Deterministic homogenization for fast-slow systems with chaotic noise

August 29, 2016

David Kelly[⋆] **Ian Melbourne**[†]

⋆ Courant Institute, New York University, NY, USA.
Email: dtkelly@cims.nyu.edu
† Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.
Email: I.Melbourne@warwick.ac.uk

Abstract

Consider a fast-slow system of ordinary differential equations of the form $\dot{x} = a(x,y)+\varepsilon^{-1}b(x,y)$, $\dot{y} = \varepsilon^{-2}g(y)$, where it is assumed that b averages to zero under the fast flow generated by g. We give conditions under which solutions x to the slow equations converge weakly to an Itô diffusion X as $\varepsilon \to 0$. The drift and diffusion coefficients of the limiting stochastic differential equation satisfied by X are given explicitly.

Our theory applies when the fast flow is Anosov or Axiom A, as well as to a large class of nonuniformly hyperbolic fast flows (including the one defined by the well-known Lorenz equations), and our main results do not require any mixing assumptions on the fast flow.

1 Introduction

Let $\{\phi_t\}_{t\geq 0}$ be a smooth, deterministic flow on a finite dimensional manifold M, with invariant ergodic probability measure μ. One should think of ϕ_t as the flow generated by an ordinary differential equation (ODE) with a chaotic invariant set $\Omega \subset M$ and μ supported on Ω. Define $y(t) = \phi_t y_0$ where the initial condition y_0 is chosen at random according to μ. Hence $y(t) = y(t, y_0)$ is a random variable on the probability space (Ω, μ); from here on we omit y_0 from the notation, as is conventional with random variables. Let $a, b : \mathbb{R}^d \times M \to \mathbb{R}^d$ be vector fields with suitable regularity assumptions. We are interested in the asymptotic behaviour of the ODE

$$\frac{dx^{(\varepsilon)}}{dt} = \varepsilon^2 a(x^{(\varepsilon)}, y) + \varepsilon b(x^{(\varepsilon)}, y) \quad , \quad x^{(\varepsilon)}(0) = \xi$$

as $\varepsilon \to 0$ and $t \to \infty$, with $\varepsilon^2 t$ remaining fixed. The initial condition $\xi \in \mathbb{R}^d$ is assumed deterministic. Due to the dependence on y_0, we interpret $x^{(\varepsilon)}$ as a random variable on Ω taking values in the space of continuous functions $C([0,T], \mathbb{R}^d)$ for some finite $T > 0$.

To make the statement of convergence precise, we define $y_\varepsilon(t) = y(\varepsilon^{-2}t)$ and x_ε as the solution to the ODE

$$\frac{dx_\varepsilon}{dt} = a(x_\varepsilon, y_\varepsilon) + \frac{1}{\varepsilon} b(x_\varepsilon, y_\varepsilon) \quad , \quad x_\varepsilon(0) = \xi \; . \tag{1.1}$$

In particular, we arrive at this equation under the rescaling $t \mapsto t/\varepsilon^2$ and setting $x_\varepsilon(t) = x^{(\varepsilon)}(t/\varepsilon^2)$. Our aim is to identify the limiting behavior of the random variable x_ε on the space of continuous functions as $\varepsilon \to 0$.

Under certain assumptions on the fast flow ϕ_t, it is known that $x_\varepsilon \to_w X$ where X is an Itô diffusion, and where $\to_w$ denotes weak convergence of random variables on the space $C([0,T],\mathbb{R}^d)$. At an intuitive level, the a term *averages* to an ergodic mean, via a law of large numbers type effect and the b term *homogenizes* to a stochastic integral, via a central limit theorem type effect. This type of problem is often referred to as *deterministic* homogenization, since the randomness is not coming from a typical stochastic process, but rather from a ergodic dynamical system with random initial condition.

Assuming rather strong mixing conditions on ϕ_t, one can show that x_ε converges weakly to the solution X of an Itô SDE

$$dX = \tilde{a}(X)dt + \sigma(X)dB \quad , \quad X(0) = \xi \tag{1.2}$$

where B is an $\mathbb{R}^d$ valued standard Brownian motion, the drift $\tilde{a} : \mathbb{R}^d \to \mathbb{R}^d$ is given by

$$\tilde{a}^i(x) = \int_\Omega a^i(x,y)d\mu(y) + \int_0^\infty \int_\Omega b(x,y)\cdot\nabla b^i(x,\phi_t y)d\mu(y)dt \tag{1.3}$$

for all $i = 1,\dots,d$ and the diffusion coefficient $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is given by

$$\sigma(x)\sigma^T(x) = \int_0^\infty \int_\Omega \big(b(x,y)\otimes b(x,\phi_t y) + (b(x,\phi_t y)\otimes b(x,y)\big)d\mu(y)dt\;. \tag{1.4}$$

The mixing assumptions required on ϕ_t are typically very strong. For instance, the above result follows from [PK74] under the assumption of phi mixing with rapidly decaying mixing coefficient ($L^{1/2}$-integrable). Such an assumption is quite reasonable in the setting of ergodic Markov processes (as intended in [PK74]). Unfortunately this is quite unreasonable for general ergodic flows. In particular, for most natural *deterministic* situations it is difficult to prove any mixing properties at all. On top of that, it is seldom clear that the formulas for $\tilde{a}$ and σ are even well-defined.

In this article, we show that for a very general class of ergodic flows, the above result holds with explicit (but sometimes more complicated) formulas for $\tilde{a}$ and σ that generalise the ones given above.

1.1 Anosov and Axiom A flows

One well-known class of fast flows to which our results apply is given by the Axiom A (uniformly hyperbolic) flows introduced by Smale [Sma67]. This includes Anosov flows [Ano67]. We do not give the precise definitions, since they are not needed for understanding the paper, but a rough description is as follows. (See [Bow75, Rue78, Sin72] for more details.) Let $\phi_t : M \to M$ be a C^2 flow defined on a compact manifold M. A flow-invariant subset $\Omega \subset M$ is uniformly hyperbolic if for all $x \in \Omega$ there exists a $D\phi_t$-invariant splitting transverse to the flow into uniformly contracting and expanding directions. The flow is Anosov if the whole of M is uniformly hyperbolic. More generally, an Axiom A flow is characterised by the property that the dynamics decomposes into finitely many hyperbolic equilibria and finitely many uniformly hyperbolic subsets $\Omega_1,\dots,\Omega_k$, called hyperbolic basic sets, such that the flow on each Ω_i is transitive (there is a dense orbit). If Ω is a hyperbolic basic set, there is a unique ϕ_t-invariant ergodic probability measure (called an equilibrium measure) associated to each Hölder function on Ω. (In the special case that Ω is an attractor, there is a distinguished equilibrium measure called the physical measure or SRB measure (after Sinai, Ruelle, Bowen).) In the remainder of the introduction, we assume that Ω is a hyperbolic basic set with equilibrium measure μ (corresponding to a Hölder potential). We exclude the trivial case where Ω consists of a single periodic orbit.

Given $b : \mathbb{R}^d \times M \to \mathbb{R}$, we define the mixed Hölder norm

$$\|b\|_{C^{\alpha,\kappa}} = \sum_{|k|\le\lfloor\alpha\rfloor} \sup_{x\in\mathbb{R}^d} \left\|D^k b(x,\cdot)\right\|_{C^\kappa} + \sum_{|k|=\lfloor\alpha\rfloor} \sup_{x,z\in\mathbb{R}^d} \frac{\left\|D^k b(x,\cdot) - D^k b(z,\cdot)\right\|_{C^\kappa}}{|x-z|^{\alpha-\lfloor\alpha\rfloor}}$$

for $\alpha \in [0,\infty)$, $\kappa \in [0,1)$, where the second summation is omitted when α is an integer. Here D^k is the differential operator acting in the x component and $\|\cdot\|_{C^\kappa}$ is the standard Hölder norm acting in the y component. If $b : \mathbb{R}^d \times M \to \mathbb{R}^m$ is vector-valued, we define $\|b\|_{C^{\alpha,\kappa}} = \sum_{i=1}^m \|b^i\|_{C^{\alpha,\kappa}}$.

We write $b \in C^{\alpha,\kappa}(\mathbb{R}^d \times M, \mathbb{R}^m)$ if $\|b\|_{C^{\alpha,\kappa}} < \infty$. Let $C_0^{\alpha,\kappa}(\mathbb{R}^d \times M, \mathbb{R}^m)$ denote the space of observables $b \in C^{\alpha,\kappa}(\mathbb{R}^d \times M, \mathbb{R}^m)$ with $\int_\Omega b(x,y)d\mu(y) = 0$ for all $x \in \mathbb{R}^d$. When $m = 1$, we write $C^{\alpha,\kappa}(\mathbb{R}^d \times M)$ instead of $C^{\alpha,\kappa}(\mathbb{R}^d \times M, \mathbb{R}^m)$ and so on. We also write $C_0^\kappa(\Omega, \mathbb{R}^m)$ to denote C^κ observables $v : \Omega \to \mathbb{R}^m$ with mean zero. We now state the main result.

Theorem 1.1. *Let $\Omega \subset M$ be a hyperbolic basic set with equilibrium measure μ. Let $\kappa > 0$ and suppose that $a \in C^{1+,0}(\mathbb{R}^d \times M, \mathbb{R}^d)$, $b \in C_0^{2+,\kappa}(\mathbb{R}^d \times M, \mathbb{R}^d)$. Then*

(i) The limit

$$\mathfrak{B}(v,w) = \lim_{n\to\infty} n^{-1} \int_\Omega \int_0^n \int_0^s v \circ \varphi_r w \circ \varphi_s dr ds d\mu,$$

exists for all $v, w \in C_0^\kappa(\Omega)$ and the resulting bilinear operator $\mathfrak{B} : C_0^\kappa(\Omega) \times C_0^\kappa(\Omega) \to \mathbb{R}$ is bounded and positive semidefinite.

(ii) The drift and diffusion coefficients given by

$$\tilde{a}^i(x) = \int a^i(x,y)d\mu(y) + \sum_{k=1}^d \mathfrak{B}(b^k(x,\cdot), \partial_k b^i(x,\cdot)) \quad , \quad i = 1,\ldots,d\ ,$$

$$(\sigma(x)\sigma^T(x))^{ij} = \mathfrak{B}(b^i(x,\cdot), b^j(x,\cdot)) + \mathfrak{B}(b^j(x,\cdot), b^i(x,\cdot)) \quad , \quad i,j = 1,\ldots,d\ ,$$

are Lipschitz.

(iii) The family of solutions x_ε to the ODE (1.1) *converges weakly in the supnorm topology to the unique solution X of the SDE*

$$dX = \tilde{a}(X)dt + \sigma(X)dB \quad , \quad X(0) = \xi \tag{1.5}$$

where B is a standard Brownian motion in $\mathbb{R}^d$.

(iv) Let $v, w \in C_0^\kappa(\Omega)$. If in addition the integral $\int_0^\infty \int_\Omega v\, w \circ \phi_t d\mu dt$ exists, then

$$\mathfrak{B}(v,w) = \int_0^\infty \int_\Omega v\, w \circ \phi_t d\mu dt\ .$$

By part (i), the $\mathfrak{B}$-terms in the formulas for $\tilde{a}$ and σ can be written as

$$\mathfrak{B}(b^k(x,\cdot),\partial_k b^i(x,\cdot)) = \lim_{n\to\infty} n^{-1} \int_\Omega \int_0^n \int_0^s b^k(x,\phi_r y)\partial_k b^i(x,\phi_s y)drdsd\mu(y)$$

and

$$\mathfrak{B}(b^i(x,\cdot),b^j(x,\cdot)) = \lim_{n\to\infty} n^{-1} \int_\Omega \int_0^n \int_0^s b^i(x,\phi_r y)b^j(x,\phi_s y)drdsd\mu(y)\ .$$

By part (iv), if the integrals $\int_0^\infty \int_\Omega b^i(x,y)b^j(x,\phi_t y)d\mu(y)dt$, $\int_0^\infty \int_\Omega b^k(x,y)\partial_k b^i(x,\phi_t y)d\mu(y)dt$ exist for all i,j,k and $x\in\mathbb{R}^d$, then the coefficients $\tilde{a}$ and σ are given by the formulas in (1.3), (1.4). In general, even for nonmixing flows ϕ_t, the bilinear operator $\mathfrak{B}$ can still be written down explicitly, in terms of the finer structure of the flow, see (2.1).

Remark 1.2. The Lipschitz statement in Theorem 1.1(ii) follows from boundedness of $\mathfrak{B}$ together with the regularity assumptions on a and b. A consequence of this is the uniqueness of the limiting diffusion X as stated in part (iii).

Remark 1.3. Since the expression defining $\sigma\sigma^T$ is symmetric and positive semidefinite, a square root σ always exists. Also, it is a standard result that the diffusion X is independent of the choice of any square root σ.

Remark 1.4. In the special case $b(x,y) = h(x)v(y)$ considered in [KM16], $\mathfrak{B}(v^i,v^j)$ is defined through its symmetric part, denoted $\frac{1}{2}\Sigma^{ij}$ and its anti-symmetric part, denoted $\frac{1}{2}D^{ij}$. One easily recovers the above via the Itô-Stratonovich correction.

1.2 Non-uniformly hyperbolic flows

For the sake of exposition, we have stated the homogenization results for fast flows that are uniformly hyperbolic. In reality, the results apply much more generally. The convergence result stated in Theorem 1.1(iii) can be recast in an abstract framework as explained in Section 2. In brief, we only require that ϕ_t satisfies an iterated central limit theorem (CLT) along with a moment control estimate. As shown in [KM16], the assumptions in Section 2 hold true for broad classes of flows which have a Poincaré map modelled by a Young tower [You98, You99]. In the case of Young towers with exponential tails, or at least superpolynomial tails, Theorem 1.1 goes through unchanged. This includes the case of Hénon-like attractors and the classical Lorenz equations.

When the Poincaré map is modelled by a Young tower with tails that decay more slowly, the assumptions in Section 2 still hold provided the return time function for the Young tower lies in L^q for some $q > 11/2$. In this case, Assumption 2.2 below holds with $p = 2(q-1)/3 > 3$ by [KM16, Proposition 7.5] and the discussion in [KM16, Section 10]. It then follows from Theorem 2.3 below that Theorem 1.1 goes through under stronger regularity assumptions (depending on the value of p) on b.

1.3 Mixing hypotheses

An important feature of the program initiated by [MS11], and continued in [GM13, KM16] and the current paper, is that in our main results we make no mixing assumptions on the fast flow ϕ_t. (The only role of mixing is to obtain simplified formulas for the drift and diffusion coefficients.)

To understand the significance of the lack of mixing assumptions, it is convenient to recall part of the history of the study of mixing for Axiom A dynamical systems. The discrete time case, namely Axiom A diffeomorphisms, is very well-understood: any hyperbolic basic set for an Axiom A diffeomorphism is mixing up to a finite cycle, and in the mixing case Hölder observables enjoy exponential decay of correlations [Bow75, Rue78, Sin72]. For Axiom A flows it is trivial to construct examples with no mixing properties: constant suspensions (more generally, suspensions where the roof function is cohomologous to a constant), do not mix. However, [BR75] asked whether mixing Axiom A flows are automatically exponentially mixing. This turned out to be false: [Pol84, Rue83] independently gave examples of mixing Axiom A flows that do not mix exponentially quickly and [Pol85] showed that Axiom A flows can mix arbitrarily slowly. Eventually [Dol98] gave examples of Anosov flows that mix exponentially quickly, but at the time of writing there are no known examples of robustly exponentially mixing Anosov flows.

On the other hand, it is well-known since [Rat73] that mixing is irrelevant for certain statistical properties. In particular, the central limit theorem holds for *all* Axiom A flows, even the nonmixing ones [MT04]. This theme has been extended significantly over the years; in particular [KM16] verified that the statistical properties required to apply rough path theory, namely the iterated CLT and moment estimates, hold for all Axiom A flows and for the nonuniformly hyperbolic flows described in Subsection 1.2. As a consequence the homogenization results in [KM16] and the current paper are independent of mixing properties of the fast flow.

For the specific classes of examples mentioned in this paper, the fast flow has a Poincaré map with good mixing properties, and this happens to be useful for establishing the iterated CLT and moment estimates at the level of the Poincaré map; these properties then lift to the flow. But this should not be confused with assuming mixing for the flow itself. Mixing remains poorly understood for flows, and moreover there are examples where mixing fails or happens arbitrarily slowly. Our results on homogenization hold regardless of these issues.

1.4 Previous results

It is only fairly recently that results on homogenization have been obtained in a fully deterministic setting with realistic assumptions on the fast dynamics. The first such results were obtained by [Dol04, Dol05] for discrete time systems where the fast dynamics is uniformly or partially hyperbolic with sufficiently fast decay of correlations. As explained in Subsection 1.3, a program to remove mixing assumptions on the fast dynamics was initiated in [MS11] where the authors prove a result on homogenization for general fast flows that are uniformly or nonuniformly hyperbolic, but under the assumption that the noise appears additively in the slow ODE, that is $b(x, y) = h(y)$. This was extended to the case of multiplicative noise $b(x, y) = h(x)v(y)$ in the scalar case $d = 1$ by [GM13] who also treated the discrete time situation. The case $b(x, y) = h(x)v(y)$ was treated in general dimensions in [KM16] (again for both discrete and continuous time). We remark that the results of the current article should carry over to the discrete time setting, but this requires additional work to incorporate the discrete time rough path theory introduced in [Kel16].

