On stochastic differential equations with piecewise smooth drift and noise coefficients.

Contents

1 Introduction

1.1 Physical Motivation

We consider stochastic differential equations arising from the large-N limit of chemical master equations, which themselves emerge from coarse-graining of Potts models. The resulting SDEs have two key features:

1. **Piecewise-smooth drift**: The deterministic dynamics has a discontinuity across a codimension-one switching manifold
2. **Multiplicative noise**: The noise coefficient depends on the state inherited from the underlying CME

Our goal is to compute the Gaussian envelope of fluctuations around deterministic orbits, particularly near and on the discontinuity surface.

1.2 Literature and Gap

The Freidlin-Wentzell theory of large deviations [1] provides the foundational framework for weak-noise asymptotics of SDEs with smooth coefficients. Extensions to piecewise-smooth systems include:

- **Chiang-Sheu** [2], [3]: Large deviations for discontinuous drift with additive noise, using occupation times and local times

- **Chen-Baule-Touchette-Just** [4], Mostly asmptotics, and the system is very simple it is piecewise constant with additive noise.
  - Useful to chaeck if my results match theirs in the limit.
- **Hill-Zanetell-Gemmer** [5]: Most probable paths via with mollification (smearing the shit out of), additive noise only

The gap: No existing work treats multiplicative noise with discontinuous diffusion coefficient and derives the Gaussian fluctuations near the switching manifold. None have ever attempted to resolve the case when we have hidden dynamics.

- this requries one to obtain estimates for higher moments of λ not just the mean

1.3 Main Contributions

1. Derivation of the switching variable dynamics via Meyer-Itō
2. Rigourous derivation of the fast Fokker-Planck equation with reflecting boundaries, via intermediate timescale
3. Averaging principle for the slow dynamics
4. Explicit formula for the Gaussian envelope including contributions from switching variable fluctuations

1.4 Literature Gap

The Freidlin-Wentzell theory of large deviations [1] deals with weak noise SDEs with smooth drift and noise coefficient.

2 Background

Definition 2.1 (Piecewise-smooth ODE).

Let $\sigma : \mathbb{R}^d \mapsto \mathbb{R}$ be a smooth function, $\epsilon > 0$, and $x \in \mathbb{R}^d$ be a deterministic processes satisfying the ODE

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \begin{cases} a^+(x) & \sigma(x) > 0 \\ a^-(x) & \sigma(x) < 0 \end{cases} \tag{1}$$

where $a^\pm : [0, \infty) \times \mathbb{R}^d \mapsto \mathbb{R}^d$ are smooth vector fields.

The switching paramter

$$\lambda = \begin{cases} 1 & \sigma(x) > 0 \\ -1 & \sigma(x) < 0 \end{cases} \tag{2}$$

$$a(x, \lambda) = \frac{1}{2}(1 + \lambda)a_+(x) + \frac{1}{2}(1 - \lambda)a_-(x) + (1 - \lambda^2)h(x) \tag{3}$$

3 Piecewise-Smooth Stochastic Differential Equations

We are interested in treating stochastic systems with weak noise whose behaviour switches on either side of a discontinuity set. However, away from discontinuity set we have sufficient smoothness in both the drift field and the noise amplitude. The following definition formalises this setup.

Definition 3.1 (Weak-Noise Piecewise-Smooth SDE).

Let $T > 0$, $t \in [0, T]$, $\epsilon > 0$, $\alpha \in [0, 1]$, $\sigma : \mathbb{R}^d \mapsto \mathbb{R}$ be a smooth function, and $x_t \in \mathbb{R}^d$ be a stochastic processes satisfying the SDE

$$\mathrm{d}x_t = a(x_t)\,\mathrm{d}t + \sqrt{\epsilon} b(x_t) \overset{\alpha}{*} \mathrm{d}W_t, \tag{4}$$

where

$$a(x) \overset{\text{def}}{=} \begin{cases} a^+(x) & \sigma(x) > 0, \\ a^-(x) & \sigma(x) < 0, \end{cases} \quad b(x) \overset{\text{def}}{=} \begin{cases} b^+(x) & \sigma(x) > 0, \\ b^-(x) & \sigma(x) < 0, \end{cases} \tag{5}$$

are piecewise smooth drift and diffusion coefficients respectively satisfying the following conditions:

1. *(A1 - Smoothnes) The constituent coefficients are sufficiently smooth* $a^\pm \in C^2(\mathbb{R}^d; \mathbb{R}^d)$ *and* $b^\pm \in C^2(\mathbb{R}^d; \mathbb{R}^{d\times m})$.
2. *(A2 - Linear Growth) We have*

$$\|a^\pm(x)\| + \|b^\pm(x)\| \leq C^\pm(1 + |x|) \tag{6}$$

for some $C > 0$, *where* $\|a^\pm(\cdot,\cdot)\|$ *is the Euclidean norm and*

$$\|b^\pm(\cdot,\cdot)\| = \sqrt{\sum_{ij} |b_{ij}(\cdot,\cdot)|^2}. \tag{7}$$

3. *(A3 - Lipshitz Continuity)*

$$\|a^\pm(x) - a^\pm(y)\| + \|b^\pm(x) - b^\pm(y)\| \leq K^\pm \; |x - y|, \tag{8}$$

for some $K^\pm > 0$.

4. *(A4 - Transversality)*

$$\|\partial_x \sigma(x)^T b^\pm(x)\| \geq M^\pm > 0. \tag{9}$$

The α *is used to control evaluation point of the stochastic integral.*

The conditions (A1-3) ensure that away from the discontinuity set, that is $\mathcal{D} = \{x \in \mathbb{R}^d \mid \sigma(x) = 0\}$, we have the existance and uniqueness of solutions. In other words away from $\mathcal{D}$, one can employ standard methods of SDE theory to analyse the dynamics, while near the discontinuity one can The stochastic integral in Eq. (4) is understood in the α−sense, i.e. with evaluation point $(1 - \alpha)x_t + \alpha x_{t+\Delta t}$. While the recasting of typical α-SDE into an Itō form is straight forward, see for example, [6] or [7], one cannot naively follow the procedure here as the noise coefficient does not have a continuous derivative.

Instead, we must fisrt employ Filippov's convex construction [8] for the drift and noise coefficient, with $\lambda \in [-1, 1]$ we define the convex combinations

$$a(x,\lambda) \overset{\text{def}}{=} \frac{1}{2}(1+\lambda)a^+(x) + \frac{1}{2}(1-\lambda)a^-(x), \tag{10}$$

and

$$b(x,\lambda) \overset{\text{def}}{=} \frac{1}{2}(1+\lambda)b^+(x) + \frac{1}{2}(1-\lambda)b^-(x), \tag{11}$$

which are smooth in λ, as well as x and t as they inherit the smoothness conditions given in Definition 3.1. The switching variable obviously depends on the state and we will regularise the definition given in Eq. (2) as

$$\lambda = \Lambda_\epsilon[\sigma(x)] \tag{12}$$

where

$$\Lambda_\epsilon(u) \stackrel{\text{def}}{=} \begin{cases} u/\epsilon & \sigma(x) \le \epsilon \\ \text{sign}[\sigma(x)] & \sigma(x) > \epsilon \end{cases} \tag{13}$$

is an auxiliary function used to control the regularisation. Notice that the regularisation implicitly defines the layer

$$\mathcal{D}_\epsilon \stackrel{\text{def}}{=} \{x \in \mathbb{R}^d \mid |\sigma(x)| \le \epsilon\}, \tag{14}$$

and which affords a precise meaning to the term dynamics near the discontinuity, i.e. when $x_t \in \mathcal{D}$. These definitions allow as to recast Eq. (4) into the Itō analgoue

$$\mathrm{d}x_t = [a(x_t, \lambda_t) + \alpha\epsilon c(x_t, \lambda_t)]\,\mathrm{d}t + \sqrt{\epsilon} b(x_t, \lambda_t)\,\mathrm{d}W_t, \tag{15}$$

where the correction term is

$$c(x, \lambda) = \sum_j J_x\left[b_j(x, \lambda)\right] b_j(x, \lambda). \tag{16}$$

with $b_j(x, \lambda)$ denoting the j^{th} column of the matrix $b(x, \lambda)$ and $J_x(\cdot)$ is the Jacobian matrix of the vector argument with respect to x. Obiously λ is itself a stochastic variable since it dependends on x_t via $\lambda_t = \Lambda_\epsilon[\sigma(x_t)]$, and, like its deterministic counterpart is dynamic on the the much faster timescale $\mathcal{O}(1/\epsilon)$. However, one cannot simply employ Itō's Lemma on $\lambda_t = \Lambda_\epsilon[\sigma(x_t)]$ as the latter is not a smooth function of x_t. Instead to study the dynamics of λ_t we must first introduce two new concepts: local time of a semi-martingale and the Meyer-Itō Theorem.

Local time of a semi-martingale x_t, denoted with $L_t^x(z)$is a measure of the "visits" of the proccess on a given value z for times up to t. It is given via Tanaka's formula which we summarise in the folowing defintion.

Definition 3.2 (Local time of a semi-martingale: Tanaka's Formula).

Let x_t be a semi-martingale in $\mathbb{R}^d$, and let $L_{t(z)}^x$ be the local time of the process at level

$$|x_t - z| = |x_0 - z| + \int_0^t \text{sign}(x_t - a)\,\mathrm{d}x_s + L_t^x(z) \tag{17}$$

where

$$\text{sign}(x) = \begin{cases} 1 & x > 0, \\ -1 & x \le 0. \end{cases} \tag{18}$$

For derivation and discussion see See Chap. 3 of [9], and Chap. IV of [10]. Secondly we require the Meyer-Itō theorem, also called the genearlised Itō's formula. We restate it here without the proof which can be found in Theorem 70, Chapter IV of [10].

Theorem 3.3 (Meyer-Itō).

Let $f : \mathbb{R}^d \mapsto \mathbb{R}$ be the difference of two convex functions, f'^{-} denote its left derivative, $\mu_{f''}$ be signed measure of the second derivative of f in the generalised function (distribution) sense, and let x_t be a semi-martingale in $\mathbb{R}^d$ then evolution of $f(x_t)$ is given by

$$f(x_t) = f(x_0) + \int_0^t f'^{-}(x_t)\,\mathrm{d}x_t + \frac{1}{2}\int_{\mathbb{R}^d} L_t^x(z)\,\mathrm{d}\mu_{f''}(z) \tag{19}$$

where $L_t^x(z)$ is the local time of x_t at z and the final integral in Eq. (19) is a Lebesque-Stieltjes integral.

