above expression is bounded by

$$\frac{\delta+|x|}{\delta}$$

and if $|x| \le l(\varepsilon) \ll \delta(\varepsilon)$,

$$\lim_{\varepsilon\to 0}\frac{1}{\delta}\mathrm{E}_zL^{\bar{X}^\varepsilon}(\varepsilon u,\theta^\delta) = 1.$$

Therefore, it follows from (5.2) and the dominated convergence theorem that

$$\frac{1}{\delta}\mathrm{E}_z(\bar{Y}_i^\varepsilon(\theta^\delta)-y_i) \to \int_{-\infty}^{\infty}\frac{b_i}{|\varphi|^2}(x,y)dx.$$

Similar calculations show that

$$\frac{1}{\delta}\mathrm{E}_z(\bar{Y}_i^\varepsilon(\theta^\delta)-y_i)(\bar{Y}_j^\varepsilon(\theta^\delta)-y_j)=\frac{1}{\delta\varepsilon^2}\mathrm{E}_z\int_0^{\theta^\delta}(\sigma\sigma^T)_{ij}(\varepsilon^{-2}\bar{X}^\varepsilon(t),\bar{Y}^\varepsilon(t))dt\to\int_{-\infty}^{\infty}\frac{(\sigma\sigma^T)_{ij}}{|\varphi|^2}(x,y)dx.$$

and a further straightforward application of Lemma 3.1 yields

$$\sup_{|x|\le l(\varepsilon)}\mathrm{E}|\bar{Y}_i^\varepsilon(\theta^\delta)-y|^3=\mathcal{O}(\delta^{3/2})=o(\delta).$$

□

Set $\gamma>0$, $l(\varepsilon)=\varepsilon^{2-\gamma}$ and $\delta(\varepsilon)=\max\{\varepsilon^{2-2\gamma},\varepsilon^{-\gamma}\sup_{|x|>l(\varepsilon)}|k(\varepsilon)|\}$ where $k(\varepsilon)$ is the convergence rate in Lemma 5.3. Lemmas 5.3-5.7 and Theorem 5.1 together with tightness imply that every subsequence of $(\bar{X}^\varepsilon,\bar{Y}^\varepsilon)$ has a further subsequence converging to a solution of the martingale problem associated to $\bar{\mathcal{L}}$. In the next section we will argue that this martingale problem is well posed which establishes the convergence of the entire sequence.

5.3 Convergence to the solution of the martingale problem in the normalized deviation case

In this section, we show that the assumptions of Theorem 5.1 are satisfied by $Z^\varepsilon=(X^\varepsilon,y,\zeta^\varepsilon)$ with $Z^\varepsilon(0)=z=(x,y,\xi)$ under the assumptions of Theorem 2.2 or Theorem 2.4. We start with a simple growth estimate on ζ^ε.

Lemma 5.8. *For any stopping time $\tau>0$ with $\mathrm{E}\tau<\infty$,*

(1) under the conditions of Theorem 2.4, we have

$$\mathrm{E}_z\sup_{0\le t\le\tau}|\zeta^\varepsilon(t)|^2\le C\left(\mathrm{E}_z e^{4||b_1||_{C^1}\tau}\right)^{1/2}(|\xi|^4+\mathrm{E}_z\tau^2)^{1/2},\tag{5.15}$$

(2) under the conditions of Theorem 2.2, we have

$$\mathrm{E}_z\sup_{0\le t\le\tau}|\zeta^\varepsilon(t)|^2\le\left(\mathrm{E}_z e^{4||b_1||_{C^1}\tau}\right)^{1/2}(|\xi|^4+\mathrm{E}_z\tau+\varepsilon^2\mathrm{E}_z\tau^2)^{1/2}.\tag{5.16}$$

Proof. Recall that under the conditions of Theorem 2.4, we have

$$d\zeta^\varepsilon(t)=\frac{1}{\varepsilon}[b_1(y(t)+\varepsilon\zeta^\varepsilon(t))-b_1(y(t))]+\frac{1}{\varepsilon}b_2\left(\varepsilon^{-1}X^\varepsilon(t),y(t)+\varepsilon\zeta^\varepsilon(t)\right)dt,$$

which implies

$$|\zeta^\varepsilon(t)|\le|\xi|+||b_1||_{C^1}\int_0^t|\zeta^\varepsilon(s)|ds+\int_0^t\varepsilon^{-1}\hat{b}(\varepsilon^{-1}X^\varepsilon(s))ds.$$

By Gronwall's lemma,

$$\sup_{0\le t\le\tau}|\zeta^\varepsilon(t)|\le\left(|\xi|+\int_0^\tau\varepsilon^{-1}\hat{b}(\varepsilon^{-1}X^\varepsilon(s))ds\right)e^{||b_1||_{C^1}\tau},$$

and therefore by the Cauchy-Schwartz inequality,

$$\mathrm{E}_z \sup_{t\in[0,\tau]} |\zeta^\varepsilon(t)|^2 \le C\left(\mathrm{E}_z e^{4\|b_1\|_{C^1}\tau}\right)^{1/2}\left(|\xi|^4 + \mathrm{E}_z\left(\int_0^\tau \varepsilon^{-1}\hat{b}(\varepsilon^{-1}X^\varepsilon(t))dt\right)^4\right)^{1/2},$$

and Lemma 3.1 and Lemma 5.2 yield (1).

To prove (2), note that under the conditions of Theorem 2.2,

$$d\zeta^\varepsilon(t) = \frac{1}{\sqrt{\varepsilon}}[b_1(y(t)+\sqrt{\varepsilon}\zeta^\varepsilon(t)) - b_1(y(t))] + \frac{1}{\sqrt{\varepsilon}}b_2\left(\varepsilon^{-1}X^\varepsilon(t), y(t)+\sqrt{\varepsilon}\zeta^\varepsilon(t)\right)dt +$$
$$+\frac{1}{\sqrt{\varepsilon}}\sigma\left(\varepsilon^{-1}X^\varepsilon(t), y(t)+\sqrt{\varepsilon}\zeta^\varepsilon(t)\right)dW(t)$$

and one can easily see as above that

$$\mathrm{E}_z \sup_{t\in[0,\tau]} |\zeta^\varepsilon(t)|^2 \le C\left(\mathrm{E}_z e^{4\|b_1\|_{C^1}\tau}\right)^{1/2}\cdot$$
$$\cdot\left(|\xi|^4 + \varepsilon^2\mathrm{E}_z\left(\int_0^\tau \varepsilon^{-1}\hat{b}(\varepsilon^{-1}X^\varepsilon(t))dt\right)^4 + \mathrm{E}_z\left(\int_0^\tau \varepsilon^{-1}\hat{\sigma}(\varepsilon^{-1}X^\varepsilon(s))ds\right)^2\right)^{1/2},$$

where we used the BDG inequality in the last term. (2) once again follows by applying Lemma 3.1 twice and Lemma 5.2. □

