L² convergence of smooth approximations of stochastic differential equations with unbounded coefficients

Sahani Pathiraja
School of Mathematics and Statistics, University of New South Wales, Sydney, New South Wales, Australia

ABSTRACT
The aim of this article is to obtain convergence in mean in the uniform topology of piecewise linear approximations of stochastic differential equations (SDEs) with C¹ drift and C² diffusion coefficients with uniformly bounded derivatives. Convergence analyses for such Wong-Zakai approximations most often assume that the coefficients of the SDE are uniformly bounded. Almost sure convergence in the unbounded case can be obtained using now standard rough path techniques, although L⁹ convergence appears yet to be established and is of importance for several applications involving Monte-Carlo approximations. We consider L² convergence in the unbounded case using a combination of traditional stochastic analysis and rough path techniques. We expect our proof technique extend to more general piecewise smooth approximations.

1. Introduction
Given a filtered probability space (Ω, ℱ, ℙ, {ℱₜ}ₜ∈[0,T]) satisfying the usual conditions and a time interval [0, T], T > 0, we are interested in a stochastic process \{Xₜ\}ₜ∈[0,T] given by

\[ Xₜ = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s, \]  

(1.1)

where \(Xₜ \in \mathbb{R}^n\), \(b : \mathbb{R}^n \to \mathbb{R}^{n}\) and \(\sigma : \mathbb{R}^n \to \mathbb{R}^{n \times r}\) satisfy appropriate regularity conditions and \(B\) is a \(r\)-dimensional Wiener process. The stochastic integrals in the rhs of (1.1) are interpreted in the Ito sense. It is often necessary in many applications to utilize smooth approximations of the Brownian motion and to consider the corresponding solutions to (1.1). To this end, let \(Π_d\) denote a partition of the time interval \([0, T]\) consisting of \(d\) units, \(d \in \mathbb{N}\) and associated filtration \(ℱ_d := \sigma(B_t : t \in Π_d)\). In the case where a sequence of approximations \((B_d^t)_{d \geq 1}\) are \(ℱ_d\)-martingales and satisfy a uniform tightness condition [1], it is known that (1.1) with \(dB_d^t\) instead of \(dB_t\) converges in distribution to (1.1)¹ [2]. Many piecewise smooth approximations do not satisfy these conditions so that the limiting process involves additional correction terms. More precisely, consider \(B^d(ω)\) a piecewise smooth approximation of the Brownian motion \(B(ω)\), with \(B\) as its limiting process, and the ODE

CONTACT Sahani Pathiraja s.pathiraja@unsw.edu.au School of Mathematics and Statistics, University of New South Wales, Sydney, NSW, Australia.

¹This holds for more general semimartingale drivers, but for the purposes of this article, we focus on the case of Wiener processes.

© 2023 Taylor & Francis Group, LLC
\[ dX_t^d = b(X_t^d)dt + \sigma(X_t^d)\dot{B}_t^d dt. \] (1.2)

In the above and throughout the article, we drop the \( \omega \) for notational ease. It is now well-known that under certain regularity conditions on the coefficients and conditions on the construction of \( B_t^d \), solutions of (1.2) do not converge to solutions of (1.1) as the mesh size \( \delta \to 0 \), but rather to a Stratonovich SDE plus an anti-symmetric drift term that depends on how the approximation is constructed, that is,

\[ dX_t = b(X_t)dt + \sigma(X_t) \circ dB_t + \sum_{n,m=1}^\nu S_{n,m}[A_nX_t, A_mX_t]dt, \] (1.3)

where the notation \( C_{ij} \) denotes the \( i,j \)-th entry of a matrix \( C \), \( \,[\cdot, \cdot] \) denotes the Lie bracket and the operator \( A_n \) is defined by \( A_n x := \sum_i \sigma_{i,n}(x) \frac{\partial}{\partial x_i} \forall n = 1, 2, \cdots r \). One important condition for the previous statement is the existence of a skew symmetric matrix \( S \) as the limit of the area process

\[ \frac{1}{2\delta} \mathbb{E} \left[ \int_0^\delta \left( B_s^d \otimes \dot{B}_s^d - \dot{B}_s^d \otimes B_s^d \right) ds \right]. \]

For reference, see Theorem 7.2 in [3] which captures much of the pioneering work on this matter for the multi-dimensional case by for example [4], [5], [6], [7], [8]. Such piecewise smooth approximation models are relevant in many areas of mathematical modelling where it is necessary to work with ODEs with random fluctuations, for example for extracting macroscopic models from detailed microscopic stochastic dynamics (e.g., [9], [10]). Second such approximations are useful because they are naturally tied to the Stratonovich interpretation of stochastic integrals, in particular, when the anti-symmetric term in (1.3) is zero, which holds in the scalar case, when the vector fields \( A_n \) commute, or when \( S \equiv 0 \) which occurs for piecewise linear approximations or mollifiers [3]. Continuity of the solutions of ODEs driven by such smooth approximations with respect to the driving path in the uniform topology is ensured by the Doss-Sussman theorem. This property is advantageous for various applications, in particular for robust filtering [11]. From a numerical approximation point of view, it is possible to utilize the many techniques available for ODEs (rather than SDEs) to then construct time-discrete approximations to (1.3). This is in contrast to more direct numerical approximation schemes applied to Ito SDEs (1.1) like the Euler-Maruyama or Milstein which converge strongly to (1.1) as the time step tends to zero, that is, without any additional correction term. The convergence analysis of Wong-Zakai approximations is complicated by the presence of anticipative integrands, so that the traditional martingale techniques cannot be directly applied as in the analysis of Euler-Maruyama and related schemes. A potential route forward is the recent work [12,13] where an Alekseev-Groebner formula of Skorokhod type is derived to obtain uniform in time \( L^p \) convergence of the Euler-Maruyama discretization scheme.

The literature on convergence analyses of piecewise smooth approximations for both SDEs and SPDEs (e.g., [14], [15], [16], [17], [18], [19]) is enormous, starting from the seminal work of [20] where convergence in probability of (1.2) to (1.3) with \( a = 0 \) (antisymmetric term) in the scalar case was established. Much of the initial work in this direction for SDEs relies on standard stochastic analysis techniques, loosely speaking, by introducing time shifts so that martingale inequalities can be applied (e.g., [21], [3]). In these works, uniform boundedness of the coefficients is crucial in order to control the extra terms introduced by the time shift.
More recently, Liu and Han [22] showed convergence in probability in the Skorokhod M1 topology of Wong-Zakai approximations of scalar SDEs driven by a general semimartingale (thereby extending the earlier work of [23] and [24]), again for uniformly bounded coefficients. Extensions to unbounded coefficients are much more restricted. Hairer and Pardoux [25] established convergence of piecewise smooth approximations to Ito SDEs with linear diffusion coefficients, when the product between the diffusion coefficient and the smoothed noise term is interpreted as a Wick product. This scenario does not correspond to the classical interpretation that renders a Stratonovich limit, which is the focus of this article. In the context of SPDEs, many Wong-Zakai-type convergence results also require boundedness of drift and diffusion coefficients (e.g., [26]) or bounded domains (e.g., [14]), although there have been extensions that partially relax this condition, for example, Gyöngy and Raúl Stinga [27] who consider linear SPDEs. Typically, relaxing the boundedness constraint on the diffusion coefficient is more challenging than the drift coefficient; many existing results still require some boundedness conditions on the diffusion coefficient even if not on the drift coefficient (see e.g., [28] for SDEs, [17] and [29] for SPDEs).