For $f \in C^2[\mathbb{R}^D, \mathbb{R}]$, Theorem 3.3 redues to Itō's Lemma, TODO

- add definition for

$$L_t^x(z) = \lim_{\delta \downarrow 0} \frac{1}{2\delta} \int_{\mathbb{R}} \mathbb{1}_{(z-\delta, z+\delta)}(z_s) \, \mathrm{d}\langle x_t \rangle \tag{20}$$

- convex functions of martingales are themselves martingales
- generalises itos theorem
  - add example about how it redues to the standard itos lemma when f has a second derivative

In order to apply Theorem 3.3, we must first write the our regulariser function as a difference of convex function which is easily done as the following lemma shows.

Lemma 3.4 (The regulariser $\Lambda_\epsilon(u)$ is DC).

The function $\Lambda_\epsilon(u)$ as defined in Eq. (13), can written as

$$\Lambda_\epsilon(u) = \varphi_\epsilon^+(u) - \varphi_\epsilon^-(u) \tag{21}$$

where $\varphi^\pm(u)$ are convex functions.

Proof.

Let

$$\psi_{+,\epsilon}(u) \stackrel{\text{def}}{=} \begin{cases} -1 & u \leq -\epsilon, \\ u/\epsilon & u > -\epsilon, \end{cases} \qquad \psi_{-,\epsilon}(u) \stackrel{\text{def}}{=} \begin{cases} 0 & u \leq \epsilon, \\ u/\epsilon - 1 & u > \epsilon \end{cases}, \tag{22}$$

which are convex in u. □

Notice that the functions $\psi^\pm$ chosen are left continuous, i.e. $\lim_{u \downarrow a} \psi_{\pm,\epsilon} = \psi_{\pm,\epsilon}(a)$, for all $a \in \mathbb{R}$. This is intentional as we must require the left derivative of $\Lambda_{\epsilon(u)}$ given by

$$\Lambda_{\epsilon(u)}'^{-} = \psi_{+,\epsilon}'^{-}(u) - \psi_{-,\epsilon}'^{-}(u) = \begin{cases} 0 & u \leq -\epsilon, \\ 1/\epsilon & \epsilon < u \leq \epsilon, \\ 0 & u > \epsilon. \end{cases} \tag{23}$$

Similarly we have second derivative as a signed measure

$$\mu_{\Lambda_\epsilon''}(u) = \frac{1}{\epsilon}\delta(u+\epsilon) - \frac{1}{\epsilon}\delta(u-\epsilon), \tag{24}$$

where $\delta(u)$ is the Dirac-delta distribution. In order to apply Meyer-Itō we also need the dynamics of scalar observable $\sigma(x_t)$ which we state in the following lemma.

Lemma 3.5 (SDE for $z_t = \sigma(x_t)$).

Let $\lambda \in [-1, 1]$, $\sigma \in C^2(\mathbb{R}^d, \mathbb{R})$, x_t be an Itō process according to Eq. (15), supplemented by the conditions in Definition 3.1, then the random variable $z_t = \sigma(x_t)$ evolves according to the SDE

$$\mathrm{d}z_t = \tilde{a}(x_t, \lambda_t)\, \mathrm{d}t + \sqrt{\epsilon}\tilde{b}(x_t, \lambda_t)\, \mathrm{d}W_t, \tag{25}$$

where

$$\tilde{a}(x, \lambda) \stackrel{\text{def}}{=} \partial_x \sigma(x)^T [a(x, \lambda) + \alpha\epsilon c(x, \lambda)] + \frac{\epsilon}{2} \operatorname{Tr}[b(x, \lambda) \partial_{xx}^2 \sigma(x) b(x, \lambda)], \tag{26}$$

and

$$\tilde{b}(x,\lambda) \stackrel{\text{def}}{=} \partial_x \sigma(x)^T b(x,\lambda)], \tag{27}$$

Proof.

This is trivial application of Itō's lemma. Since $\sigma(x_t)$ is smooth, apply Itō's lemma to obtain

$$\mathrm{d}z_t = \partial_x \sigma(x_t)^\top \mathrm{d}x_t + \mathrm{d}x_t^\top \partial^2_{xx}\sigma(x)\,\mathrm{d}x^\top, \tag{28}$$

then substritute for $\mathrm{d}x_t$ from Eq. (15) into Eq. (28) and apply Itō's product rule.

□

We are now in a position to consider the dynamics of the switching variable as an SDE.

Theorem 3.6 (SDE for the switching variable λ).

Let $\epsilon > 0$, $\sigma \in C^2[\mathbb{R}^d, \mathbb{R}]$ such that $\mathcal{D}_\epsilon = \{x \in \mathbb{R}^d \mid \sigma(x) \le \epsilon\}$ is close set, let $x_t \in \mathcal{D}_\epsilon$ evolve according to according to Eq. (15), and let $\Lambda_\epsilon(u)$ be a be a family of regularisers of the sign function as dfined in Eq. (13). then the switching variable $\lambda_t = \Lambda_\epsilon[\sigma(x_t)]$ evolves in the interval interval [−1, 1] according to the SDE

$$\mathrm{d}\lambda_t = \frac{1}{\epsilon}\mathbb{1}_{(-\epsilon,\epsilon]}[\sigma(x_t)]\tilde{a}(x_t,\lambda_t)\,\mathrm{d}t + \frac{1}{\sqrt{\epsilon}}\mathbb{1}_{(-\epsilon,\epsilon]}[\sigma(x_t)]\tilde{b}(x_t,\lambda_t)\,\mathrm{d}W_t \tag{29}$$

$$+\frac{1}{\epsilon}[\,\mathrm{d}L^z_t(-\epsilon) - \mathrm{d}L^z_t(\epsilon)], \tag{30}$$

where $\tilde{a}(x,\lambda)$ and $\tilde{b}(x,\lambda)$ are defined in Eq. (26) and Eq. (27) respectively, $\mathrm{d}L^z_t(\pm\epsilon)$ is the change in the local time of z_t at $z = \pm\epsilon$ where the evolution of z_t is given by Eq. (25).

Proof.

Since $\Lambda_\epsilon(u)$ is a difference of convex functions, whose left derivative is given in Eq. (23) and signed second derivative given as a measure given in Eq. (24), it then follows from Theorem 3.3, that for a generic random variable z_t we have

$$\Lambda_\epsilon(z_t) = \Lambda_\epsilon(z_0) + \frac{1}{\epsilon}\int_0^t \mathbb{1}_{(-\epsilon,\epsilon]}(z_t)\,\mathrm{d}z_t + \frac{\epsilon}{2}[L^z_t(-\epsilon) - L^z_t(\epsilon)]. \tag{31}$$

By letting $\lambda_t = \Lambda_\epsilon(z_t = \sigma(x_t))$, and using Lemma 3.5 we obtain

$$\lambda_t = \lambda_0 + \int_0^t \frac{1}{\epsilon}\mathbb{1}_{(-\epsilon,\epsilon]}[\sigma(x_s)]\tilde{a}(x_s,\lambda_s)\,\mathrm{d}t + \frac{1}{\sqrt{\epsilon}}\int_0^t \mathbb{1}_{(-\epsilon,\epsilon]}[\sigma(x)]\tilde{b}(x_s,\lambda_s)\,\mathrm{d}W_s \tag{32}$$

$$+\frac{1}{\epsilon}[L^z_t(-\epsilon) - L^z_t(\epsilon)]. \tag{33}$$

□

The dynamics of the full system are then represented by the coupled SDE

$$\mathrm{d}x_t = [a(x_t,\lambda_t) + \alpha\epsilon b(x_t,\lambda_t)]\,\mathrm{d}t + \sqrt{\epsilon} b(x_t,\lambda_t)\,\mathrm{d}W_t, \tag{34}$$

$$\mathrm{d}\lambda_t = \frac{1}{\epsilon}\mathbb{1}_{(-\epsilon,\epsilon]}[\sigma(x_t)]\tilde{a}(x_t,\lambda_t)\,\mathrm{d}t + \frac{1}{\sqrt{\epsilon}}\mathbb{1}_{(-\epsilon,\epsilon]}[\sigma(x_t)]\tilde{b}(x_t,\lambda_t)\,\mathrm{d}W_t \tag{35}$$

$$+\frac{1}{\epsilon}[\,\mathrm{d}L^z_t(-\epsilon) - \mathrm{d}L^z_t(\epsilon)], \tag{36}$$

where $a(x,\lambda)$ and $b(x,\lambda)$ are defined, respectively, in Eq. (10) and Eq. (11), while $\tilde{a}(x,\lambda)$ and $\tilde{b}(x,\lambda)$ are given defined in Eq. (26) and Eq. (27) respectively. The coupled system is a slow-fast stochastic system, and our goal is to obtain controlled approximation for the dynamics of the slow process by closing the dynamics of the switching variable λ_t.

4 The intermediate timescale

4.1 Necessity of an intermediate timescale.

From Eq. (29), it is evident that the switching variable λ_t evolves on a faster timescale compared to the state variable x_t when near the discontinuity set. As we have discussed in Section 2, in the deterministic setting, the standard approach is to rescale time via $t = \epsilon\tau$, take the limit $\epsilon \to 0$, and solve the resulting algebraic condition to obtain a $\lambda^* \in (-1, 1)$ which gives us our sliding mode. It is tempting to follow the same procedure here, where instead of a single value for the switching variable, we obtain the steady-state distribution $\lim_{t\to\infty} P_\infty(\lambda, t, \mid x)$. However we shall see the stochastic nature of the problem yeilds multiple objections concerning the mathematical subtleties in the scaling, the physical interpretation, and the analysis of the original problem given in Definition 3.1, that must be addressed individually.

4.1.1 Objection I: Incompatible scaling of the dynamics.

Before we attempt to rescale time must first clarify the ϵ-order of local time terms in Eq. (29) which we do in the following lemma.

Lemma 4.1.1 (Scaling of local time terms).