Lemma 5.9. *There is a $\lambda_0 > 0$ such that for $\lambda > \lambda_0$, if we choose $l(\varepsilon)$ so that $l(\varepsilon) \to 0$ and $\varepsilon^{-1}l(\varepsilon) \to +\infty$, then (5.1) is satisfied with some rate $k(\varepsilon)$ for $Z^\varepsilon(t) = (X^\varepsilon(t), y(t), \zeta^\varepsilon(t))$ with*

$$\bar{\mathcal{L}}f(x,y,\xi) = \frac{1}{2}a_\pm(x,y)f_{xx}(x,y,\xi) + b_1(y)\cdot f_y(x,y,\xi) + (\partial_y b_1(y)\cdot\xi)\cdot f_\xi(x,y,\xi).$$

Proof. We only prove the lemma under the hypothesis of Theorem 2.2, the other case being almost identical. Using Ito's formula,

$$\mathrm{E}_z\left[e^{-\lambda\tau^\varepsilon}f(Z^\varepsilon(\tau^\varepsilon)) - f(z) + \int_0^{\tau^\varepsilon} e^{-\lambda t}(\lambda f(Z^\varepsilon(t)) - \bar{\mathcal{L}}_i f(Z^\varepsilon(t)))dt\right] = I_1(z,\varepsilon) + I_2(z,\varepsilon)$$

where $I_1(z,\varepsilon)$ contains terms that can be dealt with exactly as we did with their counterparts in Lemma 5.3 and therefore

$$\sup_{z\in H_i}|I_1(z,\varepsilon)| \to 0.$$

On the other hand,

$$I_2(z,\varepsilon) = \frac{1}{2}\mathrm{E}_z\int_0^{\tau^\varepsilon} e^{-\lambda t}f_{xx}(X^\varepsilon(t), y(t), \zeta^\varepsilon(t))\left[\frac{b_1(y(t)+\sqrt{\varepsilon}\zeta^\varepsilon(t)) - b_1(y(t))}{\sqrt{\varepsilon}} - (\partial_y b_1)(y(t))\zeta^\varepsilon(t)\right]dt,$$

and because $b_1(y) \in C_b^2(\mathbb{R}^d)$, we have that

$$|I_2(z,\varepsilon)| \le C\sqrt{\varepsilon}\int_0^\infty e^{-\lambda t}\mathrm{E}_z|\zeta^\varepsilon(t)|^2 dt, \tag{5.17}$$

where the constant C depeds on $|f|_{C^2}$ and the second derivatives of b_1. By (5.16), we have for all $t > 0$

$$\mathrm{E}_z|\zeta^\varepsilon(t)|^2 \le C_1 t e^{C_2 t}$$

for some $C_1, C_2 > 0$ and therefore choosing $\lambda_0 = C_2$ makes the integral in (5.17) converge for every $\lambda > \lambda_0$ and the result is proved. $\square$

Lemma 5.10. *For each $\lambda > 0$, and $i = +, -$, there is a $u_{i,\lambda} \in D_i$ such that*

$$\bar{\mathcal{L}}_i u_{i,\lambda} = \lambda u_{i,\lambda}.$$

Proof. The operators $\mathcal{L}_\pm$ are both generators of non-degenerate diffusion processes $Z = (X, y, \zeta)$ in $H_\pm$. Define

$$\tau = \inf\{t > 0 : X(t) = 0\},$$

then it is well known that

$$u_{i,\lambda}(z) := \mathrm{E}_z e^{-\lambda\tau} \tag{5.18}$$

has the property we desire. That the requirement on the derivative is also fulfilled once again follows from a time change argument and the corresponding Brownian formula. $\square$

The statements of the following lemma are proved completely analogously to those in the previous section.

Lemma 5.11. *Property (5.2) holds for Z^ε. Also, $p_+ = p_- = \frac{1}{2}$ in the sense that*

$$\lim_{\varepsilon\to 0} \mathrm{P}_z\left(Z^\varepsilon(\tau^\varepsilon) \in H_+\right) = \lim_{\varepsilon\to 0} \mathrm{P}_z\left(Z^\varepsilon(\tau^\varepsilon) \in H_-\right) = \frac{1}{2}$$

uniformly for $(x, y, z) \in \mathbb{R}^{1+2d}$, $|x| < l(\varepsilon)$.

Lemma 5.12. *For $i, j = 1, .., n$, let*

$$\beta_i(y) = \begin{cases} \int_{-\infty}^{\infty} \frac{b_{2,i}}{|\varphi|^2}(x, y)dx & \text{if } \sigma \equiv 0, \\ 0 & \text{otherwise,} \end{cases} \qquad \alpha_{ij}(y) = \int_{-\infty}^{\infty} \frac{(\sigma\sigma^T)_{ij}}{|\varphi|^2}(x, y)dx.$$

Then if $\delta(\varepsilon)/l(\varepsilon) \to \infty$,

$$\frac{1}{\delta}\mathrm{E}_z(\zeta_i^\varepsilon(\theta^\delta) - \xi_i) \to \beta_i(y), \qquad \frac{1}{\delta}\mathrm{E}_z(\zeta_i^\varepsilon(\theta^\delta) - \xi_i)(\zeta_j^\varepsilon(\theta^\delta) - \xi_j) \to \alpha_{ij}(y) \tag{5.19}$$

as $\varepsilon \to 0$ uniformly in $|x| < l(\varepsilon)$. Moreover,

$$\mathrm{E}_z|\zeta_i^\varepsilon(\theta^\delta) - \xi|^3 = o(\delta). \tag{5.20}$$

Proof. We only prove the case $\sigma \equiv 0$, the other one being similar. First note that by (5.15) and Lemma 5.2, there is a $\delta_0 > 0$ and $M > 0$ such that

$$\mathrm{E}_z \sup_{t\in[0,\theta^\delta]} |\zeta^\varepsilon(t)|^2 \le M, \tag{5.21}$$

whenever $\delta < \delta_0$. Now write

$$\frac{1}{\delta}\mathrm{E}_z(\zeta_i^\varepsilon(\theta^\delta) - \xi_i) = \tag{5.22}$$

$$= \frac{1}{\delta\varepsilon}\mathrm{E}_z \int_0^{\theta^\delta} [b_1(y(s) + \varepsilon\zeta^\varepsilon(u)) - b_1(y(s))]\, dt + \frac{1}{\delta\varepsilon}\mathrm{E}_z \int_0^{\theta^\delta} b_2(\varepsilon^{-1}X^\varepsilon(s), y(s) + \varepsilon\zeta^\varepsilon(s))ds$$