Homogenization results for chaotic systems have many interesting physical applications, most notably in stochastic climate models [MTVE01]. For more examples, see [PS08, Section 11.8].

1.5 Outline of the article

To prove Theorem 1.1 (or more precisely Theorem 2.3, the abstract version) we reformulate the slow equation (1.1) as an ODE of the form

$$dx_\varepsilon = F(x_\varepsilon)dV_\varepsilon + H(x_\varepsilon)dW_\varepsilon \ ,$$

where V_ε and W_ε are function space valued paths that are smooth (in time) for each fixed ε. The path V_ε is a smooth approximation of a function space valued drift and the path W_ε is a smooth approximation of a function space valued Brownian motion. To be precise, we take

$$V_\varepsilon(t) = \int_0^t a(\cdot, y_\varepsilon(r))dr \quad \text{and} \quad W_\varepsilon(t) = \varepsilon^{-1}\int_0^t b(\cdot, y_\varepsilon(r))dr \ .$$

The operators $F(x), H(x)$ are Dirac distributions (evaluation maps) located at x, that is $F(x)\varphi = \varphi(x)$ for any φ in the function space and similarly for H.

Remark 1.5. Note that although F, H are both Dirac distributions, they will act on different domains, hence the different labels.

One should think of the pair $(V_\varepsilon, W_\varepsilon)$ as "noise" driving the solution x_ε. Using the theory of rough paths, we build a continuous solution map from the "noise space" into the "solution space". The "noise space" will contain not just smooth paths, but also paths of Brownian regularity (which is the type of regularity we expect from the limiting W_ε). Since the solution map is continuous, a weak convergence result for the noise processes can be lifted to a weak convergence result for the solution, via the continuous mapping theorem.

The outline of the article is as follows. In Section 2 we write the abstract formulation of Theorem 1.1; this constitutes the main result of the article. In Section 3 we give an overview of rough path theory and state the tools that will be used. In Sections 4, 5 and 6 we state and prove a localized version of the main result. In Section 7 we lift the localized result to the full result.

1.6 Notation

We write $\mathbf{E}_\mu$ for expectation with respect to μ and write $\mathbf{E}$ when referring to expectation on a generic probability space. We write for example $a \in C^{1+}$ if there exists $\alpha > 1$ such that $a \in C^\alpha$. For a normed linear space $\mathcal{B}$ we write $L(\mathcal{B}, \mathbb{R})$ for the space of bounded linear functionals on $\mathcal{B}$, with the usual norm $\|f\|_{L(\mathcal{B},\mathbb{R})} = \sup_{\|x\|_{\mathcal{B}}=1}|f(x)|$. We write $a_n \lesssim b_n$ as $n \to \infty$ if there is a constant $C > 0$ such that $a_n \leq Cb_n$ for all $n \geq 1$.

2 The abstract convergence result

We now state an abstract version of Theorem 1.1. Let $\phi_t : M \to M$ be a smooth flow on a finite dimensional manifold and suppose that $\Omega \subset M$ is a closed flow-invariant set with ergodic probability measure μ. For $v \in L^1(\Omega, \mathbb{R}^m)$ with $\int_\Omega v d\mu = 0$, we define

$$W_{v,n}(t) = n^{-1/2}\int_0^{tn} v \circ \phi_s \, ds \quad \text{and} \quad \mathbb{W}_{v,n}(t) = n^{-1}\int_0^{tn}\int_0^s v \circ \phi_r \otimes v \circ \phi_s dr ds \ .$$

By definition of the tensor product for vectors, $\mathbb{W}_{v,n}$ takes values in $\mathbb{R}^{m\times m}$. Recall that, due to the dependence on the (omitted) initial condition of the flow, $W_{v,n}(t)$ and $\mathbb{W}_{v,n}(t)$ are random variables on the probability space (Ω,μ).

For $v,w\in L^1(\Omega,\mathbb{R})$, we define

$$v_t=\int_0^t v\circ\phi_s ds \quad \text{and} \quad S_t=\int_0^t\int_0^s v\circ\phi_r w\circ\phi_s dr ds\,.$$

Fix $\kappa>0$. The abstract assumptions are as follows.

Assumption 2.1. There exists a bilinear operator $\mathfrak{B}:C_0^\kappa(\Omega)\times C_0^\kappa(\Omega)\to\mathbb{R}$ such that for every $m\geq 1$ and every $v\in C_0^\kappa(\Omega,\mathbb{R}^m)$,

$$(W_{v,n},\mathbb{W}_{v,n})\to(W_v,\mathbb{W}_v)$$

as $n\to\infty$, in the sense of finite dimensional distributions, where W_v is a Brownian motion in $\mathbb{R}^m$ and $\mathbb{W}_v$ is the process with values in $\mathbb{R}^{m\times m}$ defined by

$$\mathbb{W}_v^{ij}(t)=\int_0^t W_v^i dW_v^j+\mathfrak{B}(v^i,v^j)t\,.$$

(Here, the integral is of Itô type.)

Assumption 2.2. There exists $p>3$, and for all $v,w\in C_0^\kappa(\Omega)$ there exists $K=K_{v,w,p}>0$ such that

$$(\mathbf{E}_\mu|v_t|^{2p})^{1/(2p)}\leq Kt^{1/2} \quad \text{and} \quad (\mathbf{E}_\mu|S_t|^p)^{1/p}\leq Kt$$

for all $t\geq 0$. If the estimates hold for all $p>3$ then we say the estimates hold for $p=\infty$.

Theorem 2.3. *Suppose that Assumptions 2.1 and 2.2 hold with some $p\in(3,\infty]$ and $\kappa>0$. Suppose that $a\in C^{1+,0}(\mathbb{R}^d\times M,\mathbb{R}^d)$ and $b\in C_0^{\alpha,\kappa}(\mathbb{R}^d\times M,\mathbb{R}^d)$ for some $\alpha>2+\frac{2}{p-1}+\frac{d}{p}$. Then we have the same conclusion as Theorem 1.1(i,ii,iii).*

We now show how Theorem 1.1 follows from Theorem 2.3.

Proof of Theorem 1.1. Assumptions 2.1 and 2.2 (with $p=\infty$) are valid for hyperbolic basic sets by [KM16, Theorem 1.1] and [KM16, Proposition 7.5, Remark 7.7] respectively. Hence Theorem 1.1(i,ii,iii) follows from Theorem 2.3. Moreover, Theorem 1.1(iv) follows from [KM16, Theorem 1.1(b)]. □

Remark 2.4. In [KM16], we considered the special case where $b(x,y)=h(x)v(y)$ is a product (for some $v:M\to\mathbb{R}^e$ and $h:\mathbb{R}^d\to\mathbb{R}^{d\times e}$) under less stringent regularity conditions on b. It is easy to check that when b is a product, the method in this paper applies provided $b\in C^{\alpha,\kappa}$ for some $\alpha>2+2/(p-1)$ recovering the results of [KM16]. The only place where the additional regularity is required for general b is in the tightness estimates in Section 5 below.

Remark 2.5. A general formula for the bilinear operator $\mathfrak{B}$ in the case of (not necessarily mixing) Axiom A flows can be obtained by considering the associated suspension flow. The same is true for nonuniformly hyperbolic flows modelled as a suspension flow over a Young tower. We recall the basic definitions; further details can be found in [KM16] and references therein.

Suppose that $f : \Lambda \to \Lambda$ is a map with ergodic invariant probability measure μ. Let $r : \Lambda \to \mathbb{R}^+$ be an integrable roof function with $\bar{r} = \int_\Lambda r\, d\mu$. Define the suspension $\Lambda^r = \{(x,u) \in \Lambda \times \mathbb{R} : 0 \le u \le r(x)\}/\sim$ where $(x, r(x)) \sim (fx, 0)$. Define the suspension flow $\phi_t(x,u) = (x, u+t)$ computed modulo identifications. The measure $\mu^r_\Lambda = \mu_\Lambda \times \mathrm{Lebesgue}/\bar{r}$ is an ergodic invariant probability measure for ϕ_t.

Our results apply in particular to the case where Λ is a Young tower with return time function lying in L^q for some $q > 11/2$, and r is a bounded piecewise Hölder roof function. Since Young towers are mixing up to a finite cycle, by taking a smaller Poincaré cross-section we may suppose without loss of generality that Λ is mixing. Given $v \in C^\kappa_0(\Omega, \mathbb{R}^m)$, we define the induced observable $\tilde{v} \in L^\infty(\Lambda, \mu_\Lambda)$ by setting $\tilde{v} = \int_0^r v \circ \phi_t dt$. Similarly, we associate $\tilde{w}$ to w. By [You98, You99], $\tilde{v}$ and $\tilde{w}$ have summable decay of correlations (exponential decay when Λ is Axiom A or a Young tower with exponential tails, and decay rate $O(1/n^{q-1})$ in general), so in particular the series $\sum_{n=1}^\infty \int_\Lambda \tilde{v}\, \tilde{w} \circ f^n\, d\mu_\Lambda$ is absolutely convergent.

Moreover, as shown in [KM16, Corollary 8.1],

$$\mathfrak{B}(v,w) = (\bar{r})^{-1} \sum_{n=1}^{\infty} \int_\Lambda \tilde{v}\, \tilde{w} \circ f^n\, d\mu_\Lambda + (\bar{r})^{-1} \int_\Lambda S(v,w) d\mu_\Lambda \;, \tag{2.1}$$

where

$$S(v,w)(y) = \int_0^{r(y)} \Big(\int_0^s v(\phi_u y) du \Big) w(\phi_s y) ds \;.$$

is the iterated integral of the path $(v \circ \phi_t, w \circ \phi_t)$ along the orbit until its return to Λ.

Remark 2.6. There is a slightly simpler way of writing $\mathfrak{B}$ which gives a more geometric description of the bilinear form. We introduce the symmetric and anti-symmetric parts

$$\mathfrak{A}(v,w) = \frac{1}{2}\Big(\mathfrak{B}(v,w) + \mathfrak{B}(w,v)\Big) \quad \text{and} \quad \mathfrak{D}(v,w) = \frac{1}{2}\Big(\mathfrak{B}(v,w) - \mathfrak{B}(w,v)\Big) \;.$$

For the symmetric part, it follows from the product rule that

$$S(v,w) + S(w,v) = \Big(\int_0^r v \circ \phi_t dt \Big)\Big(\int_0^r w \circ \phi_t dt \Big) = \tilde{v}\tilde{w}$$

and hence

$$\mathfrak{A}(v,w) = \sum_{n=1}^{\infty} \frac{1}{2\bar{r}} \int_\Lambda \big(\tilde{v}\, \tilde{w} \circ f^n + \tilde{w}\, \tilde{v} \circ f^n\big) d\mu_\Lambda + \frac{1}{2\bar{r}} \int_\Lambda \tilde{v}\tilde{w} d\mu_\Lambda \;.$$

In particular the symmetric part of the bilinear form is completely determined by the cross correlations between induced observables. Similarly

$$\mathfrak{D}(v,w) = \sum_{n=1}^{\infty} \frac{1}{2\bar{r}} \int_\Lambda \big(\tilde{v}\, \tilde{w} \circ f^n - \tilde{w}\, \tilde{v} \circ f^n\big) d\mu_\Lambda + \frac{1}{2\bar{r}} \int_\Lambda \big(S(v,w) - S(w,v)\big) d\mu_\Lambda \;.$$

The advantage here is that the expression

$$\frac{1}{2}\big(S(v,w)(y) - S(w,v)(y)\big)$$

is equal (by Green's theorem) to the *signed area* traced out in $\mathbb{R}^2$ by the loop $(v(\phi_t y), w(\phi_t y))_{t=0}^{r(y)}$ (closed by the secant joining the endpoints).

In the remainder of this section, we describe some elementary properties that follow immediately from the assumptions on the fast dynamics. Firstly, we show that in Assumption 2.2 the constant K can be chosen uniformly in v, w. Define the incremental objects

$$v_{s,t} = \int_s^t v \circ \phi_r dr \quad \text{and} \quad S_{s,t} = \int_s^t \int_s^r v \circ \phi_u w \circ \phi_r du dr \ .$$

Proposition 2.7. *If the fast flow satisfies Assumption 2.2, then*

$$(\mathbf{E}_\mu |v_{s,t}|^{2p})^{1/(2p)} \lesssim \|v\|_{C^\kappa} |t-s|^{1/2} \quad \textit{and} \quad (\mathbf{E}_\mu |S_{s,t}|^p)^{1/p} \lesssim \|v\|_{C^\kappa} \|w\|_{C^\kappa} |t-s|$$

for all $s,t \geq 0$, $v,w \in C_0^\kappa(\Omega)$.

Proof. By stationarity it suffices to check the claim with $s = 0$. Consider the family of linear operators $\{L_t : C_0^\kappa(\Omega) \to L_{2p}(\Omega),\ t > 0\}$ given by $L_t v = t^{-1/2} v_t$. Since $\|L_t v\|_\infty \leq t^{1/2} \|v\|_\infty$ it is certainly the case that $L_t : C_0^\kappa(\Omega) \to L_{2p}(\Omega)$ is bounded for each t. By Assumption 2.2, for each $v \in C_0^\kappa(\Omega)$, there exists a constant $K = K_v$ such that $\|L_t v\|_{2p} \leq K_v$ for all $t > 0$. By the uniform boundedness principle, there is a uniform constant K such that $\|L_t v\|_{2p} \leq K \|v\|_{C^\kappa}$ for all $v \in C_0^\kappa(\Omega)$, $t > 0$. This establishes the desired estimate for v_t.

The estimate for S_t is proved similarly by considering the family of bilinear operators $\{B_t : C_0^\kappa(\Omega) \times C_0^\kappa(\Omega) \to L^p(\Omega),\ t > 0\}$ given by $B_t(v,w) = t^{-1} S_t$. $\square$

The next result is a collection of simple facts that will be used throughout the rest of the article.

Proposition 2.8. *If the fast flow satisfies Assumptions 2.1 and 2.2, then*

- *(a) The covariance of* W_v *is given by* $\mathbf{E} W_v^i(1) W_v^j(1) = \mathfrak{B}(v^i, v^j) + \mathfrak{B}(v^j, v^i)$ *for all* $v \in C_0^\kappa(\Omega, \mathbb{R}^m)$.
- *(b)* $\mathfrak{B}(v,v) \geq 0$ *for all* $v \in C_0^\kappa(\Omega)$.
- *(c)* $|\mathfrak{B}(v,w)| \lesssim \|v\|_{C^\kappa} \|w\|_{C^\kappa}$ *for all* $v,w \in C_0^\kappa(\Omega)$.
- *(d)* $(W_{v,n}, \mathbb{W}_{v,n}) \to_w (W_v, \mathbb{W}_v)$ *as* $n \to \infty$ *in the supnorm topology, for all* $v \in C_0^\kappa(\Omega, \mathbb{R}^m)$.

Proof. (a) It follows from Assumptions 2.1 and 2.2 that

$$\mathbf{E}_\mu W_{v,n}^i(1) W_{v,n}^j(1) \to \mathbf{E} W_v^i(1) W_v^j(1), \quad \mathbf{E}_\mu \mathbb{W}_{v,n}^{ij}(1) \to \mathbf{E} \mathbb{W}_v^{ij}(1) = \mathfrak{B}(v^i, v^j),$$

where we have used the fact that Itô integrals have zero mean. Taking expectations on both sides of the identity

$$W_{v,n}^i(1) W_{v,n}^j(1) = \mathbb{W}_{v,n}^{ij}(1) + \mathbb{W}_{v,n}^{ji}(1)$$

and letting $n \to \infty$ yields the desired result.
(b) It follows from part (a) that $\mathfrak{B}(v^i, v^i) = \frac{1}{2}\mathbf{E}W_v^i(1)^2 \geq 0$.
(c) Define S_t^{ij} using the definition of S_t but with $v = v^i$ and $w = v^j$. We note that $n^{-1}S_n^{ij} = \mathbb{W}_{v,n}^{ij}(1)$ and hence by Proposition 2.7, $\mathbf{E}_\mu|\mathbb{W}_{v,n}^{ij}(1)| \lesssim \|v^i\|_{C^\kappa}\|v^j\|_{C^\kappa}$. By Assumptions 2.1 and 2.2, $|\mathfrak{B}(v^i, v^j)| = \lim_{n\to\infty}|\mathbf{E}_\mu \mathbb{W}_{v,n}^{ij}(1)| \lesssim \|v^i\|_{C^\kappa}\|v^j\|_{C^\kappa}$.
(d) Since the limiting random variable is (almost surely) continuous, it is sufficient to prove the weak convergence result in the Skorokhod topology. But this is a simple consequence of [Bil99, Theorem 13.5], combined with the Assumptions 2.1 and 2.2. $\square$

Finally, we show that convergence as $n \to \infty$ of the sequence of processes $(W_{v,n}, \mathbb{W}_{v,n})$ implies convergence as $\varepsilon \to 0$ of the family of processes

$$W_v^{(\varepsilon)}(t) = \varepsilon \int_0^{t\varepsilon^{-2}} v \circ \phi_s \, ds, \quad \mathbb{W}_v^{(\varepsilon)}(t) = \varepsilon^2 \int_0^{t\varepsilon^{-2}} \int_0^s v \circ \phi_r \otimes v \circ \phi_s \, dr \, ds, \ \varepsilon > 0.$$

Before doing so, we need the following elementary lemma.

