View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Carolina Digital Repository

The Annals of Applied Probability
2016, Vol. 26, No. 1, 425–461
DOI: 10.1214/15-AAP1096
© Institute of Mathematical Statistics, 2016

arXiv:1402.3188v2 [math.PR] 12 Jan 2016

ROUGH PATH RECURSIONS AND DIFFUSION APPROXIMATIONS

By David Kelly[1]

University of North Carolina, Chapel Hill

In this article, we consider diffusion approximations for a general class of stochastic recursions. Such recursions arise as models for population growth, genetics, financial securities, multiplicative time series, numerical schemes and MCMC algorithms. We make no particular probabilistic assumptions on the type of noise appearing in these recursions. Thus, our technique is well suited to recursions where the noise sequence is not a semi-martingale, even though the limiting noise may be. Our main theorem assumes a weak limit theorem on the noise process appearing in the random recursions and lifts it to diffusion approximation for the recursion itself. To achieve this, we approximate the recursion (pathwise) by the solution to a stochastic equation driven by piecewise smooth paths; this can be thought of as a pathwise version of backward error analysis for SDEs. We then identify the limit of this stochastic equation, and hence the original recursion, using tools from rough path theory. We provide several examples of diffusion approximations, both new and old, to illustrate this technique.

1. Introduction. In this article, we consider the limiting behaviour for a class of stochastic recursions. These recursions are natural approximations of continuous time stochastic equations. They arise as models for fast, discretely evolving random phenomena [15, 18, 33, 40, 41] and also as numerical discretizations of continuous stochastic equations [23]. The class is similar to the rough path schemes of [3] (see also [11], Section 8.5), but more general in the sense that the noise driving the recursion is not required to be a rough path, but may be an approximation (or discretization) of a rough path.

Let $V:\mathbb{R}^e \to \mathbb{R}^{e\times d}$ and $\mathbb{V} = (\mathbb{V}_1, \ldots, \mathbb{V}_e)$ where $\mathbb{V}_\kappa:\mathbb{R}^e \to \mathbb{R}^{d\times d}$ is defined by $\mathbb{V}^{\alpha\beta}_\kappa(\cdot) = \sum_\gamma \partial^\gamma V^\beta_\kappa(\cdot) V^\alpha_\gamma(\cdot)$ for $\alpha, \beta = 1, \ldots, d$ and $V = (V^\beta_\kappa)$ for $\kappa = 1, \ldots, e$,

Received February 2014; revised December 2014.
[1]Supported by ONR Grant N00014-12-1-0257.
AMS 2000 subject classifications. 60H05, 60H10, 60H35, 39A50.
Key words and phrases. Stochastic differential equations, rough path theory, diffusion limits, nonsemi-martingale, numerical schemes.

This is an electronic reprint of the original article published by the Institute of Mathematical Statistics in *The Annals of Applied Probability*, 2016, Vol. 26, No. 1, 425–461. This reprint differs from the original in pagination and typographic detail.

$\beta = 1, \ldots, d$. For each $n \geq 0$, define $Y_k^n \in \mathbb{R}^e$ by the recursion

$$Y_{k+1}^n = Y_k^n + V(Y_k^n)\xi_k^n + \mathbb{V}(Y_k^n) : \Xi_k^n + \text{error}, \tag{1.1}$$

where $\xi_k^n \in \mathbb{R}^d$, $\Xi_k^n \in \mathbb{R}^{d\times d}$ are noise sources and we use the notation $A : B = \operatorname{trace}(AB^T) = \sum_{\alpha,\beta} A^{\alpha\beta} B^{\alpha\beta}$ to denote the matrix inner product.

Let $\mathcal{P}_n = \{\tau_k^n : k = 0, \ldots, N_n\}$ be a partition of a finite time interval $[0, T]$, which gets finer as n tends to infinity. The vector ξ_k^n should be thought of as an approximation of a random increment

$$\xi_k^n \approx X(\tau_{k+1}^n) - X(\tau_k^n), \tag{1.2}$$

where X is some given stochastic process (a semi-martingale or fractional Brownian motion, e.g.). Formally, the symbol $\approx$ means that the approximation gets better as n tends to infinity. Likewise, the matrix Ξ_k^n should be thought of as some approximation of an iterated stochastic integral

$$\Xi_k^n \approx \int_{\tau_k^n}^{\tau_{k+1}^n} (X(s) - X(\tau_k^n)) \otimes dX(s), \tag{1.3}$$

where $\otimes$ denotes the outer product and where the notion of stochastic integration (Itô, Stratonovich or otherwise) is given.

Define the path $Y^n : [0, T] \to \mathbb{R}^e$ by $Y^n(t) = Y_j^n$ where τ_j^n is the largest mesh point in $\mathcal{P}_n$ with $\tau_j^n \leq t$ [note that we could equally define $Y^n(\cdot)$ by linear interpolation, without altering the results of the article]. Our objective is to show that the path $Y^n(\cdot)$ converges to the solution of a stochastic differential equation (SDE) driven by X as n tends to infinity.

REMARK 1.1. In the case where (ξ_k^n, Ξ_k^n) are the increments of a rough path, that is, $\xi_k^n = X(\tau_{k+1}^n) - X(\tau_k^n)$ and $\Xi_k^n = \int_{\tau_k^n}^{\tau_{k+1}^n} (X(s) - X(\tau_k^n)) \otimes dX(s)$, the recursions we consider are precisely the rough path schemes defined in [3]. However, we only require that (ξ_k^n, Ξ_k^n) be *approximations* of rough paths. This means the class of recursions we consider is much more general than the class of rough path recursions and includes many natural approximations that do not fall under [3]. This fact will be illustrated by the examples below.

To see why such diffusion approximations should be possible, it is best to look at a few examples. The most common variant of (1.1) is the "first-order" recursion, where $\Xi^n = 0$, so that

$$Y_{k+1}^n = Y_k^n + V(Y_k^n)\xi_k^n + \text{error}. \tag{1.4}$$

This resembles an Euler scheme with approximated noise $\xi_k^n \approx X(\tau_{k+1}^n) - X(\tau_k^n)$. Hence, it is reasonable to believe that there should be a diffusion approximation $Y^n \Rightarrow Y$ (where $\Rightarrow$ denotes weak convergence of random variables), where Y satisfies the SDE

$$dY = V(Y) \star dX,$$

and $\star dX$ denotes some method of stochastic integration (e.g., Itô, Stratonovich or otherwise). It turns out that the choice of approximating sequence ξ_k^n of the increment $X(\tau_{k+1}^n) - X(\tau_k^n)$ has a huge influence as to what type of stochastic integral arises in the limit.

We now explore this idea with a few examples. The first four examples are first-order recursions as in (1.4) and the final two are higher order recursions, as in (1.1).

EXAMPLE 1.1 (Euler scheme). Suppose that B is a d-dimensional Brownian motion, let $\xi_k^n = B((k+1)/n) - B(k/n)$ and define the partition $\mathcal{P}_n$ with $\tau_k^n = k/n$. Then clearly Y^n defines the usual Euler–Maruyama scheme on the time window $[0,1]$. It is well known that $Y^n \Rightarrow Y$ where Y satisfies the Itô SDE

$$dY = V(Y)\,dB.$$

This creates a feeling that any Euler looking scheme, like (1.4), should produce Itô integrals. As we shall see in the next few examples, when some *correlation* is introduced to the random variables ξ_k^n, this is certainly not the case.

Less trivial recursions of the form (1.4) have shown up in the areas of population genetics [18, 41], econometric models [40], psychological learning models [33], nonlinear time series models [10] and MCMC algorithms [36], to name but a few. Here, we will list the example from [18]; our analysis follows that performed in [25].

EXAMPLE 1.2 (Population and genetics models). In [18], the authors consider the stochastic difference equation

$$Y_{k+1}^n = f(S_k^n) + \exp(g(S_k^n))Y_k^n, \tag{1.5}$$

where $f(0) = g(0) = 0$ and $\{S_k^n\}_{k\geq 0}$ is a stationary sequence of random variables with $\mathbf{E}S_k^n = \mu/n$, $\operatorname{var}(S_k^n) = \sigma^2/n$, $\operatorname{cov}(S_k^n, S_0^n) = \sigma^2 r_k/n$ and with mixing assumptions on the centered sequences $(S_k^n - \mathbf{E}S_k^n)$ and $((S_k^n)^2 - \mathbf{E}(S_k^n)^2)$. This recursion arises naturally in models for population growth and also gene selection, where the environment is evolving in a random way.

Since the equation (1.5) is linear, the solution can be written down explicitly. As a consequence, it is easy to directly identify the limiting behaviour of each term appearing in the solution, for instance with the help of Prokhorov's theorem. Alternatively, we can incorporate the problem into the scope of this article by making (1.5) look more like the recursion (1.4). We first write

$$\begin{aligned} Y_{k+1}^n = Y_k^n &+ \mathbf{E}f(S_k^n) + \mathbf{E}(\exp(g(S_k^n)) - 1)Y_k^n \\ &+ (f(S_k^n) - \mathbf{E}f(S_k^n)) + (\exp(g(S_k^n)) - \mathbf{E}\exp(g(S_k^n)))Y_k^n. \end{aligned}$$

Now if we replace g, f and exp by their second-order Taylor expansion, we obtain

$$\begin{aligned}Y_{k+1}^n = Y_k^n &+ n^{-1}\left(f_s(0)\mu + \frac{1}{2}f_{ss}(0)\sigma^2 + \mu g_s(0)Y_k^n + \frac{1}{2}(g_{ss}(0) + g_s^2(0))\sigma^2 Y_k^n\right)\\ &+ n^{-1/2}(f_s(0)n^{1/2}(S_k^n - \mathbf{E}S_k^n) + g_s(0)n^{1/2}(S_k^n - \mathbf{E}S_k^n)Y_k^n)\\ &+ \frac{n^{-1}}{2}(f_{ss}(0)n((S_k^n)^2 - \mathbf{E}(S_k^n)^2)\\ &\qquad + (g_{ss}(0) + g_s^2(0))n((S_k^n)^2 - \mathbf{E}(S_k^n)^2)Y_k^n).\end{aligned}$$

Thus, if we set

$$\begin{aligned}&V^1(y) = f_s(0)\mu + \tfrac{1}{2}f_{ss}(0)\sigma^2 + \mu g_s(0)Y_k^n + \tfrac{1}{2}(g_{ss}(0) + g_s^2(0))\sigma^2 y,\\ &V^2(y) = f_s(0) + g_s(0)y, \qquad V^3(y) = \tfrac{1}{2}(f_{ss}(0) + (g_{ss}(0) + g_s^2(0))y)\end{aligned}$$

and

$$\begin{aligned}&\xi_k^{n,1} = n^{-1}, \qquad \xi_k^{n,2} = n^{-1/2}(n^{1/2}(S_k^n - \mathbf{E}S_k^n)),\\ &\xi_k^{n,3} = n^{-1}(n((S_k^n)^2 - \mathbf{E}(S_k^n)^2)),\end{aligned}$$

then Y^n satisfies the recursion

$$Y_{k+1}^n = Y_k^n + V^1(Y_k)\xi_k^{n,1} + V^2(Y_k)\xi_k^{n,2} + V^3(Y_k)\xi_k^{n,3} + \text{error}.$$

Moreover, due to the assumptions on $S_k^n - \mathbf{E}S_k^n$ and $(S_k^n)^2 - \mathbf{E}(S_k^n)^2$, the functional central limit theorem for stationary mixing sequences implies that

$$\left(\sum_{i=0}^{\lfloor n\cdot\rfloor - 1}\xi_i^{n,2}, \sum_{i=0}^{\lfloor n\cdot\rfloor - 1}\xi_i^{n,3}\right) \Rightarrow (W_2, W_3),$$

where W_2, W_3 are Brownian motions with a computable covariance structure. Thus, we should expect a diffusion limit of the form

$$dY = V^1(Y)\,dt + (V^2, V^3)(Y) \star (dW_2, dW_3).$$

By writing down the solution explicitly, it is shown in [18] that this is indeed the case and the method of integration involves a correction term that is neither Itô nor Stratonovich.

In nonlinear scenarios, more sophisticated machinery is required [25], but this still entails quite heavy and often nonrealistic mixing assumptions on the stationary sequence. The framework of martingale problems [38] has proved quite suitable for this analysis [25]. In [24], the authors beautifully address the case where the noise is a semi-martingale sequence, using the idea of a *good* sequence of semi-martingales. The following example is taken directly from [9, 24].

EXAMPLE 1.3 (Discrete time asset pricing). Let r_k^n denote the periodic rate of return for a security with value S_k^n. It follows that

$$S_{k+1}^n = S_k^n + S_k^n r_k^n.$$

The authors consider the case where r_k^n is a semi-martingale difference sequence defined in such a way that, if M^n denotes the partial sum process

$$M^n(t) = \sum_{i=0}^{\lfloor nt \rfloor - 1} r_k^n$$

then $M^n \Rightarrow M$ where M is a semi-martingale. It is natural to expect a diffusion approximation

$$dS = S \star dM$$

with some undetermined method of integration $\star dM$. In [24], the authors provide a natural condition on M^n that ensures this method of integration is Itô type, which is clearly the most natural. Sequences M^n that satisfy this condition are called *good* semi-martingales. Thus, if M^n is good, then

$$dS = S\, dM,$$

where the integral is of Itô type. The authors also permit for a class of semi-martingales which are a reasonable perturbation of a good semi-martingale. For instance, suppose that $M^n = \tilde{M}^n + A^n$, where $\tilde{M}^n$ is a good sequence of semi-martingales and A^n is a sequence of semi-martingales with $A^n \Rightarrow 0$ as $n \to \infty$ (hence $\tilde{M}^n \Rightarrow M$). Now define $H^n(t) = \int_0^t A^n(s)\, dA^n(s)$ where the integral is of Itô type and $K^n(t) = [\tilde{M}^n, A^n]_t$ where $[\cdot, \cdot]$ denotes quadratic covariation and suppose that $(M^n, A^n, H^n, K^n) \Rightarrow (M, 0, H, K)$ as $n \to \infty$. Then $S^n \Rightarrow S$ where S satisfies the Itô SDE

$$dS = S\, dM + S\, d(H - K).$$

So formally speaking, we have $\star dM = dM + d(H - K)$. Thus, two equally reasonable approximations of M can yield two vastly different limiting diffusions. This class of perturbed semi-martingales is comprehensive enough to cover virtually every diffusion approximation where the recursion is driven by a semi-martingale sequence.

The next example is a rather important one, which unfortunately does not fit into the classes of diffusion approximations already studied in the literature. Understanding the diffusion approximation for this example is one of the main motivations of this paper.