The second term can be dealt with exactly as in (5.12) and (5.13) and it can be shown to converge to

$$\beta(y) = \int_{-\infty}^{\infty} \frac{b_2}{|\varphi|^2}(x, y)dx$$

uniformly in $|x| < l(\varepsilon)$ once one notices that

$$\mathrm{E}_z \sup_{t\in[0,\theta^\delta]} |y(t) + \varepsilon\zeta^\varepsilon(t) - y|^2 \leq C\mathrm{E}_z \sup_{t\in[0,\theta^\delta]} |y(t) - y|^2 + \varepsilon^2\mathrm{E}_z \sup_{t\in[0,\theta^\delta]} |\zeta^\varepsilon(t)|^2 = o(1)$$

by (5.21), a standard Gronwall estimate on $y(t)$, and Lemma 5.2. On the other hand, the integral in the first term in (5.22) is immediately less than or equal to

$$\frac{||b_1||_{C^1}}{\delta}\mathrm{E}_z \int_0^{\theta^\delta} |\zeta^\varepsilon(t)|dt \leq \frac{||b_1||}{\delta}\mathrm{E}_z \left[\theta^\delta \cdot \sup_{t\in[0,\theta^\delta]} |\zeta^\varepsilon(t)|\right] \leq \left(\mathrm{E}_z \sup_{t\in[0,\theta^\delta]} |\zeta^\varepsilon(t)|^2\right)^{1/2} \sqrt{\delta^{-2}\mathrm{E}_z(\theta^\delta)^2}$$

which converges to zero by (5.15) and Lemma 5.2.

The second claim in (5.19) and (5.20) can be proved similarly. $\square$

If we set $\gamma > 0$, $l(\varepsilon) = \varepsilon^{1-\gamma}$ and $\delta(\varepsilon) = \max\{\varepsilon^{1-2\gamma}, \varepsilon^{-\gamma}k(\varepsilon)\}$ then Lemmas 5.9-5.12 and Theorem 5.1 with (y, ξ) in place of y show that the possible subsequential limits are all solutions of the martingale problem associated to $\bar{\mathcal{L}}$ and the convergence will be established once we show well-posedness in the next section.

6 Uniqueness and Markov property of the martingale problem and characterization of the solution.

In this section, we prove that the martingale problems associated to the operators in Theorem 2.2, Theorem 2.4 and Theorem 2.6 are well-posed and the unique solutions satisfy the stated SDEs.

Our main tool will be the following result which is a special case of a result of Ethier and Kurtz which we reformulate for our current purposes.

Theorem 6.1 ([7, Theorem 4.4.1] Uniqueness of the Martingale Problem)**.** *Let $\mathcal{L}$ be linear and dissipative on $C_0(\mathbb{R}^{d+1})$. If $\overline{D(\mathcal{L})} = C_0(\mathbb{R}^{d+1})$ and there exists a $\lambda > 0$ such that $\overline{(\lambda - \mathcal{L})(D(\mathcal{L}))} = C_0(\mathbb{R}^{d+1})$, then the martingale problem is unique for $\mathcal{L}$. That is, if $Z, \tilde{Z}$ are processes in $C([0,\infty), \mathbb{R}^{1+d})$ and $Z(0) = \tilde{Z}(0)$ in distribution and for any $f \in D$,*

$$f(Z(t)) - \int_0^t Lf(Z(s))ds \text{ and } f(\tilde{Z}(t)) - \int_0^t Lf(\tilde{Z}(s))ds,$$

are martingales with respect to the filtration generated by the coordinate process then Z and $\tilde{Z}$ are equal in distribution. Moreover, if there is a solution Z, it is a Markov process corresponding to the semigroup on $C_0(\mathbb{R}^{d+1})$ generated by the closure of L.

In all three cases, we probabilistically construct a solution of the corresponding stochastic differential equation and conclude by the Hille-Yosida theorem that the generator of the associated Markov semigroup satisfies the hypothesis of Theorem 6.1. This together with Lemma 3.5 implies that the measure generated by this process is the unique solution of the martingale problem associated to $\bar{\mathcal{L}}$.

6.1 Long-time case

By the concrete form of β and α (as in Theorem 2.6), there is a weak solution (Y_0, W_2) to the equation

$$dY_0(t) = \beta(Y_0(t))dt + \sqrt{\alpha(Y_0(t))}dW_2(t).$$

Take an independent Brownian moton W_1 and consider the time changed process

$$\hat{Z}(t) = (\hat{X}(t), \hat{Y}(t)) = (W_1(t), Y_0(L^{W_1}(t,0)))$$

where $L^{W_1}(t,0)$ is the local time of W_1 at zero. Finally, set

$$t(s) = \int_0^s \frac{1}{a_\pm(W_1(u), \hat{Y}(u))}du$$

and let $s(t)$ denote its inverse. Let $Z(t) = \hat{Z}(s(t))$.

Lemma 6.2. *$Z(t) =: (\bar{X}^0(t), \bar{Y}^0(t))$ is a weak solution of (2.14)-(2.15).*

Proof. First note that if we introduce $V^{W_1}(t) = W_2(L^{W_1}(t,0))$ then $\hat{Y}$ satisfies

$$\hat{Y}(t) = \int_0^t \beta(\hat{Y}(s))L^{W_1}(ds,0) + \int_0^t \sqrt{\alpha(\hat{Y}(s))}dV^{W_1}(t)$$

Note that V^{W_1} is a continuous martingale and

$$\mathrm{P}\left(\sum_{i,j=1}^d \int_0^t \alpha_{ij}(\hat{Y}(s))d\langle V_j^{W_1}, V_i^{W_1}\rangle(s) < \infty\right) = 1$$

for all $t > 0$ and therefore the stochastic integral is well defined for all times. Indeed, it follows from the construction that $\left\langle V_i^{W_1}, V_j^{W_1}\right\rangle = L^{W_1}(t,0)\delta_{ij}$. We also calculate that by (2.3)

$$\sum_{i=1}^d \alpha_{ii}(y) = \int_{-\infty}^{\infty} \frac{\mathrm{Tr}(\sigma\sigma^T)}{|\varphi|^2}(x,y)dx \le c|\hat{\sigma}|^2_{L^2(\mathbb{R})}.$$