Let $x_t \in \mathbb{R}^d$ be a stochastic process given the SDE

$$\mathrm{d}x_t = a(x_t)\,\mathrm{d}t + b(x_t)\,\mathrm{d}W_t, \tag{37}$$

$$\mathbb{E}[L_t^x(a)] \leq Ct \tag{38}$$

TODO

Finish the lemma!!

Taking Eq. (20) and taking the expectation of both sides we obtain

$$\begin{aligned}
\mathbb{E}[L_t^z(a)] &= \mathbb{E}\left[\lim_{\delta\downarrow 0}\frac{1}{2\delta}\int_0^t \mathbb{1}_{(a-\delta,a+\delta)}(z_s)\,\mathrm{d}\langle z\rangle_s\right], && (39)\\
&= \lim_{\delta\downarrow 0}\frac{1}{2\delta}\int_0^t \mathbb{E}\left[\mathbb{1}_{(a-\delta,a+\delta)}\,\mathrm{d}\langle z\rangle_s\right] && (40)\\
&= \lim_{\delta\downarrow 0}\frac{1}{2\delta}\int_0^t \epsilon\mathbb{E}\left[\mathbb{1}_{(a-\delta,a+\delta)}\tilde{b}^2(x_s,\lambda_s)\right]\mathrm{d}s, && (41)\\
&= \epsilon\int_0^t \lim_{\delta\downarrow 0}\frac{1}{2\delta}\int_{a-\delta}^{a+\delta} P^{(z)}(a,s)\tilde{b}^2(x_s,\lambda_s)\,\mathrm{d}s, && (42)\\
&= \epsilon\int_0^t P^{(z)}(a,s)\tilde{b}^2(x_s,\lambda_s)\,\mathrm{d}s, && (43)
\end{aligned}$$

or equivalently in differential form

$$\mathrm{d}\mathbb{E}[L_t^z(a)] = \epsilon P^{(z)}(a,t)\tilde{b}^2(x_t,\lambda_t)\,\mathrm{d}t. \tag{44}$$

From Eq. (44) we conclude that $\mathrm{d}L^z_{t(a)} \sim \epsilon P^{(z)}(a,t)\tilde{b}^2(t,x_t,\lambda_t)\,\mathrm{d}t$. Now let us consider the rescaling $t = \epsilon^\beta \tau$ for some $\beta > 0$, then we have $\mathrm{d}t = \epsilon^\beta\,\mathrm{d}\tau$ and $\mathrm{d}W_t = \epsilon^{\beta/2}\,\mathrm{d}W_\tau$. Applying this to Eq. (29) obtain

$$O\big(\epsilon^{\beta-1}\big) : \mathbb{1}_{(-\epsilon,\epsilon]}[\sigma(x_\tau)]\tilde{a}(x_\tau,\lambda_\tau)\,\mathrm{d}\tau, \tag{45}$$

$$O\Big(\epsilon^{\frac{\beta-1}{2}}\Big) : \mathbb{1}_{(-\epsilon,\epsilon]}[\sigma(x_\tau)]\tilde{b}(x_\tau,\lambda_\tau)\,\mathrm{d}W_\tau, \tag{46}$$

$$O\big(\epsilon^{\beta}\big) : \mathrm{d}L^z_\tau(a). \tag{47}$$

The naive rescaling from the determinisitic dynamics would be to have $\beta = 1$ which would put drift and martingale terms on $\mathcal{O}(1)$, but the local time contribution would be $\mathcal{O}(\epsilon)$ and vanishes in the limit. Clearly there is no β that would bring all of these terms together on equal fotting such that under $\epsilon \to 0$ all of the features of the dynamics are maintained: balancing drift and noise suppresses the local time; preserving the local time would require a slower timescale on which drift or noise diverges. This presents a fundemental technical hurdle. We have three contributions to the dynamics of λ that operate on on incompatible scales. Any rescaling followed by $\epsilon \to 0$ necessarily discards at least one of these contributions.

4.1.2 Objection II: Loss of physical interpretation. Not a good reason

Even granting mathematical well-posedness, the $\epsilon \to 0$ limit produces an object whose physical meaning has degenerated. As $\epsilon \to 0$:

1. The layer $\mathcal{D}_\epsilon = \{x \in \mathbb{R}^d : |\sigma(x)| \le \epsilon\}$ shrinks to the codimension-1 surface $\mathcal{D} = \{x : \sigma(x) = 0\}$.
2. The switching variable $\lambda \in [-1,1]$ parametrises a convex interpolation between the vector fields $a^\pm$ and noise coefficients $b^\pm$. This interpolation only has meaning within the layer, where the dynamics transitions between the two regimes.
3. The stationary distribution $P^\epsilon_\infty(\lambda \mid x)$ converges to some limiting distribution on $[-1,1]$, but this limit lives on a domain whose connection to the original geometry has been los.

In the deterministic case, the $\epsilon \to 0$ limit yields a single value $\lambda^*(x)$ i.e. the Filippov sliding mode (see Section 2). In that case, interpretation is clear, λ^* selects the unique convex combination that keeps trajectories on the discontinuity surface.

4.1.2.1 Objection III: Incompatibility with weak-noise analysis.

The most fundamental objection concerns the purpose of the analysis. The weak-noise framework treats ϵ as the small parameter governing the asymptotic expansion. The objects of interest are beyon the typcal paths, obtained by minimising the Freidlin-Wentzell action functional or analgously the deterministic limit $\epsilon \to 0$, instead we are interested in charertising the Gaussian fluctuations around the most probable path, arising at order $\mathcal{O}(\sqrt{\epsilon})$. These phenomena are intrinsically ϵ-dependent and the noise is the object of study. The stationary distribution $P_\infty(\lambda \mid x)$ at finite ϵ encodes how noise selects among the continuum of Filippov solutions, how it smooths the transition across $\mathcal{D}$, and what the fluctuation structure looks like near the discontinuity.

Taking $\epsilon \to 0$ in the fast dynamics collapses this structure to a deterministic Filippov vector field. Thus it eliminates the noise-induced selection mechanism among sliding vectors returns us to the deterministic theory discardign the stochastic phenomena and precluding the study of fluctaions which we set out to analyse in the first place.

The weak-noise SDE is already the result of retaining terms to second order in the noise coefficient. Consistency demands that ϵ be preserved throughout the analysis, including in the treatment of the fast variable.

4.1.3 The intermediate timescale resolution.

The intermediate timescale δ satisfying

$$\epsilon \ll \delta \ll 1 \tag{48}$$

resolves all three by avoiding the $\epsilon \to 0$ limit in the layer dynamics entirely. At fixed $\epsilon > 0$:

- **Well-defined quantities**: The layer $\mathcal{D}_\epsilon$ has finite width, the boundaries $\pm\,\epsilon$ are well-separated, and all probabilistic quantities — including local times and their contributions to the stationary distribution — are well-defined without singular limits.
- **Preserved interpretation**: The switching variable λ retains its meaning as parametrising the interpolation within a layer of finite width. The stationary distribution $P_\infty(\lambda \mid x)$ describes the equilibrium of λ within this layer, with clear geometric content.
- **Retained noise structure**: The parameter ϵ appears throughout the effective slow dynamics, preserving the weak-noise structure required for Freidlin-Wentzell analysis.

The conditions on δ ensure:

- $\delta \gg \epsilon$: The fast variable λ equilibrates to $P_\infty(\lambda \mid x)$ within the δ-window (see Theorem 6.7).
- $\delta \ll 1$: The slow variable x remains approximately frozen over the δ-window (see Lemma 5.1).

The effective slow dynamics follows by averaging against the stationary distribution:

$$\bar{a}(x) = \int_{-1}^{1} a(x,\lambda) P_\infty(\lambda \mid x)\,\mathrm{d}\lambda, \quad \bar{b}(x) = \int_{-1}^{1} b(x,\lambda) P_\infty(\lambda \mid x)\,\mathrm{d}\lambda. \tag{49}$$

The resulting equation,

$$\mathrm{d}x_t = \bar{a}(x_t)\,\mathrm{d}t + \sqrt{\epsilon}\bar{b}(x_t)\,\mathrm{d}W_t, \tag{50}$$

remains a weak-noise SDE with ϵ-dependent coefficients, to which standard Freidlin-Wentzell theory applies. This is the stochastic analogue of Filippov's construction: where the deterministic theory yields a unique sliding vector field via an algebraic condition, the stochastic theory yields an averaged vector field weighted by the stationary distribution of the fast variable.

Remark

The choice $\delta = \epsilon^\alpha$ for some $\alpha \in (0,1)$ provides a concrete realisation of the ordering Eq. (48). The value of α does not affect the limiting averaged dynamics, provided the bounds in Lemma 5.1 and Theorem 6.7 hold uniformly.

5 Estimates for the dynamics on the intermediate time-scale

We want to study the stochastic dynamics for a small but non zero ϵ, so we will introduce an intermediate timescale δ suchh that

$$\epsilon \ll \delta \ll 1. \tag{51}$$

On this time scale, the dynamics of x_t is frozen but the dynamics of λ_t has equilibriated on to a steady state distribution. Typically one defines δ as a function of ϵ, e.g. $\delta(\epsilon) = \epsilon^\alpha$, for some $\alpha \in (0,1)$.

Lemma 5.1 (Slow variation of x_t in the δ-window).

Let $x_t \in \mathcal{D}_\epsilon$ and $\delta > 0$ satisfying Eq. (51), then

$$\mathbb{P}\left[\sup_{0\le s\le\delta}|x_{t+s}-x_t|>\gamma\right]\le\frac{C}{\gamma^2}(\delta^2+\epsilon\delta) \tag{52}$$

for some $C,\gamma>0$.

Proof of Lemma 5.1.