EXAMPLE 1.4 (Fast–slow systems). Let $T:\Lambda\to\Lambda$ describe a chaotic dynamical system with invariant ergodic measure μ. Define the fast–slow system

$$Y^n_{k+1}=Y^n_k+n^{-1/2}h(Y^n_k,T^k\omega)+n^{-1}f(Y^n_k,T^k\omega),$$

where $\omega\in\Lambda$ and $h,f:\mathbb{T}^e\times\Lambda\to\mathbb{T}^e$ where $\mathbb{T}$ denotes the torus $[0,2\pi)$ and h satisfies the centering condition $\int h(x,y)\mu(dy)=0$. If we assume that ω is a random variable with law μ, then the path $Y^n(\cdot)=Y^n_{\lfloor n\cdot\rfloor}$ becomes a random variable on càdlàg space. Note that the assumption $\omega\sim\mu$ simply means that the chaotic dynamical system is started in stationarity.

Fast–slow systems of this type have been considered in [5, 8, 15] and are fundamental to the understanding of naturally occurring physical systems with separated time scales [30]. Previous attempts at diffusion approximations typically involve heavy mixing assumption on the dynamical system T which are difficult to prove for most reasonable systems [25]. In [8], the author develops an alternative Itô calculus, but only in the case where T defines a partially hyperbolic dynamical system. In [15], the authors study the special case where the noise is additive. This allows them to use path-space continuity properties to lift convergence of the partial sum process to convergence of Y^n.

Let us see how fast–slow systems fit into the recursion framework (1.4). Using a Fourier expansion truncated at level d, we can replace $h(x,y)$ with the product $h(x)v(y)$ where $h:\mathbb{R}^e\to\mathbb{R}^{e\times d}$, $v:\Lambda\to\mathbb{R}^d$, $\int v(y)\mu(dy)=0$ and similarly replace $f(x,y)$ with $f(x)g(y)$. Hence, we obtain

$$Y^n_{k+1}=Y^n_k+n^{-1/2}h(Y^n_k)v(T^k\omega)+n^{-1}f(Y^n_k)g(T^k\omega).$$

This clearly satisfies the recursion (1.4) with $V=(h,f)$ and $\xi^n_k=(n^{-1/2}\times v(T^k\omega),n^{-1}g(T^k\omega))$. The limiting behaviour of the partial sums

$$W^n(t)=n^{-1/2}\sum_{i=0}^{\lfloor nt\rfloor-1}v(T^i)\quad\text{and}\quad S^n(t)=n^{-1}\sum_{i=0}^{\lfloor nt\rfloor-1}g(T^i)$$

is well understood under extremely weak conditions on the dynamical system [1, 19, 32, 42]. In particular,

$$W^n\Rightarrow W\quad\text{and}\quad S^n(t)\to t\bar{g}\qquad(\mu\text{-a.s.}),$$

where W is a multiple of Brownian motion and $\bar{g}=\int g\,d\mu$. Thus, we would expect a diffusion approximation of the form

$$dY=h(Y)\star dW+\bar{g}f(Y)\,dt.$$

In the situations that are already understood, namely partially hyperbolic dynamical systems [8] or additive noise [15], the limiting stochastic integral shown to be neither Itô nor Stratonovich type. Thus, the interpretation of the integral in a more general setting is an important problem.

The more general family of recursions defined in (1.1) (with $\Xi_k^n \neq 0$) arise when using a second-order approximation. Naturally, it is easy to find examples from numerical analysis.

EXAMPLE 1.5 (Semi-implicit numerical schemes). Let X be some stochastic process (e.g., fractional Brownian motion) and introduce the shorthand $X(s,t) = X(t) - X(s)$. Suppose we approximate a stochastic equation using a semi-implicit method of integration, for instance,

$$Y_{k+1}^n = Y_k^n + \tfrac{1}{2}(V(Y_k^n) + V(Y_{k+1}^n))X(\tau_k^n, \tau_{k+1}^n).$$

It is easy to show that Y^n satisfies (1.1) with $\xi_k^n = X(\tau_k^n, \tau_{k+1}^n)$ and $\Xi_k^n = \frac{1}{2}X(\tau_k^n, \tau_{k+1}^n) \otimes X(\tau_k^n, \tau_{k+1}^n)$. For a simple stochastic process X, like Brownian motion, it is well known that the limit of this numerical scheme is

$$dY = V(Y) \circ dX,$$

where the integral is of Stratonovich type. But for more complicated objects like fractional Brownian motion, it is not so simple [37]. Thus, studying recursions of the type (1.1) can lead to a better understanding well-posedness for numerical schemes, that is, whether they are approximating the correct continuous time limit.

EXAMPLE 1.6 (Sub-diffusion approximations). Despite the article's title, its scope is not restricted to diffusions, in particular the results also concern sub-diffusions. In [4, 39], the authors consider partial sum processes of the form

$$X^n(t) = d_n^{-1} \sum_{k=0}^{\lfloor nt \rfloor - 1} \xi_k,$$

where $\{\xi_k\}_{k \geq 0}$ is a stationary dependent sequence of random variables and d_n is some normalizing constant, such that $X^n \Rightarrow X$ in the Skorokhod topology, where X is fractional Brownian motion with some Hurst parameter $H \in (0,1)$ depending on the correlation structure. With this in mind, it is natural to consider a recursion

$$Y_{k+1}^n = Y_k^n + V(Y_k^n)\, d_n^{-1} \xi_k + \mathbb{V}(Y_k^n) : \Xi_k^n,$$

where Ξ_k^n is some approximation of an iterated integral defined using the sequence $\{\xi_k\}_{k \geq 0}$. For instance, as indicated by Example 1.5, if $\Xi_k^n = \frac{d_n^{-2}}{2} \xi_k \otimes \xi_k$ then the above recursion corresponds to a mid-point rule approximation of a stochastic integral. In particular one would expect $Y^n \Rightarrow Y$ where

$$Y(t) = Y(0) + \int_0^t V(Y(s)) \circ dX(s)$$

and where the integral is of *symmetric* type [37], which is the natural limit of the mid-point scheme. Of course, this is only a guess and it is quite possibly wrong. As we will see, the tools introduced in this article provide a natural basis for the investigation of such sub-diffusion approximations. Understanding such recursions is vastly important and could facilitate for the design of new methods for simulating stochastic differential equations driven by fractional Brownian motion.

The technique employed in this article is similar in spirit to that found in [15, 24], in that we will lift an approximation result for the noise signal into diffusion approximation for the recursion. However, in our more general scenario, where we do not assume any particular probabilistic structure on the noise, we require not just an invariance principle for the noise but also for its *iterated integral*. More precisely, define the noise signal

$$X^n(t) \stackrel{\text{def}}{=} \sum_{i=0}^{\lfloor nt \rfloor -1} \xi_i^n ,$$

which is the natural approximation of the limiting noise signal X. Moreover, define the discrete iterated integral

$$\mathbb{X}^n(t) \stackrel{\text{def}}{=} \sum_{i=0}^{\lfloor nt \rfloor -1} \sum_{j=0}^{i-1} \xi_i^n \otimes \xi_j^n + \sum_{i=0}^{\lfloor nt \rfloor -1} \Xi_i^n ,$$

which is the natural approximation of the limiting iterated integral $\int_0^t X \otimes dX$. In this paper, we shall lift a limit theorem for the discrete pair $(X^n, \mathbb{X}^n)$ into a diffusion approximation for the recursion Y^n. In essence, the limiting behaviour of X^n tells us what type of noise appears in the limiting stochastic integral and the limiting behaviour of $\mathbb{X}^n$ tells us what type of stochastic integral we are talking about. Looking back at Example 1.4, for instance, this suggest that we can interpret the integral $\star\, dW$, provided we can identify the limit of the discrete iterated integral

$$\sum_{i=0}^{\lfloor nt \rfloor -1} \sum_{j=0}^{i-1} v(T^j) \otimes v(T^i).$$

To derive this diffusion approximation technique, we use tools from *rough path theory* [29].

1.1. *Diffusion approximations using rough path theory.* For stochastic differential equations driven by piecewise smooth signals, the relationship between the noise and the solution is extremely well understood—mostly thanks to rough path theory. For the purpose of exposition, suppose that X is some piecewise smooth stochastic process and that Y solves the equation

$$Y(t) = Y(0) + \int_0^t V(Y(s))\,dX(s), \tag{1.6}$$

where the integral is defined in the Riemann–Stieltjes sense. It is well known that the map $X \mapsto Y$ is not continuous in the sup-norm topology. The theory of rough path proposes that we can build a continuous map from the noise to the solution, provided we know a bit more information about X. In particular, suppose that we can define $\mathbb{X}(t) = \int_0^t X(s) \otimes dX(s)$ where the integral is again of Riemann–Stieltjes type. Then one can show that the map $(X, \mathbb{X}) \mapsto Y$ is continuous in a topology called the ρ_γ topology (known colloquially as the *rough path topology*). This topology can be thought of as an extension of the γ-Hölder topology, defined on the space of objects similar to the pair $(X, \mathbb{X})$. The objects $(X, \mathbb{X})$ are called rough paths and the metric space of such objects is called the space of γ-Hölder rough paths.

This idea clearly has ramifications to the diffusion approximations. Indeed, suppose that Y^n solves the stochastic equation

$$Y^n(t) = Y^n(0) + \int_0^t V(Y^n(s))\,dX^n(s)$$

for some smooth stochastic process X^n and also define the iterated integral $\mathbb{X}^n(t) = \int_0^t X^n(s) \otimes dX^n(s)$. Since continuous maps preserve weak convergence, this suggests that a weak limit theorem for the pair $(X^n, \mathbb{X}^n)$ in the ρ_γ topology can be lifted to a weak limit theorem for Y^n. The general procedure can be summarized by two steps.

1. Show that $(X^n, \mathbb{X}^n) \overset{\text{f.d.d.}}{\to} (X, \mathbb{X})$, where $\overset{\text{f.d.d.}}{\to}$ denotes convergence of finite-dimensional distributions.

2. Show that the sequence is tight in the ρ_γ topology. For instance, one could use a Kolmogorov type argument, by checking estimates of the form

$$(\mathbf{E}|X^n(s,t)|^q)^{1/q} \lesssim |t-s|^\gamma \quad \text{and} \quad (\mathbf{E}|\mathbb{X}^n(s,t)|^{q/2})^{2/q} \lesssim |t-s|^{2\gamma} \tag{1.7}$$

for all $s, t \in [0,T]$, with some suitable γ and with q large enough.

Since the map: *rough path* $\mapsto$ *solution* is continuous in the rough path topology, the conclusion from these two steps is that $Y^n \Rightarrow Y$ where Y is the solution to an SDE whose form can be determined by the limit $\mathbb{X}$. For instance, suppose that X were a continuous semi-martingale and that

$$\mathbb{X}(t) = \int_0^t X(r) \circ dX(r) + \lambda t,$$

where the above integral is Stratonovich type. Then the limiting equation can be written

$$dY = V(Y) \circ dX + \lambda : \mathbb{V}(Y)\, dt.$$

This precise idea has proved useful in the areas of stochastic homogenization [27] and equations driven by random walks [2].

Unfortunately, for the recursion (1.1) the path Y^n does not satisfy a stochastic equation in the sense of rough path theory, so we cannot simply apply the above procedure.

The objective of this article is to overcome this obstacle. It turns out that the same two step procedure defined above, more or less still works. All we have to do is replace iterated integrals with their discrete counterparts and replace step 2 with the same statement *up to some resolution*. That is, we need only check the estimates (1.7) for all $s, t \in \mathcal{P}_n$, which requires no continuity at all. In checking these discrete estimates, we obtain a tightness-like result for a discrete version of the Hölder metric, defined (on càdlàg paths) by

$$\max_{s \neq t \in \mathcal{P}_n} \frac{|A(t) - A(s)|}{|t - s|^\gamma}.$$

This is of course *always* finite, since it is a maximum over a finite set, but the tightness result will tell us something about the asymptotics.

At the heart of the proof is an approximation theorem (Theorem 2.2), which we believe to be useful in its own right. The theorem allows us to approximate the recursion (1.1) with the solution to a stochastic differential equation driven by piecewise smooth paths. This approximation theorem can be thought of as a generalization of *the method of modified equations* for SDEs [43] (otherwise known as *backward error analysis* [6]). In particular, our approximation theorem has the advantage of being completely pathwise, without depending on the probabilistic properties of the stochastic process X^n whatsoever. By approximating Y^n by the solution to a genuine stochastic equation, we unlock the tools of rough path theory introduced above.

The outline of the paper is as follows. In Section 2, we sketch the main theorem of the paper. In Section 3, we list a few applications. In Section 4, we give a brief introduction to rough path theory and mention some results that are important to the present article. In Section 5, we rigourously define rough paths recursions, these are the central objects to the article. In Section 6, we derive the properties of rough path recursions that will be needed for the main theorem. In Section 7, we prove the main theorem of the article, concerning weak convergence of rough path recursions.

2. The main results and some applications. In this section, we state the main theorem, avoiding the technical definitions that will be introduced in subsequent sections. In particular, the main theorem (Theorem 2.1) can be stated and applied without requiring any knowledge of rough path theory and similarly for the approximation theorem (Theorem 2.2).

Let $\mathcal{P}_n = \{\tau_j^n : j = 0, \ldots, N_n\}$ be a partition of $[0, T]$ with mesh size $\Delta_n = \max_j |\tau_{j+1}^n - \tau_j^n|$. As stated above, one should regard $\xi_j^n \in \mathbb{R}^d$ as an approximation of the increment

$$\xi_j^n \approx X(\tau_{j+1}^n) - X(\tau_j^n). \tag{2.1}$$

Likewise, one should regard $\Xi_j^n \in \mathbb{R}^{d\times d}$ as an approximation of the iterated integral

$$\Xi_j^n \approx \int_{\tau_j^n}^{\tau_{j+1}^n} (X(s) - X(\tau_j^n)) \otimes dX(s). \tag{2.2}$$

The only consequence of this analogy is that it influences how we define the *path* corresponding to the incremental processes. Indeed, the increments can be anything at all, provided they satisfy the convergence properties stated in the theorem below. To recap, the recursions we consider in this article are of the form

$$Y_{j+1}^n = Y_j^n + V(Y_j^n)\xi_j^n + \mathbb{V}(Y_j^n) : \Xi_j^n + r_j^n, \tag{2.3}$$

where $j = 0, \ldots, N_n - 1$ and $|r_j^n| \lesssim \Delta_n^\lambda$ for some $\lambda > 1$ and the implied constant is uniform in n.