Therefore,

$$\sum_{i,j=1}^d \int_0^t \alpha_{ij}(\hat{Y}(s))d\left\langle V_j^{W_1}, V_i^{W_1}\right\rangle(s) \le c|\hat{\sigma}|^2_{L^2(\mathbb{R})}L^{W_1}(t,0).$$

which is finite with probability 1. Now performing the second time change yields

$$\begin{aligned}
\bar{X}^0(t) &= W_1(s(t)) \\
\bar{Y}^0(t) &= \int_0^t \beta(\bar{Y}^0(u))L^{W_1}(s(du),0) + \int_0^t \alpha(\bar{Y}^0(u))V^{W_1}(s(du))
\end{aligned}$$

Note that for any Borel measurable $f : \mathbb{R} \to [0, \infty)$,

$$\int_{\mathbb{R}} f(x) L^{W_1}(s(t), x) dx = \int_0^{s(t)} f(W_1(u)) du = \int_0^t f(W_1(s(u))) a_{\pm}(W_1(s(u)), \hat{Y}(s(u))) du =$$
$$= \int_0^t f(\bar{X}^0(u)) a_{\pm}(\bar{X}^0(u), \bar{Y}^0(u)) du = \int_{\mathbb{R}} f(x) L^{\bar{X}^0}(t, x) dx,$$

which implies $L^{W_1}(s(t), x) = L^{\bar{X}^0}(t, x)$. This also implies that if we let $V^{\bar{X}^0}(t) = V^{W_1}(s(t))$, then we have

$$\langle V_i^{\bar{X}^0}, V_j^{\bar{X}^0} \rangle(t) = \langle V_i^{W_1}, V_j^{W_1} \rangle(s(t)) = \delta_{ij} L^{W_1}(s(t), x) = \delta_{ij} L^{\bar{X}^0}(t, x).$$

This means that

$$\bar{Y}^0(t) = \int_0^t \beta(\bar{Y}^0(u)) L^{\bar{X}^0}(du, 0) + \int_0^t \alpha(\bar{Y}^0(u)) dV^{\bar{X}^0}(u)$$

as required. We also have that the martingale

$$W(t) := \int_0^t \frac{1}{\sqrt{a_{\pm}(\bar{X}^0(u), \bar{Y}^0(u))}} W_1(s(du))$$

has quadratic variation t and therefore by Levy's theorem it is a Brownian motion. Consequently,

$$\bar{X}^0(t) = \int_0^t \sqrt{a_{\pm}(\bar{X}^0(u), \bar{Y}^0(u))} dW(u).$$

It remains to show that the appropriate cross-variations vanish. Since $\langle W_1, W_2 \rangle \equiv 0$, it is easy to see that $\langle W_1, V_i^{W_1} \rangle_t \equiv 0$. As $s'(t) \in [c_1, c_2]$, $\langle \bar{X}^0, V_i^{\bar{X}^0} \rangle_t = \langle W_1, V_i^{W_1} \rangle_{s(t)} \equiv 0$. Moreover,

$$\langle W, V_i^{\bar{X}^0} \rangle_t = \int_0^t \frac{1}{\sqrt{a_{\pm}(\bar{X}^0(u), \bar{Y}^0(u))}} d\langle \bar{X}^0, V_i^{\bar{X}^0} \rangle_u \equiv 0.$$

□

Proof of Theorem 2.6. As a Markov process, $Z(t)$ generates a strongly continuous contraction semigroup T_t on $C_0(\mathbb{R}^{1+d})$ by $T_t f(z) = \mathrm{E}_z f(Z(t))$. From this definition and Lemma 3.5, it is easy to see that the closure of $\bar{\mathcal{L}}$ (denoted by $\bar{\mathcal{L}}$ again) is an extension of the generator of T_t and hence, by a standard semigroup fact, they are actually equal. Thus, by the Hille Yosida theorem, $\mathcal{D}(\bar{\mathcal{L}})$ is dense in $C_0(\mathbb{R}^{1+d})$ and $\mathcal{R}(\lambda - \bar{\mathcal{L}}) = C_0(\mathbb{R}^{1+d})$. Therefore by Theorem 6.1, the associated martingale problem is well-posed. It is straightforward to show using Lemma 3.5, that $Z(t)$ is actually a solution of the martingale problem. This, combined with the tightness result of Section 4 and the convergence result of Section 5, finishes the proof of Theorem 2.6. □

6.2 Normalized deviation case

Similarly to the previous case, we are going to directly construct the limiting Markov process which solves the martingale problem associated to $\mathcal{L}$. We are only going to carry out the construction, the other steps are analogous.

Lemma 6.3. *There is a process X^0 which solves*

$$dX^0(t) = \sqrt{a_\pm(X^0(t), y(t))}dW(t).$$

This X^0 solves the martingale problem associated to $\bar{\mathcal{L}}$ for functions f that do not depend on y or ζ.

Proof. Such a result is not completely obvious because of the discontinuity of the coefficients. Consider instead the system

$$\begin{aligned} d\hat{X}(s) &= dW(s) \\ d\hat{y}(s) &= \frac{b_1(\hat{y}(s))}{a_\pm(\hat{X}(s), \hat{y}(s))}ds. \end{aligned}$$

The pair $(\hat{X}, \hat{y})$ is perfectly well-defined because the formula for $\hat{y}$ is a random ODE that we can consider pathwise in the integral sense. The function $b_1(y)/a_\pm(x, y)$ is discontinuous in x , but uniformly Lipschitz continuous in y. Classical successive approximation arguments show that a unique solution exists. We then define the random time change

$$t(s) = \int_0^s \frac{1}{a_\pm(\hat{X}(r), \hat{y}(r))}dr$$

and denote its functional inverse by $s(t)$. Then we define

$$\begin{aligned} X^0(t) &= \hat{X}(s(t)), \\ y(t) &= \hat{y}(s(t)) \end{aligned}$$

which has the properties that we desire. $\square$

The following two lemmas are a straightforward consequence of Lemma 3.5.

Lemma 6.4.

1. *Let $(X^\varepsilon, y, \zeta^\varepsilon)$ satisfy the assumptions of Theorem 2.2. Let*

$$\zeta^0(t) = \int_0^t e^{\int_s^t \partial_y b_1(y(r))dr} \sqrt{\left(\int_{-\infty}^{\infty} (\sigma\sigma^T)(x, y(s))dx\right)}dV^{X^0}(s). \tag{6.1}$$

where $V^{X^0} = W_0(L^{X^0}(t, 0))$. Then the triple (X^0, y, ζ^0) is a Markov process with generator $\bar{\mathcal{L}}$.