Lemma 2.9. *Suppose that $a : \mathbb{R} \to \mathbb{R}$ is bounded on compact sets. Let $b > 0$, $T \geq 0$. If $\lim_{\varepsilon\to 0} \varepsilon^b a(\varepsilon^{-1}) = 0$, then $\lim_{\varepsilon\to 0} \varepsilon^b \sup_{t\in[0,T]} |a(t\varepsilon^{-1})| = 0$.*

Proof. The proof is standard and included just for completeness.

Fix $\delta > 0$. Choose $\varepsilon_0 > 0$ such that $\varepsilon^b a(\varepsilon^{-1}) < \delta/T^b$ for $\varepsilon < \varepsilon_0$. Now choose $\varepsilon_1 > 0$ such that $\varepsilon_1^b \sup_{t\leq\varepsilon_0^{-1}} |a(t)| < \delta$.

We show that $\sup_{t\in[0,T]} |\varepsilon^b a(t\varepsilon^{-1})| < \delta$ for all $\varepsilon < \varepsilon_1$. There are two cases. If $\varepsilon/t \geq \varepsilon_0$, then $|\varepsilon^b a(t\varepsilon^{-1})| \leq \varepsilon_1^b \sup_{t\leq\varepsilon_0^{-1}} |a(t)| < \delta$. If $\varepsilon/t < \varepsilon_0$, then $|\varepsilon^b a(t\varepsilon^{-1})| \leq T^b(\varepsilon/t)^b|a(t\varepsilon^{-1})| < \delta$. $\square$

Proposition 2.10. *If Assumption 2.1 holds, then $(W_v^{(\varepsilon)}, \mathbb{W}_v^{(\varepsilon)}) \to_w (W_v, \mathbb{W}_v)$ as $\varepsilon \to 0$ in the supnorm topology, for all $v \in C_0^\kappa(\Omega, \mathbb{R}^m)$.*

Proof. Let $n = [\varepsilon^{-2}]$. We have

$$W_v^{(\varepsilon)}(t) = \varepsilon n^{1/2} W_{v,n}(t) + \varepsilon \int_{tn}^{t\varepsilon^{-2}} v \circ \phi_s \, ds.$$

As $\varepsilon \to 0$, $\varepsilon n^{1/2} \to 1$. Also, $\| \int_{tn}^{t\varepsilon^{-2}} v \circ \phi_s \, ds\|_\infty \leq t\, \|v\|_\infty$ and hence $W_v^{(\varepsilon)} - W_{v,n} \to_w 0$.

Similarly,

$$\mathbb{W}_v^{(\varepsilon)}(t) = \varepsilon^2 n \mathbb{W}_{v,n}(t) + \varepsilon^2 \int_{tn}^{t\varepsilon^{-2}} \int_0^s v \circ \phi_r \otimes v \circ \phi_s \, dr \, ds = \varepsilon^2 n \mathbb{W}_{v,n}(t) + \varepsilon^2 A^{(\varepsilon)}(t),$$

where

$$A^{(\varepsilon)}(t) = \int_{tn}^{t\varepsilon^{-2}} \int_{tn}^s v \circ \phi_r \otimes v \circ \phi_s \, dr \, ds + \int_{tn}^{t\varepsilon^{-2}} \int_0^{tn} v \circ \phi_r \otimes v \circ \phi_s \, dr \, ds.$$

Now $|A^{(\varepsilon)}(t)| \leq t^2|v|_\infty^2 + t|v|_\infty|v_{tn}|$. By the ergodic theorem, $\varepsilon^2 v_{\varepsilon^{-2}} \to 0$ almost everywhere, and hence by Lemma 2.9 $\sup_{t\in[0,T]} \varepsilon^2|v_{t\varepsilon^{-2}}| \to 0$ almost everywhere. It follows that $\sup_{t\in[0,T]} \varepsilon^2|A^{(\varepsilon)}(t)| \to 0$ almost everywhere, and so $\mathbb{W}_v^{(\varepsilon)} - \mathbb{W}_{v,n} \to_w 0$.

Altogether, we obtain that $(W_v^{(\varepsilon)}, \mathbb{W}_v^{(\varepsilon)}) - (W_{v,n}, \mathbb{W}_{v,n}) \to_w 0$ as required. $\square$

3 Some rough path theory

In this section we formalize some of the ideas from rough path theory put forward in Section 1: namely, that one can build a continuous map from "noise space" to "solution space". This map is constructed using rough path theory [Lyo98]. The formulation of rough path theory that we employ closely follows the recent book [FH14]. Before going into the theory, we list some preliminary facts concerning tensor products of Banach spaces.

3.1 Tensor products of Banach spaces

Let $\mathcal{A}, \mathcal{B}$ be Banach spaces (over $\mathbb{R}$). The *algebraic* tensor product space $\mathcal{A} \otimes_a \mathcal{B}$ is defined as the vector space

$$\mathcal{A} \otimes_a \mathcal{B} = \operatorname{span}\{x \otimes y \mid x \in \mathcal{A}\ ,\ y \in \mathcal{B}\}\ .$$

That is, $\mathcal{A} \otimes_a \mathcal{B}$ is the space of finite sums $\sum_n x_n \otimes y_n$ for $x_n \in \mathcal{A}$, $y_n \in \mathcal{B}$. For $f \in L(\mathcal{A}, \mathbb{R})$, $g \in L(\mathcal{B}, \mathbb{R})$ we define a linear functional $f \otimes g : \mathcal{A} \otimes_a \mathcal{B} \to \mathbb{R}$ by

$$(f \otimes g) \sum_n x_n \otimes y_n = \sum_n f(x_n) g(y_n)\ . \tag{3.1}$$

A norm $\|\cdot\|_{\mathcal{A}\otimes\mathcal{B}} : \mathcal{A} \otimes_a \mathcal{B} \to \mathbb{R}_+$ is called *admissible* if

$$\|x \otimes y\|_{\mathcal{A}\otimes\mathcal{B}} = \|x\|_{\mathcal{A}} \|y\|_{\mathcal{B}} \quad \text{and} \quad \|f \otimes g\|_{L(\mathcal{A}\otimes_a\mathcal{B},\mathbb{R})} = \|f\|_{L(\mathcal{A},\mathbb{R})} \|g\|_{L(\mathcal{B},\mathbb{R})} \tag{3.2}$$

for all $x \in \mathcal{A}, y \in \mathcal{B}$ and all $f \in L(\mathcal{A}, \mathbb{R})$ and $g \in L(\mathcal{B}, \mathbb{R})$. By [LC85, Lemma 1.4], to check admissibility it is sufficient to check (3.2) with $=$ replaced by $\leq$.

For an admissible norm $\|\cdot\|_{\mathcal{A}\otimes\mathcal{B}}$ we define the tensor product space $\mathcal{A} \otimes \mathcal{B}$ as the completion of $\mathcal{A} \otimes_a \mathcal{B}$ under the norm $\|\cdot\|_{\mathcal{A}\otimes\mathcal{B}}$. Hence $(\mathcal{A} \otimes \mathcal{B}, \|\cdot\|_{\mathcal{A}\otimes\mathcal{B}})$ is a Banach space. All tensor products we consider will be constructed using an admissible norm.

The admissibility requirement guarantees that $f \otimes g \in L(\mathcal{A} \otimes_a \mathcal{B}, \mathbb{R})$ and since $\mathcal{A} \otimes_a \mathcal{B}$ is (by definition) dense in $\mathcal{A} \otimes \mathcal{B}$, $f \otimes g$ extends uniquely to an element of $L(\mathcal{A} \otimes \mathcal{B}, \mathbb{R})$.

3.2 Spaces of rough paths

In this subsection, we show how to build a "noise space" of Banach space valued paths as mentioned in Section 1.5. Recall that this should include smooth paths and also Brownian-like paths. It turns out that it is necessary also to add extra structure to the set of paths. The resulting space is called the space of *rough paths*.

Let $\mathcal{A}$ be a Banach space. For $\beta \in (\frac{1}{2}, 1)$, we define $\mathscr{C}^\beta = \mathscr{C}^\beta(\mathcal{A})$ to be the set of all continuous paths $V : [0, T] \to \mathcal{A}$ with $V(0) = 0$ and

$$|V|_{C^\beta} = \sup_{s,t} \frac{|V(s,t)|}{|t-s|^\beta} < \infty\ ,$$

where $V(s,t) = V(t) - V(s)$. The pair $(\mathscr{C}^\beta, |\cdot|_{C^\beta})$ is a Banach space.

Let $\mathcal{B}$ be a Banach space with tensor product space $\mathcal{B} \otimes \mathcal{B}$. For $\gamma \in (\frac{1}{3}, \frac{1}{2}]$, the space $\mathscr{C}^\gamma = \mathscr{C}^\gamma(\mathcal{B})$ is defined to be the set of all continuous paths $(W, \mathbb{W}) : [0, T] \to \mathcal{B} \times (\mathcal{B} \otimes \mathcal{B})$ with $(W(0), \mathbb{W}(0)) = 0$ and such that

$$\sup_{s,t} \frac{\|W(s,t)\|_{\mathcal{B}}}{|t-s|^\gamma} < \infty \quad \text{and} \quad \sup_{s,t} \frac{\|\mathbb{W}(s,t)\|_{\mathcal{B}\otimes\mathcal{B}}}{|t-s|^{2\gamma}} < \infty\ ,$$

where $W(s,t) = W(t) - W(s)$ and $\mathbb{W}(s,t) = \mathbb{W}(t) - \mathbb{W}(s) - W(s) \otimes W(s,t)$. The set $\mathscr{C}^\gamma$ is known as the set of γ-rough paths and forms a complete metric space under the metric

$$\rho_\gamma((W_1, \mathbb{W}_1), (W_2, \mathbb{W}_2)) = \sup_{s,t} \frac{\|W_1(s,t) - W_2(s,t)\|_{\mathcal{B}}}{|t-s|^\gamma} + \sup_{s,t} \frac{\|\mathbb{W}_1(s,t) - \mathbb{W}_2(s,t)\|_{\mathcal{B}\otimes\mathcal{B}}}{|t-s|^{2\gamma}} .$$

We also make use of the norm-like object

$$|\!|\!|(W, \mathbb{W})|\!|\!|_{\mathscr{C}^\gamma} = \sup_{s,t} \frac{\|W(s,t)\|_{\mathcal{B}}}{|t-s|^\gamma} + \sup_{s,t} \frac{\|\mathbb{W}(s,t)\|^{1/2}_{\mathcal{B}\otimes\mathcal{B}}}{|t-s|^\gamma} ,$$

which shows up in some estimates, but does not play any role in defining the topology.

Finally, we define the set of (β,γ)-rough paths $\mathscr{C}^{\beta,\gamma} = \mathscr{C}^\beta(\mathcal{A}) \times \mathscr{C}^\gamma(\mathcal{B})$; this is a complete metric space with the product metric.

Remark 3.1. One should think of $\mathscr{C}^{\beta,\gamma}$ as the "noise space". This space clearly contains irregular Brownian paths, in addition to smooth paths. The pair $\mathbf{W} = (W, \mathbb{W})$, when combined with the rough path topology, is what we mean by "extra structure". We view $\mathbb{W}(t)$ as a candidate for the integral $\int_0^t W \otimes dW$ and $\mathbb{W}(s,t)$ as a candidate for $\int_s^t W(s,r) \otimes dW(r)$. Note that since W is only Hölder continuous, there may be many candidates for the integral $\mathbb{W}$; hence it must be specified.

Next, we define a subspace $\mathscr{C}_g^\gamma \subset \mathscr{C}^\gamma$ known as the geometric rough paths. Let $W : [0,T] \to \mathcal{B}$ be a smooth (piecewise C^1) path and let $\mathbb{W} : [0,T] \to \mathcal{B} \otimes \mathcal{B}$ be the path of Riemann integrals

$$\mathbb{W}(t) = \int_0^t W \otimes dW = \int_0^t W \otimes \dot{W}\, dt . \tag{3.3}$$

The γ-geometric rough paths $\mathscr{C}_g^\gamma$ are defined as the closure of the set of all such smooth pairs $(W, \mathbb{W})$ in $\mathscr{C}^\gamma$.

Remark 3.2. The smoothness of W combined with the admissibility of the tensor product space ensure that $t \mapsto W(t) \otimes \dot{W}(t)$ is a piecewise continuous map and hence Riemann integrable.

3.3 Rough differential equations

Suppose that V, W are smooth and that $F : \mathbb{R}^d \to L(\mathcal{A}, \mathbb{R}^d)$, $H : \mathbb{R}^d \to L(\mathcal{B}, \mathbb{R}^d)$. Under suitable regularity assumptions on F, H, the ODE

$$X(t) = \xi + \int_0^t F(X) dV + \int_0^t H(X) dW$$

has a unique solution X. We call the map $\Phi : (V, W) \mapsto X$ the solution map. In this subsection, we show how the map Φ extends to the space of rough paths $\mathscr{C}^{\beta,\gamma}$.

For the moment, we suppose that F is C^1 and H is C^2. Recall that $\beta > \frac{1}{2}$ and $\gamma > \frac{1}{3}$ and suppose in addition that $\beta + \gamma > 1$. For $(V, W, \mathbb{W}) \in \mathscr{C}^{\beta,\gamma}$ there is a class of paths $X : [0,T] \to \mathbb{R}^d$ known as *controlled rough paths* for which one can define the integrals

$$\int_0^t F(X) dV \quad \text{and} \quad \int_0^t H(X) d\mathbf{W} ,$$

with the shorthand $\mathbf{W} = (W, \mathbb{W})$. We call X a controlled rough path if $X(s,t) = X(t) - X(s)$ has the form

$$X^i(s,t) = X_i'(s)W(s,t) + O(|t-s|^{2\gamma})$$

for all $i = 1, \ldots, d$ and $0 \le s \le t \le T$, where $X_i' \in C^\gamma([0,T], L(\mathcal{B}, \mathbb{R}))$. For a thorough treatment of controlled rough paths and their use in defining the above integrals, see [FH14, Section 4].

Since $\beta + \gamma > 1$, the dV integral is well-defined as a Young integral [You36], namely

$$\int_0^t F(X)dV = \lim_{\Delta \to 0} \sum_{[t_n, t_{n+1}] \in \Delta} F(X(t_n))V(t_n, t_{n+1})$$

where $\Delta = \{[t_n, t_{n+1}] : 0 \le n \le N-1\}$ denotes partitions of $[0,t]$. The integral is defined pathwise, for each fixed V.

The $d\mathbf{W}$ integral is defined as a compensated Riemann sum

$$\int_0^t H(X)d\mathbf{W} = \lim_{\Delta \to 0} S_\Delta \tag{3.4}$$

where

$$S_\Delta^i = \sum_{[t_n, t_{n+1}] \in \Delta} H^i(X(t_n))W(t_n, t_{n+1}) + \sum_{k=1}^d \big(X_k'(t_n) \otimes \partial_k H^i(X(t_n))\big)\mathbb{W}(t_n, t_{n+1}) \tag{3.5}$$

with Δ as above. The dual tensor product $X_i'(t_n) \otimes \partial_k H^i(X(t_n))$ is defined as in (3.1). Note that the integral is defined pathwise, for each fixed path $(W, \mathbb{W})$. In the special case where W is a Brownian path and $\mathbb{W}$ is the iterated Itô integral, $d\mathbf{W}$ becomes Itô integration.

A controlled rough path X is said to solve the RDE $dX = F(X)dV + H(X)d\mathbf{W}$ with initial condition $X(0) = \xi$ if it solves the integral equation

$$X(t) = \xi + \int_0^t F(X)dV + \int_0^t H(X)d\mathbf{W}\ ,$$

for all $t \in [0,T]$. For a thorough treatment of rough differential equations, see [FH14, Section 8]. In particular, we have the following basic result which includes existence, uniqueness and continuous dependence of solutions to RDEs.

Theorem 3.3. *Let $\gamma \in (\frac{1}{3}, \frac{1}{2}]$ and $\gamma' \in (\frac{1}{3}, \gamma)$. Suppose that $F \in C^{1+\delta'}(\mathbb{R}^d, L(\mathcal{A}, \mathbb{R}^d))$, $H \in C^{\frac{1}{\gamma}+\delta}(\mathbb{R}^d, L(\mathcal{B}, \mathbb{R}^d))$, where $\delta, \delta' > 0$,. Then there exists $\beta_* = \beta_*(\gamma, \gamma', \delta') \in (\frac{1}{2}, 1)$ such that the solution map $\Phi : \mathscr{C}^{\beta,\gamma} \to C^{\gamma'}([0,T], \mathbb{R}^d)$ given by*

$$\Phi(V, W, \mathbb{W}) = X$$

is continuous for $\beta \in (\beta_*, 1)$,

The solution map Φ is a genuine extension of the classical solution map in the sense that, if V, W are smooth paths and $\mathbb{W}$ is the iterated integral above W (as in (3.3)) then $X = \Phi(V, W, \mathbb{W})$ agrees with the solution to the classical ODE $dX = F(X)dV + H(X)dW$ with the same initial condition.