We start by bounding the squared deviation in the δ time window,

$$\mathbb{E}[|x_{t+s}-x_t|^2]=\mathbb{E}\left[\left|\int_t^{t+s}a(x_\tau,\lambda_\tau)\,\mathrm{d}\tau+\sqrt{\epsilon}\int_t^{t+s}b(x_\tau,\lambda_\tau)\,\mathrm{d}W_\tau\right|^2\right], \tag{53}$$

$$\le 2\mathbb{E}\left[\left|\int_t^{t+s}a(x_\tau,\lambda_\tau)\,\mathrm{d}\tau\right|^2\right]+2\epsilon\mathbb{E}\left[\left|\int_t^{t+s}b(x_\tau,\lambda_s)\,\mathrm{d}W_\tau\right|^2\right]. \tag{54}$$

We will bound each integral term separately, for the drift part we have

$$\mathbb{E}\left[\left|\int_t^{t+s}a(x_\tau,\lambda_\tau)\,\mathrm{d}\tau\right|^2\right]\le s\int_t^{t+s}\mathbb{E}\left[|a(x_\tau,\lambda_\tau)|^2\right]\mathrm{d}\tau, \tag{55}$$

$$\le s\int_t^{t+s}C(1+\mathbb{E}[|x_\tau|^2])\,\mathrm{d}\tau, \tag{56}$$

$$\le C's^2, \tag{57}$$

and for the martingale part we have

$$\mathbb{E}\left[\left|\int_t^{t+s}b(x_s,\lambda_s)\,\mathrm{d}W_s\right|^2\right]\le\int_t^{t+s}\mathbb{E}\left[\|b(x_s,\lambda_s)\|^2\right]\mathrm{d}s \tag{58}$$

$$\le\int_t^{t+s}C(1+\mathbb{E}[|x_s|^2])\,\mathrm{d}s \tag{59}$$

$$\le C''s. \tag{60}$$

Putting both bounds together we obtain Eq. (52), we obtain

$$\mathbb{E}[|x_{t+s}-x_t|^2]\le C(s^2+\epsilon s), \tag{61}$$

and taking the supremum over the interval we find

$$\mathbb{E}\left[\sup_{0\le s\le\delta}|x_{t+s}-x_t|^2\right]\le\sup_{0\le s\le\delta}C(s^2+\epsilon s)\le C(\delta^2+\epsilon\delta), \tag{62}$$

Using Markov's inequality on Eq. (62) yeilds

$$\mathbb{P}\left[\sup_{0\le s\le\delta}|x_{t+s}-x_t|^2>\gamma^2\right]\le\frac{1}{\gamma^2}\mathbb{E}\left[\sup_{0\le s\le\delta}|x_{t+s}-x_t|^2\right]\le\frac{1}{\gamma^2}C(\delta^2+\epsilon\delta). \tag{63}$$

□

The consequence of Lemma 5.1 becomes more apparent when we choose any mesoscopic scale $\delta(\epsilon)$ satisfying Eq. (51), e.g. $\delta(\epsilon)=\epsilon^\beta$, for some $\beta>0$ and then letting $\epsilon\to0$. The bound in Eq. (52) ensures that for any fixed γ,

$$\mathbb{P}\left[\sup_{0\le s\le\delta(\epsilon)}|x_{t+s}-x_t|>\gamma\right]\le\frac{C}{\gamma^2}(\delta^2(\epsilon)+\epsilon\delta(\epsilon))\to 0, \tag{64}$$

and therefore the slow variable x_t, with probability tending to one, remains constant on the entire interval $[t, t+\delta(\epsilon)]$. Simultaneously, we have

$$\frac{\delta(\epsilon)}{\epsilon}=\epsilon^{\beta-1}\to\infty,\quad \text{as } \epsilon\to 0, \tag{65}$$

which shows that the δ-window is arbitrarily large on the fast λ–timescale. Thus, in δ-interval, the slow variable may be regarded as fixed while the fast variable has sufficient time to equilibrate.

We now fix $x\in\mathcal{D}_\epsilon$ and consider the dynamics of the λ on the interval $[t, t+\delta]$.

Lemma 5.2 (Backward generator of the switching variable).

Let $\delta>0$ satisfying Eq. (51), let $t\in[t', t'+\delta]\subset[0,T]$ for some $t'\in[0,T-\delta]$. Let $x_t=x\in\mathcal{D}_\epsilon$ be fixed (see Lemma 5.1). Then the backward generator $\mathcal{A}_x$ of $\lambda_t\in[-1,1]$ evolving according to Eq. (29), acts on sufficently smooth test function $f:[-1,-1]\mapsto\mathbb{R}$ via

$$(\mathcal{A}_x f)(\lambda)=\frac{1}{\epsilon}\partial_\lambda f(\lambda)\Big\{\partial_x\sigma(x)^\intercal a(x,\lambda)+\frac{\epsilon}{2}\operatorname{Tr}\big[b(x,\lambda)^\intercal\partial^2_{xx}\sigma(x)b(x,\lambda)\big]\Big\} \tag{66}$$

$$+\frac{1}{2\epsilon}\partial^2_{\lambda\lambda}f(\lambda)\partial_x\sigma(x)^\intercal b(x,\lambda)b(x,\lambda)^\intercal\partial_x\sigma(x), \tag{67}$$

with the domain

$$\operatorname{Dom}(\mathcal{A}_x)=\{f\in C^2([-1,1])\mid\partial_\lambda(1)=\partial_\lambda(-1)=0\}. \tag{68}$$

Note that the generator is conditional on x.

Proof.

With $x_t=x$ fixed on the interval $t\in[t', t'+\delta]\subset[0,T]$ for some $t'\in[0,T]$ and $\delta>0$, (see Lemma 5.1). Let $f\in C^2([-1,1])$ and set an initial conditoin $\lambda_{t'}=\lambda\in[-1,1]$. Applying Itō's lemma to $f(\lambda_t)$ to yield

$$f(\lambda_t)-f(\lambda)=\int_{t'}^t\partial_\lambda f(\lambda_s)\,\mathrm{d}\lambda_s+\frac{1}{2}\int_{t'}^t\partial^2_{\lambda\lambda}f(\lambda_s)\,\mathrm{d}\langle\lambda\rangle_s, \tag{69}$$

where $\langle\lambda\rangle$ is the quadratic variation of λ_t. Substituting Eq. (29) into first the first integral Eq. (69) we obtain the decomposition

$$\int_{t'}^t\partial_\lambda f(\lambda_s)\,\mathrm{d}\lambda_s=I_t^{(1)}+I_t^{(2)}+M_t, \tag{70}$$

where

$$I_t^{(1)}\overset{\text{def}}{=}\frac{1}{\epsilon}\int_0^t\partial_\lambda f(\lambda_s)\tilde{a}(x,\lambda_s)\,\mathrm{d}s, \tag{71}$$

$$I_t^{(2)}\overset{\text{def}}{=}\frac{1}{\epsilon}\int_0^t\partial_\lambda f(\lambda_s)[\,\mathrm{d}L_s^z(-\epsilon)-\mathrm{d}L_s^z(\epsilon)], \tag{72}$$

and M_t is the martingale term arising from the stochastic integral with respect to W_t. For the quadratic variation, since we know from Lemma 4.1.1 that $\mathrm{d}L_t^z(\pm\epsilon) = O(\mathrm{d}t)$, the second integral in Eq. (69) becomes

$$\frac{1}{2}\int_{t'}^{t} \partial^2_{\lambda\lambda} f(\lambda_s)\,\mathrm{d}\langle\lambda\rangle_s = I_t^{(3)} \overset{\text{def}}{=} = \frac{1}{2\epsilon}\int_0^t \partial^2_{\lambda\lambda} f(\lambda_s)\tilde{b}(x,\lambda_s)\tilde{b}(x,\lambda_s)^\intercal\,\mathrm{d}s, \tag{73}$$

Taking expectations, the martingale term vanishes leaving

$$\mathbb{E}[f(\lambda_t) - f(\lambda)] = \mathbb{E}\left[I_t^{(1)}\right] + \mathbb{E}\left[I_t^{(2)}\right] + \mathbb{E}\left[I_t^{(3)}\right]. \tag{74}$$

Interior terms. The terms $I_t^{(1)}$ and $I_t^{(3)}$ are the drift and diffusion on the interior of the [−1, 1], i.e. for $\lambda \in (-1,1)$. Since the coefficients are continious in t, we have

$$\lim_{t\to t'} \frac{1}{t}\mathbb{E}\left[I_t^{(1)}\right] = \frac{1}{\epsilon}\partial_\lambda f(\lambda)\tilde{a}(x,\lambda), \tag{75}$$

$$\lim_{t\to t'} \frac{1}{t}\mathbb{E}\left[\mathbb{I}_t^{(3)}\right] = \frac{1}{2\epsilon}\partial^2_{\lambda\lambda} f(\lambda)\tilde{b}(x,\lambda)\tilde{b}(x,\lambda)^\intercal, \tag{76}$$

hence, the interior contribution to the generator is the right-hand side of Eq. (66) when we express $\tilde{a}$ and $\tilde{b}$ in terms of a and b using Eq. (26) and Eq. (27).

Local time term. For the local time contribution $I_t^{(2)}$, we know that when $z_t = \pm\epsilon$, we have $\lambda_t = \pm 1$, thus $\partial_\lambda f(\lambda_s) = \partial_\lambda f(\pm 1)$ and

$$\mathbb{E}\left[I_t^{(2)}\right] = \frac{1}{\epsilon}\int_0^t \mathbb{E}[\partial_\lambda f(-1)\,\mathrm{d}L_s^z(-\epsilon) - \partial_\lambda f(1)\,\mathrm{d}L_s^z(\epsilon)]. \tag{77}$$

Using also Eq. (44) (see Lemma 4.1.1), we have

$$\mathrm{d}\mathbb{E}[\partial_\lambda f(\pm 1)L_t^z(\pm\epsilon)] = \partial_\lambda f(\pm 1)\,\mathrm{d}\mathbb{E}[L_t^z(\pm\epsilon)], \tag{78}$$

$$= \epsilon\partial_\lambda f(\pm 1)P^{(z)}(\pm\epsilon, t)\tilde{b}(x,\lambda_t)\tilde{b}(x,\lambda_t)^\intercal\,\mathrm{d}t, \tag{79}$$

from which we conclude

$$\lim_{t\to t'} \frac{1}{t}\mathbb{E}\left[I_t^{(2)}\right] = \epsilon C_\pm(x,t')\partial_\lambda f(\pm 1), \tag{80}$$

where $|C_\pm(x,t')| < \infty$ are smooth x, t' dependent coefficients. combining all of the expectation of the integrals together gives