Over the last couple of decades, the theory of rough paths pioneered by Lyons [30] has been utilized to obtain convergence results for a broader class of drivers and smooth approximations (see e.g., [31], [32], [33], [34]). After defining an appropriate lift of the driving path (e.g., Stratonovich lift in the case of piecewise linear approximations), and by ensuring that the pathwise solution of the rough differential equation (RDE) coincides with the limiting SDE, convergence is easily established using continuity of the Ito-Lyons map (i.e., the solution map corresponding to the RDE, as a function of the initial condition and driver) (see e.g., Theorem 9.3 in [34]). Generally, weak convergence of smooth approximations of the driver then imply weak convergence of the (random) ODE solutions. Strong convergence results can be obtained in specific situations; for example when $b$ is $\text{Lip}(1+\epsilon)$ and $\sigma$ is $\text{Lip}(2+\epsilon)$ in the sense of Stein, for dyadic piecewise linear approximations of Brownian motion [33] (see also [34] for the non-dyadic case). A common feature of these works is the lipschitz assumption as in the Stein sense [31], [33], therefore requiring uniform boundedness of the coefficients. Kelly and Melbourne [35] use rough path techniques to obtain weak convergence under much weaker conditions ($C^{1+\epsilon}$ and $C^{2+\epsilon}$ for the drift and diffusion coefficients, respectively) and for a broad range of smooth approximations satisfying a weak invariance principle. As noted in [34], it is in general not straightforward to use rough path techniques for $L^q$ convergence, although Chapter 10 of [34] and also [36] presents some analysis in this direction in the case of Gaussian processes. To summarize, the aforementioned results apply to the case of bounded coefficients (either on both the drift and diffusion coefficient or on the diffusion coefficient only) and/or weak convergence or almost sure convergence.

Our aim is to obtain $L^2$ rates of convergence of piecewise smooth approximations of (1.1) with unbounded drift and diffusion coefficients. $L^q$ convergence of numerical approximations of SDEs is often of central importance in many Monte Carlo sampling methods such as in Markov Chain Monte Carlo [37] and in control-type sequential Monte Carlo methods which rely on empirical approximations of McKean-Vlasov SDEs [38]. Here we present a simple approach that combines elements of stochastic analysis and rough path techniques to achieve this result, specifically, focusing on nested piecewise linear approximations of the Brownian path. That is, $\Pi_d$ is a partition of the time interval $[0, T]$ with associated mesh size $\delta_d$ such that $\Pi_d \subset \Pi_{d+1}$ $\forall$ $d \in \mathbb{N}$ and $\delta_d \to 0$ as $d \to \infty$ and
\[ B^d_t = B^d_t + \frac{t-t_i}{\delta_d} (B^d_{t_{i+1}} - B^d_{t_i}), \quad \text{for } t \in (t_i, t_{i+1}). \]

We expect that our result can be extended to more general approximations using the same proof structure.

### 1.1. Statement of main result

Before stating our main result, we state the main assumptions utilized throughout the article.

**Assumption 1.1.** \( b \in C^1(\mathbb{R}^v, \mathbb{R}^v) \) and \( \sigma \in C^2(\mathbb{R}^v, \mathbb{R}^{v \times r}) \) and there exists some constant \( M_c \) such that

\[
\sup_{x \in \mathbb{R}^v, j \in \{1,2,\ldots,r\}} |\nabla b| \vee |\nabla \sigma_{.,j}| \vee |\nabla^2 \sigma_{.,j}| \leq M_c
\]

where \( \sigma_{.,j} \) denotes the \( j \)th column of the matrix \( \sigma \).

**Assumption 1.2.** The vector function \( \Sigma : \mathbb{R}^v \rightarrow \mathbb{R}^v \) with \( i \)-th entry given by

\[
\Sigma_i(x) := \sum_{k,m=1}^{r} \sum_{l=1}^{v} \sigma_{l,k}(x) \frac{\partial}{\partial x_l} \sigma_{i,m}(x)
\]

is globally Lipschitz-continuous.

These assumptions ensure well-posedness of the ODE and limiting SDE, as well as the RDE corresponding to the Stratonovich lift of \( B^d \). Assumption 1.2 in particular plays a crucial role in obtaining a pathwise control on solutions to the RDE, see Lemma 3.3.

**Theorem 1.1.** Assume (1.1) and (1.2). Suppose \( X^d_t \) is the unique solution of

\[
dX^d_t = b(X^d_t) dt + \sigma(X^d_t) dB^d_t \tag{1.4}
\]

and \( X_t \) is the unique strong solution of

\[
dX_t = b(X_t) dt + \sigma(X_t) \circ dB_t. \tag{1.5}
\]

with \( X^d_0 = X_0 = x_0 \) for some fixed but arbitrary \( x_0 \). It holds that for every \( T > 0 \),

\[
\lim_{d \to \infty} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \left| X_t - X^d_t \right|^2 \right] = 0.
\]

**Remark 1.3.** Almost sure convergence under Assumptions 1.1 and 1.2 can be obtained using continuity of the Ito-Lyons map established in [39] and standard rough path arguments.

### 2. Notation and background on rough path theory

We briefly summarize some of the necessary background on rough path theory that will be utilized in Lemma 3.3. This summary is by no means exhaustive and much of the algebraic and geometric technicalities are left out as they are not needed for understanding the main results. We refer to the extensive literature on rough paths (e.g., [34], [40], [41]) for further details.
A rough path framework allows for developing pathwise solutions to SDEs by considering a fixed realization of the driver (i.e., $B(\omega)$ for fixed $\omega$, in our case), so that the differential equation is seen rather as a controlled ODE. The irregularity of the driver means that special care is needed to define solutions to such ODEs. The fundamental insight from rough path theory is that $n$-order iterated integrals of the form $\int_{t}^{1} X_{t}^{x_{1}} \int_{s}^{x_{2}} X_{s}^{x_{2}} \cdots \int_{0}^{x_{n}} X_{0}^{x_{n}} \, \mathrm{d}x_{n}$ with $n \leq \left\lfloor \frac{1}{\alpha} \right\rfloor$ for a driver $X$ of $\alpha$-Hoelder regularity (and an apriori interpretation of these iterated integrals) is sufficient to make these differential equations well-defined in a deterministic sense. The iterated integrals or “signature” can be seen heuristically as arising from a Taylor integrals) is sufficient to make these differential equations well-defined in a deterministic sense. The iterated integrals or “signature” can be seen heuristically as arising from a Taylor expansion [42]. Since Brownian paths are $\alpha$-Hoelder regular with $\alpha \in (1/3, 1/2)$, it is only necessary to consider the second-order iterated integral taking the form $\int_{s}^{t}(B_{t_{1}} - B_{s}) \otimes dB_{t_{1}} \in \mathbb{R}^{r} \otimes \mathbb{R}^{r}$. There is still some ambiguity as to how this stochastic integral should be interpreted; a Stratonovich (Ito) interpretation corresponds to the so-called Stratonovich (Ito)-lift or enhancement. We use the notation $X := (X^{1}, X^{2})$ to denote the rough path lift of a continuous path $X : [0, T] \rightarrow \mathbb{R}^{r}$, where $X^{1} := X$ and $X^{2} : [0, T]^{2} \rightarrow \mathbb{R}^{r} \otimes \mathbb{R}^{r}$ denotes the second-order process such that $X_{s,t}^{2} := \int_{s}^{t} X_{s,t_{1}} \otimes dB_{t_{1}}$ where $X_{s,t} := X_{t} - X_{s}$. We are concerned with Brownian rough paths; these paths belong to the space of $\alpha$-Hoelder rough paths $C^{\alpha}([0, T], \mathbb{R}^{r})$ such that