Proof of Theorem 3.3. This is (a slight modification of) a standard result in rough path theory. Indeed when $V = 0$, it follows from [FH14, Theorem 8.5]. The extension to nontrivial V is a simple exercise in controlled rough paths.

To apply rough path theory in Banach spaces one typically assumes an embedding

$$L(\mathcal{B}, L(\mathcal{B}, \mathbb{R})) \hookrightarrow L(\mathcal{B} \otimes \mathcal{B}, \mathbb{R}) \ .$$

See for instance [FH14, Section 1.5]. We do not assume such an embedding. However, since we only interested in results concerning RDEs (and not general controlled rough paths) it is sufficient to assume the tensor product norm used to construct $\mathcal{B} \otimes \mathcal{B}$ is admissible. In particular, the only elements of $L(\mathcal{B}, L(\mathcal{B}, \mathbb{R}))$ required to satisfy the above embedding are of product form. That is, they are described by $(f, g)x = f(x)g$ for all $x \in \mathcal{B}$, with $f \in L(\mathcal{B}, \mathbb{R})$ and $g \in L(\mathcal{B}, \mathbb{R})$. Specifically, they are described by $(f, g) = (X'_k(t), \partial_k H(X(t)))$ where (X, X') is the controlled rough path candidate for the solution to the RDE. But clearly we can always perform such an embedding, by the identification $(f, g) \sim f \otimes g$ and by admissibility we have that $f \otimes g \in L(\mathcal{B} \otimes \mathcal{B}, \mathbb{R})$ as required. $\square$

In the remainder of the article we will use the following result which is an immediate consequence of Theorem 3.3.

Corollary 3.4. *Suppose that V_ε, W_ε are smooth paths, and that $\mathbb{W}_\varepsilon$ is the iterated integral of W_ε (as in* (3.3)*). Let $\gamma \in (\frac{1}{3}, \frac{1}{2}]$. Suppose that $F \in C^{1+}(\mathbb{R}^d, L(\mathcal{A}, \mathbb{R}^d))$ and $H \in C^{\frac{1}{\gamma}+}(\mathbb{R}^d, L(\mathcal{B}, \mathbb{R}^d))$, and that X_ε solves the ODE*

$$dX_\varepsilon = F(X_\varepsilon)dV_\varepsilon + H(X_\varepsilon)dW_\varepsilon \quad X_\varepsilon(0) = \xi \ . \tag{3.6}$$

If $(V_\varepsilon, W_\varepsilon, \mathbb{W}_\varepsilon) \to_w (V, W, \mathbb{W})$ in the $\mathscr{C}^{\beta,\gamma}$ topology for all $\beta \in (\frac{1}{2}, 1)$, then $X_\varepsilon \to_w X$ in the supnorm topology, where X solves the RDE

$$dX = F(X)dV + H(X)d\mathbf{W} \quad , \quad X(0) = \xi \ , \tag{3.7}$$

with $\mathbf{W} = (W, \mathbb{W})$.

Next, we list some properties of solutions to RDEs. Since these properties are completely standard, no proof will be given.

Proposition 3.5. *When X solves the RDE* (3.7) *we can always take $X'_k(\cdot) = H^k(X(\cdot))$ in the definition of the $d\mathbf{W}$ integral in* (3.4)*,* (3.5)*.*

Proposition 3.6. *Assume the set up of Theorem 3.3 and suppose moreover that $\mathbf{W} = (W, \mathbb{W}) \in \mathscr{C}_g^\gamma$. Then the classical chain rule*

$$\varphi(X(t)) = \varphi(X(s)) + \sum_{k=1}^{d} \int_s^t \partial_k\varphi(X)F^k(X)dV + \int_s^t \partial_k\varphi(X)H^k(X)d\mathbf{W} \ ,$$

is valid for any smooth $\varphi : \mathbb{R}^d \to \mathbb{R}$.

This result is an immediate consequence of the fact that the integrals are limits of smooth integrals, for which the chain rule holds. (The result fails for general rough paths $\mathbf{W} \in \mathscr{C}^\gamma$.)

The last result is an extension of the standard Kolmogorov continuity criterion to (smooth) rough paths, taking values in $\mathbb{R}$. A proof can be found in [Gub04, Corollary 4]. The one dimensional case turns out to be sufficient for our needs, even in the Banach space setting.

Lemma 3.7. *Let $T > 0$ and let $W_\varepsilon, \widetilde{W}_\varepsilon : [0,T] \to \mathbb{R}$ be smooth paths. Define $I_\varepsilon(s,t) = \int_s^t W_\varepsilon(s,r) d\widetilde{W}_\varepsilon(r)$. Let $p > 1$ and $\gamma \in (0, \frac{1}{2} - \frac{1}{2p})$, and suppose that M, $\tilde{M}$ are constants.*

(a) If $(\mathbf{E}|W_\varepsilon(s,t)|^{2p})^{1/(2p)} \le M|t-s|^{1/2}$ for all $\varepsilon > 0$, $s,t \in [0,T]$, then there is a constant C depending only on T, d, p, γ such that

$$\left(\mathbf{E}\left(\sup_{s,t\in[0,T]} \frac{|W_\varepsilon(s,t)|}{|t-s|^\gamma}\right)^{2p}\right)^{1/(2p)} \le CM \;, \quad \textit{for all } \varepsilon > 0 \;.$$

(b) If

$$(\mathbf{E}|W_\varepsilon(s,t)|^{2p})^{1/(2p)} \le M|t-s|^{1/2} \quad \textit{and} \quad (\mathbf{E}|\widetilde{W}_\varepsilon(s,t)|^{2p})^{1/(2p)} \le \tilde{M}|t-s|^{1/2}$$

and

$$(\mathbf{E}|I_\varepsilon(s,t)|^p)^{1/p} \le M\tilde{M}|t-s|$$

for all $\varepsilon > 0$, $s,t \in [0,T]$, then there is a constant C depending only on T, d, p, γ such that

$$\left(\mathbf{E}\left(\sup_{s,t\in[0,T]} \frac{|I_\varepsilon(s,t)|}{|t-s|^{2\gamma}}\right)^{p}\right)^{1/p} \le CM\tilde{M} \;, \quad \textit{for all } \varepsilon > 0 \;.$$

4 The localized convergence result

In this section, we state the *localized* version of Theorem 2.3.

Theorem 4.1. *Suppose that Assumptions 2.1 and 2.2 hold with some $p \in (3,\infty]$ and $\kappa > 0$. Suppose that $a \in C^{1+,0}(\mathbb{R}^d \times M, \mathbb{R}^d)$ and $b \in C_0^{\alpha,\kappa}(\mathbb{R}^d \times M, \mathbb{R}^d)$ for some $\alpha > 2 + \frac{2}{p-1} + \frac{d}{p}$. Moreover, suppose that a, b have compact support in the sense that there exists $E > 0$ such that $a(x,y) = b(x,y) = 0$ for any $|x| > E$ and $y \in M$. Then the conclusions from Theorem 2.3 hold.*

The proof is split into several steps:

1. In the remainder of this section, we reformulate x_ε into a rough path framework and show that x_ε solves a ODE of the form (3.6).
2. In Section 5, we use the theory from Section 3 to show that $x_\varepsilon \to_w X$ where X is defined by an RDE of the form (3.7).
3. In Section 6, we show that the RDE in step 2 can be re-written as the desired Itô SDE.

The proof of Theorem 4.1, which is a simple combination of the above facts, can be found at the end of Section 6. Then in Section 7, we show how Theorem 2.3 follows from Theorem 4.1.

4.1 The rough path reformulation of the fast-slow system

We define $C^\theta(\mathbb{R}^d, \mathbb{R}^d)$ to be the vector space of continuous functions $u : \mathbb{R}^d \to \mathbb{R}^d$ with components $u^1, \dots, u^d \in C^\theta(\mathbb{R}^d)$. This is a Banach space with norm $\|u\|_{C^\theta} = \sum_{i=1}^d \|u^i\|_{C^\theta}$.

Since $\alpha > 2 + 2/(p-1) + d/p$, we can choose $\theta > 2 + 2/(p-1)$ such that $\alpha > \theta + d/p$. For the reformulation described in step 1 above, we take $\mathcal{A}$ and $\mathcal{B}$ to be the Holder spaces $\mathcal{A} = C^{1+}(\mathbb{R}^d, \mathbb{R}^d)$, $\mathcal{B} = C^\theta(\mathbb{R}^d, \mathbb{R}^d)$. For $\varepsilon > 0$, we define the smooth paths

$$V_\varepsilon(t) = \int_0^t a(\cdot, y_\varepsilon(r))dr \quad \text{and} \quad W_\varepsilon(t) = \varepsilon^{-1} \int_0^t b(\cdot, y_\varepsilon(r))dr, \quad t \in [0, T] .$$

Proposition 4.2. *If a and b are as in Theorem 4.1, then V_ε and W_ε take values in $\mathcal{A}$ and $\mathcal{B}$ respectively for each $\varepsilon > 0$, $t \in [0, T]$.*

Proof. By definition, $\|W_\varepsilon(t)\|_{\mathcal{B}} = \sum_{i=1}^d \|\varepsilon^{-1} \int_0^t b^i(\cdot, y_\varepsilon(r))dr\|_{C^\theta}$. But

$$\sup_x \left| \int_0^t D_x^k b^i(\cdot, y_\varepsilon(r))dr \right| \le t \sup_x \sup_{y \in \Omega} |D_x^k b^i(x, y)|$$

and similarly

$$\sup_{x,x'} \frac{|\int_0^t D_x^k b^i(x, y_\varepsilon(r))dr - \int_0^t D_x^k b^i(x', y_\varepsilon(r))dr|}{|x - x'|^{\lfloor\theta\rfloor - \theta}} \le t \sup_{x,x'} \sup_{y \in \Omega} \frac{|D_x^k b^i(x, y) - D_x^k b^i(x', y)|}{|x - x'|^{\lfloor\theta\rfloor - \theta}} .$$

By definition of the Holder norm, it follows that

$$\|W_\varepsilon(t)\|_{\mathcal{B}} \le \varepsilon^{-1} t \sum_{i=1}^d \sup_{y \in \Omega} \left\|b^i(\cdot, y)\right\|_{C^\theta} = \varepsilon^{-1} t \, \|b\|_{C^{\theta,0}}$$

which is finite by the assumption on b. Similarly, $\|V_\varepsilon(t)\|_{\mathcal{A}} \le t\,\|a\|_{C^{1+,0}} < \infty$. ☐

For $x \in \mathbb{R}^d$, we define the multidimensional Dirac distribution operator $H : \mathbb{R}^d \to L(C^\theta(\mathbb{R}^d, \mathbb{R}^d), \mathbb{R}^d)$ by setting $H(x)(u) = u(x)$. It is easily shown that

Proposition 4.3. *$H \in C^\theta(\mathbb{R}^d, L(C^\theta(\mathbb{R}^d, \mathbb{R}^d), \mathbb{R}^d))$ for all $\theta \ge 0$.*

In this way, we obtain operators $F : \mathbb{R}^d \to L(\mathcal{A}, \mathbb{R}^d)$ and $H : \mathbb{R}^d \to L(\mathcal{B}, \mathbb{R}^d)$. The following result states that the above definitions are sufficient to reformulate (1.1) in the rough path framework of Corollary 3.4.

Lemma 4.4. *Suppose that a and b are as in Theorem 4.1 and define $F, H, V_\varepsilon, W_\varepsilon$ as above. Then the solution x_ε to the ODE* (1.1) *satisfies the ODE*

$$dx_\varepsilon = F(x_\varepsilon)dV_\varepsilon + H(x_\varepsilon)dW_\varepsilon \quad , \quad x_\varepsilon(0) = \xi . \tag{4.1}$$

Moreover, $F \in C^{1+}(\mathbb{R}^d; L(\mathcal{A}, \mathbb{R}^d))$ and $H \in C^\theta(\mathbb{R}^d; L(\mathcal{B}, \mathbb{R}^d))$ where $\theta > 2 + \frac{2}{p-1}$.

Proof. The regularity of F and H follows immediately from Proposition 4.3 and the choice of $\mathcal{A}, \mathcal{B}$.

For each fixed y, the function $x \mapsto a(x,y)$ is by assumption in $\mathcal{A}$ and so the operation $F(x)a(\cdot,y) = a(x,y)$ is well-defined. Hence for fixed x,t,

$$F(x)\frac{dV_\varepsilon(t)}{dt} = F(x)a(\cdot, y_\varepsilon(t)) = a(x, y_\varepsilon(t)) \ .$$

Similarly, $H(x)\frac{dW_\varepsilon(t)}{dt} = \varepsilon^{-1}b(x, y_\varepsilon(t))$. It follows that

$$\frac{dx_\varepsilon(t)}{dt} = a(x_\varepsilon(t), y_\varepsilon(t)) + \varepsilon^{-1}b(x_\varepsilon(t), y_\varepsilon(t)) = F(x_\varepsilon(t))\frac{dV_\varepsilon(t)}{dt} + H(x_\varepsilon(t))\frac{dW_\varepsilon(t)}{dt} \ .$$

In the incremental form, we have precisely (4.1). □

4.2 Tensor product of Holder spaces

As preparation for the application of rough path theory in Section 5, we define the tensor product $\mathcal{B} \otimes \mathcal{B}$ for the Hölder space $\mathcal{B} = C^\theta(\mathbb{R}^d, \mathbb{R}^d)$.

First we consider the scalar situation. Define $C^{\theta,\theta}(\mathbb{R}^d \times \mathbb{R}^d)$ to be the space of continuous functions $u : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ with bounded norm

$$\|u\|_{C^{\theta,\theta}} = \sum_{|k|\le\lfloor\theta\rfloor} \sup_x \left\|D_x^k u(x,\cdot)\right\|_{C^\theta} + \sum_{|k|=\lfloor\theta\rfloor} \sup_{x,x'} \frac{\left\|\delta_{x,x'}D_x^k u(x,\cdot)\right\|_{C^\theta}}{|x-x'|^{\theta-\lfloor\theta\rfloor}} \ , \tag{4.2}$$

with the shorthand $\delta_{x,x'}u(x,z) = u(x,z) - u(x',z)$, where the second summation is omitted if θ is an integer. Expanding the inner norm, we obtain

$$\begin{aligned}\|u\|_{C^{\theta,\theta}} = &\sum_{|k|\le\lfloor\theta\rfloor,|l|\le\lfloor\theta\rfloor} \sup_{x,z} |D_x^k D_z^l u(x,z)| + \sum_{|k|=\lfloor\theta\rfloor,|l|\le\lfloor\theta\rfloor} \sup_{x,x',z} \frac{|\delta_{x,x'}D_x^k D_z^l u(x,z)|}{|x-x'|^{\theta-\lfloor\theta\rfloor}} \\ &+ \sum_{|k|\le\lfloor\theta\rfloor,|l|=\lfloor\theta\rfloor} \sup_{x,z,z'} \frac{|\delta_{z,z'}D_x^k D_z^l u(x,z)|}{|z-z'|^{\theta-\lfloor\theta\rfloor}} + \sum_{|k|,|l|=\lfloor\theta\rfloor} \sup_{x,x',z,z'} \frac{|\delta_{x,x'}\delta_{z,z'}D_x^k D_z^l u(x,z)|}{(|x-x'||z-z'|)^{\theta-\lfloor\theta\rfloor}} \ .\end{aligned}$$

Here, we use the shorthand $\delta_{z,z'}u(x,z) = u(x,z) - u(x,z')$ and $\delta_{x,x'}\delta_{z,z'}u(x,z) = u(x,z) - u(x',z) - u(x,z') + u(x',z')$. It follows that we could equally define the norm in (4.2) with the roles of x and z reversed.

Let $\iota : C^\theta(\mathbb{R}^d) \otimes_a C^\theta(\mathbb{R}^d) \hookrightarrow C^{\theta,\theta}(\mathbb{R}^d \times \mathbb{R}^d)$ denote the embedding

$$\iota\Big(\sum_n u_n \otimes v_n\Big)(x,z) = \sum_n u_n(x)v_n(z).$$

Define the tensor product norm $\|\cdot\|_{C^\theta\otimes C^\theta}$ by setting $\|\sum_n u_n \otimes v_n\|_{C^\theta\otimes C^\theta} = \|\iota(\sum_n u_n \otimes v_n)\|_{C^{\theta,\theta}}$, and take the completion to obtain the tensor product space $(C^\theta(\mathbb{R}^d) \otimes C^\theta(\mathbb{R}^d), \|\cdot\|_{C^\theta\otimes C^\theta})$.