2. *Let $(X^\varepsilon, y, \zeta^\varepsilon)$ satisfy the assumptions of Theorem 2.4. Let*

$$\zeta^0(t) = \int_0^t e^{\int_s^t \partial_y b_1(y(r))dr} \sqrt{\left(\int_{-\infty}^{\infty} b(x, y(s))dx\right)} b(X^0(s), y(s)) L^{X^0}(ds, 0) \qquad (6.2)$$

where the above integral is the Riemann-Stieltjes integral with respect to the increasing function $t \mapsto L^{X^0}(t, 0)$. Then the triple (X^0, y, ζ^0) is a Markov process with generator $\bar{\mathcal{L}}$.

Proof of Theorem 2.2 and Theorem 2.4. Analogous to the proof of Theorem 2.6 in the previous subsection. □

7 Remarks and further directions

There are several interesting further questions one can ask in relation with (2.1)-(2.2).

- In some sense, the original result of [17] and the results in this paper carry out the first few steps of the program contained in ([10], Chapter 7) when the fast process is positively recurrent. In particular, a result similar to Theorem 2.2 was proved by Khasminskii ([15]) and one simlar to Theorem 5 was proven again by Borodin ([5]) sharpening a result of Khasminskii ([14]) that in turn was inspired by the non-rigorous result of Stratonovich ([28]). The next step in this program would be to study large deviation estimates of $(X^\varepsilon, Y^\varepsilon)$ from their respective limit in [17].

- Along the same lines, it is expected that one can generalize Theorem 2.6 to the case when there are several conserved quantities of that are conserved by the limiting motion in [17]. Namely, if H is such a quantity and (X^0, Y^0) is the limit of $(X^\varepsilon, Y^\varepsilon)$ in distribution then $H(Y^\varepsilon(t))$ will converge to the constant process process $H(y)$. uniformly on compact time intervals in probability. It is natural to conjecture that this process needs to be considered on time-scales of order ε^{-2} in order to see any non-trivial behavior. The corresponding result for positive recurrent averaging was carried out by Borodin and Freidlin ([6]). The result is expected to be a combination of external averaging due to the null recurrent fast process and internal averaging inside the over the level sets of H. In the case of Theorem 2.6, $H(y) = y$ and the level sets are single points, so this internal averaging does not take place.

- A different direction is to replace the fast motion with a more general null recurrent process or a process that is only neighborhood recurrent. For example, by considering a situation when the fast process is driven by a Bessel process of order $n \in (1, 2]$, one can hope to study the asymptotic behavior of a three dimensional diffusive stochastic dynamical system the dynamics of which is perturbed in a narrow tube.

- The results in this paper also imply certain results on partial differential equations through the well known representation formulas. For example, consider the Cauchy problem for $u^\varepsilon : \mathbb{R}^{1+d} \times \mathbb{R}_+ \to \mathbb{R}$

$$\begin{aligned}\partial_t u^\varepsilon = &\frac{1}{2}\left|\varphi\left(\varepsilon^{-1}x, y\right)\right|^2 \partial_{xx} u^\varepsilon + \left[b_1(y) + b_2\left(\varepsilon^{-1}x, y\right)\right] \partial_y u^\varepsilon + \\ &+ \frac{1}{2}\sum_{i,j=1}^d (\sigma\sigma^T)_{ij}\left(\varepsilon^{-1}x, y\right) \partial_{y_i y_j} u^\varepsilon + \sum_{i=1}^d \sum_{j=1}^k \sigma_{ij}(\varepsilon^{-1}x, y)\varphi_j(\varepsilon^{-1}x, y)\partial_{x y_i} u^\varepsilon\end{aligned}$$

with initial condition $u^\varepsilon(x,y,0) = f_\varepsilon(x,y)$. It is well known that

$$u^\varepsilon(x,y,t) = \mathrm{E}_{(x,y)} f_\varepsilon(X^\varepsilon(t), Y^\varepsilon(t))$$

is the unique solution of this Cauchy problem. In this setting, if $f_\varepsilon(x,y) = f(y)$ then we have by Lemma 3.2 that u^ε converges to $u^0(x,y,t) = f(y(t))$ uniformly on compact sets which solves the transport equation

$$\partial_t u^0 = b_1(y)\partial_y u^0, \qquad u^0(x,y,t) = f(y).$$

On the other hand if $b_1 \equiv 0$, $f_\varepsilon(x,y) = f(\varepsilon x, y)$, then Theorem 2.6 implies that $u^\varepsilon(x,y,t/\varepsilon^2)$ converges to $\hat{u}^0(x,y,t) = \mathrm{E}_{(x,y)} f(\bar{X}^0, \bar{Y}^0)$. Formally, this function solves

$$\partial_t \hat{u}^0 = \frac{1}{2} a_\pm(x,y)\partial_{xx}\hat{u}^0 + \delta(x)\left[\beta(y)\partial_y \hat{u}^0 + \frac{1}{2}\alpha(y)\partial_{yy}\hat{u}^0\right], \qquad \hat{u}^0(0,x,y) = f(x,y), \quad (7.1)$$

where $\delta(x)$ is the Dirac-delta distribution at x=0, even though the solution theory of such an equation is non-trivial. The case when there is only a generalized drift, i.e. $\alpha \equiv 0$, has been studied (see e.g. [25], but see also [3]) but to our knowledge, there have been no successful attempts to make sense directly of (7.1) in the general case. Another approach is to note that $Y(t)$ is a fractional diffusion. Indeed, in the case of e.g. $a_\pm \equiv 1$, and $f(x,y) \equiv f(y)$, we have (see [26] or the more general [1])

$$\partial_t^{1/2} \hat{u}^0 = \beta(y)\partial_y \hat{u}^0 + \frac{1}{2}\alpha(y)\partial_{yy}\hat{u}^0, \qquad \hat{u}^0(0,y) = f(y), \quad (7.2)$$

where ∂_y^β is the Caputo-derivative, i.e.

$$\partial_t^\beta f(t) = \frac{1}{\Gamma(1-\beta)} \int_0^\infty f'(t-r) r^{-\beta} dr$$

This latter non-local equation expresses the averaged (non-Markovian) dynamics of the Y-process (which is trivial when $a_\pm \equiv 1$ and it is expected that a similar equation can be written down for the general case. It would also be interesting to investiate the relation between (7.1) and (7.2).

Similarly, the results in this paper can be also used to study other initial-boundary problems through the usual representation formulas.

Appendix A Martingale problem - SDE equivalence in the general case

In this section we establish that there is a one to one correspondence between the solutions of the martingale problem associated to the operator $\bar{\mathcal{L}}$ and the weak solutions of the corresponding SDEs under very general circumstances. Even though this result is not needed for the specific results of this paper, we believe it is of independent interest and useful for future studies. For convenience, we prove the case $p_+ = p_- = 1/2$ case rigorously and then outline the necessary changes for the case when p_+ and p_- are constants but not necessarily equal.