$$|X|_{\alpha} := \sup_{0 \leq s < t \leq T} \max \left\{ \frac{|X_{s,t}^{1}|}{|t - s|^{\alpha}}, \frac{|X_{s,t}^{2}|}{|t - s|^{2\alpha}} \right\}$$

is finite and Chen’s relation is satisfied. For an $\alpha$-Hoelder continuous path $X$ taking values in $\mathbb{R}^{m}$, we define

$$|X|_{\alpha} := \sup_{0 \leq s < t \leq T} \frac{|X_{s,t}|}{|t - s|^{\alpha}}.$$  

Throughout the article, we denote by $B$ the Stratonovich lift of $B$, which is a geometric rough path. These have the important property that there exists a sequence of smooth (or piecewise smooth) rough paths converging uniformly on $[0, T]$ in the rough path topology to the lifted path (see e.g., Proposition 2.5 in [34]). Likewise, we denote by $B^{d}$ the Stratonovich lift of a piecewise smooth approximation $B^{d}$.

It is classical that the solution of (1.2) coincides with the solution (when it exists and is unique) of the following rough differential equation (RDE) driven by $B^{d}$

$$\hat{X}_{t}^{d} = \hat{X}_{0} + \int_{0}^{t} b(\hat{X}_{s}^{d}) \, \mathrm{d}s + \int_{0}^{t} \sigma(\hat{X}_{s}^{d}) B_{s}^{d},$$

(see e.g., Section 9 in [39]). Roughly speaking, the integrand must be “smooth enough” to counteract the irregularity of the driver. Several concepts of solutions to RDEs have been developed, for example solution in the sense of Lyons [30], Gubinelli’s controlled paths formulation [44], Davie’s formulation in terms of a Taylor expansion or Euler discretization [45], and also more recently the coordinate independent definition from Bailleul [46]. There exist equivalencies between various solution concepts [47], for example solutions in the sense of Lyons and Davie are equivalent when $\sigma$ is $\gamma$-Hoelder continuous with $\alpha(2 + \gamma) > 1$, see [39]. Here we consider solutions in the sense of Davie [48] that is where $\hat{X}^{d}$ is a continuous

\footnote{the notation $\rho_{i}$ is used to refer to the $i$th index of a permutation of the set $\{1, 2, \cdots r\}$}
path from $[0, T] \to \mathbb{R}^v$ of finite $p$-variation such that for some constant $L$,
\[
\left| \hat{X}_t^d - \hat{X}_s^d - \int_s^t b(\hat{X}_u) du - \sigma(\hat{X}_s) \mathbf{B}_{s,t}^{d,1} - \Sigma(\hat{X}_s) \mathbf{B}_{s,t}^{d,2} \right| \leq L|t - s|^{\theta} \quad \forall \ 0 \leq s < t \leq T,
\]
where $\theta := 3\alpha$. With a slight abuse of notation, we denote by $\sigma$ a linear vector field from $\mathbb{R}^v \to \mathcal{L}(\mathbb{R}', \mathbb{R}^v)$ and $\Sigma(x)$ is also a linear vector field from $\mathbb{R}^v \to \mathcal{L}(\mathbb{R}' \otimes \mathbb{R}', \mathbb{R}^v)$.

3. Proof of Theorem 1.1

The proof involves utilizing localization arguments to extend well-known convergence results in the case of $C^2_b$ coefficients (i.e., the space of twice continuously differentiable uniformly bounded functions with bounded derivatives) as in [3] to unbounded coefficients. The main goal is to close the commutative diagram in Figure 1. We do so by establishing uniform in the approximation parameter $d$ and pointwise in the localization parameter $n$ convergence results which thereby permit the interchange of limits in $n$ and $d$ using the Moore-Osgood theorem. We also develop uniform in $d$ pathwise moment bounds on $X^d$ in Lemma 3.4 using rough path techniques, which plays a crucial role in establishing uniform convergence.

We start by considering a localized form of the limiting SDE in Itô form. For all $n \in \mathbb{N}$, let $S_n := \{x \in \mathbb{R}^v \mid |x|^2 \leq n\}$. Consider $X^n_t$ which is the unique strong solution of
\[
dX^n_t = b^n(X^n_t)dt + \Sigma^n(X^n_t)dt + \sigma^n(X^n_t)dB_t
\]
where
\[
b^n(x) = \begin{cases} b(x), & x \in S_n \\ f_b(x), & x \notin S_n \end{cases}
\]
where $f_b \in C^1_b(\mathbb{R}^v, \mathbb{R}^v)$ is chosen such that $b^n \in C^1_b(\mathbb{R}^v, \mathbb{R}^v)$. The functions $\sigma^n$ and $\Sigma^n$ are constructed in a similar manner. Consider also the following ODE utilizing the smooth approximation of the Brownian motion as described above
\[
dX^{d,n}_t = b^n(X^{d,n}_t)dt + \sigma^n(X^{d,n}_t)dB^d_t.
\]
We will assume throughout that $X^0 = X^{d,n}_0 = x_0$ where $x_0$ is fixed but arbitrary. The following lemma utilizes standard results to establish convergence along the clockwise route starting from $X^{d,n}$ in Figure 1.

Lemma 3.1. For every $T > 0$ and $X^{d,n}_t$ satisfying (3.1) and $X_t$ satisfying (1.5),
\[
\lim_{n \to \infty} \lim_{d \to \infty} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X^{d,n}_t - X^n_t|^2 + \sup_{0 \leq t \leq T} |X^n_t - X_t|^2 \right] = 0.
\]

Figure 1. Limit diagram. Convergence is always $L^2$ with respect to the uniform metric over $[0, T]$. 
Proof. Since the coefficients of (3.1) and (1.5) are continuously differentiable and uniformly bounded with bounded derivatives, we have from Theorem 7.2 in [3] that for every $T > 0$,

$$\lim_{d \to \infty} E \left[ \sup_{0 \leq t \leq T} |X_{t}^{d,n} - X_{t}^{n}|^2 \right] = 0, \quad \forall \ n \in \mathbb{N}. \quad (3.4)$$

Since the coefficients $b^n, \sigma^n, b, \sigma$ are globally Lipschitz continuous, we have from standard localization arguments that

$$\lim_{n \to \infty} E \left[ \sup_{0 \leq t \leq T} |X_{t}^{n} - X_{t}^{0}|^2 \right]$$

which gives the desired result.