We now define the *rough step-function* $(X^n, \mathbb{X}^n)$ corresponding to the increments ξ_j^n, Ξ_j^n. If τ_k^n is the largest grid point in $\mathcal{P}_n$ such that $\tau_k^n \le t$ then

$$X^n(t) = \sum_{j=0}^{k-1} \xi_j^n \quad \text{and} \quad \mathbb{X}^n(t) = \sum_{i=0}^{k-1}\sum_{j=0}^{i-1} \xi_i^n \otimes \xi_j^n + \sum_{i=0}^{k-1} \Xi_i^n. \tag{2.4}$$

We similarly define the incremental paths

$$X^n(s,t) = \sum_{j=l}^{k-1} \xi_j^n \quad \text{and} \quad \mathbb{X}^n(s,t) = \sum_{i=l}^{k-1}\sum_{j=l}^{i-1} \xi_i^n \otimes \xi_j^n + \sum_{i=l}^{k-1} \Xi_i^n, \tag{2.5}$$

where τ_l^n is the largest grid point in $\mathcal{P}_n$ such that $\tau_l^n \le s$. It is easy to check that this is the natural choice, given the motivation (2.1) and (2.2). The main theorem is as follows.

THEOREM 2.1. *Let Y^n satisfy (2.3) and let $(X^n, \mathbb{X}^n)$ be càdlàg paths defined by (2.4). Suppose that $(X^n, \mathbb{X}^n) \overset{f.d.d.}{\to} (X, \mathbb{X})$ where X is a continuous* *semi-martingale and* $\mathbb{X}$ *is of the form*

$$\mathbb{X}(t) = \int_0^t X(r) \otimes \circ\, dX(r) + \nu t,$$

where the integral is defined in the Stratonovich sense and $\nu \in \mathbb{R}^{d\times d}$. *Suppose that the pair* $(X^n, \mathbb{X}^n)$ *satisfy the estimates*

$$\begin{aligned} (\mathbf{E}|X^n(\tau_j^n, \tau_k^n)|^q)^{1/q} &\lesssim |\tau_j^n - \tau_k^n|^\gamma \quad \textit{and} \\ (\mathbf{E}|\mathbb{X}^n(\tau_j^n, \tau_k^n)|^{q/2})^{2/q} &\lesssim |\tau_j^n - \tau_k^n|^{2\gamma} \end{aligned} \tag{2.6}$$

for all $\tau_j^n, \tau_k^n \in \mathcal{P}_n$ *where* $q > 0$, $\gamma \in (1/3 + q^{-1}, 1/2]$ *and the implied constant is uniform in* n. *Then* $Y^n \Rightarrow Y$ *in the sup-norm topology, where* Y *satisfies the SDE*

$$dY = V(Y) \circ dX + \nu : \mathbb{V}(Y)\, dt.$$

REMARK 2.1. Although we only require $q > 0$ it is clear from $\gamma \in (1/3 + q^{-1}, 1/2]$ that we always have $q > 6$.

REMARK 2.2. If the estimates (2.6) hold for $\gamma = 1/2$ and all $q \geq 1$, then the condition $V \in C^3$ can be relaxed to $V \in C^{2+}$. This follows using the standard techniques of (p, q) rough paths (see [28] and [14], Chapter 12). Additional details will be given in Remark 7.1.

REMARK 2.3. The result naturally extends to the case with an additional "drift" vector field $W \in C^{1+}(\mathbb{R}^e; \mathbb{R}^e)$

$$Y_{j+1}^n = Y_j^n + V(Y_j^n)\xi_j^n + \mathbb{V}(Y_j^n) : \Xi_j^n + W(Y_j^n)(\tau_{j+1}^n - \tau_j^n) + r_j^n.$$

In this setting, the limiting SDE is given by

$$dY = V(Y) \circ dX + (\nu : \mathbb{V}(Y) + W(Y))\, dt.$$

This is a more natural way to treat the problem introduced in Example 1.4. As with Remark 2.2, this extension is a standard application of (p, q) rough paths.

The next result is not so much a theorem as it is a guide for other theorems. It applies to situations where the noise driving the limiting equation is not a semi-martingale, such as the sub-diffusions encountered in Example 1.6.

META THEOREM 2.1. *In the same context as above. Suppose that* $(X^n, \mathbb{X}^n) \overset{f.d.d.}{\to} (X, \mathbb{X})$ *where* X *is some continuous stochastic process and*

$$\mathbb{X}(t) = \int_0^t X(r) \star dX(r),$$

where $\star dX$ denotes some constructible method of integration. Suppose moreover that $(X^n, \mathbb{X}^n)$ satisfy the estimates (2.6). Then $Y^n \Rightarrow Y$ where Y satisfies the stochastic equation

$$Y(t) = Y(0) + \int_0^t V(Y(s)) \star dX(s).$$

Theorem 2.1 and Meta Theorem 2.1 will be proved in Section 7. The proof of the meta theorem indicates what we mean by a "constructible method of integration".

Finally, the main tool used to derive the results above is the approximation theorem, which should be thought of as a pathwise version of backward error analysis (or the method of modified equations) [6, 43]. The rate estimate depends on the *discrete* γ-Hölder norm $C_{\gamma,n}$ which is the smallest number such that

$$|X^n(\tau_j^n, \tau_k^n)| \le C_{\gamma,n} |\tau_j^n - \tau_k^n|^\gamma \quad \text{and} \quad |\mathbb{X}^n(\tau_j^n, \tau_k^n)| \le C_{\gamma,n}^2 |\tau_j^n - \tau_k^n|^{2\gamma} \tag{2.7}$$

for all $\tau_j^n, \tau_k^n \in \mathcal{P}_n$. Since this number can be achieved by taking the maximum over a finite set, it is clear that each $C_{\gamma,n}$ is finite, regardless of the path $(X^n, \mathbb{X}^n)$. We will always need some kind of asymptotic estimate on $C_{\gamma,n}$ to make use of the approximation theorem.

THEOREM 2.2. *Suppose that $Y^n(\cdot)$ is the path defined by the recursion (2.3) and that the pair $(X^n, \mathbb{X}^n)$ defined by (2.4) satisfy the estimates (2.6) for some q, γ as in Theorem 2.1. Then for each n we can find a pair of piecewise smooth paths $(\tilde{X}^n, \tilde{Z}^n) : [0, T] \to \mathbb{R}^d \times \mathbb{R}^{d \times d}$ such that if $\tilde{Y}^n$ solves*

$$\tilde{Y}^n(t) = \tilde{Y}^n(0) + \int_0^t V(\tilde{Y}^n(s))\, d\tilde{X}^n(s) + \int_0^t \mathbb{V}(\tilde{Y}^n(s)) : d\tilde{Z}^n(s), \tag{2.8}$$

where the integrals are of Riemann–Stieltjes type, then

$$\|\tilde{Y}^n - Y^n\|_\infty \lesssim K_{\gamma,n} \Delta_n^{3\gamma-1}, \tag{2.9}$$

where $\|\cdot\|_\infty$ denotes the sup-norm and where the constant $K_{\gamma,n} = 1 \wedge C_{\gamma,n}^4$, where $C_{\gamma,n}$ is the constant defined in (2.7).

The proof of Theorem 2.2 is contained in Section 6. We will give one simple example to illustrate the idea behind this approximation theorem.

EXAMPLE 2.1. Suppose that B is a Brownian motion and that

$$Y_{k+1}^n = Y_k^n + V(Y_k^n)(B(\tau_{k+1}^n) - B(\tau_k^n)).$$

It is easy to check that, for almost every Brownian path, the constant $C_{\gamma,n}$ defined in (2.7) is bounded uniformly in n, for any $\gamma < 1/2$. It follows that we can find an equation driven by smooth paths, with solution $\tilde{Y}^n$ such that

$$\|\tilde{Y}^n - Y^n\|_\infty \lesssim \Delta_n^{3\gamma-1}$$

for any $\gamma < 1/2$.

3. Some applications. We will now discuss some potential applications for the tools introduced above.

3.1. *Random walk recursions.* We start with a quite trivial and well known result, with the sole intention of illustrating how Theorem 2.1 should be used. A continuous time version of this example can be found in [2]. It should be said that the following can easily be deduced from either [24] or [2].

Suppose that $\{\xi_i\}_{i\geq 0}$ is an i.i.d. sequence of random variables taking values in $\mathbb{R}^d$ with $\mathbf{E}\xi_i = 0$ and $\mathbf{E}\xi_i \otimes \xi_i = D$, with $D \in \mathbb{R}^{d\times d}$. We will consider the random walk recursion

$$Y_{k+1}^n = Y_k^n + n^{-1/2} V(Y_k^n)\xi_k$$

with associated partition $\mathcal{P}^n$ with $\tau_k^n = k/n$. If we define the path $Y^n(\cdot) = Y_{\lfloor n\cdot\rfloor}^n$ then it is well known that $Y^n \Rightarrow Y$ in càdlàg space (with sup-norm topology), where Y solves the SDE

$$dY = V(Y) D^{1/2}\, dW,$$

where W is standard Brownian motion on $\mathbb{R}^d$. The following lemma illustrates how to prove this using Theorem 2.1. First, we define the rough step function

$$X^n(t) = n^{-1/2} \sum_{i=0}^{\lfloor nt\rfloor -1} \xi_i, \qquad \mathbb{X}^n(t) = n^{-1} \sum_{i=0}^{\lfloor nt\rfloor -1} \sum_{j=0}^{i-1} \xi_j \otimes \xi_i.$$

LEMMA 3.1. *Suppose that $\mathbf{E}|\xi_0|^q < \infty$ for some $q > 6$. Then the pair $(X^n, \mathbb{X}^n)$ satisfy the assumptions of Theorem 2.1 with $X = D^{1/2}W$ and $\nu = -\frac{1}{2}D$. In particular $Y^n \Rightarrow Y$ where*

$$dY = V(Y) D^{1/2}\, dW,$$

where the integral is of Itô type.

REMARK 3.1. The moment condition on $\mathbf{E}|\xi_0|^q$ is much stronger than required by more traditional solutions to the problem. This is due to the fact that the conclusion of the theorem is actually *stronger* than most traditional versions. In particular, we could actually show that Y^n converges in (a discrete version) of the rough path topology, which is much stronger than the sup-norm topology. We will not pursue such statements in this article.

PROOF OF LEMMA 3.1. From Donsker's theorem, we already know that $X^n \Rightarrow X = D^{1/2}W$. To identify the limit of $\mathbb{X}^n$ we simply write it as a stochastic integral. In particular, we see that

$$\mathbb{X}^n(t) = \int_0^t X^n(s-) \otimes dX^n(s),$$

where the integral is of left-Riemann type (hence Itô type). That is,

$$\int_0^t Y(s-)\,dZ(s) = \lim \sum Y(s_i)(Z(s_{i+1}) - Z(s_i)),$$

where $\{s_i\}$ is a partition of $[0,t]$ and the limit is taken as the maximum of $s_{i+1} - s_i$ tends to zero. The theory of good semi-martingales [24] provides a class of semi-martingale sequences for which the limit of a sequence of Itô integrals is an Itô integral.

Since the partial sum process X^n is clearly a martingale with respect to the filtration generated by the sequence $\{\xi_i\}_{i=0}^{\lfloor nt\rfloor-1}$, we can appeal to [24], Theorem 2.2. In particular, since the quadratic variation

$$[X^n, X^n]_t = n^{-1} \sum_{i=0}^{\lfloor nt\rfloor-1} \xi_i \otimes \xi_i,$$

we have that $\mathbf{E}[X^n, X^n]_t = D\lfloor nt\rfloor/n$ which is of course bounded uniformly in n. Thus, X^n is good and [24], Theorem 2.2, immediately tells us that

$$(X^n, \mathbb{X}^n) \Rightarrow (X, \mathbb{X})$$

in the sup-norm topology, where

$$\mathbb{X}(t) = \int_0^t X(s) \otimes dX(s) = \int_0^t X(s) \otimes \circ\, dX(s) - \frac{1}{2} Dt,$$

where the integrals are of Itô and Stratonovich type, respectively, and we have converted between them in the usual way. This is of course stronger than the finite-dimensional distribution result which we required, but the tools from [24] make it quite easy to prove.

All that remains is to obtain the discrete tightness estimates (2.6) with $q > 6$ and $\gamma = 1/2$. Since X^n is a martingale, we can apply the Burkholder–Davis–Gundy (BDG) inequality

$$\begin{aligned}\mathbf{E}|X^n(j/n, k/n)|^q &\lesssim \mathbf{E}|[X^n, X^n]_{j/n,k/n}|^{q/2} \\ &= n^{-q/2}\mathbf{E}\left|\sum_{i=j}^{k-1} \xi_i \otimes \xi_i\right|^{q/2} \le n^{-q/2}\mathbf{E}\left(\sum_{i=j}^{k-1} |\xi_i|^2\right)^{q/2}.\end{aligned}$$

By the Hölder inequality, we see that

$$\sum_{i=j}^{k-1} |\xi_i|^2 \le (k-j)^{1-q/2} \left(\sum_{i=j}^{k-1} |\xi|^q\right)^{2/q}.$$

It follows that

$$\mathbf{E}|X^n(j/n,k/n)|^q \lesssim n^{-q/2}(k-j)^{q/2-1}\sum_{i=j}^{k-1}\mathbf{E}|\xi_i|^q$$

$$= n^{-q/2}(k-j)^{q/2-1}\sum_{i=j}^{k-1}\mathbf{E}|\xi_0|^q = \mathbf{E}|\xi_0|^q(k/n-j/n)^{q/2}.$$

Since $\mathbb{X}^n$ is a stochastic integral (or martingale transform) it too is a martingale and hence we can again apply the BDG inequality. A similar argument yields

$$\mathbf{E}|\mathbb{X}^n(j/n,k/n)|^{q/2} \lesssim \mathbf{E}|\xi_0|^q(k/n-j/n)^q.$$

And since $q>6$, the interval $(1/3+q^{-1},1/2]$ is nonempty, so we do indeed satisfy the requirements of Theorem 2.1. It follows that $Y^n \Rightarrow Y$ where

$$dY = V(Y)\circ D^{1/2}\,dW - \tfrac{1}{2}D:\mathbb{V}(Y)\,dt$$

and we obtain the required expression by converting Stratonovich to Itô. □

3.2. *Fast–slow systems.* Instead of showing how the tools can be used on each of the examples given in the Introduction, we concentrate on the fast–slow systems, since it is the least understood. The tools of this article are applied to fast–slow systems in a companion paper [21] (see also [20]), to yield new results for fast–slow systems. The dynamical system theory required is slightly too involved to be included in this paper, thus we will only sketch the ideas behind the result.