Theorem A.1. *Let $\bar{\mathcal{L}}$ be the operator on $C_0(\mathbb{R}^{1+d})$ described in Section 2 with $p_+ = p_- = 1/2$. Then the distribution of a process $Z(t)$ solves the martingale problem associated to $\bar{\mathcal{L}}$ if and only if there is a Brownian motion W and a martingale V^{Z_1} such that $(Z(t), W, V^{Z_1})$ is a weak solution of (2.8)-(2.9).*

Proof. Using Lemma 3.5, it is straightforward to see that for every weak solution, the distribution of Z solves the martingale problem associated to $\bar{\mathcal{L}}$.

Now we prove the opposite direction. Assume that (X, Y) is a Martingale solution to the operator $\bar{\mathcal{L}}$ given by

$$\bar{\mathcal{L}}f(x, y) = \frac{1}{2}A(x, y)f_{xx}(x, y) + \sum_{k=1}^{d} B_k(x, y)f_{y_k}(x, y)$$

whose domain consists of sufficiently regular functions subjected to the boundary conditions

$$\frac{1}{2}f_x(0+, y) - \frac{1}{2}f_x(0-, y) + \sum_{k=1}^{d} \beta_k(y)f_{y_i}(0, y) + \frac{1}{2}\sum_{k,j=1}^{d} \alpha_{kj}(y)f_{y_k y_j}(y) = 0. \tag{A.1}$$

Let $\hat{\mathcal{D}}(\bar{\mathcal{L}})$ be defined the same as $\mathcal{D}(\bar{\mathcal{L}})$ but without requiring its members to decay at infinity. This means that it consists of those continuous functions twice continuously differentiable in x and y with the x derivative possibly being discontinuous at $x = 0$. Since

$$f(X(t), Y(t)) - f(X(0)) - \int_0^t \bar{\mathcal{L}}f(X(s), Y(s))ds,$$

is a continuous martingale for every $f \in \mathcal{D}(\bar{\mathcal{L}})$, it is not hard to show that it is a local martingale for all $f \in \hat{\mathcal{D}}(\mathcal{L})$.

First, we claim that the process $Y(t) - \int_0^t B(X(s), Y(s))ds$ does not move unless $X(t) = 0$. To see this, let $|x| > \delta > 0$ and $\tau = \inf\{t > 0 : |X(t)| \leq \delta\}$. Then by the martingale problem with a smooth function satisfying

$$f(x, y) = \begin{cases} y_k & |x| > \delta \\ 0 & |x| < \delta/2 \end{cases},$$

we have that

$$Y_k(\tau \wedge t) - \int_0^{\tau \wedge t} B_k(X(s), Y(s))ds$$

is a local martingale. Applying the same procedure with y_k^2 shows that this local martingale has quadratic variation 0 and therefore does not move. The boundary conditions are not relevant for this stopped process because we stop the process before it reaches the boundary. To study the motion at the boundary, we note that for any $k \in \{1, \dots, d\}$, the function

$$f(x, y) = y_k - |x|\beta_k(y)$$

satisfies the boundary conditions (A.1). Because (X, Y) solves the martingale problem,

$$Y_k(t) - |X(t)|\beta_k(Y(t)) - y_k - |x|\beta_k(y) - \int_0^t B_k(X(s), Y(s))ds- \tag{A.2}$$
$$- \int_0^t \sum_{j=1}^{d} |X(s)|\frac{\partial \beta_k}{\partial y_j}(Y(s))B_j(X(s), Y(s))ds$$

is a local martingale. Note that because $f(x,y) = x$ satisfies the boundary conditions, we know that $X(t)$ is also a local martingale. Therefore by the Meyer-Tanaka formula,

$$|X(t)| = |x| + L^X(t,0) + \text{ local martingale.}$$

Next, we observe that for any partition $0 = t_0 < t_1 < .. < t_{N+1} = t$,

$$|X(t)|\beta_k(Y(t)) - |x|\beta_k(y) = \\ = \sum_{i=0}^{N} \beta_k(Y(t_i))(|X(t_{i+1})| - |X(t_i)|) + \sum_{i=0}^{N} |X(t_{i+1})| \left(\beta_k(Y(t_{i+1})) - \beta_k(Y(t_i))\right).$$

As the partition gets finer,

$$\lim_{\|P\|\to 0} \sum_{i=0}^{N} \beta_k(Y(t_i))(|X(t_{i+1})| - |X(t_i)|) = \int_0^t \beta_k(Y(s))L^X(ds,0) + \text{ local martingale.}$$

in $L^2(\Omega)$. As for the second sum,

$$\sum_{i=0}^{N} |X(t_{i+1})| \left(\beta_k(Y(t_{i+1})) - \beta_k(Y(t_i))\right) \\ = \sum_{i=0}^{N} |X(t_{i+1})| \left(\beta_k\left(\tilde{Y}(t_{i+1}) + \int_0^{t_{i+1}} B(X(s),Y(s))ds\right) - \beta_k\left(\tilde{Y}(t_i) + \int_0^{t_i} B(X(s),Y(s))ds\right)\right)$$

where

$$\tilde{Y}(t) = Y(t) - \int_0^t B(Y(s))ds,$$

which does not move unless $X(t) = 0$ with probability one. Given any partition, P of $[0,t]$, we can refine the partition (randomly) in the following manner. For any $I = (t_i, t_{i+1}) \in P$,

1. If there exists $s \in I$ such that $X(s) = 0$ but $X(t_{i+1}) \neq 0$, we split the interval, I, at the random time
$$\tau = \sup\{s \in (t_i, t_{i+1}) : X(s) = 0\}$$