As discussed in Section 2, the solution of the RDE (2.1) coincides with $X_{t}^{d}$, so that we can now work with the rich toolbox offered by rough path theory to obtain a pathwise moment bound on $X_{t}^{d}$, uniformly in $d$ (Lemma 3.3). An analog to Lemma 3.3 using stochastic analysis techniques in the case of bounded coefficients can be found in Lemma 7.2 in [3], where the now standard approach of applying a time shift to obtain martingales has proven difficult to adapt to the case of unbounded coefficients. The rough path framework allows to side-step this difficulty, particularly since dealing with RDE solutions in the sense of Davie means we can avoid having to directly control the integrals driven by $B_{d}^{t}$.

First, existence and uniqueness of solutions (in the sense of Davie, and equivalently in the sense of Proposition 3 and Theorem 1 for driftless RDEs in [39]). We obtain the following bound on the solution to (2.1) in the small time horizon $\mu := T^{\alpha} \leq K^{\alpha}$ by a straightforward extension of the boundedness of solutions to driftless RDEs in Proposition 2 of [39], where the crucial point is to utilize the sewing lemma to obtain an expression for $L$. The proof of the following Lemma can be found in Appendix.

**Lemma 3.2. Boundedness of solution to (2.1) in the short time horizon.** Under Assumptions 1.1 and 1.2, and with $\alpha \in (1/3, 1/2)$ for Brownian motion and recalling the notation $\mu := T^{\alpha}$, $\theta := 3\alpha$, we have that

$$|\dot{X}_{t}^{d}|_{\alpha} \leq C_{15}(\mu) + C_{16}(\mu)|x_{0}|, \quad (3.5)$$

where

$$C_{15}(\mu) := C_{13} \left( C_{14} \mu^{2} |\nabla \Sigma|_{\infty} \left( (1 + \mu^{2})|B_{d}^{\alpha}|_{\alpha}^2 + |B_{d}^{\alpha}|_{\alpha} \right) + |\nabla \sigma|_{\infty}^{2} |B_{d}^{\alpha}|_{\alpha}^{\alpha} |\sigma(0)| \right)$$

$$+ |\sigma(0)| \left( |B_{d}^{\alpha}|_{\alpha} + |B_{d}^{\alpha}|_{\alpha} |\nabla \sigma|_{\infty}^{\alpha} \mu + \mu^{\frac{1}{\alpha}-1} \right)$$

$$C_{16}(\mu) := C_{13} |\nabla \sigma|_{\infty} \left( C_{14} \mu^{2} |\nabla \sigma|_{\infty}^{2} |B_{d}^{\alpha}|_{\alpha}^2 + |B_{d}^{\alpha}|_{\alpha} + \mu |\nabla \sigma|_{\infty} |B_{d}^{\alpha}|_{\alpha} + \mu^{\frac{1}{\alpha}-1} \right)$$

$$C_{13} := \frac{1}{(1 - 3M)(1 - K_{2})}$$

$$C_{14} := \frac{K}{1 - K_{2}}$$

for a small enough time horizon $\mu$ satisfying

$$|B_{d}^{\alpha}|_{\alpha} |\nabla \Sigma|_{\infty} \mu^{2} \leq M \quad (3.6a)$$
\[ |B_d|_\alpha \nabla \sigma |_{\infty} \mu \leq M \] (3.6b)
\[ |\nabla b|_{\infty} \mu^{\frac{1}{\alpha}} \leq M \] (3.6c)
\[ |\nabla \sigma|_{\infty} |B_d|_\alpha \mu \leq \frac{K_2}{K} \] (3.6d)
\[ \mu^2 (C_4(\mu) + C_3(\mu)) \leq \frac{K_2 (1 - 3M)(1 - K_2)}{K} \] (3.6e)

with
\[ C_3(\mu) := |\nabla \Sigma|_{\infty} \left( (1 + \mu^2) |B_d|_\alpha^2 + |B_d|_\alpha \right) \]
\[ C_4(\mu) := |\nabla \sigma|_{\infty}^2 |B_d|_\alpha^2 \mu \]

and where \( 0 < M < \frac{1}{3} \), \( 0 < K_2 < 1 \) and \( K = f(\theta) \).

**Lemma 3.3. Uniform estimates on the smooth approximation** Under Assumptions 1.1 and 1.2, and when \( x_0 \) is a fixed value, it holds that

\[ \sup_{d \in \mathbb{N}} \mathbb{E} \left[ \sup_{t \in [0,T]} |X^d_t|^2 \right] < \infty \]

**Proof.** Before extending to the large time horizon, we require a precise expression for \( K^* \) based on the conditions in (3.6). In particular,

\[ \mu \leq \min \left( \frac{1}{|B_d|_\alpha^{1/2}}, \frac{1}{|B_d|_\alpha}, 1 \right) C(M, |\nabla \Sigma|_{\infty}, |\nabla b|_{\infty}, |\nabla \sigma|_{\infty}, \alpha) \] (3.7)

ensures \( \mu \) satisfies (3.6a)–(3.6d). To analyze the final condition (3.6e), let \( K_3 := \frac{K_2 (1 - K_2)}{K} (1 - 3M) \) and \( a := |\nabla \sigma|_{\infty} \) and \( \beta := |\nabla \Sigma|_{\infty} \). We then require

\[ \mu^3 a^2 |B_d|_\alpha^2 + \beta \left( |B_d|_\alpha^2 + |B_d|_\alpha \right) \mu^2 + \beta |B_d|_\alpha^2 \mu^4 \leq K_3. \] (3.8)

It is not difficult to see that (3.8) holds in the case \( \mu \geq 1 \) when

\[ \left( a^2 |B_d|_\alpha^2 + \beta \left( 2 |B_d|_\alpha^2 + |B_d|_\alpha \right) \right) \mu^4 \leq K_3 \]

and in the case \( \mu < 1 \) when

\[ \left( a^2 |B_d|_\alpha^2 + \beta \left( 2 |B_d|_\alpha^2 + |B_d|_\alpha \right) \right) \mu^2 \leq K_3. \]

Together, these conditions imply that we require

\[ \mu \leq \min \left( \left( \frac{K_3}{(a^2 + 2\beta)|B_d|_\alpha^2 + |B_d|_\alpha} \right)^{1/2}, \left( \frac{K_3}{(a^2 + 2\beta)|B_d|_\alpha^2 + |B_d|_\alpha} \right)^{1/4} \right). \]