We will restrict our attention to the fast–slow system

$$Y_{k+1}^n = Y_k^n + n^{-1/2}h(Y_k^n)v(T^k\omega),$$

the general case is treated in [21]. Setting $V=h$, $\xi_k^n = n^{-1/2}v(T^k)$ and $\Xi_k^n=0$ we see that the rough step function is defined by

$$X^n(t) = n^{-1/2}\sum_{i=0}^{\lfloor nt\rfloor-1} v(T^i) \quad\text{and}\quad \mathbb{X}^n(t) = n^{-1}\sum_{i=0}^{\lfloor nt\rfloor-1}\sum_{j=0}^{i-1} v(T^i)\otimes v(T^j).$$

We will also introduce the sigma algebra $\mathcal{M}$ which is whatever sigma algebra we chose to go with the measure space (Λ,μ).

PROPOSITION 3.1. *Under "sufficient" mixing conditions on T, the pair $(X^n,\mathbb{X}^n)$ satisfy the assumptions of Theorem 2.1 with $X=D^{1/2}W$ where*

$$D^{\alpha\beta} = \int v^\alpha v^\beta\,d\mu + \sum_{j=1}^{\infty}\left(\int v^\alpha v^\beta(T^j)\,d\mu + \int v^\alpha(T^j)v^\beta\,d\mu\right)$$

and

$$\nu^{\alpha\beta} = -\frac{1}{2}\int v^\alpha v^\beta d\mu + \frac{1}{2}\sum_{j=1}^{\infty}\Big(\int v^\alpha v^\beta(T^j)\, d\mu - \int v^\alpha(T^j) v^\beta\, d\mu\Big).$$

In particular, $Y^n \Rightarrow Y$ *where*

$$dY = h(Y)D^{1/2} \circ dW + \nu : \mathbb{H}(Y)\, dt,$$

where $\mathbb{H}$ *is defined precisely as* $\mathbb{V}$*, but in terms of* h*.*

Sketch of proof. To identify the limit of the pair $(X^n, \mathbb{X}^n)$, we proceed similarly to the random walk recursion case, namely identify the limit of X^n and then lift it to $\mathbb{X}^n$. To identify the limit of X^n, we will use a martingale central limit theorem on the *time reversal* of the partial sum process X^n.

By applying the *natural extension* of a dynamical system, we can assume without loss of generality that the map T is invertible. Now, fix a time window $[0, L]$ on which we will identify the limit of X^n. By stationarity, we have that

$$X^n(t) = n^{-1/2}\sum_{i=0}^{nt-1} v(T^i) \overset{\text{dist}}{=} n^{-1/2}\sum_{i=0}^{nt-1} v(T^{i-nL}).$$

By setting $i = nL - k$, the above equals

$$\sum_{k=nT-nt}^{nT} v(T^{-k}) = \sum_{k=1}^{nT} v(T^{-k}) - \sum_{k=1}^{n(T-t)} v(T^{-k}).$$

Now, if we define the *backward time* partial sum $X^n_-(t) = n^{-1/2}\sum_{i=1}^{\lfloor nt\rfloor} v(T^{-i})$ then the above calculation show that

$$X^n(t) = X^n_-(L) - X^n_-(L-t).$$

Under "sufficient" mixing conditions on the dynamical system, one can show that X^n_- is a martingale with respect to the backward time filtration $\mathcal{F}_t = T^{\lfloor nt\rfloor -1}\mathcal{M}$. Thus, using the central limit theorem for ergodic stationary L^2 martingale difference sequences [31] it follows that $X^n_- \Rightarrow X$ where $X = D^{1/2}W$. And thus,

$$X^n \Rightarrow X(L) - X(L-t) \overset{\text{dist}}{=} X(t),$$

using the time reversal property of Brownian motion. Now that we know the limiting behaviour of X^n, we can use the tools from [24] to identify the limiting behaviour of $\mathbb{X}^n$. In particular, Theorem 2.2 from [24] allows us to identify the limit of integrals against the martingale X^n_-, so all we have to do is rewrite $\mathbb{X}^n$ in backward time, so that it becomes an integral driven by dX_-^n (plus corrections). Using this idea, we show that $(X^n, \mathbb{X}^n) \Rightarrow (X, \mathbb{X})$ where

$$\mathbb{X}(t) = \int_0^t X(s) \otimes \circ\, dX(s) + \nu t,$$

where the integral is of Stratonovich type. All that remains is to prove the discrete tightness estimates. To do so, we again write the pair $(X^n, \mathbb{X}^n)$ in terms of the martingale X_-^n and stochastic integrals driven by dX_-^n. Since these are both martingales, we can apply the BDG inequality and the tightness estimates follow (somewhat) easily, using the ergodic properties and stationarity of T. It follows from Theorem 2.1 that $Y^n \Rightarrow Y$ where Y solves the SDE

$$dY = h(Y) \circ dW + \nu : \mathbb{H}(Y)\, dt. \qquad \square$$

It is important to note that although the diffusion approximation is essentially a consequence of the martingale central limit theorem, those martingale sequences only appear in *backward time*. In particular, the fast–slow system cannot be written as an equation driven by a semi-martingale, so the theory of good semi-martingales cannot be applied directly to the fast–slow system. The advantage of Theorem 2.1 is that even though it is not possible to apply martingale limit theory to the recursion, it is quite easy to do so to the *noise* processes $(X^n, \mathbb{X}^n)$.

REMARK 3.2. In the companion paper [21], the details are far more complicated than we present above. For example, the backward time object X_-^n is not in fact a good martingale, but rather a reasonable perturbation of a good martingale, as described in Example 1.3. This makes matters more interesting.

3.3. *Connection to [24].* We will now briefly comment on the connection between Theorem 2.1 and the tools introduced in [24]. Given a partition $\mathcal{P}^n$, consider the recursion

$$Y_{k+1}^n = Y_k^n + V(Y_k^n)\xi_k^n,$$

where the increments ξ^n are defined in such a way that the step-function $X^n(t) = \sum_{i=0}^{k-1} \xi_i^n$ were a semi-martingale with respect to some given sequence of filtrations, the random walk recursion provides a nice example. It follows that the path $Y^n(\cdot)$ solves the equation

$$Y^n(t) = Y^n(0) + \int_0^t V(Y^n(s-))\, dX^n(s),$$

where the integral is of Itô type, as defined Section 3.1. Suppose moreover that $X^n = M^n + A^n$ where M^n is a good sequence of semi-martingales and $A^n \Rightarrow 0$. Also define

$$H^n(t) = [M^n, A^n]_t, \qquad K^n(t) = \int_0^t A^n(s) \otimes dA^n(s),$$

where the integral is of Itô type. Suppose moreover that

$$(X^n, A^n, H^n, K^n) \Rightarrow (X, 0, H, K) \tag{3.1}$$

in the sup-norm topology, where the limits are continuous semi-martingales. Then [24], Theorem 5.1, states that $Y^n \Rightarrow Y$ where

$$dY = V(Y)\,dX + \mathbb{V} : d(H - K) \tag{3.2}$$

and the integrals are of Itô type.

Let us see how this fits into Theorem 2.1. It is not hard to see that the assumption (3.1) implies the assumption $(X^n, \mathbb{X}^n) \Rightarrow (X, \mathbb{X})$. For instance, since

$$\begin{aligned}
\mathbb{X}^n(t) &= \int_0^t X^n(s-) \otimes dX^n(s) \\
&= \int_0^t M^n(s-) \otimes dM^n(s) + \int_0^t M^n(s-) \otimes dA^n(s) \\
&\quad + \int_0^t A^n(s-) \otimes dM^n(s) + \int_0^t A^n(s-) \otimes dA^n(s) \\
&= \int_0^t M^n(s-) \otimes dM^n(s) + A^n(t)M^n(t) - [M^n, A^n]_t \\
&\quad + \int_0^t A^n(s-) \otimes dA^n(s) \\
&= \int_0^t M^n(s-) \otimes dM^n(s) + A^n(t)M^n(t) - H^n(t) + K^n(t).
\end{aligned}$$

Combining the fact that M^n is good with (3.1) we see that $(X^n, \mathbb{X}^n) \Rightarrow (X, \mathbb{X})$ where

$$\begin{aligned}
\mathbb{X}(t) &= \int_0^t M(s) \otimes dM(s) - H(t) + K(t) \\
&= \int_0^t M(s) \otimes \circ\, dM(s) + \frac{1}{2}[M, M]_t - H(t) + K(t).
\end{aligned}$$

Taking the tightness estimates for granted, we see that in the case where $\frac{1}{2}[M, M]_t - H(t) + K(t) = \nu t$, Theorem 2.1 reproduces the diffusion approximation (3.2). It is quite possible to extend the Theorem 2.1 so that it only requires, for instance, $[M,M]-H+K$ to be of bounded variation, which would yield a result closer to that of [24], but we do not pursue this here.

3.4. *Numerical schemes.* Several recent articles have used rough path ideas to study numerical schemes for stochastic equations. To name a few, [3, 7, 12, 13] are all concerned with similar but typically higher order schemes than (2.3). In [7], the authors also use the idea that a recursion can be approximated by an RDE, but only for much higher order Milstein-type schemes. In a recent preprint [35], the authors consider Euler-type schemes, again by approximating the recursion with a genuine RDE.

The recursion considered in this article handles most numerical schemes for SDEs, provided the driving noise is a random path with Hölder exponent $\gamma > 1/3$. It is easy to see that the Euler scheme

$$Y_{j+1}^n = Y_j^n + V(Y_j^n)X(j/n,(j+1)/n)$$

fits into the framework of (2.3). Using nothing more than a Taylor expansion, one can also show that another typical numerical scheme, the semi-implicit scheme, fits into (2.3). This is defined by

$$Y_{j+1}^n = Y_j^n + \theta V(Y_j^n)X(j/n,(j+1)/n) + (1-\theta)V(Y_{j+1}^n)X(j/n,(j+1)/n),$$

where $\theta \in [0,1]$. When $\theta = 1$ this is of course the (forward) Euler scheme, when $\theta = 1/2$ this is the Stratonovich mid-point scheme and when $\theta = 0$ this is the backward Euler scheme. In [23], one can find a plethora of schemes that also fit into the class of recursions defined by (2.3), using similar arguments to that given below.

In the context of numerical schemes, we see two key areas where the ability to identify weak limits is beneficial.

1. Well-posedness of numerical schemes. When the noise is not a semimartingale, it may not be clear whether a limit exists and if it does—how it should be interpreted. Theorem 2.1 provides a quick criterion for this situation. In particular, since $X^n = X$ one need only identify the limit of $\mathbb{X}^n$. If the limit of $\mathbb{X}^n$ corresponds to a reasonable type of integral (it should correspond to the method of integration used by the numerical scheme) then the limiting equation can be interpreted in the sense of that integral.

2. Numerical schemes that depend on an approximation of the noise, rather than the exact distribution. Such situations arise if the noise is difficult to simulate and must instead be approximated, a common scenario when Gaussianity is not present. One also encounters this situation in the context of *stochastic climate modeling*, where ocean–atmosphere equations are driven by an under-resolved source of noise with persistent correlations in time [34] and also in *data assimilation*, where a perturbation of a stochastic observation is fed into the numerical simulation of a forecast model. The article [22] contains a brief overview of the latter idea.

Finally, the approximation theory above clearly has applications to determining the pathwise order of numerical schemes. For example, suppose that Y^n is defined by the Euler scheme

$$Y^n_{j+1} = Y^n_j + V(Y^n_j)X(j/n,(j+1)/n).$$

Since X does not depend on n the weak limit Y is determined by the weak limit $\mathbb{X}$ of

$$\mathbb{X}^n(t) = \sum_{j=0}^{\lfloor nt\rfloor -1} X(0,j/n) \otimes X(j/n,(j+1)/n).$$

Using Theorem 2.2 as well as the tools from rough path theory (Lemma 4.2) it is easy to show that

$$\sup_{k=0,\dots,n} |Y^n(\tau^n_k) - Y(\tau^n_k)| \le K_{\gamma,n}\Big(\Delta_n^{3\gamma-1} + \sup_{k=0,\dots,n} |\mathbb{X}^n(\tau^n_k) - \mathbb{X}(\tau^n_k)|^{\theta}\Big),$$

where $K_{\gamma,n}$ only depends on n through the discrete Hölder norm of $\mathbb{X}^n$. If X were Brownian motion, then one can trivially calculate moments of $|\mathbb{X}^n(\tau^n_k) - \mathbb{X}(\tau^n_k)|$ exactly, thus obtaining a rate of convergence is simple. However, obtaining the *optimal* rate of convergence is slightly more subtle. The topic of convergence rates will not be discussed further in this article but is the subject of a future article.

4. A taste of rough path theory.

In this section, we will serve an appetizer in rough path theory. For the full course, we recommend [11], which is closely aligned with the exposition below.

4.1. *Space of rough paths.*

A rough path has two components to its definition, an algebraic one and an analytic one. The algebraic component ensures that the objects $X, \mathbb{X}$ do indeed behave like the increments they hope to imitate. The analytic component describes the Hölder condition that is required to construct solution maps. In the definition below, we *always* require that the exponent $\gamma > 1/3$. We use the notation $T^2(\mathbb{R}^d) = \mathbb{R}^d \oplus \mathbb{R}^{d\times d}$ for the step-2 tensor product algebra.

DEFINITION 4.1. We say that $\mathbf{X} : [0,T]\times[0,T] \to T^2(\mathbb{R}^d)$ is a *rough path* if for $\mathbf{X} = (X, \mathbb{X})$

$$\begin{aligned} &X(s,t) = X(s,u) + X(u,t) \quad \text{and} \\ &\mathbb{X}(s,t) = \mathbb{X}(s,u) + \mathbb{X}(u,t) + X(s,u)\otimes X(u,t) \end{aligned} \tag{4.1}$$

for all $s,u,t \in [0,T]$. These are known as *Chen's relations*. If moreover we have that

$$|X(s,t)| \lesssim |t-s|^{\gamma} \quad \text{and} \quad |\mathbb{X}(s,t)| \lesssim |t-s|^{2\gamma} \tag{4.2}$$

for all $s,t \in [0,T]$ then $\mathbf{X}$ is a *γ-Hölder rough path*. The set of γ-Hölder rough paths will be denoted $\mathcal{C}^\gamma([0,T];\mathbb{R}^d)$. Every rough path defines a path $\mathbf{X}:[0,T] \to T^2(\mathbb{R}^d)$ by setting $\mathbf{X}(t) = \mathbf{X}(0,t)$. Likewise, we could equally have defined rough paths as paths $\mathbf{X}:[0,T] \to T^2(\mathbb{R}^d)$ and then simply taken Chen's relations as a definition for $\mathbf{X}(s,t)$. This identification between paths and increments will be used frequently throughout the article.