Combining all of the intergral expecations with Eq. (74) and dividing by t gives

$$\lim_{t\to t'} \frac{1}{t}\mathbb{E}[f(\lambda_t) - f(\lambda)] = (\mathcal{A}_x f)(\lambda) + C_+(x,t')\partial_\lambda f(1) - C_-(x,t')\partial_\lambda f(-1), \tag{81}$$

where $\mathcal{A}_x$ is the interior differential operator defined in Eq. (66). By definition of the generator of a Markov process, and equivalently by Dynkin's formula, the infinitesimal generator must satisfy

$$\lim_{t\downarrow 0} \frac{1}{t}\mathbb{E}[f(\lambda_t) - f(\lambda)] = (\mathcal{A}_x f)(\lambda) \tag{82}$$

without any additional boundary terms. In view of Eq. (81), this is possible if and only if

$$C^+(x,t')\partial_\lambda f(1) - C^-(x,t')\partial_\lambda f(-1) = 0, \tag{83}$$

Since $C^{\pm}(x, t')$ are non-zero for $x \in \mathcal{D}_\epsilon$ due to the accumulation of local time at the boundaries, the only way this can hold for all $x \in \mathcal{D}_\epsilon$ is to impose the Neumann boundary conditions

$$\partial_\lambda f(1) = \partial_\lambda f(-1) = 0. \tag{84}$$

This characterises precisely the domain of $\mathcal{A}_x$ claimed in the lemma. □

From the backward generator it is then possible to obtain the forward generator using the L^2-adjoint relation

$$\int_{-}^{1} P_t(\lambda \mid x)(\mathcal{A}_x f_t)(\lambda)\,\mathrm{d}\lambda = \int_{-}^{1} (A_x^* P_t)(\lambda\mid x) f_t(\lambda)\,\mathrm{d}\lambda, \quad \forall f_t \in \mathrm{Dom}(\mathcal{A}_x) \tag{85}$$

where $t = [t', t+\delta] \subset [0, T]$, and $P_t(\lambda \mid x)$ is the occupation probability density of the switching variable λ. Going forwards we will drop the x notation in favour of $P(\lambda, t)$. The forward generator is sumarised in the following lemma.

Lemma 5.3 (Forward generator of the switching variable).

Let $\delta > 0$ satisfying Eq. (51), let $t \in [t', t' + \delta] \subset [0, T]$ for some $t' \in [0, T - \delta]$. Let $x_t = x \in \mathcal{D}_\epsilon$ be fixed (see Lemma 5.1). Then the forward generator $\mathcal{A}_x^*$ of $\lambda_t \in [-1, 1]$ evolving according to Eq. (29), acts on sufficently smooth probability density $P_t : [-1, -1] \mapsto [0, \infty)$ via

$$(\mathcal{A}_x^* P_t)(\lambda) = -\frac{1}{\epsilon}\partial_\lambda \Big(P_t(\lambda)\Big\{ \partial_x \sigma(x_t)^T a(x, \lambda) + \frac{\epsilon}{2}\operatorname{Tr}\big[b(x, \lambda)^T \partial_{xx}^2 \sigma(x) b(x, \lambda)\big]\Big\}\Big) \tag{86}$$

$$+\frac{1}{2\epsilon}\partial_{\lambda\lambda}^2 \big[P_t(\lambda)\partial_x \sigma(x_t)^T b(x, \lambda) b(x, \lambda)^T \partial_x \sigma(x_t)\big], \tag{87}$$

with the domain

$$\mathrm{Dom}(\mathcal{A}_x^*) = \{P_t \in C^2([-1, 1]; [0, \infty)) \mid J_t(\pm 1) = 0\}, \tag{88}$$

where

$$J_t(\lambda) = P_t(\lambda)\Big\{\partial_x \sigma(x_t)^T a(x, \lambda) + \frac{\epsilon}{2}\operatorname{Tr}\big[b(x, \lambda)^T \partial_{xx}^2 \sigma(x) b(x, \lambda)\big]\Big\} \tag{89}$$

$$-\frac{1}{2}\partial_\lambda \big[P_t(\lambda)\partial_x \sigma(x_t)^T b(x, \lambda) b(x, \lambda)^T \partial_x \sigma(x_t)\big], \tag{90}$$

is the scaled probability current, i.e. $J_t(\lambda)/\epsilon$ would be the probability current of the process.

Proof.

Let $P_t(\lambda)$ denote the occupation density of the switching variable $\lambda_t \in [-1, 1]$, conditioned on a frozen value of $x_t = x \in \mathcal{D}_\epsilon$. Inserting the backward generator from Lemma 5.2 into Eq. (85) we obtain

$$\frac{1}{\epsilon}\int_{-1}^{1} P_t(\lambda)\partial_\lambda f(\lambda)\tilde{a}(x, \lambda) + \frac{1}{2\epsilon}\int_{-1}^{1} P_t(\lambda)\partial_{\lambda\lambda}^2 f(\lambda)\tilde{b}(x, \lambda)\tilde{b}(x, \lambda)^\intercal \tag{91}$$

$$= \int_{-1}^{1} f_t(\lambda)(\mathcal{A}_x^* P_t)(\lambda) \tag{92}$$

$$(\mathcal{A}_x f)(\lambda) = \frac{1}{\epsilon}\partial_\lambda f(\lambda)\Big\{\partial_x \sigma(x_t)^\intercal a(x, \lambda) + \frac{\epsilon}{2}\operatorname{Tr}\big[b(x, \lambda)^\intercal \partial_{xx}^2 \sigma(x) b(x, \lambda)\big]\Big\} \tag{93}$$

$$+\frac{1}{2\epsilon}\partial_{\lambda\lambda}^2 f(\lambda)\partial_x \sigma(x_t)^\intercal b(x, \lambda) b(x, \lambda)^\intercal \partial_x \sigma(x_t). \tag{94}$$

We treat the drift and diffusion contributions seperately.

Drift contribution. Integration by parts gives

$$\frac{1}{\epsilon}\int_{-1}^{1} \partial_\lambda f(\lambda) P_t(\lambda)\tilde{a}(x,\lambda)\,\mathrm{d}\lambda = \frac{1}{\epsilon} f(\lambda) P_t(\lambda)\tilde{a}(x,\lambda)\Big|_{-1}^{1} \tag{95}$$

$$-\frac{1}{\epsilon}\int_{-1}^{1} f(\lambda)\partial_\lambda [P_t(\lambda)\tilde{a}(x,\lambda)]\,\mathrm{d}\lambda, \tag{96}$$

Diffusion contribution. Employing intrgration by parts twice yields

$$\frac{1}{2\epsilon}\int_{-1}^{1} \partial^2_{\lambda\lambda} f(\lambda) P_t(\lambda)\tilde{b}\tilde{b}^\top\,\mathrm{d}\lambda = \frac{1}{2\epsilon}\partial_\lambda f(\lambda)\left[P_t(\lambda)\tilde{b}\tilde{b}^\top\right] - \frac{1}{2\epsilon} f(\lambda)\partial_\lambda\left[P_t(\lambda)\tilde{b}\tilde{b}^\top\right]\Big|_{-1}^{1} \tag{97}$$

$$+\frac{1}{2\epsilon}\int_{-1}^{1} f(\lambda)\partial^2_{\lambda\lambda}\left[P_t(\lambda)\tilde{b}\tilde{b}^\top\right]\mathrm{d}\lambda, \tag{98}$$

where the arguments (x, λ) are dropped in the notation of $\tilde{a}$ and $\tilde{b}$ for clarity. Since $f \in \mathrm{Dom}(\mathcal{A}) = \{f \in C^2([-1,1]) \mid \partial_\lambda f(\pm 1) = 0\}$, all boundary terms proportional to $\partial_\lambda f(\pm 1)$ vanish. The remaining boundary terms must also vanish to respect conservation of probability (i.e. zero probability flux through $\lambda = \pm 1$), giving the boundary condtion

$$\left[P_t(\lambda)\tilde{a}(x_t,\lambda) - \frac{1}{2}\partial_\lambda\left(P_t(\lambda)\tilde{b}(x,\lambda)\tilde{b}(x,\lambda)^\top\right)\right]_{-1}^{1} = 0. \tag{99}$$

Using Eq. (95), Eq. (97), and enforcing Eq. (99), we identify the forward operator as

$$(\mathcal{A}^*_x P_t)(\lambda) = -\frac{1}{\epsilon}\partial_\lambda[P_t(\lambda)\tilde{a}(x,\lambda)] + \frac{1}{2\epsilon}\partial^2_{\lambda\lambda}\left[P_t(\lambda)\tilde{b}(x,\lambda)\tilde{b}(x,\lambda)^\top\right]. \tag{100}$$

This is also called the Fokker–Planck or Kolmogorov forward operator associated with the dynamics of the switching variable λ_t conditional on $x_t = x \in \mathcal{D}_\epsilon$. Substituting the definitions Eq. (26) Eq. (27) yields the defnition give in the lemma. □

Remark

Since $P_t(\lambda)$ is the one dimensional occupation probability density with zero flux boundary conditions, and since $J_t(\pm 1) = 0$ must always be satisfied, we have no current for all $\lambda \in [-1, 1]$ giving us the condtion

$$J_t(\lambda) = P_t(\lambda)\left\{\partial_x\sigma(x_t)^\top a(x,\lambda) + \frac{\epsilon}{2}\operatorname{Tr}\left[b(x,\lambda)^\top \partial^2_{xx}\sigma(x) b(x,\lambda)\right]\right\} \tag{101}$$

$$-\frac{1}{2}\partial_\lambda\left[P_t(\lambda)\partial_x\sigma(x_t)^\top b(x,\lambda)b(x,\lambda)^\top\partial_x\sigma(x_t)\right] = 0. \tag{102}$$

This ofcourse also means detailed balance is satified,.

TODO

fix this text introducing the linear bounds

Since Theorem 6.3 relies on bounding the diffusion coefficient, in order for us to to apply it we must bound

$$\tilde{d}(x,\lambda) \stackrel{\text{def}}{=} \tilde{b}(x,\lambda)\tilde{b}(x,\lambda)^\mathsf{T} = \partial_x \sigma(x)^\mathsf{T} b(x,\lambda) b(x,\lambda)^\mathsf{T} \partial_x \sigma(x). \tag{103}$$

Indeed $\tilde{d}$ as well as the coefficients $\tilde{a}$, $\tilde{b}$ can all b can all be bounded from above as the inherit the conditions of a and b as laid out in Definition 3.1. We summarise these in the next lemma as they will then be used in the later results.

Lemma 5.4 (Bounds on the coefficients).