Proposition 4.5. *The tensor product norm $\|\cdot\|_{C^\theta\otimes C^\theta}$ is admissible.*

Proof. By an obvious factorization we have that $\|u\otimes v\|_{C^\theta\otimes C^\theta} = \|u\|_{C^\theta}\|v\|_{C^\theta}$ for every $u, v \in C^\theta(\mathbb{R}^d)$. It remains to check that $\|f\otimes g\|_{L(C^\theta\otimes_a C^\theta,\mathbb{R})} \le \|f\|_{L(C^\theta,\mathbb{R})}\|g\|_{L(C^\theta,\mathbb{R})}$ for all $f, g \in L(C^\theta, \mathbb{R})$. Notice that

$$\big|(f\otimes g)\sum_n u_n \otimes v_n\big| = \big|\sum_n f(u_n)g(v_n)\big| = \big|f\big(\sum_n u_n g(v_n)\big)\big| \tag{4.3}$$

$$\leq \|f\|_{L(C^\theta,\mathbb{R})} \big\| \sum_n u_n g(v_n) \big\|_{C^\theta} \,.$$

But

$$\big\| \sum_n u_n g(v_n) \big\|_{C^\theta} = \sum_{|k|\leq \lfloor\theta\rfloor} \sup_x \big| \sum_n D_x^k u_n(x) g(v_n) \big| + \sum_{|k|=\lfloor\theta\rfloor} \sup_{x,x'} \frac{|\sum_n \delta_{x,x'} D_x^k u_n(x) g(v_n)|}{|x-x'|^{\theta-\lfloor\theta\rfloor}} \,.$$

For each fixed x,

$$\begin{aligned}\big|\sum_n D_x^k u_n(x) g(v_n)\big| &= \big| g\big(\sum_n D_x^k u_n(x) v_n\big)\big| \leq \|g\|_{L(C^\theta,\mathbb{R})} \big\| \sum_n D_x^k u_n(x) v_n \|_{C^\theta} \\ &\leq \|g\|_{L(C^\theta,\mathbb{R})} \sum_n |D_x^k u_n(x)| \, \|v_n\|_{C^\theta} \;\;,\end{aligned}$$

and similarly $|\sum_n \delta_{x,x'} D_x^k u_n(x) g(v_n)| \leq \|g\|_{L(C^\theta,\mathbb{R})} \sum_n |\delta_{x,x'} D_x^k u_n(x)| \, \|v_n\|_{C^\theta}$. Substituting this back into (4.3), we obtain

$$\big|(f\otimes g) \sum_n u_n \otimes v_n\big| \leq \|f\|_{L(C^\theta,\mathbb{R})} \|g\|_{L(C^\theta,\mathbb{R})} \big\| \iota\big(\sum_n u_n \otimes v_n\big)\|_{C^{\theta,\theta}} \,.$$

Hence $\|f\otimes g\|_{L(C^\theta\otimes C^\theta,\mathbb{R})} \leq \|f\|_{L(C^\theta,\mathbb{R})} \|g\|_{L(C^\theta,\mathbb{R})}$. ☐

Next, we define the tensor product $\mathcal{B}\otimes\mathcal{B} = C^\theta(\mathbb{R}^d,\mathbb{R}^d)\otimes C^\theta(\mathbb{R}^d,\mathbb{R}^d)$ to be the space of $d\times d$ "matrices" with entries in $C^\theta(\mathbb{R}^d)\otimes C^\theta(\mathbb{R}^d)$, endowed with the norm $\|\sum_n u_n\otimes v_n\|_{C^\theta\otimes C^\theta} = \sum_{i,j=1}^d \|\sum_n u_n^i \otimes v_n^j\|_{C^\theta\otimes C^\theta}$.

Corollary 4.6. *$C^\theta(\mathbb{R}^d,\mathbb{R}^d)\otimes C^\theta(\mathbb{R}^d,\mathbb{R}^d)$ is a Banach space with admissible tensor product norm $\|\cdot\|_{C^\theta\otimes C^\theta}$.*

Proof. Completeness is an immediate consequence of the completeness of $C^\theta(\mathbb{R}^d)\otimes C^\theta(\mathbb{R}^d)$. Admissibility of $\|\cdot\|_{C^\theta\otimes C^\theta}$ is proved by a calculation similar to the one in Proposition 4.5. ☐

5 Convergence to the RDE

The objective of this section is to use Corollary 3.4 to characterize the $\varepsilon \to 0$ limit of the solution x_ε for the fast-slow ODE (1.1) as the solution to an RDE.

We suppose throughout that Assumptions 2.1 and 2.2 are valid with $p > 3$ and $\kappa > 0$, and that a, b are as in Theorem 4.1. Define $V_\varepsilon : [0,T] \to \mathcal{A}$ and $W_\varepsilon : [0,T] \to \mathcal{B}$ as in Section 4.1; in particular, $\mathcal{A} = C^{1+}(\mathbb{R}^d,\mathbb{R}^d)$ and $\mathcal{B} = C^\theta(\mathbb{R}^d,\mathbb{R}^d)$ where $\theta > 2 + 2/(p-1)$ and $\alpha > \theta + d/p$. Define $\mathcal{B}\otimes\mathcal{B}$ as in Section 4.2, and the iterated integral $\mathbb{W}_\varepsilon : [0,T] \to \mathcal{B}\otimes\mathcal{B}$ by

$$\mathbb{W}_\varepsilon(t) = \int_0^t W_\varepsilon \otimes dW_\varepsilon = \varepsilon^{-2} \int_0^t \int_0^r b(\cdot, y_\varepsilon(u)) \otimes b(\cdot, y_\varepsilon(r)) du dr \,.$$

The integral is well defined by Remark 3.2.

Let $\mathbf{W}_\varepsilon = (W_\varepsilon, \mathbb{W}_\varepsilon) : [0,T] \to \mathcal{B}\times(\mathcal{B}\otimes\mathcal{B})$. Define $\bar{a}\in\mathcal{A}$ by

$$\bar{a} = \int_\Omega a(\cdot, y) d\mu(y) \,. \tag{5.1}$$

We now state the main result of this section.

Theorem 5.1. *The family $\{x_\varepsilon\}_{\varepsilon>0}$ is tight in $C([0,T];\mathbb{R}^d)$. Moreover, every limit point X satisfies an RDE of the form*

$$dX = F(X)\bar{a}dt + H(X)d\mathbf{W} \quad , \quad X(0) = \xi \; , \tag{5.2}$$

where $\mathbf{W}$ is a limit point of $\{\mathbf{W}_\varepsilon\}_{\varepsilon>0}$ in $\mathscr{C}^\gamma$ for all $\gamma \in (\frac{1}{3}, \frac{1}{2} - \frac{1}{2p})$.

Remark 5.2. Evidently it suffices to prove Theorem 5.1 for γ arbitrarily close to $\gamma^* = \frac{1}{2} - \frac{1}{2p}$. By Lemma 4.4, H is C^θ where $\theta > 2 + \frac{2}{p-1} = \frac{1}{\gamma^*}$. Hence $\theta > \frac{1}{\gamma}$ for γ close to γ^* ensuring that H has the regularity required in Corollary 3.4.

The second aim of this section is to characterise the finite-dimensional distributions of the limit points of $\mathbf{W}_\varepsilon$. This is done in Lemma 5.12.

To control the tightness of $\mathbf{W}_\varepsilon$, we make use of Besov spaces described in Subsection 5.1. In Subsection 5.2, we prove tightness of $(V_\varepsilon, \mathbf{W}_\varepsilon)$ and deduce tightness of x_ε. In Subsection 5.3, we complete the proof of Theorem 5.1 and characterize the limit points of $(V_\varepsilon, \mathbf{W}_\varepsilon)$.

5.1 Besov spaces

Let $s > 0$ and fix (arbitrarily) an integer $m > s$. The classical Besov space $B_p^s = B_p^s(\mathbb{R}^d)$ can be defined (for all $p \geq 1$) as the set of all L_p functions $u : \mathbb{R}^d \to \mathbb{R}$ such that

$$\|u\|_{B_p^s} = \left(\|u\|_{L_p}^p + \int_{|\sigma|\leq 1} |\sigma|^{-sp-d} \|\Delta_\sigma^m u\|_{L_p}^p \, d\sigma \right)^{1/p} < \infty$$

where

$$\Delta_\sigma u(x) = u(x+\sigma) - u(x) \quad \text{and} \quad \Delta_\sigma^{l+1} = \Delta_\sigma \circ \Delta_\sigma^l$$

and $\|\cdot\|_{L_p}$ is the standard L_p norm on $\mathbb{R}^d$. For more details, see [Tri85, Tri06].

Remark 5.3. The classical Besov spaces $B_{p,q}^s$ typically come with two indices of integrability and norm $\|u\|_{B_{p,q}^s} = \left(\|u\|_{L_p}^q + \int_{|\sigma|\leq 1} |\sigma|^{-sq-d} \|\Delta_\sigma^m u\|_{L_p}^q \, d\sigma \right)^{1/q}$. In this article, we always take $p = q$. Hence our Besov spaces B_p^s are really the same as the Slobodeckij spaces, when $s \neq \mathbb{N}$. The norm we employ is not the most standard choice but is well known to be equivalent to the usual Besov norm [Tri85, Section 2.5.12].

For $\kappa \in [0,1)$, we also introduce a norm on functions $u = u(x,y)$ that are B_p^s in the x variable and C^κ in the y variable:

$$\|u\|_{B_p^s;C^\kappa} = \left(\int \|u(x,\cdot)\|_{C^\kappa}^p \, dx + \int_{|\sigma|\leq 1} \sigma^{-sp-d} \left(\int \|\Delta_\sigma^m u(x,\cdot)\|_{C^\kappa}^p \, dx \right) d\sigma \right)^{1/p} ,$$

with Δ_σ^m acting only in the x component.

Lemma 5.4. *We have the embeddings*

$$\|u\|_{C^\theta} \lesssim \|u\|_{B_p^{\theta+d/p}}$$

and

$$\|u\|_{C^{\theta,\theta}} \lesssim \left(\int \|u(x,\cdot)\|^p_{B_p^{\theta+d/p}}\, dx + \int_{|\sigma|\le 1} |\sigma|^{-p\theta-2d} \int \left\|\Delta^m_{x,\sigma} u(x,\cdot)\right\|^p_{B_p^{\theta+d/p}}\, dx d\sigma \right)^{1/p} .$$

Proof. The first estimate can be found in [Tri85, Section 2.7.1]. The second estimate is obtained by applying the first estimate in the x coordinate (for each fixed z) and then in the z coordinate (for each fixed x). □

Lemma 5.5. *If $u \in C^{\alpha,\kappa}(\mathbb{R}^d \times M)$ and has compact support (in the sense of Theorem 4.1) then $\|u\|_{B^s_p;C^\kappa} < \infty$ for any $s < \alpha$.*

Proof. Since $m > s$ is arbitrary, it suffices to take $m = \lceil s \rceil$. The $\|u\|_{L_p}$ part is obviously finite for any $p \ge 1$, since u is bounded and has compact support. Hence it suffices to bound the semi-norm part of the Besov norm. We claim that

$$\sup_{x\in\mathbb{R}^d} \|\Delta^m_\sigma u(x,\cdot)\|_{C^\kappa} \le \|u\|_{C^{\alpha,\kappa}} |\sigma|^\alpha . \tag{5.3}$$

In this case

$$\int_{|\sigma|\le 1} |\sigma|^{-sp-d} \int \left\|\Delta^m_{x,\sigma} u(x,\cdot)\right\|^p_{C^\kappa} dx d\sigma \le \|u\|^p_{C^{\alpha,\kappa}} \int_{|\sigma|\le 1} |\sigma|^{-sp-d} |\sigma|^{\alpha p} d\sigma < \infty$$

as required, since $\alpha > s$.

All that is left is to prove the inequality (5.3). By the chain rule, we have that

$$u(x+\sigma,y) - u(x,y) = \int_0^1 D_x u(x+(1-\zeta)\sigma, y) d\zeta \cdot \sigma .$$

Repeating this identity, and writing $\delta_{y,y'}u(x,y) = u(x,y) - u(x,y')$ we obtain

$$\begin{aligned} &\Delta^{m-1}_\sigma u(x,y) - \Delta^{m-1}_\sigma u(x,y') \\ &\quad = \int_{[0,1]^{m-1}} D^{m-1}_x \delta_{y,y'} u(x + ((m-1) - (\zeta_1 + \cdots + \zeta_{m-1})\sigma, y) d\zeta \cdot \sigma^{m-1} . \end{aligned}$$

It follows easily that

$$\begin{aligned} |\Delta^m_\sigma u(x,y) - \Delta^m_\sigma u(x,y')| &\le \sup_x |D^{m-1}\delta_{y,y'} u(x+\sigma,y) - D^{m-1}\delta_{y,y'} u(x,y)||\sigma|^{m-1} \\ &\le \|u\|_{C^{m-1+\zeta,\kappa}} |\sigma|^{m-1+\zeta} |y-y'|^\kappa . \end{aligned}$$

This proves (5.3). □

5.2 Tightness of $(V_\varepsilon, \mathbf{W}_\varepsilon)_{\varepsilon>0}$ and $(x_\varepsilon)_{\varepsilon>0}$

Firstly, we estimate $\|V_\varepsilon(s,t)\|_{\mathcal{A}}$.

Lemma 5.6. *We have that $\sup_{\varepsilon>0} \|V_\varepsilon(s,t)\|_{\mathcal{A}} \lesssim \|a\|_{C^{\eta,0}} |t-s|$ uniformly over Ω.*

Proof. Without loss, we suppose that a is real-valued. Write $V_\varepsilon(s,t;x) = \int_s^t a(x, y_\varepsilon(r))dr$. We have

$$|D_x^k V_\varepsilon(s,t;x)| \le \left\|D_x^k a(x,\cdot)\right\|_{C^0} |t-s|, \quad |\delta_{x,x'} D_x^k V_\varepsilon(s,t;x)| \le \left\|\delta_{x,x'} D_x^k a^i(x,\cdot)\right\|_{C^0} |t-s|$$

and hence

$$\begin{aligned}
\|V_\varepsilon(s,t)\|_{\mathcal{A}} &= \sum_{|k|\le\lfloor\eta\rfloor} \sup_x |D_x^k V_\varepsilon(s,t;x)| + \sum_{|k|=\lfloor\eta\rfloor} \sup_{x,x'} \frac{|\delta_{x,x'} D_x^k V_\varepsilon(s,t;x)|}{|x-x'|^{\eta-\lfloor\eta\rfloor}} \\
&\le \Bigg(\sum_{|k|\le\lfloor\eta\rfloor} \sup_x \left\|D_x^k a(x,\cdot)\right\|_{C^0} + \sum_{|k|=\lfloor\eta\rfloor} \sup_{x,x'} \frac{\left\|\delta_{x,x'} D_x^k a(x,\cdot)\right\|_{C^0}}{|x-x'|^{\eta-\lfloor\eta\rfloor}} \Bigg) |t-s| \\
&= \|a\|_{C^{\eta,0}} |t-s| \ ,
\end{aligned}$$

as required. ∎

We now obtain an analogous estimate for $\mathbf{W}_\varepsilon$. Again, we may suppose without loss that b is real-valued. First, we introduce the notation

$$\Delta_\sigma^m W_\varepsilon(s,t;x) = \varepsilon^{-1} \int_s^t \Delta_\sigma^m b(x, y_\varepsilon(r))dr \ ,$$

for $m \ge 0$, where the operator Δ_σ^m is omitted when $m = 0$. Similarly, we write

$$\Delta_\sigma^m \Delta_{\sigma'}^{m'} \mathbb{W}_\varepsilon(s,t;x,x') = \varepsilon^{-2} \int_s^t \int_s^r \Delta_\sigma^m b(x, y_\varepsilon(u)) \Delta_{\sigma'}^{m'} b(x', y_\varepsilon(r))dudr \ .$$

Proposition 5.7. *(a)* $\mathbf{E}_\mu\bigg(\sup_{s,t} \dfrac{|\Delta_\sigma^m W_\varepsilon(s,t;x)|}{|t-s|^\gamma} \bigg)^{2p} \lesssim \|\Delta_\sigma^m b(x,\cdot)\|_{C^\kappa}^{2p}$.

(b) $\mathbf{E}_\mu\bigg(\sup_{s,t} \dfrac{|\Delta_{x,\sigma}^m \Delta_{x',\sigma'}^{m'} \mathbb{W}_\varepsilon(s,t;x,x')|}{|t-s|^{2\gamma}} \bigg)^{p} \lesssim \left\|\Delta_{x,\sigma}^m b(x,\cdot)\right\|_{C^\kappa}^p \left\|\Delta_{x',\sigma'}^{m'} b(x',\cdot)\right\|_{C^\kappa}^p$.

Proof. Recall from the introduction that $y_\varepsilon(t) = y(t\varepsilon^{-2}) = \phi_{t\varepsilon^{-2}}(y_0)$ where ϕ is the underlying fast flow and $y_0 \in \Omega$ is the initial condition. Hence by change of variables,

$$\Delta_\sigma^m W_\varepsilon(s,t;x)(y) = \varepsilon \int_{s\varepsilon^{-2}}^{t\varepsilon^{-2}} \Delta_\sigma^m b(x, \phi_r y)dr.$$

But $\Delta_\sigma^m b(x,\cdot) \in C_0^\kappa(\Omega, \mathbb{R})$ for each x, σ, so by Proposition 2.7,

$$(\mathbf{E}_\mu |\Delta_\sigma^m W_\varepsilon(s,t;x)|^{2p})^{1/(2p)} \lesssim \|\Delta_\sigma^m b(x,\cdot)\|_{C^\kappa} |t-s|^{1/2},$$

uniformly in s,t,x,σ,ε. Hence part (a) follows from the Kolmogorov criterion, Lemma 3.7(a). Part (b) is proved almost identically using Lemma 3.7(b). ∎

Lemma 5.8. *We have that* $\sup_{\varepsilon>0} \mathbf{E}_\mu \|\!|\mathbf{W}_\varepsilon\|\!|_{\mathscr{C}^\gamma}^{2p} < \infty$ *for any* $\gamma \in (\frac{1}{3}, \frac{1}{2} - \frac{1}{2p})$.