2. Otherwise, we do not refine I.

Call the randomly refined partition $\tilde{P}$. This partition has the property that if $[t_i, t_{i+1}] \in \tilde{P}$, then either $X(t_{i+1}) = 0$ or $\tilde{Y}(t_{i+1}) = \tilde{Y}(t_i)$. Switching to this random family of partitions,

$$\lim_{\|\tilde{P}\|\to 0} \sum_{k=0}^{N} |X(t_{i+1})| \left(\beta_k\left(\tilde{Y}(t_i+1) + \int_0^{t_{i+1}} B(Y(s))ds\right) - \beta_k\left(\tilde{Y}(t_i) + \int_0^{t_i} B(Y(s))ds\right)\right) \\ = \int_0^t \sum_{j=1}^{d} |X(s)| \frac{\partial \beta_k}{\partial y_j}(Y(s))B_j(Y(s))ds$$

in $L^2(\Omega)$. In this way, we have proven that

$$|X(t)|\beta_k(Y(t)) - |x|\beta_k(y) \overset{a.s}{=} \int_0^t \beta_k(Y(s))L^X(ds,0) + \int_0^t \sum_{j=1}^d |X(s)|\frac{\partial \beta_k}{\partial y_j}(Y(s))B_j(X(s),Y(s))ds \tag{A.3}$$

and by (A.2),

$$M(t) := Y(t) - y - \int_0^t B(X(s),Y(s))ds - \int_0^t \beta(Y(s))L^X(ds,0)$$

is a local martingale. By the Ito formula, for any $f \in C^2(\mathbb{R}^{d+1})$,

$$\begin{aligned} f(X(t),Y(t)) &= f(x,y) + \sum_{k=1}^d \int_0^t f_{y_k}(X(s),Y(s))B_k(X(s),Y(s))ds \\ &+ \sum_{k=1}^d \int_0^t f_{y_k}(X(s),Y(s))\beta_k(Y(s))L^X(ds,0) + \frac{1}{2}\int_0^t f_{xx}(X(s),Y(s))d\langle X\rangle_s \\ &+ \frac{1}{2}\sum_{i,j=1}^d \int_0^t f_{y_iy_j}(X(s),Y(s))d\langle M_i,M_j\rangle_s + \sum_{j=1}^d \int_0^t f_{xy}(X(s),Y(s))d\langle X,M_j\rangle_s \\ &+ \text{local martingale}. \end{aligned} \tag{A.4}$$

We compare this to the martingale problem. Let $f \in C^3(\mathbb{R}^{d+1})$ and define

$$g(x,y) = f(x,y) - \sum_{k=1}^d |x|\beta_k(y)f_{y_k}(0,y) - \frac{1}{2}\sum_{i,j=1}^d |x|\alpha_{ij}(y)f_{y_iy_j}(0,y)$$

This g satisfies the boundary conditions (A.1) and therefore

$$g(X(t),Y(t)) - \int_0^t \mathcal{L}g(X(s),Y(s))ds$$

is a martingale. That is,

$$\begin{aligned} f(X(t),Y(t)) &= f(x,y) + \sum_{k=1}^d \left(|X(t)|\beta_k(Y(t))f_{y_k}(0,Y(t)) - |x|\beta_k(y)f_{y_k}(0,y)\right) \\ &- \sum_{k,j=1}^d \int_0^t |X(s)|\frac{\partial}{\partial y_j}\left(\beta_k(Y(s))f_{y_k}(0,Y(s))\right)B_j(X(s),Y(s))ds \\ &+ \sum_{i,j=1}^d \left(|X(t)|\alpha_{ij}(Y(t))f_{y_iy_j}(0,Y(t)) - |x|\alpha_{ij}(y)f_{y_iy_j}(0,y)\right) \\ &- \sum_{i,j,k=1}^d \int_0^t |X(s)|\frac{\partial}{\partial y_k}\left(\alpha_{ij}(Y(s))f_{y_iy_j}(X(s),Y(s))\right)B_k(X(s),Y(s))ds \qquad \text{(A.5)} \\ &+ \sum_{k=1}^d \int_0^t f_{y_k}(X(s),Y(s))B_k(X(s),Y(s))ds + \frac{1}{2}\sum_{i,j=1}^d \int_0^t A(X(s),Y(s))f_{xx}(X(s),Y(s))ds \\ &+ \text{local martingale}. \end{aligned} \tag{A.6}$$

By (A.3), it follows that

$$
\begin{aligned}
&|X(t)|\beta_k(Y(t))f_{y_k}(0,Y(t)) - |x|\beta_k(y)f_{y_k}(0,y) \\
&- \sum_{j=1}^{d}\int_0^t |X(s)|\frac{\partial}{\partial y_j}\left(\beta_k(Y(s))f_{y_k}(0,Y(s))\right) B_j(X(s),Y(s))ds \\
&= \int_0^t \beta_k(Y(s))f_{y_k}(X(s),Y(s))L^X(ds,0)
\end{aligned}
$$

and

$$
\begin{aligned}
&|X(t)|\alpha_{ij}(Y(t))f_{y_iy_j}(0,Y(t)) - |x|\alpha_{ij}(y)f_{y_iy_j}(0,y) \\
&- \sum_{k=1}^{d}\int_0^t |X(s)|\frac{\partial}{\partial y_k}\left(\alpha_{ij}(Y(s))f_{y_iy_j}(X(s),Y(s))\right) B_k(X(s),Y(s))ds \\
&= \int_0^t \alpha_{ij}(X(s),Y(s))f_{y_iy_j}(X(s),Y(s))L^X(ds,0).
\end{aligned}
$$

Then by comparing (A.4) and (A.5) we can conclude that

$$
\begin{aligned}
\langle X, M_j\rangle_t &= 0, \\
\langle X\rangle_t &= \int_0^t A(X(s),Y(s))ds, \\
\langle M_i, M_j\rangle_t &= \int_0^t \alpha_{ij}(Y(s))L^X(ds,0).
\end{aligned}
$$

Finally, we define

$$
W(t) := \int_0^t A^{-1/2}(X(s),Y(s))dX(s), \qquad V^X(t) := \int_0^t \alpha^{-1/2}(X(s),Y(s))dM(s).
$$

By Levy's theorem, $W(t)$ is a one-dimensional Brownian-motion. Similarly, if

$$
L^{-1}(s) = \inf\{t > 0 : L^X(t,0) = s\},
$$

then by Levy's theorem

$$
W_0(s) := V^X(L^{-1}(s))
$$

is a d-dimensional Wiener process, and $V^X(t) = W_0(L^X(t,0))$. Finally, we can conclude that $(X(t),Y(t))$ is the weak solution of

$$
\begin{cases}
dX(t) = \sqrt{A(X(t),Y(t))}dW(t), \\
dY(t) = B(X(t),Y(t))dt + \beta(Y(t))L^X(dt,0) + \sqrt{\alpha(Y(t))}dV^X(t).
\end{cases}
\tag{A.7}
$$