> ⓘ ***TODO***
>
> *add all of the bounds*

Let $\lambda \in [-1, 1]$ and satisfies the bound

$$\|a(x,\lambda)\| + \|b(x,\lambda)\| \leq C(1+|x|) \tag{104}$$

for some C

$$\tilde{b} \geq \widetilde{M}\ \|\partial_x \sigma(x)\| \tag{105}$$

$$\tilde{d} = \tilde{b}\tilde{b}^T \geq \widetilde{M}^2\ \|\partial_x \sigma(x)\|^2 \tag{106}$$

Proof.

$$\|a(x,\lambda)\| + \|b(x,\lambda)\| = \frac{1}{2}(1+\lambda)[\|a_+(x)\| + \|b_+(x)\|] + \frac{1}{2}(1-\lambda)[\|a_-(x)\| + \|b_-(x)\|] \tag{107}$$

$$\leq \frac{1}{2}[(1+\lambda)C_+ + (1-\lambda)C_-](1+|x|) \tag{108}$$

$$\leq (C_+ + C_-)(1+|x|) \tag{109}$$

> ⓘ **TODO**
>
> complete the proof! check notebook

□

Lemma 5.5 (Invariant Measure).

Let $x_t = x \in \mathcal{D}_\epsilon$ be fixed (see Lemma 5.1), and let the forward generator of A_x^* of λ_t be define in Eq. (86), then the invariant measure $P_\infty(\lambda)$ satifies $(\mathcal{A}_x^* P_\infty)(\lambda) = 0$ and is given by

$$P_\infty(\lambda) = \frac{R(x)}{\tilde{d}(x,\lambda)} \exp\left(\int_{-1}^{\lambda} \frac{\tilde{a}(x,\nu)}{\tilde{d}(x,\nu)}\,\mathrm{d}\nu\right), \tag{110}$$

where

$$R(x) = \left[\int_{-1}^{1} \frac{1}{\tilde{d}(x,\lambda)} \exp\left(\int_{-1}^{\lambda} \frac{\tilde{a}(x,\nu)}{\tilde{d}(x,\nu)}\,\mathrm{d}\nu\right) \mathrm{d}\lambda\right]^{-1}, \tag{111}$$

is an x-dependent normalisation constant, $\tilde{a}(x,\lambda)$, and $\tilde{b}(x,\lambda)$ are defined in Eq. (26) and Eq. (103), respectively.

Proof.

The proof is a direct trivial calulation. We have

$$(\mathcal{A}^* P_\infty)(\lambda) = -\frac{1}{\epsilon}\partial_\lambda[P_\infty(\lambda)\tilde{a}(x,\lambda)] + \frac{1}{2\epsilon}\partial^2_{\lambda\lambda}\left[P_\infty(\lambda)\tilde{d}(x,\lambda)\right] = 0, \tag{112}$$

which immediately gives us the ordinary differential equation

$$P_\infty(\lambda)\tilde{a}(x,\lambda) = \frac{1}{2}\partial_\lambda\left[P_\infty(\lambda)\tilde{d}(x,\lambda)\right]. \tag{113}$$

It follows then that

$$\frac{\partial_\lambda P_\infty(\lambda)}{P_\infty(\lambda)} = \frac{2\tilde{a}(x,\lambda) - \partial_\lambda \tilde{d}(x,\lambda)}{\tilde{d}(x,\lambda)}, \tag{114}$$

which after intergrating both sides with respect to λ we obtain

$$\ln P_\infty(\lambda) = 2\int_{-1}^{\lambda} \frac{\tilde{a}(x,\nu)}{\tilde{d}(x,\nu)}\,\mathrm{d}\nu - \ln \tilde{d}(x,\lambda) + \ln R(x), \tag{115}$$

where $R(x)$ is an integration constant. Enforcing normalisation on the invariant density gives Eq. (111).

□

Lemma 5.6 (Bounds on the Invariant Measure).

For all $x \in \mathcal{D}_\epsilon$ fixed and $\|\tilde{a}(x,\lambda)\| \leq \tilde{C}_1(x)$ and $\|\tilde{b}(x,\lambda)\| \geq \tilde{C}_1(x) > 0$, set $\tilde{C}_{12}(x) = \frac{\tilde{C}_1(x)}{\tilde{C}_2(x)}$

$$\frac{\tilde{C}_{12}(x)\exp\left[-\tilde{C}_{12}(x)(1+4\,|\lambda|)\right]}{\sinh\left(\tilde{C}_{12}(x)\right)} \leq P_\infty(\lambda) \leq \frac{\tilde{C}_{12}(x)\exp\left[\tilde{C}_{12}(x)(1+4\,|\lambda|)\right]}{\sinh\left(\tilde{C}_{12}(x)\right)} \tag{116}$$

Proof.

> **TODO**
>
> proof in notebook add it in.

$$\frac{\alpha\,\mathrm{e}^{-2\alpha(1+4\,|\lambda|)}}{\sinh\left(\frac{\alpha}{2}\right)} \leq P_\infty(\lambda) \leq \frac{\alpha\,\mathrm{e}^{2\alpha(1+4\,|\lambda|)}}{\sinh\left(\frac{\alpha}{2}\right)} \tag{117}$$

□

6 Norm tings

Before we proceed it is usefull to introduce a defnition for the L^2 space that we will be working in, but first let us consider the general defintion of an L^2 inner product and norms on real functions.

Definition 6.1 (L^2-norm).

Let $(\Omega, \mathcal{F}, \mu)$ be a measurable space, where Ω is the set, $\mathcal{F}$ is the sigma algebra of Ω, and μ the measure. Then for any $\mathcal{F}$-measurable $f, g : \Omega \mapsto \mathbb{R}$, the L^2 inner product is defined as

$$\langle f, g\rangle_{L^2_\mu} \stackrel{\text{def}}{=} \int_\Omega f(\omega)g(\omega)\,\mathrm{d}\mu(\omega), \tag{118}$$

which induces the L^2-norm

$$\| f \|_{L^2_\mu} \stackrel{\text{def}}{=} \langle f, f \rangle^{\frac{1}{2}}_{L^2_\mu} = \left(\int_\Omega f^2(\omega) \, \mathrm{d}\mu(\omega) \right)^{\frac{1}{2}}. \tag{119}$$

Corollary 6.1.1 (L^2-bound to supremum bound).

It is straight forward to bound an L^2-norm using a supremum bound via the argument

$$\| f \|_{L^2_\mu} = \left(\int_\Omega f^2(\omega) \, \mathrm{d}\mu(\omega) \right)^{\frac{1}{2}} = \left(\int_\Omega |f(\omega)|^2 \, \mathrm{d}\mu(\omega) \right)^{\frac{1}{2}}, \tag{120}$$

$$\leq \sup_{x \in \Omega} |f(\omega)| \left(\int_\Omega \mathrm{d}\mu(\omega) \right)^{\frac{1}{2}}, \tag{121}$$

$$\leq \sqrt{\mu(\Omega)} \sup_{x \in \Omega} |f(\omega)|. \tag{122}$$

Since we have a probability density we will use the notation $\langle . \mid . \rangle_{L^2_{P_t}}$ and $\| \, . \, \|_{L^2_{P_t}}$ where $P_t \in \mathrm{Dom}(\mathcal{A}^*_x)$.

Lemma 6.2 (Symmetry of $\mathcal{A}_x$).

Let $x \in \mathcal{D}_\epsilon$ be fixed, $\mathcal{A}_x$ be the backward generator given in Lemma 5.2, and $P_\infty(\lambda \mid x)$ satisfy $\mathcal{A}^*_x P_\infty = 0$ with $J_\infty(\pm 1) = 0$. Then $\mathcal{A}_x$ is symetric in $L^2_{P_\infty}$, that is satisfying the ralation

$$\langle \mathcal{A}_x f, g \rangle_{L^2_{P_\infty}} = \langle f, \mathcal{A}_x g \rangle_{L^2_{P_\infty}}, \tag{123}$$

for all $f, g \in \mathrm{Dom}(\mathcal{A}_x)$.

Proof.

We proceed by substituting Eq. (66) into the left hand side of Eq. (123) gives,

$$\int_{-1}^{1} (\mathcal{A}_x f)(\lambda) g(\lambda) P_\infty(\lambda) = I_1 + I_2 \tag{124}$$

where

$$I_1 \stackrel{\text{def}}{=} \int_{-1}^{1} \partial_\lambda f(\lambda) \tilde{a}(x, \lambda) g(\lambda) P_\infty(\lambda) \, \mathrm{d}\lambda, \tag{125}$$

$$I_2 \stackrel{\text{def}}{=} \frac{1}{2} \int_{-1}^{1} \partial^2_{\lambda\lambda} f(\lambda) \tilde{d}(x, \lambda) g(\lambda) P_\infty, \mathrm{d}\lambda \tag{126}$$

are, respectively, the drift and diffusion contributions which we treat separately.