Combining the above with (3.7) means that (3.6a)–(3.6e) are satisfied when

\[ \mu \leq \min \left( \frac{1}{(a^2 + 2\beta)|B_d|_\alpha^2 + |B_d|_\alpha}^{1/2}, \frac{1}{(a^2 + 2\beta)|B_d|_\alpha^2 + |B_d|_\alpha}^{1/4}, \frac{1}{|B_d|_\alpha^{1/2}}, \frac{1}{|B_d|_\alpha}, 1 \right) \times C(M, |\nabla \Sigma|_{\infty}, |\nabla b|_{\infty}, |\nabla \sigma|_{\infty}, K, K_2, \alpha). \]
Again, considering the cases $|B^d|_\alpha \leq 1$ and $|B^d|_\alpha > 1$ separately we obtain the following expression

$$K^* = \frac{1}{1 + |B^d|_\alpha^{\alpha}} C^*(M, |\nabla \sigma|_\infty, |\nabla b|_\infty, |\nabla \sigma|_\infty, K, K_2, \alpha).$$

(3.9)

Furthermore, (3.6b) and (3.6c) imply that

$$C_{16}(\mu) \leq C_{13} |\nabla \sigma|_\infty \left( |B^d|_\alpha + C_{14} M^2 + M + C_{17} \right)$$

$$:= C_{18} |B^d|_\alpha + C_{19}$$

where $C_{17} := \left( \frac{M}{|\nabla \sigma|_\infty} \right)^{1-\alpha}$. Likewise, (3.6a), (3.6b), and (3.6c) imply

$$C_{15}(\mu) \leq C_{13} \left( (C_{14} M + |\sigma(0)|) |B^d|_\alpha + 1 + C_{14} M \left( 1 + \frac{M}{|\nabla \sigma|_\infty} \right) 
+ C_{14} |\sigma(0)| M^2 + |\sigma(0)|(M + C_{17}) \right)$$

$$:= C_{20} |B^d|_\alpha + C_{21}.$$  

Therefore, when $\mu \leq K^*$, we have that

$$|\hat{X}^d|_\alpha \leq C_{20} |B^d|_\alpha + C_{21} + |x_0| \left( C_{18} |B^d|_\alpha + C_{19} \right).$$

Then by Proposition 7 in [39],

$$\sup_{t \in [0,T]} |\hat{X}^d| \leq R(T) |x_0| + R(T) \frac{C_{20} |B^d|_\alpha + C_{21}}{C_{18} |B^d|_\alpha + C_{19}},$$

(3.10)

with

$$R(T) = \exp \left( C_{16}(\mu) \left( 1 + \frac{1}{K^*} \right)^{1-\alpha} \right) \exp(\max\{T, \mu\})$$

$$\leq \exp \left( C_{16}(\mu) \left( 1 + \frac{1}{K^*} \right) \right) \exp(\max\{T, \mu\}),$$

since $\frac{1}{3} < \alpha < \frac{1}{2}$. Furthermore,

$$\frac{C_{20} |B^d|_\alpha + C_{21}}{C_{18} |B^d|_\alpha + C_{19}} \leq \frac{C_{20}}{C_{18}} + \frac{C_{21}}{C_{19}}$$

$$:= C_{22}$$

and also using (3.9) we have

$$\exp \left( C_{16}(\mu) \left( 1 + \frac{1}{K^*} \right) \right) \leq \exp(C_{2}^* C_{19}) \cdot \exp \left( C_{2}^* C_{18} |B^d|_\alpha \right),$$

where $C_{2}^* := \frac{C_{2}+1}{C_{2}}$. Finally, we can conclude using (3.10) that
\[
\sup_{t \in [0, T]} |\tilde{X}_t^d|^2 \leq \exp \left( 2C_2^*C_{19} \exp(2 \max\{T, \mu\}) (|x_0| + C_{22})^2 \right) \exp \left( 2C_2^*C_{18}|B^d|_\alpha \right)
\]
\[
\leq \exp \left( 2C_2^*C_{19} + 2 \max\{T, \mu\} + 2C_2^*C_{18}|B^d|_\alpha \right) (|x_0| + C_{22})^2
\]
\[
:= C_{24} \exp(2C_2^*C_{18}|B^d|_\alpha)
\]

where \(C_2^*, C_{18}, C_{19}, C_{22}\) are positive constants depending only on \(x_0, \sigma(0), M, |\nabla \Sigma|_\infty, |\nabla b|_\infty, |\nabla \sigma|_\infty, K, K_2, \alpha\). Finally, for nested piecewise linear approximations, Theorem 13.19 in [40] gives us that
\[
\exp(C_2^*C_{18}|B^d|_\alpha) \leq \exp(C_2^*C_{18}M_g) < \infty \quad \forall \ d \in \mathbb{N}
\]
where \(M_g\) is a positive random variable with Gaussian tails independent of \(d\) and \(M_g < \infty\) almost surely. Taking expectation of both sides gives the desired result.

We are now ready to show that solutions of localized ODEs converge to the solution of the ODE of interest, uniformly in the approximation parameter.

**Lemma 3.4. Uniform convergence in \(d\).** For every \(T > 0\) and \(X_t^{d,n}\) satisfying (3.1) and \(X_t^d\) satisfying (1.4),
\[
\lim_{n \to \infty} \sup_{d \in \mathbb{N}} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_t^{d,n} - X_t^d|^2 \right] = 0 \quad (3.11)
\]