Proof. By Lemma 5.4, $\|W_\varepsilon(s,t)\|_{\mathcal{B}} \lesssim \|W_\varepsilon(s,t)\|_{B_p^{\theta+d/p}}$ and hence

$$\sup_{s,t} \frac{\|W_\varepsilon(s,t)\|_{\mathcal{B}}}{|t-s|^\gamma} \lesssim \sup_{s,t} \frac{1}{|s-t|^\gamma} \Big(\int |W_\varepsilon(s,t;x)|^p dx + \int_{|\sigma|\le 1} |\sigma|^{-\theta p - 2d} \int |\Delta_\sigma^m W_\varepsilon(s,t;x)|^p dx d\sigma \Big)^{1/p} .$$

Taking the supremum inside the integrals and using the inequality $(x+y)^{1/p} \le x^{1/p} + y^{1/p}$,

$$\sup_{s,t} \frac{\|W_\varepsilon(s,t)\|_{\mathcal{B}}}{|t-s|^\gamma} \le \Big(\int \Big(\sup_{s,t} \frac{|W_\varepsilon(s,t;x)|}{|t-s|^\gamma} \Big)^p dx \Big)^{1/p} + \Big(\int_{|\sigma|\le 1} |\sigma|^{-\theta p - 2d} \int \Big(\sup_{s,t} \frac{|\Delta_\sigma^m W_\varepsilon(s,t;x)|}{|t-s|^\gamma} \Big)^p dx d\sigma \Big)^{1/p} = \sum_{k=1}^2 \Big(\int c_k^p dz \Big)^{1/p} , \tag{5.4}$$

where $dz = dx$ or $dz = |\sigma|^{-\theta p - 2d} dx d\sigma$ respectively and c_k^p denotes the corresponding integrands. Applying the triangle inequality, first for L_{2p} and then for L_2,

$$\Big\| \sum_k \Big(\int c_k^p dz \Big)^{1/p} \Big\|_{L_{2p}(d\mu)} \le \sum_k \Big\| \Big(\int c_k^p dz \Big)^{1/p} \Big\|_{L_{2p}(d\mu)} = \sum_k \Big(\mathbf{E}_\mu \Big(\int c_k^p dz \Big)^2 \Big)^{1/(2p)}$$
$$= \sum_k \Big(\Big\| \int c_k^p \Big\|_{L_2(d\mu)} \Big)^{1/p} \le \sum_k \Big(\int \|c_k^p\|_{L_2(d\mu)} \, dz \Big)^{1/p} = \sum_k \Big(\int \Big(\mathbf{E}_\mu c_k^{2p} \Big)^{1/2} dz \Big)^{1/p} .$$

Substituting into (5.4) and applying Proposition 5.7 to each term, we obtain

$$\Big(\mathbf{E}_\mu \Big(\sup_{s,t} \frac{\|W_\varepsilon(s,t)\|_{\mathcal{B}}}{|t-s|^\gamma} \Big)^{2p} \Big)^{1/(2p)} \le \Big(\int \Big(\mathbf{E}_\mu \Big(\sup_{s,t} \frac{|W_\varepsilon(s,t;x)|}{|t-s|^\gamma} \Big)^{2p} \Big)^{1/2} dz \Big)^{1/p} + \Big(\int \Big(\mathbf{E}_\mu \Big(\sup_{s,t} \frac{|\Delta_\sigma^m W_\varepsilon(s,t;x)|}{|t-s|^\gamma} \Big)^{2p} \Big)^{1/2} dz \Big)^{1/p}$$
$$\lesssim \Big(\int \|b(x,\cdot)\|_{C^\kappa}^p \, dz \Big)^{1/p} + \Big(\int \|\Delta_\sigma^m b(x,\cdot)\|_{C^\kappa}^p \, dz \Big)^{1/p} \lesssim \|b\|_{B_p^{\theta+d/p};C^\kappa} < \infty , \tag{5.5}$$

where the last inequality follows from Lemma 5.5 (since $\theta + d/p < \alpha$.)

We now use the same method to estimate the $\mathbb{W}_\varepsilon$ term. Just as above, via Lemma 5.4 we have

$$\sup_{s,t} \frac{\|\mathbb{W}_\varepsilon(s,t)\|_{\mathcal{B}\otimes\mathcal{B}}}{|t-s|^{2\gamma}} = \sup_{s,t} \|\mathbb{W}_\varepsilon(s,t;\cdot,\cdot)\|_{C^{\theta,\theta}} \le \Big(\int \Big(\sup_{s,t} \frac{|\mathbb{W}_\varepsilon(s,t;x,x')|}{|t-s|^{2\gamma}} \Big)^p dz \Big)^{1/p}$$
$$+ \Big(\int \Big(\sup_{s,t} \frac{|\Delta_{x,\sigma}^m \mathbb{W}_\varepsilon(s,t;x,x')|}{|t-s|^{2\gamma}} \Big)^p dz \Big)^{1/p} + \Big(\int \Big(\sup_{s,t} \frac{|\Delta_{x',\sigma'}^m \mathbb{W}_\varepsilon(s,t;x,x')|}{|t-s|^{2\gamma}} \Big)^p dz \Big)^{1/p}$$
$$+ \Big(\int \Big(\sup_{s,t} \frac{|\Delta_{x,\sigma}^m \Delta_{x',\sigma'}^m \mathbb{W}_\varepsilon(s,t;x,x')|}{|t-s|^{2\gamma}} \Big)^p dz \Big)^{1/p} ,$$

where dz is variously $dxdx'$, $|\sigma|^{-\theta p-2d}dxdx'd\sigma$, $|\sigma'|^{-\theta p-2d}dxdx'd\sigma'$ or $|\sigma\sigma'|^{-\theta p-2d}dxdx'd\sigma d\sigma'$. We apply $\mathbf{E}_\mu$ to the left hand side, using the triangle inequality to take the L_1 norm inside the sums and integrals. Applying Proposition 5.7(b) to each term, we obtain

$$\begin{aligned}
\mathbf{E}_\mu \sup_{s,t} \frac{\|\mathbb{W}_\varepsilon(s,t)\|_{\mathcal{B}\otimes\mathcal{B}}}{|t-s|^{2\gamma}} &\lesssim \Big(\int \|b(x,\cdot)\|^p_{C^\kappa}\, \|b(x',\cdot)\|^p_{C^\kappa}\, dz\Big)^{1/p} \qquad (5.6)\\
&+ \Big(\int \big\|\Delta^m_{x,\sigma} b(x,\cdot)\big\|^p_{C^\kappa}\, \|b(x',\cdot)\|^p_{C^\kappa}\, dz\Big)^{1/p} + \Big(\int \|b(x,\cdot)\|^p_{C^\kappa}\, \Big\|\Delta^{m'}_{x',\sigma'} b(x',\cdot)\Big\|^p_{C^\kappa}\, dz\Big)^{1/p}\\
&+ \Big(\int \big\|\Delta^m_{x,\sigma} b(x,\cdot)\big\|^p_{C^\kappa}\, \Big\|\Delta^{m'}_{x',\sigma'} b(x',\cdot)\Big\|^p_{C^\kappa}\, dz\Big)^{1/p} .\\
&= \Big\{\Big(\int \|b(x,\cdot)\|^p_{C^\kappa}\, dx\Big)^{1/p} + \Big(\int_{|\sigma|\le 1} |\sigma|^{-\theta p-2d} \int \big\|\Delta^m_{x,\sigma} b(x,\cdot)\big\|^p_{C^\kappa}\, dxd\sigma\Big)^{1/p}\Big\}\\
&\times \Big\{\Big(\int \|b(x',\cdot)\|^p_{C^\kappa}\, dx'\Big)^{1/p} + \Big(\int_{|\sigma'|\le 1} |\sigma'|^{-\theta p-2d} \int \big\|\Delta^m_{x',\sigma} b(z,\cdot)\big\|^p_{C^\kappa}\, dx'd\sigma'\Big)^{1/p}\Big\}\\
&\lesssim \|b\|^2_{B_p^{\theta+d/p};C^\kappa} < \infty ,
\end{aligned}$$

where the last inequality follows from Lemma 5.5. Combining (5.5) and (5.6), we obtain the required estimate for $\mathbf{W}_\varepsilon$. □

Finally, we have the claimed tightness result.

Corollary 5.9. *(a) The family $(V_\varepsilon, \mathbf{W}_\varepsilon)_{\varepsilon>0}$ is tight in $\mathscr{C}^{\beta,\gamma}$ for any $\beta \in (\frac{1}{2}, 1)$, $\gamma \in (\frac{1}{3}, \frac{1}{2} - \frac{1}{2p})$.*

(b) The family $(x_\varepsilon)_{\varepsilon>0}$ is tight in $C([0,T], \mathbb{R}^d)$.

Proof. We first show that $\mathbf{W}_\varepsilon = (W_\varepsilon, \mathbb{W}_\varepsilon)$ is tight in $\mathscr{C}^\gamma$. Let $R > 2$, $\gamma' \in (\gamma, \frac{1}{2} - \frac{1}{2p})$, and let $B_R \subset \mathscr{C}^\gamma$ be the ball of radius R in the $\rho_{\gamma'}$ metric. By a standard Arzela-Ascoli argument (for instance, see [FV10, Chapter 5]) one can show that B_R is sequentially compact with respect to ρ_γ and hence compact in $\mathscr{C}^\gamma$. Since $\rho_{\gamma'}(\mathbf{W}_\varepsilon, 0) \le \|\mathbf{W}_\varepsilon\|_{\mathscr{C}^{\gamma'}} + \|\mathbf{W}_\varepsilon\|^2_{\mathscr{C}^{\gamma'}}$ and $R > 2$,

$$\mu\big(\mathbf{W}_\varepsilon \notin B_R\big) \le \mu\big(\|\mathbf{W}_\varepsilon\|_{\mathscr{C}^{\gamma'}} \ge (R/2)^{1/2}\big) .$$

Hence by Markov's inequality and Lemma 5.8,

$$\mu\big(\mathbf{W}_\varepsilon \notin B_R\big) \le 2^p \mathbf{E}_\mu \|\mathbf{W}_\varepsilon\|^{2p}_{\mathscr{C}^{\gamma'}}/R^p \lesssim R^{-p} .$$

This proves tightness of $\mathbf{W}_\varepsilon$. An analogous, but simpler, argument using Lemma 5.6 shows that V_ε is tight in $\mathscr{C}^\beta$, concluding the proof of part (a).

For part (b), let x_{ε_k} be a subsequence. By part (a), we can apply Prokhorov's theorem to $(V_{\varepsilon_k}, \mathbf{W}_{\varepsilon_k})$. Hence passing to a subsubsequence, there exists $(V, \mathbf{W}) \in \mathscr{C}^{\beta,\gamma}$ such that $(V_{\varepsilon_k}, \mathbf{W}_{\varepsilon_k}) \to_w (V, \mathbf{W})$ in the $\mathscr{C}^{\beta,\gamma}$ topology. By Lemma 4.4 and Corollary 3.4, $x_{\varepsilon_k} \to_w X$ in $C([0,T], \mathbb{R}^d)$ where X satisfies the RDE (4.1) driven by $(V, \mathbf{W})$. It follows that $\{x_\varepsilon\}_{\varepsilon>0}$ is weakly precompact in $C([0,T], \mathbb{R}^d)$. Since $C([0,T], \mathbb{R}^d)$ is Polish, we can apply Prokhorov's theorem to deduce that $\{x_\varepsilon\}_{\varepsilon>0}$ is tight. □

5.3 Characterization of limits of $(V_\varepsilon, \mathbf{W}_\varepsilon)_{\varepsilon>0}$ and $(x_\varepsilon)_{\varepsilon>0}$

We begin by describing the limit of V_ε.

Lemma 5.10. *Define the deterministic element $V \in C^1([0,T],\mathcal{A})$ given by $V(t) = \bar{a}t$ where $\bar{a} \in \mathcal{A}$ is defined in* (5.1). *Then $V_\varepsilon \to V$ in probability in $\mathscr{C}^\beta$ for any $\beta \in (\frac{1}{2}, 1)$,*

Proof. Let $\pi \in L(\mathcal{B},\mathbb{R})$. Then $\pi V_\varepsilon(t) = \varepsilon^2 \int_0^{t\varepsilon^{-2}} (\pi a) \circ \phi_s \, ds$. By ergodicity of μ, it follows from the ergodic theorem that $\pi V_\varepsilon(1) \to \pi V(1)$ almost surely. By Lemma 2.9, $\pi V_\varepsilon \to \pi V$ almost surely, and hence in probability, in $C([0,T],\mathbb{R})$.

Suppose for contradiction that V_ε fails to converge weakly to V in $\mathscr{C}^\beta$. By Corollary 5.9, the family V_ε is tight in $\mathscr{C}^\beta$, so there is a subsequence such that $V_{\varepsilon_k} \to_w Z$ in $\mathscr{C}^\beta$ where the random process Z differs from V. In particular, $\pi V_{\varepsilon_k}(t_0) \to_w \pi Z(t_0)$ in $\mathbb{R}$ for any $\pi \in L(\mathcal{A},\mathbb{R})$ and any $t_0 \in [0,T]$. Hence $\pi Z(t_0)$ has the same distribution as $\pi V(t_0)$ and so $\mathbf{P}\big(\pi Z(t_0) = \pi \bar{a} t_0\big) = 1$. Since π is arbitrary, it follows that $\mathbf{P}\big(Z(t_0) = \bar{a}t_0\big) = 1$. But Z is continuous, so $Z = V$ with probability one, giving the desired contradiction. □

Proof of Theorem 5.1. We have shown in Corollary 5.9 that $\{x_\varepsilon\}_{\varepsilon>0}$ is tight. Let X be a limit point, with $x_{\varepsilon_k} \to_w X$ in $C([0,T];\mathbb{R}^d)$. By Lemma 5.8, we can pass to a subsubsequence for which $(V_{\varepsilon_k}, \mathbf{W}_{\varepsilon_k})$ converges weakly in $\mathscr{C}^{\beta,\gamma}$. Denote the limit by $(V, \mathbf{W})$. By Lemma 4.4 and Corollary 3.4, X solves an RDE of the form (4.1) driven by $(V, \mathbf{W})$. By Lemma 5.10, $V(t) = \bar{a}t$ completing the proof. □

Finally, we obtain a partial (see Remark 5.14) characterization of the limit points of $\mathbf{W}_\varepsilon$ in terms of their finite dimensional distributions. For each fixed $\pi \in L(\mathcal{B},\mathbb{R}^m)$, let $(\mathbf{P}^\pi, \Omega^\pi, \mathcal{G}^\pi)$ be a probability space endowed with a filtration $\{\mathcal{G}^\pi_t\}_{t\geq 0}$ rich enough to support Brownian motion. We define a stochastic process $(B_\pi, \mathbb{B}_\pi) : [0,T] \to \mathbb{R}^m \times \mathbb{R}^{m\times m}$ on the probability space $(\mathbf{P}^\pi, \Omega^\pi, \mathcal{G}^\pi)$, where B_π is a $\mathbb{R}^m$-valued $\mathcal{G}^\pi_t$ - Brownian motion with covariance

$$\mathbf{E}^\pi B^i_\pi(1) B^j_\pi(1) = \mathfrak{B}(\pi^i b, \pi^j b) + \mathfrak{B}(\pi^j b, \pi^i b) \tag{5.7}$$

and $\mathbb{B}_\pi$ is defined by

$$\mathbb{B}^{ij}_\pi(t) = \int_0^t B^i_\pi dB^j_\pi + \mathfrak{B}(\pi^i b, \pi^j b)t \tag{5.8}$$

where the integral is of Itô type. The filtration $\mathcal{G}^\pi_t$ does not appear in the sequel and is present only to ensure that the Itô integral can indeed be constructed.) Notice that this is precisely the structure that arises under Assumption 2.1.

Remark 5.11. Here $\pi^i b$ denotes the observable $y \mapsto \pi^i b(\cdot, y)$, with π^i acting on b as a function of x. By the regularity assumptions on b, it is easy to check that $\pi^i b \in C^\kappa_0(\Omega,\mathbb{R})$ and lies in the domain of $\mathfrak{B}$ (this calculation is done explicitly in Lemma 5.12). Moreover, by Proposition 2.8, the covariance matrix of B_π is a symmetric, positive semi-definite matrix. This guarantees existence of the Brownian motion B_π and hence the pair $(B_\pi, \mathbb{B}_\pi)$.

For $\pi \in L(\mathcal{B},\mathbb{R}^m)$ we define $\pi \otimes \pi \in L(\mathcal{B},\mathbb{R}^{m\times m})$ by $(\pi\otimes\pi)^{ij} = \pi^i \otimes \pi^j$, where $\pi^i \otimes \pi^j$ is (as usual) the dual tensor product.

Lemma 5.12. *Let $\pi \in L(\mathcal{B}, \mathbb{R}^m)$ for some $m \in \mathbb{N}$. As $\varepsilon \to 0$,*

$$(\pi W_\varepsilon, (\pi \otimes \pi)\mathbb{W}_\varepsilon) \to (B_\pi, \mathbb{B}_\pi)$$

in the sense of finite dimensional distributions of stochastic processes.