We will make use of two metric spaces of rough paths. First, $\mathcal{C}^\gamma([0,T];\mathbb{R}^d)$ is a metric space when furnished with the metric

$$\rho_\gamma(\mathbf{X},\tilde{\mathbf{X}}) = \sup_{s\neq t\in[0,T]} \frac{|X(s,t)-\tilde{X}(s,t)|}{|s-t|^\gamma} + \sup_{s\neq t\in[0,T]} \frac{|\mathbb{X}(s,t)-\tilde{\mathbb{X}}(s,t)|}{|s-t|^{2\gamma}}.$$

It is easy to check the interpolation inequality

$$\rho_\alpha \leq \rho_\beta^{\alpha/\beta}\rho_0^{1-\alpha/\beta} \tag{4.3}$$

for any $0 \leq \alpha \leq \beta$. We also make use of the related γ-Hölder "norm"

$$|||\mathbf{X}|||_\gamma = \sup_{s\neq t\in[0,T]} \frac{|X(s,t)|}{|s-t|^\gamma} + \sup_{s\neq t\in[0,T]} \frac{|\mathbb{X}(s,t)|^{1/2}}{|s-t|^\gamma},$$

which is by definition finite on $\mathcal{C}^\gamma([0,T];\mathbb{R}^d)$. Clearly, we have that

$$|X(s,t)| \leq |||\mathbf{X}|||_\gamma |s-t|^\gamma \quad \text{and} \quad |\mathbb{X}(s,t)| \leq |||\mathbf{X}|||_\gamma^2 |s-t|^{2\gamma}$$

for all $s,t \in [0,T]$ and $\mathbf{X} = (X,\mathbb{X}) \in \mathcal{C}^\gamma([0,T];\mathbb{R}^d)$.

The second metric space we make use of is the set of continuous rough paths $\mathbf{X}:[0,T] \to T^2(\mathbb{R}^d)$ endowed with the uniform metric $\|\cdot\|_\infty$. By this, we simply mean the sup-norm defined on functions with range $\mathbb{R}^d \oplus \mathbb{R}^{d\times d}$ (with the ordinary Euclidean norm on the range). It is easy to see that this topology is equivalent to that generated by ρ_0.

REMARK 4.1. There is a good reason for using $|||\cdot|||_\gamma$ in addition to ρ_γ. This is due to the relationship between the Euclidean norm on and the Carnot–Caratheodory norm, defined on homogeneous groups. This will be utilised in Section 6.

4.2. *Rough differential equations.* For $\mathbf{X} \in \mathcal{C}^\gamma([0,T];\mathbb{R}^d)$ and $V \in C_b^2$, there is a class of paths $Y:[0,T] \to \mathbb{R}^e$ known as *controlled rough paths*, for which one can define the integral $\int_0^t V(Y)\,d\mathbf{X}$. We say that Y is an X-controlled rough path if $Y(s,t) = Y(t) - Y(s)$ has the form

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

for all $i = 1,\dots,e$ and $0 \leq s \leq t \leq T$, where $Y_i':[0,T] \to \mathbb{R}^{e\times d}$ is a γ-Hölder path. For a thorough treatment of controlled rough paths and their use in defining the above integrals, see [11], Section 4.

The integral is defined as a compensated Riemann sum

$$\int_0^t V(Y)\, d\mathbf{X} = \lim_{\mathcal{P}\to 0} S_{\mathcal{P}},$$

where

$$S^i_{\mathcal{P}} = \sum_{[t_k,t_{k+1}]\in\mathcal{P}} V^i(Y(t_k))X(t_k,t_{k+1}) + \sum_{j=1}^{e} (Y'_j(t_k)\otimes\partial_j V^i(Y(t_k))) : \mathbb{X}(t_k,t_{k+1})$$

and $\mathcal{P}$ denotes a partition of $[0,t]$. Note that the integral is defined pathwise, for each $\mathbf{X}\in\mathcal{C}^\gamma([0,T];\mathbb{R}^d)$.

A controlled rough path Y is said to solve the RDE $dY = V(Y)\, d\mathbf{X}$ with initial condition $Y(0)=\eta$ if it solves the integral equation

$$Y(t) = \eta + \int_0^t V(Y)\, d\mathbf{X}$$

for all $t\in[0,T]$. In this case, we write $Y=\Phi(\mathbf{X})$. When required, we write $Y=\Phi(\mathbf{X};V,\eta,s)$ to denote the solution to $Y(t)=\eta+\int_s^t V(Y)\, d\mathbf{X}$ with $t\geq s$. For a thorough treatment of RDEs, see [11], Section 8.

We will now state a few basis results concerning RDEs, proofs can be found in [3, 11, 16].

Proposition 4.1. *If $V\in C_b^3$ and $\mathbf{X}\in\mathcal{C}^\gamma([0,T];\mathbb{R}^d)$ then for each initial condition ξ and any $s<T$, there exists a unique global solution $Y=\Phi(\mathbf{X};V,\xi,s)$. Moreover,*

$$Y(t) = Y(s) + V(Y(s))X(s,t) + \mathbb{V}(Y(s)) : \mathbb{X}(s,t) + R(s,t), \tag{4.4}$$

for all $s,t\in[0,T]$, where $|R(s,t)|\lesssim(1\wedge\|\mathbf{X}\|_\gamma^3)|s-t|^{3\gamma}$.

Remark 4.2. To see that the remainder scales in this particular way, see the proof of [3], Lemma 3.4.

Lemma 4.1. *If $Y=\Phi(\mathbf{X};V,\eta)$ and $\tilde{Y}=\Phi(\mathbf{X};V,\tilde{\eta})$, then $|Y(t)-\tilde{Y}(t)|\lesssim \|\mathbf{X}\|_\gamma|Y(s)-\tilde{Y}(s)|$ for any $s\leq t\leq T$, where the implied constant depends only on T,V.*

Lemma 4.2. *Suppose that $V\in C_b^3$ and $\mathbf{X},\tilde{\mathbf{X}}\in\mathcal{C}^\gamma([0,T];\mathbb{R}^d)$ satisfying $\rho_\gamma(\mathbf{X},0),\rho_\gamma(\tilde{\mathbf{X}},0)\leq M$. Then, on any time window $[0,T]$, the map $\Phi(\cdot)$ satisfies the following local Lipschitz estimate*

$$\|\Phi(\mathbf{X})-\Phi(\tilde{\mathbf{X}})\|_\infty \leq C_M\rho_\gamma(\mathbf{X},\tilde{\mathbf{X}}),$$

where C_M depends only on M and T.

The next lemma is a slight modification of a result in [14], hence we only sketch the proof.

LEMMA 4.3. *Let $0 \leq \gamma < \alpha$. Then the ball*

$$B_{R,\alpha} = \{\mathbf{X} \in \mathcal{C}^{\alpha}([0,T];\mathbb{R}^d) : \|\mathbf{X}\|_{\alpha} \leq R\}$$

is compact in the space $\mathcal{C}^{\gamma}([0,T];\mathbb{R}^d)$.

PROOF. The proof is a standard modification of a similar statement found in [14]. Fix a sequence $\{\mathbf{X}^n\} \subset B_{R,\alpha}$. Use Arzela–Ascoli to find a subsequence that converges in the sup-norm topology. Use the interpolation (4.3) between $\mathcal{C}^{\alpha} \subset \mathcal{C}^{\gamma} \subset \mathcal{C}^0$ to show that this subsequence also converges in $\mathcal{C}^{\gamma}$. Since $\mathcal{C}^{\gamma}$ is a metric space, sequential compactness implies compactness. □

The final result, which is a direct corollary of [14], Theorem 17.3, allows us to translate RDE solutions to Stratonovich SDEs.

LEMMA 4.4. *Suppose that $V \in C_b^3$, $\mathbf{X} = (X, \mathbb{X}) \in \mathcal{C}^{\gamma}([0,T];\mathbb{R}^d)$ and let $Y = \Phi(\mathbf{X}; V)$. Suppose that X is a continuous semi-martingale and that $\mathbb{X}$ can be written*

$$\mathbb{X}^{\alpha\beta}(t) = \int_0^t X^{\beta}(s) \circ dX^{\alpha}(s) + \nu^{\alpha\beta} t, \tag{4.5}$$

for $\alpha, \beta = 1, \dots, d$. where the integral on the right-hand side is defined in the Stratonovich sense and where $\nu \in \mathbb{R}^{d\times d}$. Then Y satisfies the SDE

$$dY = V(Y) \circ dX + \mathbb{V}(Y) : \nu\, dt. \tag{4.6}$$

REMARK 4.3. In all of the results of this section, if in addition to $\gamma > 1/3$, it is also known that γ can be taken arbitrarily close to $1/2$ (as for Brownian rough paths), then the condition $V \in C^3$ can be relaxed to $V \in C^{2+}$. For instance, the versions of the results in any of [3, 14, 16] will adhere to this.

5. Rough path recursions. In this section, we introduce *rough path recursions*, driven by *rough step functions*. Before proceeding with the definitions, we must introduce some assumptions and terminology.

5.1. *Partitions of* $[0,T]$. Fix an interval $[0,T]$ and let $\mathcal{P}_n = \{\tau^n_k : k = 0,\dots,N_n\}$ be a partition of $[0,T]$, that is $0 = \tau^n_0 \le \tau^n_1 \le \cdots \le \tau^n_{N_n} = T$. We also introduce the *mesh size* $\Delta_n = \max_k |\tau^n_{k+1} - \tau^n_k|$. For all the partitions considered in this article, we will assume that

$$\Delta_n \to 0 \qquad \text{as } n\to\infty \quad \text{and} \quad \sup_{n\ge 1} N_n \Delta_n < \infty. \tag{5.1}$$

The first assumption is obviously natural, the second condition is effectively saying that the largest bin the partition does not shrink too much slower than the smallest bin in the partition. Given some $u \in [0,T]$, we will also use the notation $\tau^n(u)$ to denote the largest mesh point $\tau^n(u) \in \mathcal{P}_n$ with $\tau^n(u) \le u$. It follows from (5.1) that $\tau^n(u) \to u$ as $n \to \infty$.

5.2. *Rough path recursions.* We will now define rough step functions and rough path recursions rigorously.

Definition 5.1 (Rough step functions). Fix a partition $\mathcal{P}_n$ of $[0,T]$ and suppose $\xi^n_j \in \mathbb{R}^d$ and $\Xi^n_j \in \mathbb{R}^{d\times d}$ for all $j = 0,\dots,N_n - 1$. The *rough step function* above the increments (ξ^n, Ξ^n) is a path $\mathbf{X}^n = (X^n, \mathbb{X}^n) : [0,T] \to T^2(\mathbb{R}^d)$ defined by

$$X^n(t) = \sum_{j=0}^{k-1} \xi^n_j \quad \text{and} \quad \mathbb{X}^n(t) = \sum_{i=0}^{k-1}\sum_{j=0}^{i-1} \xi^n_i \otimes \xi^n_j + \sum_{i=0}^{k-1} \Xi^n_i,$$

where $\tau^n_k = \tau^n(t)$. We similarly define the incremental paths

$$X^n(s,t) = \sum_{j=l}^{k-1} \xi^n_j \quad \text{and} \quad \mathbb{X}^n(s,t) = \sum_{i=l}^{k-1}\sum_{j=l}^{i-1} \xi^n_i \otimes \xi^n_j + \sum_{i=l}^{k-1} \Xi^n_i,$$

where $\tau^n_k = \tau^n(t)$ and $\tau^n_l = \tau^n(s)$. We will often employ the shorthand $X^n(\tau^n_j, \tau^n_k) = X^n_{j,k}$ and $\mathbb{X}^n(\tau^n_j, \tau^n_k) = \mathbb{X}^n_{j,k}$. We define the discrete γ-Hölder "norm" $\|\!|\cdot|\!\|_{\gamma,n}$ by

$$\|\!|\mathbf{X}^n|\!\|_{\gamma,n} \overset{\text{def}}{=} \max_{\tau^n_j,\tau^n_k \in \mathcal{P}_n} \frac{|X^n_{j,k}|}{|\tau^n_j - \tau^n_k|^\gamma} + \max_{\tau^n_j,\tau^n_k \in \mathcal{P}_n} \frac{|\mathbb{X}^n_{j,k}|^{1/2}}{|\tau^n_j - \tau^n_k|^\gamma}.$$

In particular, we see that

$$|X^n(\tau^n_j, \tau^n_k)| \le \|\!|\mathbf{X}^n|\!\|_{\gamma,n} |\tau^n_j - \tau^n_k|^\gamma \quad \text{and} \quad |\mathbb{X}^n(\tau^n_j, \tau^n_k)| \le \|\!|\mathbf{X}^n|\!\|^2_{\gamma,n} |\tau^n_j - \tau^n_k|^{2\gamma}$$

for all mesh points $\tau^n_j, \tau^n_k \in \mathcal{P}_n$. Since it is a maximum over a finite set, the discrete Hölder norm is finite for every fixed n. It will only play a role in an asymptotic sense.

DEFINITION 5.2 (Rough path recursions). Fix a sequence of partitions $\{\mathcal{P}_n\}_{n\geq 1}$. A *rough path recursion* $\{Y^n\}_{n\geq 1}$ with $Y^n:[0,T]\to\mathbb{R}^e$ is defined by $Y^n(t)=Y^n_j$, where $\tau^n_j=\tau^n(t)$ and satisfying the recursion

$$Y^n_{j+1}=Y^n_j+V(Y^n_j)\xi^n_j+\mathbb{V}(Y^n_j):\Xi^n_j+r^n_j, \tag{5.2}$$

for all $j=0,\dots,N_n-1$ with $Y^n_0=\eta$ and arbitrary $\xi^n_j\in\mathbb{R}^d$ and $\Xi^n_j\in\mathbb{R}^{d\times d}$. The remainder is assumed to satisfy the estimate

$$|r_j|\lesssim(1\wedge\|\mathbf{X}^n\|^3_{\gamma,n})\Delta_n^{3\gamma}, \tag{5.3}$$

where the implied constant is uniform in n and $\mathbf{X}^n$ is the rough step function over the increments (ξ^n,Ξ^n). We will use the notation $Y^n=\Phi^n(\mathbf{X}^n)$.

REMARK 5.1. The estimate (5.3) is picked as it seems to be the most naturally occurring upper bound in applications. However, at the expense of a few extra constraints we could equally use $|r^n_j|\lesssim(1\wedge\phi(C_n))\Delta_n^\lambda$ for any increasing function $\phi:[0,\infty)\to[0,\infty)$ and $\lambda>1$.