**Proof.** For each fixed \(d\), define
\[
\tau_n^d := \inf\{t \geq \delta_d | |X_t^{d,n}|^2 \notin S_n \cup |X_t^{d}|^2 \notin S\} - \delta_d
\]
which is a stopping time with respect to \(\mathcal{F}_t\). Define
\[
S_T^{d,n} := \sup_{0 \leq t \leq T} |X_t^{d,n} - X_t^d|^2.
\]
It holds that
\[
\mathbb{E}[S_T^{d,n}] = \mathbb{E}[\mathbf{1}_{\tau_n^d > T} S_T^{d,n}] + \mathbb{E}[\mathbf{1}_{\tau_n^d \leq T} S_T^{d,n}]
\]
and by H"older inequality,
\[
\mathbb{E}[\mathbf{1}_{\tau_n^d \leq T} S_T^{d,n}] \leq \mathbb{E}[(\mathbf{1}_{\tau_n^d \leq T})^p]^{1/p} \mathbb{E}[(S_T^{d,n})^q]^{1/q}
\]
\[
= (\mathbb{P}(\tau_n^d \leq T)^{1/p} \mathbb{E}[(S_T^{d,n})^q]^{1/q},
\]
where \(\frac{1}{q} + \frac{1}{p} = 1\). For the case \(\tau_n^d > T\), we start by considering the Stratonovich step 2 lift of \(B_t^d\) as in Lemma 3.3. We can again work with the RDEs whose solutions coincide with the ODEs. Continuity of the Ito-Lyons map under Assumptions 1.1 and 1.2 is established in Theorem 1 in [39], which implies that for all \(t < \tau_n^d\),
\[
|X_t^{d,n} - X_t^d| = 0 \quad \text{a.s.}
\]
since both \(X_t^{d,n}\) and \(X_t^d\) are driven by the same \(B_t^d\) and \(X_0^{d,n} = X_0^d\). Furthermore, it follows directly that \(\mathbb{E}[S_T^{d,n}] = 0\) when \(\tau_n^d > T\) for every fixed \(n, d\) (and uniformly in \(d\)). Therefore, we have that for fixed \(n\),
\[
\sup_{d \in \mathbb{N}} \mathbb{E}[S_T^{d,n}] \leq \sup_{d \in \mathbb{N}} (\mathbb{P}(\tau_n^d \leq T))^{1/p} \mathbb{E}[(S_T^{d,n})^q]^{1/q}.
\]
Lemma 3.3 clearly holds for \(X_{t}^{d,n}\) uniformly in \(n\), therefore we have that
\[
\mathbb{E}[(S_{dn})^{q}]^{1/q} \leq \left( \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_{t}^{d,n}|^{2q} \right] + \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_{t}^{d}|^{2q} \right] \right)^{1/q}
\]
\[
< \infty
\]
uniformly in \(n\) and \(d\). Furthermore, by standard arguments we have that
\[
\mathbb{P}(\tau_{n}^{d} \leq T) = \mathbb{P} \left( \sup_{t \in [0,T]} |X_{t}^{d}| > n \right)
\]
\[
= \mathbb{E} \left[ \mathbb{1}_{\{\sup_{t \in [0,T]} |X_{t}^{d}| > n \}} \right]
\]
\[
\leq \frac{1}{n} \mathbb{E} \left[ \sup_{t \in [0,T]} |X_{t}^{d}|^{2} \right]
\]
Then taking sup wrt \(d\) on both sides in the above, and using \(\sup_{d \in \mathbb{N}} \mathbb{E} \left[ \sup_{t \in [0,T]} |X_{t}^{d}|^{2} \right] < \infty\) from Lemma 3.3 implies
\[
\lim_{n \to \infty} \sup_{d \in \mathbb{N}} \mathbb{P}(\tau_{n}^{d} \leq T))^{1/p} = 0.
\]
Combining all gives
\[
\lim_{n \to \infty} \sup_{d \in \mathbb{N}} \mathbb{E}[S_{dn}] \leq \left( \lim_{n \to \infty} \sup_{d \in \mathbb{N}} \mathbb{P}(\tau_{n}^{d} \leq T))^{1/p} \right) \left( \lim_{n \to \infty} \sup_{d \in \mathbb{N}} \mathbb{E}[(S_{dn})^{q}]^{1/q} \right)
\]
\[
= 0.
\]
since both limits are finite.

The final ingredient is the pointwise convergence statement, as stated in the following lemma.

Lemma 3.5. **Pointwise convergence in** \(n\). For every \(T > 0\) and \(X_{t}^{d,n}\) satisfying (3.2) and \(X_{t}^{n}\) satisfying (3.1),
\[
\lim_{d \to \infty} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_{t}^{d,n} - X_{t}^{n}|^{2} \right] = 0 \quad \forall n \in \mathbb{N}
\]
(3.12)

**Proof.** Follows trivially from Theorem 7.2 in [3] since for every fixed \(n\), \(b^{n}\), and \(\sigma^{n}\) are uniformly bounded.

Since \(\delta_{2}(X, Y) := \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_{t} - Y_{t}|^{2} \right] \) is a distance metric on the space of \(L^{2}\) integrable processes, to which all processes considered here belong, we can apply the Moore-Osgood Theorem (see Theorem A.1) together with Lemmas 3.5, 3.4, and 3.1 to obtain

---

3Here we are working with equivalence classes
\[
\lim_{n \to \infty} \lim_{d \to \infty} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_{t}^{d,n} - X_{t}^{n}|^2 + \sup_{0 \leq t \leq T} |X_{t}^{n} - X_{t}|^2 \right]
\]
\[
= \lim_{d \to \infty} \lim_{n \to \infty} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_{t}^{d,n} - X_{t}^{d}|^2 + \sup_{0 \leq t \leq T} |X_{t}^{d} - X_{t}|^2 \right]
\]
\[
= 0.\]
Furthermore, since uniform convergence implies pointwise convergence, we have from Lemma 3.4 that for any fixed \(d \in \mathbb{N}\),
\[
\lim_{n \to \infty} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_{t}^{d,n} - X_{t}^{d}|^2 \right] = 0
\]
which then gives the desired result. This concludes the proof of Theorem 1.1.

**Acknowledgments**

The author is grateful to Wilhelm Stannat and Sebastian Reich for helpful feedback on this work, as well as anonymous reviewer for their insightful suggestions which has improved this manuscript.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This research has been partially funded by Deutsche Forschungsgemeinschaft (DFG)- SFB1294/1 - 318763901.

**References**

[1] Jakubowski, A., Mémin, J., Pages, G. (1989). Convergence en loi des suites d’intégrales stochastiques sur l’espace D1 de Skorokhod. *Probab. Th. Rel. Fields*. 81(1):111–137. doi:10.1007/BF00343739.

[2] Kurtz, T. G., Protter, P. (1991). Characterizing the weak convergence of stochastic integrals. In M. Barlow & N. Bingham (Eds.), *Stochastic Analysis: Proceedings of the Durham Symposium on Stochastic Analysis, 1990* (London Mathematical Society Lecture Note Series, pp. 255-260). Cambridge: Cambridge University Press.

[3] Ikeda, N., Watanabe, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd ed. Japan: North Holland Publishing Company.

[4] Sussmann, H. J. (1991). Limits of the Wong-Zakai type with a modified drift term. *Stoch. Anal.* 475–493. Academic Press.

[5] Mcshane, E. J. (1972). Stochastic differential equations and models of random processes. *Proc. Sixth Berkeley Symp. Math. Stat. Probab*. 3:263–294.

[6] Stroock, D. W., Varadhan, S. R. (1972). On the support of diffusion processes with applications to the strong maximum principle. *Proc. Sixth Berkeley Symp. Math. Stat. Probab*. 3(638):333–359.

[7] Nakao, S., Yamato, Y. (1976). Approximation theorem on stochastic differential equations. In Ito, K., ed. *Proceedings of the International Symposium on SDE*, Kyotop. 283–296.

[8] Ikeda, N., Nakao, S., Yamato, Y. (1977). A class of approximations of Brownian motion. *Publ. Res. Inst. Math. Sci*. 13(1):285–300. doi:10.2977/prims/1195190109.
[9] Huisinga, W., Schütte, C., Stuart, A. M. (2003). Extracting macroscopic stochastic dynamics: Model problems. *Comm. Pure Appl. Math.* 56(2):234–269. doi:10.1002/cpa.10057.

[10] Pavliotis, G. A., Stuart, A. M. (2008). *Multiscale Methods: Averaging and Homogenization*. Springer.

[11] Clark, J., Crisan, D. (2005). On a robust version of the integral representation formula of nonlinear filtering. *Probab. Theory Relat. Fields.* 133(1):43–56. doi:10.1007/s00440-004-0412-5.

[12] Del Moral, P., Singh, S. (2020). A backward Itô-Ventzell formula with an application to stochastic interpolation. *C.R. Math.* 358(7):0–6.

[13] Del Moral, P., Singh S. S. (2022). Backward Itô-Ventzell and stochastic interpolation formulae. *Stochastic Processes and their Applications* 154:197–250.