Proof. we have

$$(\pi W_\varepsilon, (\pi\otimes\pi)\mathbb{W}_\varepsilon)(t) = \left(\varepsilon^{-1} \int_0^t (\pi b)(y_\varepsilon(r))dr, \varepsilon^{-2} \int_0^t \int_0^r (\pi b)(y_\varepsilon(u))\otimes(\pi b)(y_\varepsilon(r))dudr\right) .$$

Now

$$\begin{aligned}|(\pi b)(y) - (\pi b)(z)| = |\pi(b(\cdot, y) - b(\cdot, z))| &\leq \|\pi\|_{L(\mathcal{B},\mathbb{R}^m)} \|b(\cdot, y) - b(\cdot, z)\|_{\mathcal{B}} \\ &\leq \|\pi\|_{L(\mathcal{B},\mathbb{R}^m)} \|b\|_{C^{\theta,\kappa}} |y - z|^\kappa .\end{aligned}$$

Similarly, $|(\pi b)(y)| \leq \|\pi\|_{L(\mathcal{B},\mathbb{R}^m)} \|b\|_{C^{\theta,0}}$. Hence $\pi b \in C_0^\kappa(\Omega, \mathbb{R}^m)$ and the desired convergence follows from Proposition 2.10. □

Remark 5.13. Clearly, we can equally characterize the distribution of $(\pi_1 W, (\pi_2 \otimes \pi_3)\mathbb{W})$ using this result, where each $\pi_i : \mathcal{B} \to \mathbb{R}^{m_i}$. Simply set $\pi = (\pi_1, \pi_2, \pi_3)$ and then project out the unnecessary components.

Remark 5.14. It would be natural to combine the tightness of $\{\mathbf{W}_\varepsilon\}_{\varepsilon>0}$ with the convergence of finite dimensional distributions of $\mathbf{W}_\varepsilon$ obtained in Lemma 5.12 to obtain a weak limit theorem for $\{\mathbf{W}_\varepsilon\}_{\varepsilon>0}$. We avoid this here since showing that the finite dimensional distributions from Lemma 5.12 actually separate measures on $\mathscr{C}^\gamma$ is a non-trivial task. Moreover, we gain nothing by doing so since, as shown in Lemma 6.1 below, all limit points X agree.

6 Characterizing the RDE as a Diffusion

In this section we complete the proof of Theorem 4.1. The final ingredient is the following.

Lemma 6.1. *Let $\mathbf{W}$ be any limit point of $\{\mathbf{W}_\varepsilon\}_{\varepsilon>0}$ and let X be the solution to the RDE* (5.2) *driven by $\mathbf{W}$. Then X is a weak solution to the SDE* (1.5).

Before proceeding with the proof of Lemma 6.1, we need two technical results concerning the following filtration. For each $\pi \in \mathcal{L}(\mathcal{B}, \mathbb{R})$, let $\{\mathcal{F}_t^\pi\}_{t\geq 0}$ be the filtration generated by the stochastic process $\pi W(t)$ and let $\mathcal{F}_t = \bigvee_{\pi\in\mathcal{L}(\mathcal{B},\mathbb{R})} \mathcal{F}_t^\pi$. Note that $\{\mathcal{F}_t\}_{t\geq 0}$ is indeed a filtration due to the properties of the sigma-algebra join.

Lemma 6.2. *Let $\mathbf{W} = (W, \mathbb{W})$ be any limit point of $\{\mathbf{W}^\varepsilon\}_{\varepsilon>0}$. For each $0 \leq s \leq t$ and $\pi_1, \pi_2, \pi_3 \in \mathcal{L}(\mathcal{B}, \mathbb{R})$, we have that $\pi_1 W(s, t)$ and $(\pi_2 \otimes \pi_3)\mathbb{W}(s, t)$ are independent of $\mathcal{F}_s$.*

Proof. Let $A_s^\rho = \{\rho W(s) \in \Gamma\}$ for $\rho \in \mathcal{L}(\mathcal{B}, \mathbb{R})$ and some Borel measurable $\Gamma \subset \mathbb{R}$. Define the system

$$\mathcal{P}_s = \left\{\bigcap_{i=1}^m A_s^{\rho_i} : \rho_i \in \mathcal{L}(\mathcal{B}, \mathbb{R}), m \geq 1\right\} .$$

It is easy to see that this is a pi-system [Wil91] with $\mathcal{P}_s \subset \mathcal{F}_s$ and hence that $\mathcal{F}_s = \sigma(\mathcal{P}_s)$. Thus, it suffices to show that $\pi_1 W(s,t)$ and $(\pi_2 \otimes \pi_3)\mathbb{W}(s,t)$ are independent from sets of the form $\cap_{i=1}^m A_s^{\rho_i}$.

From Lemma 5.12 we see that $\pi_1 W(s,t)$ is independent of $\cap_{i=1}^m A_s^{\rho_i}$. Indeed, one can simply take $\pi = (\pi_1, \rho_1, \dots, \rho_m)$ and deduce that $(\pi_1, \rho_1, \dots, \rho_m)W(t)$ is a Brownian motion in $\mathbb{R}^{m+1}$, and certainly $\pi_1 W(s,t)$ is independent of $(\rho_1 W(s), \dots, \rho_m W(s))$ and hence independent of $\cap_{i=1}^m A_s^{\rho_i}$. The result for $(\pi_2 \otimes \pi_3)\mathbb{W}(s,t)$ follows similarly, using the independence properties of Itô integral increments. $\square$

Lemma 6.3. *The solution X to the RDE* (5.2) *is adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$.*

Proof. We will show that $X(t)$ is $\mathcal{F}_t$-measurable for some arbitrary t. We proceed by approximating X with a Euler-type discretization. To this end, define the partition $0 = t_0 \leq t_1 \leq \dots \leq t_N = t$ with $t_{n+1} - t_n = t/N$. Let $X_n \in \mathbb{R}^d$ satisfy the recurrence relation

$$\begin{aligned} X_{n+1}^i = X_n^i + \int_\Omega a^i(X_n, y) d\mu(y)(t_{n+1} - t_n) + H^i(X_n)W(t_n, t_{n+1}) \\ + \sum_{k=1}^d \big(H^k(X_n) \otimes \partial_k H^i(X_n)\big)\mathbb{W}(t_n, t_{n+1}) \ , \end{aligned}$$

for $n = 0, \dots, N$, $i = 1, \dots, d$ with $X_0^i = X^i(0)$. By an inductive argument on n, we see that X_n is $\mathcal{F}_t$ measurable for each n, since all terms appearing on the RHS of the recursion are either $\mathcal{F}_t$ measurable or are of the form $\pi W(t_n, t_{n+1})$ where $\pi \in \mathcal{L}(\mathcal{B}, \mathbb{R})$ is $\mathcal{F}_t$ measurable. Moreover, by [Dav07, Theorem 3.3], we have that $X_N \to X(t)$ as $N \to \infty$ where the convergence is pathwise. As pathwise limits preserve measurability, it follows that $X(t)$ is $\mathcal{F}_t$ measurable. $\square$

Remark 6.4. Note that [Dav07, Theorem 3.3] assumes the rough paths take values in finite dimensional spaces, but as with Theorem 3.3 (of this article) the proof carries through verbatim for Banach spaces, provided the tensor product is admissible.

Proof of Lemma 6.1. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a smooth function and let $\mathcal{L}$ be the generator of the SDE (1.5), given by

$$\mathcal{L}\varphi(x) = \sum_{i=1}^d \tilde{a}^i(x)\partial_i \varphi(x) + \sum_{i,j=1}^d \frac{1}{2}(\sigma\sigma^T)^{ij}(x)\partial_{ij}^2 \varphi(x)$$

with

$$\tilde{a}^i(x) = \int_\Omega a^i(x,y) d\mu(y) + \sum_{k=1}^d \mathfrak{B}(b^k(x,\cdot), \partial_k b^i(x,\cdot))$$

and

$$(\sigma\sigma^T)^{ij}(x) = \mathfrak{B}(b^i(x,\cdot), b^j(x,\cdot)) + \mathfrak{B}(b^j(x,\cdot), b^i(x,\cdot)) \ .$$

By Proposition 2.8(b), $\sigma\sigma^T(x)$ is symmetric positive semidefinite, for each $x \in \mathbb{R}^d$. We show that X solves the martingale problem associated with $\mathcal{L}$.

To this end, let $\{\mathcal{F}_t\}_{t\geq 0}$ be the filtration defined above. Note that, in order to apply [EK86, Theorem 5.3.3] we require that $\mathcal{F}_t^X \subset \mathcal{F}_t$ where $\{\mathcal{F}_t^X\}_{t\geq 0}$ is the filtration generated by X. This inclusion is guaranteed by Lemma 6.3. Thus, by [SV06, Theorem 4.5.2] (or more precisely, [EK86, Theorem 5.3.3]) it suffices to show that

$$\varphi(X(t)) - \varphi(X(s)) - \int_s^t \mathcal{L}\varphi(X(r))dr$$

is an $\mathcal{F}_t$-martingale.

Since $\mathbf{W} \in \mathscr{C}_g^\gamma$ it follows from the chain rule for RDEs, Proposition 3.6, that

$$\begin{aligned}\varphi(X(t)) = \varphi(X(s)) &+ \sum_{i=1}^d \int_s^t \partial_i\varphi(X(r)) \int_\Omega a^i(X(r),y)d\mu(y)dr \\ &+ \sum_{i=1}^d \int_s^t \partial_i\varphi(X)H^i(X)d\mathbf{W}\ ,\end{aligned} \tag{6.1}$$

with the equality holding pathwise, where we have used the identity $F(x)\bar{a} = \int_\Omega a(x,y)d\mu(y)$. Using (6.1) together with the "divergence-form" of $\mathcal{L}$,

$$\mathcal{L}\varphi(x) = \sum_{i=1}^d \int_\Omega a^i(x,y)d\mu(y)\partial_i\varphi(x) + \sum_{i,k=1}^d \mathfrak{B}\big(b^k(x,\cdot), \partial_k\{b^i(x,\cdot)\partial_i\varphi(x)\}\big)\ ,$$

we reduce to showing that for each $i = 1,\dots,d$,

$$\mathbf{E}(S_i|\mathcal{F}_s) = \sum_{k=1}^d \mathbf{E}\Big(\int_s^t G_{ik}(X(r))dr \,\big|\, \mathcal{F}_s\Big) \tag{6.2}$$

for all $s \leq t \leq T$, where

$$\begin{aligned}S_i &= \int_s^t \partial_i\varphi(X)H^i(X)d\mathbf{W}\ ,\\ G_{ik}(x) &= \mathfrak{B}\big(b^k(x,\cdot), \partial_k\{b^i(x,\cdot)\partial_i\varphi(x)\}\big)\ .\end{aligned}$$

Since we prove this for each fixed $i = 1,\dots,d$, we will from here on drop i from the notation, instead working with S and G_k.

By definition of the rough integral in (3.5), using Proposition 3.5, we see that $S = \lim_{\Delta\to 0} S_\Delta$ where the limit is defined pathwise and

$$\begin{aligned}S_\Delta = \sum_{[t_n,t_{n+1}]\in\Delta} &\partial_i\varphi(X(t_n))H^i(X(t_n))W(t_n,t_{n+1})\\ &+ \sum_{k=1}^d \big(H^k(X(t_n)) \otimes \partial_k\{\partial_i\varphi(X(t_n))H^i(X(t_n))\}\big)\mathbb{W}(t_n,t_{n+1})\end{aligned}$$

and $\Delta = \{[t_n,t_{n+1}] : 0 \leq n \leq N-1\}$ denotes partitions of $[s,t]$.

Next, we define

$$M_\Delta = S_\Delta - \sum_{[t_n,t_{n+1}]\in\Delta} \sum_{k=1}^{d} G_k(X(t_n))\Delta t_n \ ,$$

where $\Delta t_n = t_{n+1} - t_n$. It follows directly from the regularity of b, X, φ that the map

$$t \mapsto \big(b^k(X(t),\cdot), \partial_k\{b^i(X(t),\cdot)\partial_i\varphi(X(t))\}\big) \ \ , \ \ [0,T] \to C^\kappa(\Omega) \times C^\kappa(\Omega)$$

is continuous. By Proposition 2.8(c), $t \mapsto G_k(X(t))$ is continuous and hence Riemann integrable. In particular, $\lim_{\Delta\to 0}(S_\Delta - M_\Delta) = \int_s^t G_k(X(r))dr$ almost surely. Hence,

$$\begin{aligned}
\mathbf{E}(S|\mathcal{F}_s) &= \mathbf{E}(\lim_{\Delta\to 0} M_\Delta|\mathcal{F}_s) + \mathbf{E}(\lim_{\Delta\to 0}(S_\Delta - M_\Delta)|\mathcal{F}_s) \\
&= \mathbf{E}(\lim_{\Delta\to 0} M_\Delta|\mathcal{F}_s) + \sum_{k=1}^{d} \mathbf{E}\bigg(\int_s^t G_k(X(r))dr \,\Big|\, \mathcal{F}_s\bigg) \ .
\end{aligned}$$

Thus proving (6.2) reduces to showing that $\mathbf{E}(\lim_{\Delta\to 0} M_\Delta|\mathcal{F}_s) = 0$. We claim that M_Δ is square integrable uniformly in $|\Delta| \le 1$ and that $\mathbf{E}(M_\Delta|\mathcal{F}_s) = 0$ for each $|\Delta| \le 1$. Then by convergence of first moments, $\mathbf{E}(\lim_{|\Delta|\to 0} M_\Delta|\mathcal{F}_s) = \lim_{|\Delta|\to 0}\mathbf{E}(M_\Delta|\mathcal{F}_s) = 0$, completing the proof.

It remains to verify the claim. For each $x \in \mathbb{R}^d$ let us define the projections $\pi^k(x) = (\pi_1(x), \pi_2^k(x), \pi_3^k(x)) : \mathcal{B} \to \mathbb{R}^3$ by

$$\pi_1(x) = \partial_i\varphi(x)H^i(x), \quad \pi_2^k(x) = H^k(x), \quad \pi_3^k(x) = \partial_k\{\partial_i\varphi(x)H^i(x)\}.$$

Recall that i is fixed; hence we omit it from the notation. As in (5.7) (5.8), we also introduce the $\mathbb{R}^3$ valued Brownian motion $B_{\pi^k} = (B^1_{\pi^k}, B^2_{\pi^k}, B^3_{\pi^k})$ and the corresponding $\mathbb{R}^{3\times 3}$ valued Itô integral $\mathbb{B}_{\pi^k} = (\mathbb{B}^{j\ell}_{\pi^k})_{j,\ell=1,2,3}$.