REMARK 5.2. Although we require that Y^n be constant in between mesh points, it is easy to see that all the properties of rough path recursions discussed in the sequel are still true if we assume that Y^n is defined by a reasonable interpolation between mesh points. For example, even though the solution to an RDE satisfies the recursion (5.2), it is not a rough path recursion. However, it can be well approximated by a rough path recursion, and any convergence results for rough path recursion easily imply convergence results for the associated RDE solution. As a more general course of action, we could have defined a rough path recursion to any path satisfying (5.2) as well as

$$|Y^n(t)-Y^n(\tau^n(t))|\lesssim D_n\Delta_n^\mu,$$

for some sequence of constants D_n and $\mu>0$. It is clear that, assuming the right conditions on D_n and μ, all the statements made in the sequel regarding rough path recursions are unaltered if we were to adopt this more general definition.

6. Properties of rough path recursions. In this section, we discuss some useful properties of rough step functions and their associated rough path recursions. The main result, Lemma 6.3, states that every rough path recursion can be approximated arbitrarily well by the solution to a rough differential equation. At the heart of this result is the fact that for every rough step function $\mathbf{X}^n$, one can find a genuine rough path $\tilde{\mathbf{X}}^n$ that agrees with $\mathbf{X}^n$ on $\mathcal{P}_n$. This is the content of Lemma 6.2. Before stating the theorems, we must introduce some terminology associated with *geometric* rough paths. For a more detailed exposition of this material, see [14] and also [11], Section 2.2.

We first define $G^2(\mathbb{R}^d)$, the step-2 nilpotent Lie group, by $G^2(\mathbb{R}^d) = \exp(\mathcal{G}^2(\mathbb{R}^d))$ where $\mathcal{G}^2(\mathbb{R}^d)$ is the step-2 Lie algebra over $\mathbb{R}^d$ and where exp is the tensor exponential. In particular, $g \in G^2(\mathbb{R}^d)$ if and only if

$$g = \exp(a_\alpha e_\alpha + b_{\beta,\kappa}[e_\beta, e_\kappa]) = a_\alpha e_\alpha + (\tfrac{1}{2} a_\beta a_\kappa + b_{\beta,\kappa} - b_{\kappa,\beta}) e_\beta \otimes e_\kappa,$$

where $\{e_\alpha\}$ denotes the canonical basis of $\mathbb{R}^d$ and we employ the Einstein summation convention. It is easy to see that every element $A \in T^2(\mathbb{R}^d)$ can be decomposed into

$$A = g + z,$$

where $g \in G^2(\mathbb{R}^d)$ and $z \in \mathrm{Sym}(\mathbb{R}^{d\times d})$. The pair $(G^2(\mathbb{R}^d), \otimes)$ forms a group. This group has a *homogeneous metric* known as the Carnot–Caratheodory metric, defined using geodesic paths. To make this precise, we first define the *signature* of a smooth path. Let $BV([0,1];\mathbb{R}^d)$ be the space of paths $\Gamma : [0,1] \to \mathbb{R}^d$ with bounded variation. For $\Gamma \in BV([0,1];\mathbb{R}^d)$, the signature is defined by

$$\mathbf{S}(\Gamma)(s,t) = \left(\Gamma(t) - \Gamma(s), \int_s^t (\Gamma(u) - \Gamma(s)) \otimes d\Gamma(u)\right),$$

where the integral is constructed in the Riemann–Stieltjes sense. The Carnot–Caratheodory (CC) norm is defined by

$$\|g\|_{\mathrm{CC}} \stackrel{\text{def}}{=} \inf\left\{\int_0^1 |d\Gamma| : \Gamma \in BV([0,1];\mathbb{R}^d) \text{ and } g = \mathbf{S}(\Gamma)(0,1)\right\}.$$

The following result (which is a refinement of Chow's theorem) shows that the norm is well defined (see [14], Theorem 7.32, for a simple proof).

Lemma 6.1. *If $g \in G^2(\mathbb{R}^d)$ then there exists a path $\Gamma : [u,v] \to \mathbb{R}^d$ with $|\dot\Gamma| = $ const such that $g = \mathbf{S}(\Gamma)(u,v)$ and $\|g\|_{\mathrm{CC}} = (v-u)|\dot\Gamma|$.*

The CC norm can also be "compared" with the usual Euclidean norms in the following way. Suppose that $g \in G^2(\mathbb{R}^d)$ can be decomposed into $g = g_1 + g_2$ where $g_1 \in \mathbb{R}^d$ and $g_2 \in \mathbb{R}^{d\times d}$, then we have the comparison

$$|g_1| + |g_2|^{1/2} \lesssim \|g\|_{\mathrm{CC}} \lesssim |g_1| + |g_2|^{1/2}. \tag{6.1}$$

This comparison will be useful in the sequel.

Lemma 6.2. *Let $\{\mathbf{X}^n\}_{n\geq 1}$ be a rough step function on a partition $\{\mathcal{P}_n\}_{n\geq 1}$ and let $\gamma \in (1/3, 1/2]$. For each n, there exists $\tilde{\mathbf{X}}^n \in \mathcal{C}^\gamma([0,T];\mathbb{R}^d)$ with*

$$\tilde{\mathbf{X}}^n(\tau_j^n) = \mathbf{X}^n(\tau_j^n) \qquad \text{for all } \tau_j^n \in \mathcal{P}_n.$$

Moreover, we have that $\|\tilde{\mathbf{X}}^n\|_\gamma \lesssim \|\mathbf{X}^n\|_{\gamma,n}$ where the implied constant is uniform in n.

Proof. We will start by constructing $\tilde{\mathbf{X}}^n(s,t)$ for fixed $s,t \in [\tau_j^n, \tau_{j+1}^n]$ with $s \le t$. First, note that since $\mathbf{X}_{j,j+1}^n \in T^2(\mathbb{R}^d)$, we have the decomposition

$$\mathbf{X}_{j,j+1}^n = g(\tau_j^n, \tau_{j+1}^n) + z(\tau_j^n, \tau_{j+1}^n),$$

where $g(\tau_j^n, \tau_{j+1}^n) \in G^2(\mathbb{R}^d)$ and $z(\tau_j^n, \tau_{j+1}^n) \in \mathrm{Sym}(\mathbb{R}^{d\times d})$ is defined by

$$z(\tau_j^n, \tau_{j+1}^n) = \tfrac{1}{2}(\mathbb{X}_{j,j+1}^{n,\alpha\beta} + \mathbb{X}_{j,j+1}^{n,\beta\alpha} - X_{j,j+1}^{n,\alpha} X_{j,j+1}^{n,\beta}) e_\alpha \otimes e_\beta.$$

We will now define

$$\tilde{\mathbf{X}}^n(s,t) = g(s,t) + z(s,t).$$

First, define $z(s,t)$ by a simple linear interpolation

$$z(s,t) = \frac{t-s}{\tau_{j+1}^n - \tau_j^n} z(\tau_j^n, \tau_{j+1}^n).$$

Now, to define $g(s,t)$, we know from Lemma 6.1 that there exists a path $\Gamma : [\tau_j^n, \tau_{j+1}^n] \to \mathbb{R}^d$ such that $|\dot{\Gamma}| = \text{const}$ and $g(\tau_j^n, \tau_{j+1}^n) = S(\Gamma)(\tau_j^n, \tau_{j+1}^n)$ and $\|g(\tau_j^n, \tau_{j+1}^n)\|_{\mathrm{CC}} = (\tau_{j+1}^n - \tau_j^n)|\dot{\Gamma}|$. We set $g(s,t) = \mathbf{S}(\Gamma)(s,t)$.

We will now define $\tilde{\mathbf{X}}^n(s,t)$ for arbitrary $s,t \in [0,T]$ with $s \le t$. Suppose without loss of generality that $s \le \tau_j^n \le \tau_k^n \le t$ where $\tau_{j-1}^n = \tau^n(s)$ and $\tau_k^n = \tau^n(t)$. Then we define $\tilde{\mathbf{X}}^n(s,t) = (\tilde{X}^n, \tilde{\mathbb{X}}^n)(s,t)$ using Chen's relations

$$\tilde{X}^n(s,t) = \tilde{X}^n(s, \tau_j^n) + X_{j,k}^n + \tilde{X}^n(\tau_k^n, t)$$

and

$$\begin{aligned}\tilde{\mathbb{X}}^n(s,t) = {}& \tilde{\mathbb{X}}^n(s,\tau_j^n) + \mathbb{X}_{j,k}^n + \tilde{\mathbb{X}}^n(\tau_k^n, t) \\ & + X_{j,k}^n \otimes \tilde{X}^n(\tau_k^n, t) + \tilde{X}^n(s, \tau_j^n) \otimes \tilde{X}^n(\tau_j^n, t).\end{aligned}$$

We will now check that $\tilde{\mathbf{X}}^n$ satisfies the requirements of the theorem. First, we will show that Chen's relations hold. It is easy to see that Chen's relations hold when restricted to the interval $s,u,t \in [\tau_j^n, \tau_{j+1}^n]$. Indeed, since $g(s,t)$ is a signature and since $z(s,t)$ is an increment, both objects individually obey Chen's relation. Thus, if we write $g = g_1 + g_2$, with $g_1 \in \mathbb{R}^d$ and $g_2 \in \mathbb{R}^d \otimes \mathbb{R}^d$ then

$$\tilde{\mathbf{X}}^n = (\tilde{X}^n, \tilde{\mathbb{X}}^n) = (g_1, g_2 + z).$$

Therefore,

$$\tilde{X}^n(s,t) = g_1(s,t) = g_1(s,u) + g_1(u,t) = \tilde{X}^n(s,u) + \tilde{X}^n(u,t)$$

and

$$
\begin{aligned}
\mathbb{X}^n(s,t) &= g_2(s,t) + z(s,t) \\
&= g_2(s,u) + g_2(u,t) + g_1(s,u) \otimes g_1(u,t) + z(s,u) + z(u,t) \\
&= (g_2 + z)(s,u) + (g_2 + z)(u,t) + g_1(s,u) \otimes g_1(u,t) \\
&= \tilde{\mathbb{X}}^n(s,u) + \tilde{\mathbb{X}}^n(u,t) + \tilde{X}^n(s,u) \otimes \tilde{X}^n(u,t),
\end{aligned}
$$

as required. Now, for arbitrary $s \le u \le t$ it follows immediately from the construction that $(\tilde{X}^n, \tilde{\mathbb{X}}^n)$ satisfies Chen's relations.

Using the shorthand $C_n = \|\mathbf{X}^n\|_{\gamma,n}$, we will now prove that $\|\tilde{\mathbf{X}}^n\|_\gamma \lesssim C_n$. First, suppose that $s, t \in [\tau^n_j, \tau^n_{j+1}]$ with $s \le t$. Then using the comparison (6.1) and the construction of g, we have that

$$
\begin{aligned}
|\tilde{X}^n(s,t)| = |g_1(s,t)| &\le |g_1(s,t)| + |g_2(s,t)|^{1/2} \\
&\lesssim \|g(s,t)\|_{\mathrm{CC}} = \frac{(t-s)}{(\tau^n_{j+1} - \tau^n_j)} \|g(\tau^n_j, \tau^n_{j+1})\|_{\mathrm{CC}}
\end{aligned}
$$

and again by (6.1) we have that

$$
\begin{aligned}
\|g(\tau^n_j, \tau^n_{j+1})\|_{\mathrm{CC}} &\lesssim |g_1(\tau^n_j, \tau^n_{j+1})| + |g_2(\tau^n_j, \tau^n_{j+1})|^{1/2} \\
&\lesssim |X^n_{j,j+1}| + (|\mathbb{X}^n_{j,j+1}| + |X^n_{j,j+1}|^2)^{1/2} \\
&\lesssim C_n(\tau^n_{j+1} - \tau^n_j)^\gamma.
\end{aligned} \tag{6.2}
$$

It follows that

$$
|\tilde{X}^n(s,t)| \lesssim C_n \frac{(t-s)}{(\tau^n_{j+1} - \tau^n_j)^{1-\gamma}} = C_n (t-s)^\gamma \left(\frac{t-s}{\tau^n_{j+1} - \tau^n_j} \right)^{1-\gamma} \le C_n (t-s)^\gamma,
$$

where in the last inequality we use the fact that $\frac{t-s}{\tau^n_{j+1} - \tau^n_j} \le 1$. By a similar argument, we can show that

$$
|\tilde{\mathbb{X}}^n(s,t)| \lesssim \left(\frac{t-s}{\tau^n_{j+1} - \tau^n_j} \right)^2 (|X^n_{j,j+1}|^2 + |\mathbb{X}^n_{j,j+1}|) + \left(\frac{t-s}{\tau^n_{j+1} - \tau^n_j} \right) |\mathbb{X}^n_{j,j+1}| \tag{6.3}
$$

and hence

$$
|\tilde{\mathbb{X}}^n(s,t)| \lesssim C_n^2 (t-s)^{2\gamma}.
$$

Now suppose $s, t \in [0,T]$ with $s \le \tau^n_j \le \tau^n_k \le t$ as above. By Chen's relations, we have that

$$
\begin{aligned}
|\tilde{X}^n(s,t)| &\le |\tilde{X}^n(s, \tau^n_j)| + |X^n_{j,k}| + |\tilde{X}^n(\tau^n_k, t)| \\
&\lesssim C_n(|\tau^n_j - s|^\gamma + |\tau^n_j - \tau^n_k|^\gamma + |\tau^n_k - t|^\gamma) \le C_n |t-s|^\gamma
\end{aligned}
$$

and

$$\begin{aligned}|\tilde{\mathbb{X}}^n(s,t)| &\le |\tilde{\mathbb{X}}^n(s,\tau_j^n)| + |\mathbb{X}_{j,k}^n| + |\tilde{\mathbb{X}}^n(\tau_k^n,t)| \\ &\qquad + |X_{j,k}^n \otimes \tilde{X}^n(\tau_k^n,t)| + |\tilde{X}^n(s,\tau_j^n)\otimes \tilde{X}^n(\tau_j^n,t)| \\ &\lesssim C_n^2(|\tau_j^n - s|^{2\gamma} + |\tau_j^n - \tau_k^n|^{2\gamma} + |\tau_k^n - t|^{2\gamma} \\ &\qquad + |\tau_j^n - \tau_k^n|^{\gamma}|\tau_k^n - t|^{\gamma} + |\tau_j^n - s|^{\gamma}|\tau_j^n - t|^{\gamma}) \lesssim C_n|t-s|^{2\gamma}.\end{aligned}$$

This completes the proof. $\square$

We now have all the tools we need to prove the main result of this section, namely that rough path recursions can be well approximated by the solution to an RDE.

Lemma 6.3. *Let $Y^n = \Phi^n(\mathbf{X}^n)$. Let $\tilde{\mathbf{X}}^n$ be any rough path satisfying the conditions of Lemma 6.2 and let $\tilde{Y}^n = \Phi(\tilde{\mathbf{X}}^n)$. Then*

$$\|\tilde{Y}^n - Y^n\|_\infty \lesssim (1 \wedge \|\mathbf{X}^n\|_{\gamma,n}^4)\Delta_n^{3\gamma-1},$$

for any $\gamma \in (1/3, 1/2]$, where the implied constant is uniform in n.