[14] Hairer, M., Pardoux, É. (2015). A Wong-Zakai theorem for stochastic PDEs. *J. Math. Soc. Japan.* 67(4):1551–1604.

[15] Gyongy, I., Shmatkov, A. (2006). Rate of convergence of Wong–Zakai approximations for stochastic partial differential equations. *Appl. Math. Optim.* 54(3):341–341. doi:10.1007/s00245-006-1001-z.

[16] Diehl, J., Friz, P. K., Stannat, W. (2014). Stochastic partial differential equations: A rough path view on weak solutions via Feynman–Kac. Annales de la Faculté des sciences de Toulouse : Mathématiques, Serie 6, Vol. 26, pp. 911–947. doi:10.5802/afst.1556.

[17] Tessitore, G., Zacchini, J. (2006). Wong-Zakai approximations of stochastic evolution equations. *J. Evol. Equ.* 6(4):621–655. doi:10.1007/s00028-006-0280-9.

[18] Hu, Y. (2002). An approximation for the Zakai equation. *Appl. Math. Optim.* 45(1):23–44. doi:10.1007/s00245-001-0024-8.

[19] Brzeźniak, Z., Flandoli, E. (1995). Almost sure approximation of Wong-Zakai type for stochastic partial differential equations. *Stoch. Process. Appl.* 55(2):329–358. doi:10.1016/0304-4149(94)00037-T.

[20] Eugene, W., Moshe, Z. (1965). On the relation between ordinary and stochastic differential equations. *Int. J. Eng. Sci.* 2(2):213–229. doi:10.1016/0020-7225(65)90045-5.

[21] Gyöngy, I., Michaletzky, G. (2004). On Wong–Zakai approximations with δ-martingales. *Proc. R. Soc. Lond. A.* 460(2041):309–324. doi:10.1098/rspa.2003.1244.

[22] Liu, X., Han, G. (2020). A Wong-Zakai approximation of stochastic differential equations driven by a general semimartingale. *Discr. Contin. Dyn. Syst. Ser. B.* 25(12):1–10.

[23] Protter, P. (1985). Approximations of solutions of stochastic differential equations driven by semimartingales. *The Annals of Probability* 13(3):716–743.

[24] Kurtz, T., Pardoux, E., Protter, P. (1995). Stratonovich stochastic differential equations driven by general semimartingales. Annales de l'I.H.P. Probabilités et statistiques, Vol. 31, pp. 351–377.

[25] Gyöngy, I., Stinga, P. R. (2013). Rate of convergence of wong-zakai approximations for stochastic partial differential equations. In Dalang, R. C., Dozzi, M., Russo, F., eds. *Seminar on Stochastic Analysis, Random Fields and Applications VII*. Basel; Basel: Springer, p. 95–130.

[26] Ben Ammou, B. K., Lanconelli, A. (2019). Rate of convergence for Wong–Zakai-type approximations of Itô stochastic differential equations. *J. Theor. Probab.* 32(4):1780–1803. doi:10.1007/s10959-018-0837-x.

[27] Nakayama, T. (2019). Wong-Zakai approximations with convergence rate for stochastic partial differential equations. *Stochastic Analysis and Applications*, 36, 832–857. DOI:10.1080/07362994.2018.1471402.

[28] Le Jan, Y., Lejay, A. (2013). A new approach to the Wong–Zakai approximation of stochastic differential equations driven by a general semimartingale. *Ann. Probab.* 41(5):3003–3039.

[29] Nakayama, T. (2014). Wong-Zakai approximations for stochastic partial differential equations driven by general semimartingales. *J. Math. Sci. Univ. Tokyo.* 21(1):245–311.

[30] Lyons, T. J. (1998). Differential equations driven by rough signals. *Rev. Mat. Iberoam.* 14(2):215–310.

[31] Friz, P., Obenauser, H. (2009). Rough path limits of the Wong–Zakai type with a modified drift term. *J. Funct. Anal.* 256(10):3236–3256. doi:10.1016/j.jfa.2009.02.010.
[32] Lejay, A. (2006). Stochastic differential equations driven by processes generated by divergence form operators 1: A Wong-Zakai theorem. ESAIM: Probability and Statistics, Vol. 10, pp. 356–379.

[33] Coutin, L., Friz, P., Victoir, N. (2007). Good rough path sequences and applications to anticipating stochastic calculus. Ann. Probab. 35(3):1172–1193.

[34] Friz, P., Hairer, M. (2014). A Course on Rough Paths with an Introduction to Regularity Structures. Springer.

[35] Kelly, D., Melbourne, I. (2016). Smooth approximation of stochastic differential equations. Ann. Probab. 44(1):479–520.

[36] Friz, P., Victoir, N. (2010). Differential equations driven by Gaussian signals. Annales de l’institut Henri Poincare (B) Probability and Statistics 46(2):369–413.

[37] Hutzenthaler, M., Jentzen, A., Kloeden, P. E. (2012). Strong convergence of an explicit numerical method for sdes with nonglobally lipschitz continuous coefficients. Ann. Appl. Probab. 22(4):1611–1641.

[38] Pathiraja, S., Reich, S., Stannat, W. (2020). McKean-Vlasov SDEs in nonlinear filtering. SIAM J. Control and Optimization, 59(6), 4188–4215. https://doi.org/10.1137/20m1355197

[39] Lejay, A. (2012). Global solutions to rough differential equations with unbounded vector fields. Lect. Notes Math. 2046:215–246.

[40] Lyons, T., Qian, Z. (2002). System Control and Rough Paths. Oxford University Press.

[41] Lejay, A. (2003). An introduction to rough paths. In Azéma, J., Ledoux, M., Emery, M., Yor, M., eds., Lectures Notes in Mathematics. Springer, p. 1–59.

[42] Gubinelli, M. (2004). Controlling rough paths. J. Funct. Anal. 216(1):86–140. doi:10.1016/j.jfa.2004.01.002.

[43] Davie, A. M. (2008). Differential equations driven by rough paths: An approach via discrete approximation. Appl. Math. Res. eXpress. 2008(1):1–31.

[44] Bailleul, I. (2015). Flows driven by rough paths. Rev. Mat. Iberoam. 31(3):901–934. doi:10.4171/RMI/858.

[45] Bailleul, I., Catellier, R. (2018). Non-explosion criteria for rough differential equations driven by unbounded vector fields. Annales de la Faculté des sciences de Toulouse : Mathématiques, Serie 6, Vol. 29(3), pp. 721–759. doi:10.5802/afst.1644.

[46] Davie, A. M. (2010). Individual path uniqueness of solutions of stochastic differential equations. In Crisan, D., ed. Stochastic Analysis. Berlin: Springer-Verlag.

[47] Coutin, L. (2012). Rough paths via sewing Lemma. ESAIM: PS. 16:479–526. doi:10.1051/ps/2011108.