Drift term. Integration by parts gives

$$I_1 = f(\lambda) \tilde{a}(x, \lambda) g P_\infty(\lambda) |^1_{-1} - \int_{-1}^{1} f(\lambda) \partial_\lambda [\tilde{a}(x, \lambda) g(\lambda) P_\infty(\lambda)] \, \mathrm{d}\lambda. \tag{127}$$

Diffusion term. Let $\tilde{d}(x, \lambda) \stackrel{\text{def}}{=} \tilde{b}(x, \lambda) \tilde{b}(x, \lambda)^\intercal$. Integrating by parts twice yeilds

$$I_2 = \frac{1}{2} \partial_\lambda f(\lambda) d(x, \lambda) g(\lambda) P_\infty(\lambda) |^1_{-1} - \frac{1}{2} \int_{-1}^{1} \partial_\lambda f(\lambda) \partial_\lambda [d(x, \lambda) g(\lambda) P_\infty(\lambda)] \, \mathrm{d}\lambda, \tag{128}$$

$$= -\frac{1}{2} f(\lambda)\partial_\lambda[d(x,\lambda)g(\lambda)P_\infty(\lambda)]|_{-1}^{1} + \frac{1}{2}\int_{-1}^{1} f(\lambda)\partial^2_{\lambda\lambda}[d(x,\lambda)g(\lambda)P_\infty(\lambda)]\,\mathrm{d}\lambda, \tag{129}$$

where the boundary term on the first line vanishes as $\partial_\lambda f(\pm 1) = 0$. Considering only the boundary terms from Eq. (127) and Eq. (128),

$$f(\lambda)\left\{\tilde{a}(x,\lambda)g(\lambda)P_\infty(\lambda) - \frac{1}{2}\partial_\lambda\left[\tilde{d}(x,\lambda)g(\lambda)P_\infty(\lambda)\right]\right\}\Bigg|_{-1}^{1}, \tag{130}$$

$$= f(\lambda)\left\{\tilde{a}(x,\lambda)g(\lambda)P_\infty(\lambda) - \frac{1}{2}\partial_\lambda g(\lambda)\tilde{d}(x,\lambda)g(\lambda)P_\infty(\lambda) - g(\lambda)\partial_\lambda\left[\tilde{d}(x,\lambda)P_\infty(\lambda)\right]\right\}\Bigg|_{-1}^{1}, \tag{131}$$

$$= f(\lambda)g(\lambda)\left\{\tilde{a}(x,\lambda)P_\infty(\lambda) - \partial_\lambda\left[\tilde{d}(x,\lambda)P_\infty(\lambda)\right]\right\}\Big|_{-1}^{1} = 0, \tag{132}$$

where we have used $\partial_\lambda g(\pm 1) = 0$ and the condition in Eq. (101). Expanding the integrand of the first integral gives

$$f(\lambda)\partial_\lambda[g(\lambda)\tilde{a}(x,\lambda)P_\infty(\lambda)] = f(\lambda)[\partial_\lambda g(\lambda)\tilde{a}(x,\lambda)P_\infty(\lambda) + g(\lambda)\partial_\lambda[\tilde{a}(x,\lambda)P_\infty(\lambda)]], \tag{133}$$

similarlay for the second integrand

$$\frac{1}{2}f(\lambda)\partial^2_{\lambda\lambda}[d(x,\lambda)g(\lambda)P_\infty(\lambda)] = \frac{1}{2}f(\lambda)\{\partial^2_{\lambda\lambda}g(\lambda)\tilde{d}(x,\lambda)P_\infty(\lambda) \tag{134}$$

$$+2\partial_\lambda g(\lambda)\partial_\lambda\left[\tilde{d}(x,\lambda)P_\infty(\lambda)\right] \tag{135}$$

$$+g(\lambda)\partial^2_{\lambda\lambda}\left[\tilde{d}(x,\lambda)P_\infty(\lambda)\right]\}. \tag{136}$$

The coeffcients of $g(\lambda)$ vanish due to the steady-state condition on $P_\infty(\lambda)$, ie. $(\mathcal{A}_x^* P_\infty)(\lambda) = 0$, leaving

$$\frac{1}{2}f(\lambda)\partial^2_{\lambda\lambda}[d(x,\lambda)g(\lambda)P_\infty(\lambda)] - f(\lambda)\partial_\lambda[g(\lambda)\tilde{a}(x,\lambda)P_\infty(\lambda)] \tag{137}$$

$$= f(\lambda)\left\{\partial_\lambda g(\lambda)\left[\tilde{a}(x,\lambda) + \frac{1}{2}\partial^2_{\lambda\lambda}g(\lambda)\tilde{d}(x,\lambda)\right]P_\infty(\lambda)\right. \tag{138}$$

$$\left.-2\partial_\lambda g(\lambda)\left[\tilde{a}(x,\lambda)P_\infty(\lambda) - \frac{1}{2}\partial_\lambda\left[\tilde{d}(x,\lambda)P_\infty(\lambda)\right]\right]\right\}. \tag{139}$$

The second term vanishes due to the zero current condition leaving only $(\mathcal{A}_x g)(\lambda)$ in the integrand, thus

$$\int_{-1}^{1} (\mathcal{A}_x f) g P_\infty\,\mathrm{d}\lambda = \int_{-1}^{1} f(\mathcal{A}_x g) P_\infty\,\mathrm{d}\lambda. \tag{140}$$

□

TODO

add text here to say we need to introduce this theorem without proof, for the proof see ….

Theorem 6.3 (Poincaré inequality).

Let $P : [-1,1] \to (0,\infty)$ be a probability density and $d : [-1,1] \to (0,\infty)$ satisfy $d(\lambda) \geq C_1 > 0$ and $P(\lambda) \geq C_2 > 0$ for all $\lambda \in [-1,1]$. Then for all $f \in C^1([-1,1])$ with $\int_{-1}^{1} f(\lambda)P(\lambda)\,\mathrm{d}\lambda = 0$

$$\int_{-1}^{1} f^2(\lambda)P(\lambda)\,\mathrm{d}\lambda \leq \frac{1}{\kappa}\int_{-1}^{1} [\partial_\lambda f(\lambda)]^2 d(\lambda)P(\lambda)\,\mathrm{d}\lambda, \tag{141}$$

where $\kappa = (C_1C_2)/2$.

Proof. By the fudnemental thorem of calculus we have

$$f(a) - f(b) = \int_a^b \partial_\lambda f(\lambda)\,\mathrm{d}\lambda. \tag{142}$$

Multiplying both sides by $P_\infty(b)$ and integrating over the support gives

$$\int_{-1}^{1} f(a)P(b)\,\mathrm{d}b - \int_{-1}^{1} P(b)f(b)\,\mathrm{d}b = \int_{-1}^{1} P(b)\int_a^b \partial_\lambda f(\lambda)\,\mathrm{d}\lambda\,\mathrm{d}b, \tag{143}$$

where the first integral on the left hand side simplifies due to the normalisastion of probability, while the second vanishes to zero due to $\mathbb{E}[f(\lambda)] = 0$, giving the relation

$$f(a) = \int_{-1}^{1} P(b)\int_a^b \partial_\lambda f(\lambda)\,\mathrm{d}\lambda\,\mathrm{d}b. \tag{144}$$

Taking the absolute value, squaring and employing Caychy-Schwarz twice gives us the bound

$$\begin{aligned}
|\ f(a)|^2 &= \left(\int_{-1}^{1} P(b)\left|\int_a^b \partial_\lambda f(\lambda)\,\mathrm{d}\lambda\right|\mathrm{d}b\right)^2 && (145)\\
&\leq \left(\int_{-1}^{1} P(b)\,\mathrm{d}b\right)\left(\int_{-1}^{1} P(b)\int_a^b |\partial_\lambda f(\lambda)\ |^2\,\mathrm{d}\lambda\,\mathrm{d}b\right), && (146)\\
&\leq \left(\int_{-1}^{1} P(b)\int_{-1}^{1} |\partial_\lambda f(\lambda)\ |^2\,\mathrm{d}\lambda\,\mathrm{d}b\right), && (147)\\
&\leq 2\int_{-1}^{1} [\partial_\lambda f(\lambda)]^2\,\mathrm{d}\lambda. && (148)
\end{aligned}$$

To obtain the desired bound from Eq. (145), we simply multiply both by $P(a)$ and integrate over the interval to get

$$\begin{aligned}
\int_{-1}^{1} |f(a)|^2\ P(a)\,\mathrm{d}a = \int_{-1}^{1} f^2(a)P(a)\,\mathrm{d}a &\leq 2\int_{-1}^{1} [\partial_\lambda f(\lambda)]^2\,\mathrm{d}\lambda, && (149)\\
&\leq 2\int_{-1}^{1} \frac{[\partial_\lambda f(\lambda)]^2}{d(\lambda)P(\lambda)} d(\lambda)P(\lambda)\,\mathrm{d}\lambda, && (150)\\
&\leq \frac{2}{C_1C_2}\int_{-1}^{1} [\partial_\lambda f(\lambda)]^2 d(\lambda)P(\lambda)\,\mathrm{d}\lambda, && (151)
\end{aligned}$$

and letting $\kappa = (C_1C_2)/2$ yields Eq. (141). □

Lemma 6.4 (Bounding zero-mean observables).

Let $\mathcal{A}_x$ *be the backward generator defined in Eq. (66) in Lemma 5.2,* $\mathcal{A}_x^*$ *be its adjoint defined in Eq. (86), let* $P_\infty(\lambda)$ *be the steady-state probability density such that* $(\mathcal{A}_x^* P_\infty)(\lambda) = 0$*, then for all* $f \in \mathrm{Dom}(\mathcal{A})_x$ *satisfying*

$$\int_{-1}^{1} f(\lambda) P_\infty(\lambda) \, \mathrm{d}\lambda = 0, \tag{152}$$

we have the inequality

$$\| f \|^2_{L^2_{P_\infty}} \le \frac{1}{\kappa} \langle -\mathcal{A}_x f \mid f \rangle_{L^2_{P_\infty}} \tag{153}$$

Proof.

We need only show that

$$\langle -\mathcal{A}_x f \mid f \rangle_{L^2_{P_\infty}} = \frac{1}{\kappa} \int_{-1}^{1} [\partial_\lambda f(\lambda)]^2 \tilde{d}(x, \lambda) P_\infty(\lambda) \, \mathrm{d}\lambda, \tag{154}$$

since from Lemma 5.4 that $|\tilde{d}(x, \lambda)|$ and similarly we know that $P_\infty(\lambda)$ is bounded from below from Lemma 5.6, therefore we can direcly apply Theorem 6.3.

$$-\int_{-1}^{1} f(\lambda) \left[\partial_\lambda f(\lambda) \tilde{a}(x, \lambda) + \frac{1}{2} \partial^2_{\lambda\lambda} f(\lambda) \tilde{d}(x, \lambda)\right] P_\infty(\lambda) \, \mathrm{d}\lambda = \int_{-1}^{1} d(x, \lambda) [\partial_\lambda f(\lambda)]^2 P_\infty(\lambda) \, \mathrm{d}\lambda \tag{155}$$

then the rest follows via Theorem 6.3

TODO

in notebook complete it.

□

Lemma 6.5 (Dense sets in $L^2([-1, 1])$).

There exists a set $G \subset \mathrm{Dom}(\mathcal{A}_x)$ that are dense in L^2

Proof.

□

Lemma 6.6 (zero mean observables in L^2).