Proof. We will again use the shorthand $C_n = \|\mathbf{X}^n\|_{\gamma,n}$. For any $t \in [0,T]$, we have that

$$|\tilde{Y}^n(t) - Y^n(t)| \le |\tilde{Y}^n(t) - \tilde{Y}^n(\tau_k^n)| + |\tilde{Y}^n(\tau_k^n) - Y^n(\tau_k^n)|,$$

where $\tau_k^n = \tau^n(t)$, and hence $Y^n(t) = Y^n(\tau_k^n)$. It follows from (4.4) that

$$\tilde{Y}^n(t) = \tilde{Y}^n(\tau_k^n) + V(\tilde{Y}^n(\tau_k^n))\tilde{X}^n(\tau_k^n,t) + \mathbb{V}(\tilde{Y}^n(\tau_k^n)) : \tilde{\mathbb{X}}^n(\tau_k^n,t) + R(\tau_k^n,t),$$

where $|R(\tau_k^n,t)| \lesssim \|\tilde{\mathbf{X}}^n\|_\gamma^3 (t-\tau_j^n)^{3\gamma} \le C_n^3\Delta_n^{3\gamma}$. Since $\gamma \le 1/2$, it follows that

$$|\tilde{Y}^n(t) - \tilde{Y}^n(\tau_k^n)| \lesssim (1 \wedge C_n^3)\Delta_n^{\gamma} \lesssim (1\wedge C_n^4)\Delta_n^{3\gamma-1}.$$

To estimate $|\tilde{Y}^n(\tau_k^n) - Y^n(\tau_k^n)|$ we need some new terminology. For each $l \le k$, define $Z_k^{(l)}$ by

$$Z_k^{(l)} \stackrel{\text{def}}{=} \Phi(\tilde{\mathbf{X}}^n; V, Y_l^n, \tau_l^n)(\tau_k^n).$$

That is, $Z_k^{(l)} = \hat{Y}(\tau_k^n)$, where $\hat{Y}$ is the unique solution to the RDE driven by $\tilde{\mathbf{X}}^n$ with vector field V, initialised at time τ_l^n with initial condition Y_l^n [recalling that $Y_l^n = Y^n(\tau_l^n)$, as in Definition 5.2]. In particular, we have that $\tilde{Y}^n(\tau_k^n) = Z_k^{(0)}$, $Y_k^n = Z_k^{(k)}$ and

$$Z_{k+1}^{(k)} = Y_k^n + V(Y_k^n)\tilde{X}^n(\tau_k^n,\tau_{k+1}^n) + \mathbb{V}(Y_k^n)\tilde{\mathbb{X}}^n(\tau_k^n,\tau_{k+1}^n) + R(\tau_k^n,\tau_{k+1}^n) \tag{6.4}$$

for any k. It follows that

$$|\tilde{Y}^n(\tau_k^n) - Y^n(\tau_k^n)| = |Z_k^{(0)} - Z_k^{(k)}| \le \sum_{l=0}^{k-1} |Z_k^{(l)} - Z_k^{(l+1)}|. \tag{6.5}$$

But since

$$Z_k^{(l)} = \Phi(\tilde{\mathbf{X}}^n; V, Y_l^n, \tau_l^n)(\tau_k^n) = \Phi(\tilde{\mathbf{X}}^n; V, Z_{l+1}^{(l)}, \tau_{l+1}^n)(\tau_k^n)$$

and $Z_k^{(l+1)} = \Phi(\tilde{\mathbf{X}}^n; V, Z_{l+1}^{(l+1)}, \tau_{l+1}^n)(\tau_k^n)$, it follows from Lemma 4.1 that

$$|Z_k^{(l)} - Z_k^{(l+1)}| \lesssim (1 \wedge \|\tilde{\mathbf{X}}^n\|_\gamma)|Z_{l+1}^{(l)} - Z_{l+1}^{(l+1)}| \le (1 \wedge C_n)|Z_{l+1}^{(l)} - Z_{l+1}^{(l+1)}|. \tag{6.6}$$

By (6.4), we have

$$\begin{aligned}
&Z_{l+1}^{(l)} - Z_{l+1}^{(l+1)} \\
&\quad = Y_l^n + V(Y_l^n)\tilde{X}^n(\tau_l^n, \tau_{l+1}^n) + \mathbb{V}(Y_l^n) : \tilde{\mathbb{X}}^n(\tau_l^n, \tau_{l+1}^n) + R(\tau_l^n, \tau_{l+1}^n) - Y_{l+1}^n \\
&\quad = Y_l^n + V(Y_l^n)X_{l,l+1}^n + \mathbb{V}(Y_l^n) : \mathbb{X}_{l,l+1}^n + R(\tau_l^n, \tau_{l+1}^n) - Y_{l+1}^n \\
&\quad = R(\tau_l^n, \tau_{l+1}^n) - r_l^n,
\end{aligned}$$

where in the last line we have used the fact that $\tilde{\mathbf{X}}^n$ agrees with $\mathbf{X}^n$ on $\mathcal{P}_n$, as well as the recursive definition of the rough path scheme Y^n. Hence, we have that

$$|Z_{l+1}^{(l)} - Z_{l+1}^{(l+1)}| \lesssim 1 \wedge (\|\tilde{\mathbf{X}}^n\|_\gamma^3 + C_n^3)\Delta_n^{3\gamma} \lesssim (1 \wedge C_n^3)\Delta_n^{3\gamma}.$$

It follows from (6.5) and (6.6) that

$$|\tilde{Y}^n(\tau_k^n) - Y^n(\tau_k^n)| \lesssim (1 \wedge C_n^4)N_n\Delta_n^{3\gamma} \lesssim (1 \wedge C_n^4)\Delta_n^{3\gamma-1},$$

where in the last inequality we have used the assumption $\sup_n N_n\Delta_n < \infty$. This completes the proof. $\square$

Proof of Theorem 2.2. All that is required is to show that $\tilde{Y}^n = \Phi(\tilde{\mathbf{X}}^n)$ solves (2.8) where $\tilde{\mathbf{X}}^n = (\tilde{X}^n, \tilde{\mathbb{X}}^n) \in \mathcal{C}^\gamma([0,T]; \mathbb{R}^d)$ is derived in Lemma 6.2. By definition and by construction of $\tilde{\mathbf{X}}^n$ we have that

$$\tilde{Y}^n(t) = \tilde{Y}^n(s) + V(\tilde{Y}^n(s))\tilde{X}^n(s,t) + \mathbb{V}(\tilde{Y}^n(s)) : \tilde{\mathbb{X}}^n(s,t) + o(|t-s|),$$

where $\tilde{X}^n$ is a piecewise smooth path (obtained from the signature realizing g) and

$$\tilde{\mathbb{X}}^n(s,t) = \int_s^t \tilde{X}^n(s,r) \otimes d\tilde{X}^n(r) + \tilde{Z}^n(t) - \tilde{Z}^n(s),$$

where the integral is of Riemann–Stieltjes type and where $\tilde{Z}^n$ is constructed by concatenating the increments $z(s,t)$, in particular $\tilde{Z}^n$ is piecewise Lipschitz. By [14], Theorem 12.14, it follows that $\tilde{Y}^n$ satisfies (2.8). Note that [14], Theorem 12.14, is basically Lemma 4.4 but under the assumption that the driving path is piecewise smooth rather than a semi-martingale. $\square$

6.1. *Discrete Kolmogorov criterion.* In Section 7, we will employ the standard method of lifting weak convergence in the sup-norm topology to weak convergence in some γ-Hölder topology, using a tightness condition. In the *continuous time* setting (which we cannot use), the Kolmogorov–Lamperti criterion [14, 26] is the usual method for checking this tightness condition. The following is a slight modification of a version of the criterion found in Corollary A12 [14].

LEMMA 6.4. *Let $\mathbf{X}^n = (X^n, \mathbb{X}^n)$ define a sequence of rough paths. Suppose that*

$$(\mathbf{E}|X^n(s,t)|^q)^{1/q} \lesssim |t-s|^\alpha \quad \textit{and} \quad (\mathbf{E}|\mathbb{X}^n(s,t)|^{q/2})^{2/q} \lesssim |t-s|^{2\alpha} \tag{6.7}$$

for each $s, t \in [0,T]$, uniformly in $n \geq 1$. Then

$$\sup_{n\geq 1} \mathbf{E}\|\mathbf{X}^n\|_\gamma^q < \infty$$

for any $\gamma \in (0, \alpha - q^{-1})$. In particular, we have that

$$\sup_{n\geq 1} \mathbf{P}(\|\mathbf{X}^n\|_\gamma > M) \to 0 \qquad \textit{as } M \to \infty. \tag{6.8}$$

And moreover $\{\mathbf{X}^n\}_{n\geq 1}$ is tight in the ρ_γ topology for every $\gamma \in (0, \alpha - q^{-1})$.

PROOF. In the case of geometric rough paths [where $\mathbf{X}^n$ takes valued in $G^2(\mathbb{R}^d)$], the result is simply Corollary A12 of [14]. To extend the result to general rough paths, one simply applies the Garcia–Rodemich–Rumsey interpolation result to the components X and $\mathbb{X}$ individually. This argument can be found in [16], Corollary 4. $\square$

Obviously, this result cannot be used directly on rough step functions, since step functions have no hope of satisfying the Kolmogorov estimates. Fortunately, a discrete version of the above result turns out to be equally as useful. We define the *discrete tightness condition* as

$$\sup_{n\geq 1} \mathbf{P}(\|\mathbf{X}^n\|_{\gamma,n} > M) \to 0 \qquad \text{as } M \to \infty. \tag{6.9}$$

This essentially says that the rough step functions are "Hölder continuous," provided we do not look at them too closely (i.e., near the jumps). We will now show that the discrete tightness criterion can likewise be checked using a discrete version of the continuous Kolmogorov criterion. In particular, we need only check the estimate on the partition $\mathcal{P}_n$.

LEMMA 6.5. *Suppose that*

$$(\mathbf{E}|X^n(\tau_j^n, \tau_k^n)|^q)^{1/q} \lesssim |\tau_j^n - \tau_k^n|^\alpha \quad \textit{and} \quad (\mathbf{E}|\mathbb{X}^n(\tau_j^n, \tau_k^n)|^{q/2})^{2/q} \lesssim |\tau_j^n - \tau_k^n|^{2\alpha}$$

for each $\tau_j^n, \tau_k^n \in \mathcal{P}_n$ uniformly in $n \geq 1$, for some $\alpha \in (0, 1/2]$. Then the discrete tightness condition (6.9) holds for any $\gamma \in (0, \alpha - q^{-1})$.

PROOF. The idea behind the proof is to replace $\mathbf{X}^n$ with the genuine rough path $\tilde{\mathbf{X}}^n$ constructed in Lemma 6.2, which, as you recall, agrees with $\mathbf{X}^n$ on $\mathcal{P}_n$. Since

$$\|\!|\mathbf{X}^n|\!\|_{\gamma,n} = \|\!|\tilde{\mathbf{X}}^n|\!\|_{\gamma,n} \le \|\!|\tilde{\mathbf{X}}^n|\!\|_{\gamma},$$

to prove the discrete tightness condition (6.9) it is sufficient to check the Hölder estimate (6.7) for the process $\tilde{\mathbf{X}}^n$ and apply Lemma 6.4. Hence, we need only verify that

$$\mathbf{E}|\tilde{X}^n(s,t)|^q \lesssim |s-t|^{q\alpha} \quad \text{and} \quad \mathbf{E}|\tilde{\mathbb{X}}^n(s,t)|^{q/2} \lesssim |t-s|^{q\alpha}, \tag{6.10}$$

holds for each $s,t \in [0,T]$, uniformly in $n \ge 1$.

Assume without loss of generality that $s,t \in [0,T]$ and $\tau_{j-1}^n < s < \tau_j^n$ and $\tau_k^n \le t < \tau_{k+1}^n$ (note that the case $s,t \in [\tau_j, \tau_{j+1}]$ is essentially a sub-argument of the arguments below). From Chen's relations, we know that

$$\mathbf{E}|\tilde{X}^n(s,t)|^q \lesssim \mathbf{E}|\tilde{X}^n(s,\tau_j^n)|^q + \mathbf{E}|X_{j,k}^n|^q + \mathbf{E}|\tilde{X}^n(\tau_k^n,t)|^q. \tag{6.11}$$

But from (6.2), we see that

$$\begin{aligned}
\mathbf{E}|\tilde{X}(s,\tau_j^n)|^q &\lesssim (\mathbf{E}|X_{j-1,j}^n|^q + \mathbf{E}|\mathbb{X}_{j-1,j}^n|^{q/2})\left(\frac{\tau_j^n - s}{\tau_j^n - \tau_{j-1}^n}\right)^q \\
&\lesssim (\tau_j^n - \tau_{j-1}^n)^{q\alpha}\left(\frac{\tau_j^n - s}{\tau_j^n - \tau_{j-1}^n}\right)^q \\
&= (\tau_j^n - s)^{q\alpha}\left(\frac{\tau_j^n - s}{\tau_j^n - \tau_{j-1}^n}\right)^{q-q\alpha} \le (\tau_j^n - s)^{q\alpha} \le (t-s)^{q\alpha}.
\end{aligned}$$

By assumption, we have that

$$\mathbf{E}|X_{j,k}^n|^q \lesssim (\tau_k^n - \tau_j^n)^{q\alpha} \lesssim (t-s)^{q\alpha}.$$

The remaining term in (6.11) can be bounded similarly. By Chen's relations (and Hölder's inequality), we also have that

$$\begin{aligned}
&\mathbf{E}|\tilde{\mathbb{X}}^n(s,t)|^{q/2} \\
&\quad\lesssim \mathbf{E}|\tilde{\mathbb{X}}^n(s,\tau_j^n)|^{q/2} + \mathbf{E}|\mathbb{X}_{j,k}^n|^{q/2} + \mathbf{E}|\tilde{\mathbb{X}}^n(\tau_k^n,t)|^{q/2} \\
&\qquad + (\mathbf{E}|X_{j,k}^n|^q \mathbf{E}|\tilde{X}^n(\tau_k^n,t)|^q)^{1/2} + (\mathbf{E}|\tilde{X}^n(s,\tau_j^n)|^q \mathbf{E}|\tilde{X}^n(\tau_j^n,t)|^q)^{1/2}.
\end{aligned} \tag{6.12}$$

But from (6.3) we have that

$$\begin{aligned}
\mathbf{E}|\tilde{\mathbb{X}}^n(s,\tau_j^n)|^{q/2} &\lesssim \left(\frac{t-s}{\tau_{j+1}^n - \tau_j^n}\right)^q (\mathbf{E}|X_{j,j+1}^n|^q + \mathbf{E}|\mathbb{X}_{j,j+1}^n|^{q/2}) \\
&\quad + \left(\frac{t-s}{\tau_{j+1}^n - \tau_j^n}\right)^{q/2} \mathbf{E}|\mathbb{X}_{j,j+1}^n|^{q/2}.
\end{aligned}$$

As above, it follows that

$$\mathbf{E}|\tilde{\mathbb{X}}^n(s,\tau_j^n)|^{q/2} \lesssim (\tau_j^n - s)^{q\alpha} \leq (t-s)^{q\alpha}.$$

The other terms in (6.12) can be bounded similarly. This completes the proof. □

REMARK 6.1. The discrete criterion differs from the continuous case in the assumption $\alpha \leq 1/2$, which was not required in the continuous case. However, this assumption only becomes a restriction when the diffusion approximation is driven by a path with Hölder exponent $\gamma > 1/2$. Of course, one can always resolve the problem by treating the path as having the weaker Hölder exponent. On the other hand, in these higher regularity situations the iterated integrals become unnecessary and a much simpler theory of Young integration (with much weaker assumptions) would suffice.