### Appendix A

**Theorem A.1. Moore-Osgood Theorem.** Let \((M, d_M)\) be a metric space and \((m_{d,n})_{d,n \in \mathbb{N}}\) be a sequence in \(M\) where \(\mathbb{N} = \mathbb{N} \cup \infty\). If

\[
(i) \quad \lim_{n \to \infty} \sup_{d \in \mathbb{N}} d_M(m_{d,n}, m_{d,\infty}) = 0 \quad \text{and} \\
(ii) \quad \lim_{d \to \infty} d_M(m_{d,n}, m_{\infty,n}) = 0 \quad \text{for all} \quad n \in \mathbb{N},
\]

then the joint limit \(\lim_{d,n \to \infty} m_{n,d}\) exists. In particular, it holds that

\[
\lim_{d,n \to \infty} m_{n,d} = \lim_{n \to \infty} m_{\infty,n} = \lim_{d \to \infty} m_{d,\infty}.
\]
A1. Proof of Lemma 3.2

Define

\[ D(s, t) := \int_s^t b(\hat{X}_u^d)du - \sigma (\hat{X}_u^d)B_{s,t}^{d,1} - \Sigma (\hat{X}_u^d)B_{s,t}^{d,2}. \]

Using the fact that \( b \) is Lipschitz, we have that

\[
\left| \int_s^t b(\hat{X}_u^d)du \right| \leq \int_s^t |b(\hat{X}_u^d) - b(x_0)|du + \int_s^t |b(x_0)|du \\
\leq |\nabla b|_\infty \int_s^t |\hat{X}_u^d - x_0|du + |b(x_0)|(t - s) \\
\leq |\nabla b|_\infty |\hat{X}_u^d|_\alpha \int_s^t u^\alpha du + |b(x_0)|(t - s) \\
\leq ((|\nabla b|_\infty |\hat{X}_u|_\alpha) \mu + |b(x_0)|)(t - s)^\alpha \mu^{\frac{1}{\alpha} - 1}
\]

Similarly, using the fact that \( \sigma \) is Lipschitz,

\[
|\sigma (\hat{X}_s^d)||B_{s,t}^{d,1}| \leq |\nabla \sigma|_\infty \frac{|\hat{X}_s^d - x_0|}{s^\alpha}|B_{s,t}^{d,1}|(t - s)^\alpha + |\sigma (x_0)||B_{s,t}^{d,1}|(t - s)^\alpha \\
\leq |\nabla \sigma|_\infty |\hat{X}_s^d|_\alpha \mu|B_{s,t}^{d,1}|(t - s)^\alpha + |\sigma (x_0)||B_{s,t}^{d,1}|(t - s)^\alpha \mu
\]

By similar arguments, we have

\[
|\Sigma (\hat{X}_s^d)||B_{s,t}^{d,2}| \leq |\Sigma (\hat{X}_s^d)| - \Sigma (x_0) \frac{|B_{s,t}^{d,2}|}{(t - s)^{2\alpha}}(t - s)^{2\alpha} + |\Sigma (x_0)| \frac{|B_{s,t}^{d,2}|}{(t - s)^{2\alpha}}(t - s)^{2\alpha} \\
\leq |\nabla \Sigma|_\infty |\hat{X}_s^d|_\alpha \mu^2|B_{s,t}^{d,2}|(t - s)^\alpha + |\Sigma (x_0)||\nabla \sigma|_\infty |B_{s,t}^{d,2}|(t - s)^\alpha \mu
\]

Then using the existence of a solution in the sense of Davie (2.2) and the above estimates, we have that

\[
|\hat{X}_t^d|_\alpha \leq L\mu^2 + C_{61}(\mu)|\hat{X}_t^d|_\alpha + C_{62}(\mu, x_0) + C_7(\mu, x_0) + C_8(\mu)|\hat{X}_t^d|_\alpha + C_9(\mu)|\hat{X}_t^d|_\alpha
\]

where

\[
C_7(\mu, x_0) := |B_{t,x}^d\sigma (x_0)|(1 + |\nabla \sigma|_\infty \mu) \\
C_8(\mu) := |B_{t,x}^d|\nabla \Sigma|_\infty \mu^2 \\
C_9(\mu) := |B_{t,x}^d|\nabla \sigma|_\infty \mu \\
C_{61}(\mu) := |\nabla b|_\infty \mu^{\frac{1}{\alpha}} \\
C_{62}(\mu, x_0) := |b(x_0)|\mu^{\frac{1}{\alpha} - 1}.
\]

If we consider a short time horizon such that

\[
C_8(\mu) \leq M, \quad C_9(\mu) \leq M, \quad C_{61}(\mu) \leq M
\]

we have that for the case \( M < \frac{1}{3} \),

\[
|X^d|_\alpha \leq \frac{1}{1 - 3M} \left( L\mu^2 + C_{62}(\mu, x_0) + C_7(\mu, x_0) \right) \quad \text{(A.2)}
\]

Consider

\[
D(s, r, t) := D(s, t) - D(s, r) - D(r, t)
\]
which by Lemma 1 in [39] is an almost additive functional and can be bounded using similar arguments as developed above for $D(s, t)$. The sewing lemma (see e.g., Lemma 4.1 in [49]) implies that there exists a universal constant $K$ depending only on $\theta$ such that

$$L \leq \sup_{0 \leq s < t \leq T} \frac{\hat{X}_t^d - \hat{X}_s^d - D(s, t)}{(t-s)^\theta} \leq K \sup_{0 \leq s < r < t \leq T} \frac{|D(s, r, t)|}{(t-r)^\theta} \quad (A.3)$$

and by the same arguments as in the proof of Proposition 2 in [39], which are unaffected by the presence of the drift in the RDE, we have for the small time horizon satisfying

$$KC_5(\mu) \leq K_2 \quad (A.4)$$

that

$$L' \leq \frac{K}{1 - K_2} (C_{10}(\mu) + (C_4(\mu) + C_3(\mu))|\hat{X}_d|^\alpha + C_{12}(\mu)|x_0|)$$

where $0 < K_2 < 1$ and

$$C_3(\mu) := |\nabla \Sigma|_\infty((1 + \mu^2)|B^d|_\alpha^2 + |B^d|_\alpha)$$
$$C_4(\mu) := |\nabla \sigma|^2_\infty|B^d|_\alpha^2 \mu$$
$$C_5(\mu) := |\nabla \sigma|_\infty|B^d|_\alpha \mu$$
$$C_{11}(\mu) := |\nabla \sigma|^2_\infty|B^d|_\alpha^2$$
$$C_{12}(\mu) := C_{11}|\nabla \sigma|_\infty$$
$$C_{10}(\mu) := C_3(\mu) + |\sigma(0)|C_{11}(\mu).$$

Substituting into (A.2) gives

$$|X^d|_\alpha \leq \frac{1}{1 - 3M} \left( \frac{K}{1 - K_2} \mu^2 \left( C_{10}(\mu) + (C_4(\mu) + C_3(\mu))|X^d|_\alpha + C_{12}(\mu)|x_0| \right) + C_7(\mu, x_0) + C_{62}(\mu, x_0) \right)$$

Again restricting to a short-time horizon, we have that if for example

$$\frac{1}{1 - 3M} \frac{K}{1 - K_2} \mu^2(C_4(\mu) + C_3(\mu)) \leq K_2 \quad (A.5)$$

then (3.5) holds. Condition (3.6) easily follows from combining (A.1), (A.4), and (A.5).