Let $\mathcal{A}_x$ be the backward generator defined in Eq. (66) in Lemma 5.2, $\mathcal{A}^*_x$ be its adjoint defined in Eq. (86), let $P_\infty(\lambda)$ be the steady-state probability density such that $(\mathcal{A}^*_x P_\infty)(\lambda) = 0$, then for all $g \in L^2([-1, 1]; \mathbb{R})$ satisfying

$$\int_{-1}^{1} g(\lambda) P(\lambda) \, \mathrm{d}\lambda = 0 \quad \forall P \in \mathrm{Dom}(\mathcal{A}^*_x), \tag{156}$$

we have the inequality

$$\| g \|^2_{L^2_{P_\infty}} \le \frac{1}{\kappa} \langle -\mathcal{A}_x g \mid g \rangle_{L^2_{P_\infty}} \tag{157}$$

Proof.

By Lemma 6.5, there exisits a sequence smooth function in $f_n \in \mathrm{Dom}(\mathcal{A}_x)$ such that

$$\lim_{n \to 0} \| f_n - g \| = 0, \tag{158}$$

We can approximte any g in $L^2([-1, 1])$, by a sequance of smooth $f \in \mathrm{Dom}(\mathcal{A}_x)$, hence we can extend the Lemma 6.4 to generic g

TODO

finish proof.

□

Theorem 6.7 (Exponential mixing of the switching variable).

Let $x \in \mathcal{D}_\epsilon$ be fixed, let $P_t \in \mathrm{Dom}(\mathcal{A})^*_x$ represent the occupation probability density of λ conditioned on x, let P_∞ be the invariant density, i.e $(\mathcal{A}^*_x P_\infty)(\lambda) = 0$ which is uniformly bounded from below by $P_\infty(\lambda) \geq C_1 > 0$ and let the coefficient $|\tilde{d}(x, \lambda)| \geq C_2 > 0$ defined in Eq. (103), be uniformaly bounded from below. Then for any $t \in [0, T]$, and for any measurable $A \subset [-1, 1]$, there exisits $\kappa > 0$ and $C > 0$ such that

$$\left| \int_A P_t(\lambda \mid x)\, \mathrm{d}\lambda - \int_A P_\infty(\lambda \mid x)\, \mathrm{d}\lambda \right| \leq C\, \mathrm{e}^{-\kappa t}\,. \tag{159}$$

Proof.

To aid in the proof we define $\xi_t(\lambda) \stackrel{\text{def}}{=} P_t(\lambda) - P_\infty(\lambda)$, and $\zeta_t(\lambda) \stackrel{\text{def}}{=} \xi_t(\lambda)/P_\infty(\lambda)$, where we have dropped the conditional argument in the noation. Clearly $\zeta_t(\lambda)$ has

$$\begin{aligned}
\left| \int_A [P_t(\lambda) - P(\lambda)]\, \mathrm{d}\lambda \right| &= \left| \int_A \zeta_t(\lambda) P_\infty(\lambda)\, \mathrm{d}\lambda \right|, && (160)\\
&\leq \int_{-1}^{1} |\zeta_t(\lambda)|\; P_\infty(\lambda)\, \mathrm{d}\lambda, && (161)\\
&\leq \left(\int_{-1}^{1} \zeta_t^2(\lambda) P_\infty(\lambda)\, \mathrm{d}\lambda \right)^{1/2} \left(\int_{-1}^{1} P_\infty(\lambda)\, \mathrm{d}\lambda \right)^{1/2}, && (162)\\
&\leq \left(\int_{-1}^{1} \zeta_t^2(\lambda) P_\infty(\lambda)\, \mathrm{d}\lambda \right)^{1/2}, && (163)\\
&= \|\, \zeta_t \,\|_{L^2_{P_\infty}}\,. && (164)
\end{aligned}$$

Then it only remains to bound $\|\, \zeta_t \,\|_{L^2_{P_\infty}}$

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \|\, \zeta_t \,\|^2_{L^2_{P_\infty}} &= 2 \int_{-1}^{1} \zeta_t(\lambda) \partial_t \zeta_t(\lambda) P_\infty(\lambda)\, \mathrm{d}\lambda, && (165)\\
&= -2 \langle -\mathcal{A}_x \zeta_t \mid \zeta_t \rangle_{L^2_{P_\infty}}, && (166)\\
&\leq -2\kappa \|\, \zeta_t \,\|^2_{L^2_{P_\infty}}, \quad \text{(by Lemma 6.6)}. && (167)
\end{aligned}$$

Then by Gronwall's inequality,

$$\|\, \zeta_t \,\|_{L^2_{P_\infty}} \leq \|\, \zeta_0 \,\|_{L^2_{P_\infty}}\, \mathrm{e}^{-\kappa t}, \tag{168}$$

defining $C = \|\, \zeta_0 \,\|_{L^2_{P_\infty}}$ yields the relation in the lemma. □

Corollary 6.7.1 (Bound on expectations).

The bound on the differences between probability measures obtained in Theorem 6.7 affords us a further bound, namely on the differences in expectations via

$$\left|\mathbb{E}[f(\lambda_t)] - \int_{-1}^{1} f(\lambda) P_\infty(\lambda)\,\mathrm{d}\lambda\right| = \left|\int_{-1}^{1} f(\lambda)\zeta_t(\lambda) P_\infty(\lambda)\,\mathrm{d}\lambda\right|, \tag{169}$$

$$= \left|\langle f \mid \zeta_t \rangle_{L^2_{P_\infty}}\right|, \tag{170}$$

$$\leq \| f \|_{L^2_{P_\infty}} \| \zeta_t \|_{L^2_{P_\infty}}, \tag{171}$$

$$\leq 2 \| f \|_{L^2_{P_\infty}} \mathrm{e}^{-\kappa t}, \tag{172}$$

$$\leq 2 \sup_{\lambda\in[-1,1]} | f(\lambda) | \mathrm{e}^{-\kappa t} . \tag{173}$$

The final supremum bound is of course only meaningful when we have bounded f on the interval $\lambda \in [-1, 1]$.

In summary we have estabilshed the existance of an intermediate timescale δ such that $\epsilon \ll \delta \ll 1$ where the following are satisfied:

- bounds on the coefficents,
- justification of the transversality condition
- $\mathbb{E}[\sup_s |x_{t+s} - x_t|^2] \to 0$ as $\epsilon \to 0$, that is the slow varible remains fixed.
- Exponentila mixing via Poincare inequality where the poincare constant is obtained from the lower bounds on the invariant measure and $\tilde{d}$ bounds
- mixing bounds on the expectations of observables

7 Averaging Principle

We are now ready to introduce the averaging principle for the slow dynamics

Definition 7.1 (The Reduced SDE).

Let $x_t \in \mathcal{D}_\epsilon$ be a solution of the piecewise-smooth SDE given in Definition 3.1, and let $\lambda \in [-1, 1]$ be the switching variable parametrising the convex interpolation

$$a(t, x, \lambda) = \frac{1}{2}(1 + \lambda)a^+(t, x) + \frac{1}{2}(1 - \lambda)a^-(t, x), \tag{174}$$

$$b(t, x, \lambda) = \frac{1}{2}(1 + \lambda)b^+(t, x) + \frac{1}{2}(1 - \lambda)b^-(t, x), \tag{175}$$

between the piecewise-smooth coefficients $a^\pm$ and $b^\pm$. Let $P_{\mathrm{ss}}(\lambda \mid x)$ denote the stationary distribution of the switching variable conditional on x, satisfying $\mathcal{A}_x^* P_{\mathrm{ss}} = 0$ with zero-flux boundary conditions (see Lemma 5.3). The reduced SDE is

$$\mathrm{d}\overline{x}_t = [\overline{a}(\overline{x}_t) + \alpha\epsilon\overline{c}(\overline{x}_t)]\,\mathrm{d}t + \sqrt{\epsilon}\overline{b}(\overline{x}_t)\,\mathrm{d}W_t, \tag{176}$$

where the averaged coefficients are

$$\overline{a}(t, x) = \int_{-1}^{1} a(t, x, \lambda) P_{\mathrm{ss}}(\lambda \mid x)\,\mathrm{d}\lambda, \tag{177}$$

$$\overline{b}(t, x) = \int_{-1}^{1} b(t, x, \lambda) P_{\mathrm{ss}}(\lambda \mid x)\,\mathrm{d}\lambda, \tag{178}$$

$$\overline{c}(t, x) = \int_{-1}^{1} c(t, x, \lambda) P_{\mathrm{ss}}(\lambda \mid x)\,\mathrm{d}\lambda, \tag{179}$$

and

$$c(x,\lambda) = \sum_j J_x\left[b_j(x,\lambda)\right] b_j(x,\lambda) \tag{180}$$

is the Itō correction arising from the α*-interpretation of the stochastic integral, with* $b_j(x,\lambda)$ *denoting the* j*-th column of* $b(x,\lambda)$ *and* $J_x(\cdot)$ *the Jacobian with respect to* x*.*

Unsurprisingly, without any hidden term in the dynamics we require only the mean from the distribution $P_\infty(\lambda)$,

Bibliography

[1] M. I. Freidlin and A. D. Wentzell, *Random perturbations.* Springer, 1998, pp. 15–43.

[2] T.-S. Chiang and S.-J. Sheu, “Large deviation of diffusion processes with discontinuous drift and their occupation times,” *The Annals of Probability*, vol. 28, no. 1, pp. 140–165, 2000.

[3] T.-S. Chiang and S.-J. Sheu, “Small perturbation of diffusions in inhomogeneous media,” in *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 2002, pp. 285–318.

[4] Y. Chen, A. Baule, H. Touchette, and W. Just, “Weak-noise limit of a piecewise-smooth stochastic differential equation,” *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 88, no. 5, p. 52103, 2013.

[5] K. Hill, J. Zanetell, and J. A. Gemmer, “Most Probable Transition Paths in Piecewise-Smooth Stochastic Differential Equations,” *Physica D: Nonlinear Phenomena*, vol. 439, p. 133424, Nov. 2022, doi: 10.1016/j.physd.2022.133424.

[6] B. K. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Sixth edition, sixth corrected printing. in Universitext. Berlin Heidelberg New York Dordrecht London: Springer, 2013.

[7] C. W. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed., no. 13. in Springer Series in Synergetics. Berlin Heidelberg: Springer, 2009.

[8] A. F. Filippov, *Differential Equations with Discontinuous Righthand Sides: Control Systems*, vol. 18. Springer Science & Business Media, 2013.

[9] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus.* springer, 2014.

[10] P. E. Protter, *Stochastic differential equations.* Springer, 2012.