□

Remark A.2. *In the case when $p_+(y)$ and $p_-(y)$ are not necessarily equal, one direction is still given by Lemma 3.5. The main difference in proving the other direction is that now $X(t)$ is not a local martingale. Rather, by applying the martingale problem to the function $f(x,y) = x - (p_+(y) - p_-(y))|x|$ (which requires sufficient regularity from $p_\pm$), we see that*

$$X(t) - (p_+(Y(t)) - p_-(Y(t))|X(t)| \tag{A.8}$$

is a local martingale. The difficulty lies in the fact that we do not a priori know that $X(t)$ is a semimartingale and therefore we cannot immediately apply the Tanaka-Meyer formula. However, let $\tau_0^\delta = 0$, $\sigma_0^\delta = \inf\{t \geq 0 | X(t) = 0\}$ and recursively define

$$\tau_i^\delta = \inf\{t > \sigma_{i-1}^\delta | |X(t)| = \delta\}, \qquad \sigma_i^\delta = \inf\{t > \tau_i^\delta | X(t) = 0\}$$

and we can write

$$X(t) - X(0) = \sum_{i=0}^{\infty} \left[X(t \wedge \sigma_i^\delta) - X(t \wedge \tau_i^\delta)\right] + \sum_{i=1}^{\infty} \left[X(t \wedge \tau_i^\delta) - X(t \wedge \sigma_{i-1}^\delta)\right].$$

It can be shown similarly as in Lemma 2.2 in [8] that as $\delta \downarrow 0$, the first sum converges to a local martingale while the second sum converges to a monotone process that only changes when $X(t) = 0$. More precisely, this second term converges to $\int_0^t (p_+(Y(s)) - p_-(Y(s)))dL^X(s)$. This shows that X is a semimartingale and (A.8) implies that $X(t) - \int_0^t (p_+(Y(s)) - p_-(Y(s)))dL^X(s)$ is a local martingale. The rest of the proof goes through with minor modifications.

References

[1] B. Baeumer and M.M. Meerschaert, *Stochastic solutions for fractional cauchy problems*, Fractional Calculus and Applied Analysis **4** (2001), no. 4, 481–500.

[2] R.F. Bass and E. Pardoux, *Uniqueness for diffusions with piecewise constant coefficients*, Probability Theory and Related Fields **76** (1987), no. 4, 557–572.

[3] S. Blei and H.-J. Engelbert, *One-dimensional stochastic differential equations with generalized and singular drift*, Stochastic Processes and their Applications **123** (2013), no. 12, 4337–4372.

[4] N.N. Bogolyubov, *On some statistical methods in mathematical physics*, Izdat. Akad. Nauk Ukr. SSR, Kiev (1945).

[5] A Borodin, *A limit theorem for solutions of differential equation with random right-hand side*, Theory Probab. Appl. **22** (1977), no. 3, 482–497.

[6] A.N. Borodin and M.I. Freidlin, *Fast oscillating random perturbations of dynamical systems with conservation laws*, Annales de l'IHP Probabilités et statistiques, vol. 31, 1995, pp. 485–525.

[7] S. N. Ethier and T. G. Kurtz, *Markov processes: characterization and convergence*, vol. 282, John Wiley & Sons, 2009.

[8] M. I. Freidlin and S.-J. Sheu, *Diffusion processes on graphs: stochastic differential equations, large deviation principle*, Probability theory and related fields **116** (2000), no. 2, 181–220.

[9] M. I. Freidlin and A. D. Wentzell, *Diffusion processes on graphs and the averaging principle*, The Annals of probability (1993), 2215–2245.

[10] ———, *Random perturbations of dynamical systems*, vol. 260, Springer Science & Business Media, 2012.

[11] I.I. Gikhman and A.V. Skorokhod, *Stochastic differential equations and their applications. naukova dumka, kiev*, 1982.

[12] M. Hairer and C. Manson, *Periodic homogenization with an interface: the multi-dimensional case*, The Annals of Probability **39** (2011), no. 2, 648–682.

[13] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*, vol. 113, Springer Science & Business Media, 2012.

[14] R. Z. Kasiminskii, *A limit theorem for the solutions of differential equations with random right-hand sides*, Theory Probab. Appl. **11**, no. 3, 390–406.

[15] R. Z. Khasminskii, *On stochastic processes defined by differential equations with a small parameter*, Theory Probab. Appl. **11** (1966), no. 2, 211–228.

[16] ———, *On averaging principle for ito stochastic differential equations*, Kybernetika, Czechoslovakia **4** (1968), no. 3, 260–279.

[17] R.Z. Khasminskii and N. N. Krylov, *On averaging principle for diffusion processes with null-recurrent fast component*, Stochastic Processes and their applications **93** (2001), no. 2, 229–240.

[18] R.Z. Khasminskii and G. Yin, *On averaging principles: an asymptotic expansion approach*, SIAM journal on mathematical analysis **35** (2004), no. 6, 1534–1560.

[19] Y. Kifer, *Some recent advances in averaging*, Modern dynamical systems and applications (2004), 385–403.

[20] E. M. LaBolle, J. Quastel, G. E. Fogg, and J. Gravner, *Diffusion processes in composite porous media and their numerical integration by random walks: Generalized stochastic differential equations with discontinuous coefficients*, Water Resources Research **36** (2000), no. 3, 651–662.

[21] A. Lejay, *On the constructions of the skew brownian motion*, Probab. Surv **3** (2006), 413–466.

[22] Zs. Pajor-Gyulai and M. A. Salins, *On dynamical systems perturbed by a null-recurrent fast motion: The continuous coefficient case with independent driving noises*, Journal of Theoretical Probability (2015), 1–17.

[23] G.C Papanicolaou, D. Stroock, and S.R.S. Varadhan, *Martingale approach to some limit theorems*, Duke Turbulence Conference (Duke Univ., Durham, NC, 1976), Paper, vol. 6, 1977.

[24] G. Peskir, *A change-of-variable formula with local time on surfaces*, Séminaire de probabilités XL, Springer, 2007, pp. 70–96.

[25] N.I. Portenko, *Diffusion processes with generalized drift coefficients*, Theory of Probability & Its Applications **24** (1979), no. 1, 62–78.

[26] A.I. Saichev and G.M. Zaslavsky, *Fractional kinetic equations: Solutions and applications*, Chaos **7** (1997), 753–764.

[27] A. V. Skorokhod, *Asymptotic methods in the theory of stochastic differential equations*, vol. 78, American Mathematical Soc., 2009.

[28] R. L. Stratonovich, *Conditional markov processes and their application to the theory of optimal control*, Modern Analytic and Computational Methods in Science and Mathematics **7** (1968).

[29] T. Ueno, *A survey on the markov process on the boundary of multidimensional diffusion*, Proc. 5th Berkeley Symp. Math. Statist. and Prob. II, Part, vol. 2, 1967, pp. 111–130.

[30] A.D. Venttsel', *On boundary conditions for multidimensional diffusion processes*, Theory of Probability & Its Applications **4** (1959), no. 2, 164–177.

[31] ———, *On lateral conditions for multidimensional diffusion processes*, Teor. Verojatnost. i Primenen **4** (1959), 172–185.