We can therefore write

$$M_\Delta = \sum_{[t_n,t_{n+1}]\in\Delta} (M_\Delta^{n+1} - M_\Delta^n)$$

where

$$M_\Delta^{n+1} = M_\Delta^n + \pi_1(X(t_n))W(t_n,t_{n+1}) + \sum_{k=1}^{d} \big(\pi_2^k(X(t_n)) \otimes \pi_3^k(X(t_n))\big)\mathbb{W}(t_n,t_{n+1}) - G_k(X(t_n))\Delta t_n.$$

Note that

$$\begin{aligned}
&\sup_x \|\pi_1(x)b\|_{C^\kappa} \lesssim \|b\|_{C^{0,\kappa}} \ \ , \quad \sup_x \big\|\pi_2^k(x)b\big\|_{C^\kappa} \lesssim \|b\|_{C^{0,\kappa}} \ \ , \\
&\sup_x \big\|\pi_3^k(x)b\big\|_{C^\kappa} \lesssim \|b\|_{C^{1,\kappa}} \ \ .
\end{aligned} \tag{6.3}$$

Define the discrete time filtration $F_n = \mathcal{F}_{t_n}$ for $n = 0,\dots,N$. Observe that

$$\mathbf{E}(M_\Delta^{n+1} - M_\Delta^n|F_n) = \mathbf{E}\Big(\pi_1(X(t_n))W(t_n,t_{n+1}) + \sum_{k=1}^{d} \big(\pi_2^k(X(t_n)) \otimes \pi_3^k(X(t_n))\big)\mathbb{W}(t_n,t_{n+1})|F_n\Big)$$

$$- \sum_{k=1}^{d} \mathbf{E}(G_k(X(t_n))\Delta t_n | \mathcal{F}_n) \ .$$

But since $X(t_n)$ is F_n measurable, and $\pi_1(x)W(t_n, t_{n+1})$ is independent of F_n for each fixed x by Lemma 6.2,

$$\begin{aligned}
\mathbf{E}\big(\pi_1(X(t_n))W(t_n,t_{n+1})|F_n\big) &= \mathbf{E}\big(\pi_1(x)W(t_n,t_{n+1})|F_n\big)|_{x=X(t_n)} \\
&= \mathbf{E}\big(\pi_1(x)W(t_n,t_{n+1})\big)|_{x=X(t_n)} \\
&= \mathbf{E}\big(B^1_{\pi^k}(t_n,t_{n+1})\big)|_{x=X(t_n)} = 0
\end{aligned}$$

where we have used Lemma 5.12 to characterize the distribution of $\pi_1(x)W(t_n, t_{n+1})$. Likewise, we have that

$$\begin{aligned}
&\mathbf{E}\big(\pi_2^k(X(t_n)) \otimes \pi_3^k(X(t_n))\big)\mathbb{W}(t_n,t_{n+1})|F_n\big) \\
&\quad = \mathbf{E}\left(\mathbb{B}^{23}_{\pi^k}(t_n,t_{n+1})\right)|_{x=X(t_n)} + \mathfrak{B}(\pi_2^k(x)b, \pi_3^k(x)b)|_{x=X(t_n)} \\
&\quad = \mathfrak{B}\big(b^k(x,\cdot), \partial_k\{b^i(x,\cdot)\partial_i\varphi(x)\}\big)|_{x=X(t_n)} = G_k(X(t_n)) \ .
\end{aligned}$$

Thus M^n_Δ is an F_n-martingale. Moreover, due to the independence of the increments from F_n, we have

$$\begin{aligned}
&\mathbf{E}\big((M^{n+1}_\Delta - M^n_\Delta)^2|F_n\big) \\
&= \mathbf{E}\left(\left(\pi_1(X(t_n))W(t_n,t_{n+1}) + \sum_{k=1}^{d}\left(\pi_2^k(X(t_n)) \otimes \pi_3^k(X(t_n))\right)\mathbb{W}(t_n,t_{n+1}) - G_k(X(t_n))\Delta t_n\right)^2 | \ F_n\right) \\
&= \mathbf{E}\left(\left(\pi_1(x)W(t_n,t_{n+1}) + \sum_{k=1}^{d}\left(\pi_2^k(x) \otimes \pi_3^k(x)\right)\mathbb{W}(t_n,t_{n+1}) - G_k(x)\Delta t_n\right)^2\right)|_{x=X(t_n)} \\
&\lesssim \mathbf{E}\left(\left(\pi_1(x)W(t_n,t_{n+1})\right)^2|_{x=X(t_n)} + \sum_{k=1}^{d}\mathbf{E}\left(\left(\pi_2^k(x) \otimes \pi_3^k(x)\right)\mathbb{W}(t_n,t_{n+1}) - G_k(x)\Delta t_n\right)^2|_{x=X(t_n)} \ .
\end{aligned}$$

Using Lemma 5.12 combined with the identity $G_k(x) = \mathfrak{B}(\pi_2^k(x)b, \pi_3^k(x)b)$, the above formula simplifies to

$$\mathbf{E}\big((M^{n+1}_\Delta - M^n_\Delta)^2|F_n\big) \lesssim \mathbf{E}\left(B^1_{\pi^k}(t_n,t_{n+1})\right)^2|_{x=X(t_n)} + \sum_{k=1}^{d}\mathbf{E}\left(\mathbb{B}^{23}_{\pi^k}(t_n,t_{n+1})\right)^2|_{x=X(t_n)} \ .$$

But by Lemma 5.12, Proposition 2.8(c) and (6.3), we have

$$\begin{aligned}
\mathbf{E}\left(B^1_{\pi^k}(t_n,t_{n+1})\right)^2|_{x=X(t_n)} &= 2\mathfrak{A}(\pi_1(X(t_n))b, \pi_1(X(t_n))b)\Delta t_n \\
&\lesssim \sup_x \|\pi_1(x)b\|^2_{C^\kappa(\Omega,\mathbb{R})}\,\Delta t_n \lesssim \|b\|^2_{C^{1,\kappa}}\,\Delta t_n \lesssim \Delta t_n.
\end{aligned}$$

where we use the shorthand $\mathfrak{A}(v,w) = \frac{1}{2}(\mathfrak{B}(v,w) + \mathfrak{B}(w,v))$. Also, by the Itô isometry, again using Proposition 2.8(c),

$$\mathbf{E}\left(\mathbb{B}^{23}_{\pi}(t_n,t_{n+1})\right)^2|_{x=X(t_n)}$$

$$= \big(\mathfrak{A}(\pi_2(X(t_n))b, \pi_3(X(t_n))b)\big)^2 \int_{t_n}^{t_{n+1}} (r - t_n)dr \lesssim \|b\|_{C^{1,\kappa}}^2 (\Delta t_n)^2 \lesssim |\Delta|\Delta t_n \ .$$

It follows that for $|\Delta| \le 1$,

$$\mathbf{E}\big((M_\Delta^{n+1} - M_\Delta^n)^2 | F_n\big) \lesssim \Delta t_n \ .$$

In particular, $\{M_\Delta^n\}_{n=0}^N$ is an L^2-martingale, with L^2 norm bounded uniformly in $|\Delta| \le 1$. Moreover $M_\Delta = M_\Delta^N$, so this completes the verification of the claim. ☐

Proof of Theorem 4.1. By Theorem 5.1, we see that $x_\varepsilon \to_w X$ along subsubsequences where X solves the RDE (5.2). By Lemma 6.1, X is a weak solution to the SDE (1.5). In particular all subsequences converge to the same limit. The formula for $\mathfrak{B}(v, w)$ follows easily by taking $\mathbf{E}_\mu$ in Assumption 2.1 and applying Assumption 2.2 to obtain convergence of the mean. This completes the proof. ☐

7 Localization

In this section, we lift the localized convergence result Theorem 4.1 to the full convergence result Theorem 2.3.

Let $\eta_R : \mathbb{R}^d \to [0, 1]$ be a smooth cutoff function with

$$\eta_R(x) = \begin{cases} 1 & \text{for } |x| \le R \\ 0 & \text{for } |x| \ge 2R \end{cases} \ .$$

Let a, b satisfy the assumptions of Theorem 2.3 and define $a_R(x, y) = a(x, y)\eta_R(x)$, $b_R(x, y) = b(x, y)\eta_R(x)$. Clearly a_R, b_R satisfy all the requirements of Theorem 4.1. In particular, if we let $x_{\varepsilon,R}$ denote the solution to (1.1) with a, b replaced by a_R, b_R then Theorem 4.1 states that $x_{\varepsilon,R} \to_w X_R$ where X_R satisfies the SDE (1.5) with a, b replaced with a_R, b_R. The following result is the final ingredient required to complete the localization argument.

Lemma 7.1. *Let X_R be the Itô diffusion defined by*

$$dX_R = \tilde{a}_R(X_R)dt + \sigma_R(X_R)dB \quad , \quad X_R(0) = \xi \ ,$$

where the drift and diffusion coefficients $\tilde{a}_R : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma_R : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are given by

$$\tilde{a}_R^i(x) = \int a_R^i(x, y)d\mu(y) + \sum_{k=1}^d \mathfrak{B}(b_R^k(x, \cdot), \partial_k b_R^i(x, \cdot)) \quad , \quad i = 1, \ldots, d \ ,$$

$$(\sigma_R(x)\sigma_R^T(x))^{ij} = \mathfrak{B}(b_R^i(x, \cdot), b_R^j(x, \cdot)) + \mathfrak{B}(b_R^j(x, \cdot), b_R^i(x, \cdot)) \quad , \quad i, j = 1, \ldots, d \ .$$

Then $X_R \to_w X$ in the supnorm topology, as $R \to \infty$.

Proof. Firstly, it is clear that the martingale problem associated with X is well posed. Indeed, from [SV06, Theorem 6.3.4], it is sufficient to obtain the Lipschitz estimate

$$|\tilde{a}(x) - \tilde{a}(z)| + |(\sigma\sigma^T)(x) - (\sigma\sigma^T)(z)| \lesssim |x - z| \ . \tag{7.1}$$

But this is immediate from Proposition 2.8(c) and the regularity of a and b.

By [SV06, Theorem 11.1.4], to prove convergence it is sufficient to show that the coefficients $\tilde{a}_R$ and $\sigma_R\sigma_R^T$ converge uniformly on compact sets to $\tilde{a}$ and $\sigma\sigma^T$ respectively. But $\tilde{a}_R(x) = \tilde{a}(x)$ and $\sigma_R\sigma_R^T(x) = \sigma\sigma^T(x)$ for all $|x| \le R$. Hence, by taking R sufficiently large, convergence on compact sets is immediate. □

Proof of Theorem 2.3. We now show that $x_\varepsilon \to_w X$ in the supnorm topology, as $\varepsilon \to 0$. Fix a closed set $U \subset C([0,T],\mathbb{R}^d)$. By the portmanteau lemma, it suffices to show that

$$\limsup_{\varepsilon\to 0} \mu(x_\varepsilon \in U) \le \mathbf{P}(X \in U) \; . \tag{7.2}$$

For $R > |\xi|$, we let $x_{\varepsilon,R}$ be the solution to (1.1) with a, b replaced by a_R, b_R. By uniqueness and continuity of solutions to ODEs, for each fixed ε, either $x_\varepsilon(t) = x_{\varepsilon,R}(t)$ for all $0 \le t \le T$ or $\sup_{t\le T} |x_{\varepsilon,R}(t)| \ge R$. Thus we have

$$\mu(x_\varepsilon \in U) \le \mu(x_{\varepsilon,R} \in U) + \mu\big(\sup_{t\le T} |x_{\varepsilon,R}(t)| \ge R\big) \; .$$

But, since a_R, b_R satisfy the requirements of Theorem 4.1, for each fixed R we have that $x_{\varepsilon,R} \to_w X_R$ in the supnorm topology as $\varepsilon \to 0$. Since $x \mapsto \sup_{t\le T}|x(\cdot)|$ is a continuous function in the supnorm topology, it follows from the portmanteau lemma that

$$\begin{aligned}\limsup_{\varepsilon\to 0} \mu(x_\varepsilon \in U) &\le \limsup_{\varepsilon\to 0} \mu(x_{\varepsilon,R} \in U) + \limsup_{\varepsilon\to 0} \mu\big(\sup_{t\le T} |x_{\varepsilon,R}(t)| \ge R\big) \\ &\le \mathbf{P}(X_R \in U) + \mathbf{P}\big(\sup_{t\le T} |X_R(t)| \ge R\big) \; .\end{aligned}$$

Taking $\limsup_{R\to\infty}$ on both sides and using Lemma 7.1 (and again the portmanteau lemma),

$$\limsup_{\varepsilon\to 0} \mu(x_\varepsilon \in U) \le \mathbf{P}(X \in U) + \limsup_{R\to\infty} \mathbf{P}\big(\sup_{t\le T} |X_R(t)| \ge R\big) \; .$$

But X_R solves the SDE (1.5) with coefficients $\tilde{a}_R$ and σ_R that are, by an argument identical to (7.1), Lipschitz and bounded. It follows from [Mao07, Theorem 2.4.4] that

$$\mathbf{E} \sup_{t\le T} |X_R(t)| \le K \; , \tag{7.3}$$

where K depends only on T, ξ and $\sup_{x\in\mathbb{R}^d}(|\tilde{a}_R(x)| \vee |\sigma_R(x)|)$. By Proposition 2.8(c),

$$\begin{aligned}|\sigma_R(x)|^2 &\le \sum_{i,j=1}^d 2|\mathfrak{B}(b_R^i(x,\cdot,), b_R^j(x,\cdot,))| \\ &\lesssim \sum_{i,j=1}^d \big\|b_R^i(x,\cdot,)\big\|_{C^\kappa} \big\|b_R^j(x,\cdot,)\big\|_{C^\kappa} \lesssim \|b_R\|^2_{C^{0,\kappa}} \lesssim \|b\|^2_{C^{0,\kappa}}\end{aligned}$$

and we can similarly bound $\sup_x |\tilde{a}_R(x)|$ uniformly in R. It follows that the constant K in (7.3) can be chosen uniformly in R. Thus

$$\limsup_{R\to\infty} \mathbf{P}\big(\sup_{t\le T} |X_R(t)| \ge R\big) \le \limsup_{R\to\infty} \mathbf{E} \sup_{t\le T} |X_R(t)|/R = 0 \; ,$$

which proves (7.2). □

Acknowledgements

The research of DK was supported by ONR grant N00014-12-1-0257. IM was supported in part by the European Advanced Grant *StochExtHomog* (ERC AdG 320977). We are grateful to the referee for helpful comments that led to clarifications of some of the arguments.

References

[Ano67] D. V. ANOSOV. Geodesic flows on closed Riemannian manifolds of negative curvature. *Trudy Mat. Inst. Steklov.* **90**, (1967), 209.

[Bil99] P. BILLINGSLEY. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second ed., 1999. A Wiley-Interscience Publication.

[Bow75] R. BOWEN. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*. Lecture Notes in Mathematics, Vol. 470. Springer-Verlag, Berlin-New York, 1975.

[BR75] R. BOWEN and D. RUELLE. The ergodic theory of Axiom A flows. *Invent. Math.* **29**, (1975), 181–202.

[Dav07] A. M. DAVIE. Differential equations driven by rough paths: an approach via discrete approximation. *Appl. Math. Res. Express.* , no. 2(2007).

[Dol98] D. DOLGOPYAT. On the decay of correlations in Anosov flows. *Ann. of Math. (2)* **147**, (1998), 357–390.

[Dol04] D. DOLGOPYAT. Limit theorems for partially hyperbolic systems. *Trans. Amer. Math. Soc.* **356**, no. 4, (2004), 1637–1689 (electronic).

[Dol05] D. DOLGOPYAT. Averaging and invariant measures. *Mosc. Math. J.* **5**, (2005), 537–576, 742.

[EK86] S. N. ETHIER and T. G. KURTZ. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. URL `http://dx.doi.org/10.1002/9780470316658`. Characterization and convergence.

[FH14] P. FRIZ and M. HAIRER. *A course on rough paths*. Springer (to appear), 2014.

[FV10] P. FRIZ and N. VICTOIR. *Multidimensional stochastic processes as rough paths*, vol. 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.

[GM13] G. A. GOTTWALD and I. MELBOURNE. Homogenization for deterministic maps and multiplicative noise. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **469**, no. 2156, (2013), 20130201, 16.

[Gub04] M. GUBINELLI. Controlling rough paths. *J. Funct. Anal.* **216**, no. 1, (2004), 86–140.

[Kel16] D. KELLY. Rough path recursions and diffusion approximations. *Ann. Appl. Probab.* **26**, no. 1, (2016), 425–461.

[KM16] D. KELLY and I. MELBOURNE. Smooth approximation of stochastic differential equations. *Ann. Probab.* **44**, no. 1, (2016), 479–520.

[LC85] W. A. LIGHT and E. W. CHENEY. *Approximation theory in tensor product spaces*, vol. 1169 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1985.

[Lyo98] T. LYONS. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana* **14**, no. 2, (1998), 215–310.

[Mao07] X. MAO. *Stochastic differential equations and applications*. Woodhead publishing, 2 ed., 2007.

[MS11] I. MELBOURNE and A. M. STUART. A note on diffusion limits of chaotic skew-product flows. *Nonlinearity* **24**, no. 4, (2011), 1361–1367.

[MT04] I. MELBOURNE and A. TÖRÖK. Statistical limit theorems for suspension flows. *Israel J. Math.* **144**, (2004), 191–209.

[MTVE01] A. J. MAJDA, I. TIMOFEYEV, and E. VANDEN EIJNDEN. A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math.* **54**, no. 8, (2001), 891–974.

[PK74] G. C. PAPANICOLAOU and W. KOHLER. Asymptotic theory of mixing stochastic ordinary differential equations. *Comm. Pure Appl. Math.* **27**, (1974), 641–668.

[Pol84] M. POLLICOTT. A complex Ruelle-Perron-Frobenius theorem and two counterexamples. *Ergodic Theory Dynam. Systems* **4**, (1984), 135–146.

[Pol85] M. POLLICOTT. On the rate of mixing of Axiom A flows. *Invent. Math.* **81**, (1985), 413–426.

[PS08] G. A. PAVLIOTIS and A. M. STUART. *Multiscale methods*, vol. 53 of *Texts in Applied Mathematics*. Springer, New York, 2008. Averaging and homogenization.

[Rat73] M. RATNER. The central limit theorem for geodesic flows on n-dimensional manifolds of negative curvature. *Israel J. Math.* **16**, (1973), 181–197.

[Rue78] D. RUELLE. *Thermodynamic formalism*, vol. 5 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Co., Reading, Mass., 1978. The mathematical structures of classical equilibrium statistical mechanics.

[Rue83] D. RUELLE. Flows which do not exponentially mix. *C. R. Acad. Sci. Paris* **296**, (1983), 191–194.

[Sin72] Y. G. SINAĬ. Gibbs measures in ergodic theory. *Russ. Math. Surv.* **27**, (1972), 21–70.

[Sma67] S. SMALE. Differentiable dynamical systems. *Bull. Amer. Math. Soc.* **73**, (1967), 747–817.

[SV06] D. W. STROOCK and S. R. S. VARADHAN. *Multidimensional diffusion processes*. Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1997 edition.

[Tri85] H. TRIEBEL. *Theory of function spaces*. Birkhäuser, 1985.

[Tri06] H. TRIEBEL. *Theory of function spaces. III*, vol. 100 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 2006.

[Wil91] D. WILLIAMS. *Probability with martingales*. Cambridge University Press, 1991.

[You36] L. C. YOUNG. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.* **67**, no. 1, (1936), 251–282.

[You98] L.-S. YOUNG. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math. (2)* **147**, no. 3, (1998), 585–650.

[You99] L.-S. YOUNG. Recurrence times and rates of mixing. *Israel J. Math.* **110**, (1999), 153–188.