7. Convergence of rough path schemes. We can now prove the main result of the article.

THEOREM 7.1. *Suppose that* $\mathbf{X}^n \stackrel{f.d.d.}{\to} \mathbf{X}$ *and that* $\mathbf{X}^n$ *satisfies the discrete tightness condition (6.9) for some* $\gamma \in (1/3, 1/2]$. *Then* $Y^n = \Phi^n(\mathbf{X}^n) \Rightarrow \Phi(\mathbf{X})$ *in the sup-norm topology.*

PROOF. First, let $\tilde{\mathbf{X}}^n$ be the γ-Hölder rough path constructed in Lemma 6.2 and let $\tilde{Y}^n = \Phi(\tilde{\mathbf{X}}^n)$. To prove the theorem, it is sufficient to first show that $\|Y^n - \tilde{Y}^n\|_\infty \to 0$ in probability and second show that $\tilde{Y}^n \Rightarrow Y$ in the sup-norm topology, hence we will proceed as such.

As usual, we use the shorthand $C_n = \|\!|\mathbf{X}^n|\!\|_{\gamma,n}$. From Lemma 6.3, it follows that

$$\|Y^n - \tilde{Y}^n\|_\infty \lesssim (1 \wedge C_n^4)\Delta_n^{3\gamma-1}.$$

Hence,

$$\begin{aligned}\mathbf{P}(\|Y^n - \tilde{Y}^n\|_\infty > \delta) &\leq \mathbf{P}(C(1 \wedge C_n^4)\Delta_n^{3\gamma-1} > \delta) \\ &= \mathbf{P}(1 \wedge \|\!|\mathbf{X}^n|\!\|_{\gamma,n} > (C^{-1}\delta\Delta_n^{1-3\gamma})^{1/4}).\end{aligned}$$

But since $\Delta_n^{1-3\gamma} \to \infty$ as $n \to \infty$, we see that for any arbitrarily large $M > 0$

$$\begin{aligned}\lim_{n\to\infty} \mathbf{P}(\|Y^n - \tilde{Y}^n\|_\infty > \delta) &\leq \limsup_{n\to\infty} \mathbf{P}(\|\!|\mathbf{X}^n|\!\|_{\gamma,n} > C^{-1}\delta^{1/4}\Delta_n^{(1-3\gamma)/4}) \\ &\leq \limsup_{n\to\infty} \mathbf{P}(\|\!|\mathbf{X}^n|\!\|_{\gamma,n} > M).\end{aligned}$$

Finally, by taking $M \to \infty$, it follows from the discrete tightness condition that

$$\lim_{n\to\infty} \mathbf{P}(\|Y^n - \tilde{Y}^n\|_\infty > \delta) = 0.$$

Now we prove that $\tilde{Y}^n \Rightarrow Y$ in the sup-norm topology. Due to the continuity of the map Φ, as stated in Lemma 4.2, it is sufficient to prove that $\tilde{\mathbf{X}}^n \Rightarrow \mathbf{X}$ in the ρ_γ topology. It is therefore sufficient to first show that $\tilde{\mathbf{X}}^n \stackrel{\text{f.d.d.}}{\to} \mathbf{X}$ and second that $\{\tilde{\mathbf{X}}^n\}_{n\geq 1}$ is tight in the ρ_γ topology.

First, due to the regularity of $\tilde{\mathbf{X}}^n$ between mesh points, it is easy to see that $\|\mathbf{X}^n(t) - \tilde{\mathbf{X}}^n(t)\|_\infty \lesssim (1 \vee C_n^2)\Delta_n^\kappa$ for some $\kappa > 0$. Hence, by an argument similar to that found at the start of the proof, it follows from the discrete tightness condition that $\|\mathbf{X}^n - \tilde{\mathbf{X}}^n\|_\infty \to 0$ in probability. And since by assumption $\mathbf{X}^n \stackrel{\text{f.d.d.}}{\to} \mathbf{X}$ it follows that $\tilde{\mathbf{X}}^n \stackrel{\text{f.d.d.}}{\to} \mathbf{X}$. We will now move onto the tightness argument. From Lemma 6.2, we have the estimate $\|\tilde{\mathbf{X}}^n\|_\gamma \lesssim \|\mathbf{X}^n\|_{\gamma,n}$. It follows that

$$\mathbf{P}(\|\tilde{\mathbf{X}}^n\|_\gamma > M) \leq \mathbf{P}(C\|\mathbf{X}^n\|_{\gamma,n} > M),$$

and the tightness of $\tilde{\mathbf{X}}^n$ in the ρ_γ topology follows from the discrete tightness condition. This completes the proof. $\square$

We can now prove the theorems introduced in Section 2. They are both immediate corollaries.

Proof of Theorem 2.1. By the discrete Kolmogorov criterion (Lemma 6.5), we obtain the discrete tightness criterion, and hence can apply Theorem 7.1. To identify the limit $Y = \Phi(\mathbf{X})$, we simply apply Lemma 4.4. $\square$

Remark 7.1. To prove the result with the relaxed assumption described in Remark 2.2, one simply replaces Lemma 4.2 with the sharper version [14], Theorem 12.10, and the remaining argument is identical. To prove the result with additional drift, as described in Remark 2.3, we again replace Lemma 4.2 with [14], Theorem 12.10, but now we must use the (p,q)-rough path $(\tilde{\mathbf{X}}^n, t)$ with $(p,q) = (2-\kappa, 1)$ and arbitrarily small $\kappa > 0$. The remaining argument is identical.

Proof of Meta Theorem 2.1. The proof is completely identical to the proof above, but we still need to "interpret" the limit $Y = \Phi(\mathbf{X})$. This is a fairly nonrigorous statement and, therefore, has a fairly nonrigorous proof. We are merely sketching an idea that would apply more rigorously in concrete situations.

By definition, the limit Y solves the RDE

$$Y(t) = \int_0^t V(Y(s))\, d\mathbf{X}(s). \tag{7.1}$$

It is a general *heuristic* that if $\mathbb{X}$ is constructed using some known construction then the integral in (7.1) is constructed similarly. For instance, suppose there is some "method of integration," which is a bilinear operator

$$\int_0^t A \star dB \stackrel{\text{def}}{=} \mathcal{I}(A,B)(t)$$

for two continuous paths $A, B \in C([0,T];\mathbb{R})$ satisfying the obvious condition

$$\mathcal{I}(1,A)(t) = A(t).$$

Now suppose that $\mathbb{X}$ is defined by

$$\mathbb{X}^{\alpha\beta}(t) = \mathcal{I}(X^\alpha, X^\beta)(t),$$

for each $\alpha, \beta = 1, \ldots, d$. Then, using the theory of controlled rough paths [16, 17], it can be shown that Y solves (7.1) if and only if Y is a fixed point of the equation

$$Y(t) = Y(0) + \int_0^t V(Y(s)) \star dX(s).$$

The assumptions on $\mathcal{I}$ are generic enough to include virtually any reasonable construction of an integration map (for integrators with Hölder exponent $\gamma > 1/3$). $\square$

Acknowledgements. David Kelly would like to thank I. Melbourne for introducing the motivating problem and P. Friz for several constructive comments.

REFERENCES

[1] BÁLINT, P. and MELBOURNE, I. (2008). Decay of correlations and invariance principles for dispersing billiards with cusps, and related planar billiard flows. *J. Stat. Phys.* **133** 435–447. MR2448631

[2] BREUILLARD, E., FRIZ, P. and HUESMANN, M. (2009). From random walks to rough paths. *Proc. Amer. Math. Soc.* **137** 3487–3496. MR2515418

[3] DAVIE, A. M. (2007). Differential equations driven by rough paths: An approach via discrete approximation. *Appl. Math. Res. Express. AMRX* Art. ID abm009, 40. MR2387018

[4] DAVYDOV, JU. A. (1970). The invariance principle for stationary processes. *Teor. Verojatnost. i Primenen.* **15** 498–509. MR0283872

[5] DE SIMOI, J. and LIVERANI, C. (2014). The martingale approach after Varadhan and Dolgopyat. Preprint. Available at arXiv:1402.0090.

[6] DEBUSSCHE, A. and FAOU, E. (2012). Weak backward error analysis for SDEs. *SIAM J. Numer. Anal.* **50** 1735–1752. MR2970763

[7] DEYA, A., NEUENKIRCH, A. and TINDEL, S. (2012). A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 518–550. MR2954265

[8] Dolgopyat, D. (2004). Limit theorems for partially hyperbolic systems. *Trans. Amer. Math. Soc.* **356** 1637–1689 (electronic). MR2034323

[9] Duffie, D. and Protter, P. (1992). From discrete to continuous time finance: Weak convergence of the financial gain process. *Math. Finance* **2** 1–15.

[10] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York. MR1964455

[11] Friz, P. and Hairer, M. (2014). *A Course on Rough Paths*. Springer, Berlin.

[12] Friz, P. and Riedel, S. (2014). Convergence rates for the full Gaussian rough paths. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 154–194. MR3161527

[13] Friz, P. and Victoir, N. (2008). Euler estimates for rough differential equations. *J. Differential Equations* **244** 388–412. MR2376201

[14] Friz, P. K. and Victoir, N. B. (2010). *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. *Cambridge Studies in Advanced Mathematics* **120**. Cambridge Univ. Press, Cambridge. MR2604669

[15] Gottwald, G. A. and Melbourne, I. (2013). Homogenization for deterministic maps and multiplicative noise. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **469** 20130201, 16. MR3064670

[16] Gubinelli, M. (2004). Controlling rough paths. *J. Funct. Anal.* **216** 86–140. MR2091358

[17] Gubinelli, M. (2010). Ramification of rough paths. *J. Differential Equations* **248** 693–721. MR2578445

[18] Guess, H. A. and Gillespie, J. H. (1977). Diffusion approximations to linear stochastic difference equations with stationary coefficients. *J. Appl. Probab.* **14** 58–74. MR0428489

[19] Holland, M. and Melbourne, I. (2007). Central limit theorems and invariance principles for Lorenz attractors. *J. Lond. Math. Soc. (2)* **76** 345–364. MR2363420

[20] Kelly, D. and Melbourne, I. (2014). Deterministic homogenization for fast slow systems with chaotic noise. Available at arXiv:1409.5748.

[21] Kelly, D. and Melbourne, I. (2015). Smooth approximations of stochastic differential equations. *Ann. Probab.* To appear.

[22] Kelly, D. T. B., Law, K. J. H. and Stuart, A. M. (2014). Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time. *Nonlinearity* **27** 2579–2604. MR3265724

[23] Kloeden, P. E. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. *Applications of Mathematics (New York)* **23**. Springer, Berlin. MR1214374

[24] Kurtz, T. G. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19** 1035–1070. MR1112406

[25] Kushner, H. J. and Huang, H. (1981). On the weak convergence of a sequence of general stochastic difference equations to a diffusion. *SIAM J. Appl. Math.* **40** 528–541. MR0614747

[26] Lamperti, J. (1962). On convergence of stochastic processes. *Trans. Amer. Math. Soc.* **104** 430–435. MR0143245

[27] Lejay, A. and Lyons, T. (2005). On the importance of the Lévy area for studying the limits of functions of converging stochastic processes. Application to homogenization. In *Current Trends in Potential Theory*. *Theta Ser. Adv. Math.* **4** 63–84. Theta, Bucharest. MR2243956

[28] Lejay, A. and Victoir, N. (2006). On (p, q)-rough paths. *J. Differential Equations* **225** 103–133. MR2228694

[29] Lyons, T. J. (1998). Differential equations driven by rough signals. *Rev. Mat. Iberoam.* **14** 215–310. MR1654527

[30] Majda, A. J., Timofeyev, I. and Vanden Eijnden, E. (2001). A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math.* **54** 891–974. MR1829529

[31] McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2** 620–628. MR0358933

[32] Melbourne, I. and Nicol, M. (2009). A vector-valued almost sure invariance principle for hyperbolic dynamical systems. *Ann. Probab.* **37** 478–505. MR2510014

[33] Norman, M. F. (1974). Markovian learning processes. *SIAM Rev.* **16** 143–162. MR0343372

[34] Penland, C. and Ewald, B. D. (2008). On modelling physical systems with stochastic models: Diffusion versus Lévy processes. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **366** 2457–2476. MR2415440

[35] Perkowski, N. and Promel, D. J. (2013). Pathwise stochastic integrals for model free finance. Preprint. Available at arXiv:1311.6187.

[36] Roberts, G. O., Gelman, A. and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. MR1428751

[37] Russo, F. and Vallois, P. (1993). Forward, backward and symmetric stochastic integration. *Probab. Theory Related Fields* **97** 403–421. MR1245252

[38] Stroock, D. W. and Varadhan, S. R. S. (2006). *Multidimensional Diffusion Processes*. Springer, Berlin. Reprint of the 1997 edition. MR2190038

[39] Taqqu, M. S. (1974/75). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31** 287–302. MR0400329

[40] Vervaat, W. (1979). On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. in Appl. Probab.* **11** 750–783. MR0544194

[41] Watterson, G. A. (1964). The application of diffusion theory to two population genetic models of Moran. *J. Appl. Probab.* **1** 233–246. MR0168048

[42] Young, L.-S. (1998). Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math. (2)* **147** 585–650. MR1637655

[43] Zygalakis, K. C. (2011). On the existence and the applications of modified equations for stochastic differential equations. *SIAM J. Sci. Comput.* **33** 102–130. MR2783188

Courant Institute of
Mathematical Sciences
New York University
New York, New York 10012
E-mail: dtbkelly@gmail.com
URL: http://www.dtbkelly.com