

FROM ROUGH PATH ESTIMATES TO MULTILEVEL MONTE CARLO

CHRISTIAN BAYER, PETER K. FRIZ, SEBASTIAN RIEDEL, AND JOHN SCHOENMAKERS

ABSTRACT. New classes of stochastic differential equations can now be studied using rough path theory (e.g. Lyons et al. [LCL07] or Friz–Hairer [FH14]). In this paper we investigate, from a numerical analysis point of view, stochastic differential equations driven by Gaussian noise in the aforementioned sense. Our focus lies on numerical implementations, and more specifically on the saving possible via multilevel methods. Our analysis relies on a subtle combination of pathwise estimates, Gaussian concentration, and multilevel ideas. Numerical examples are given which both illustrate and confirm our findings.

1. INTRODUCTION

We consider implementable schemes for large classes of stochastic differential equations (SDEs)

$$(1) \quad dY_t = V_0(Y_t) dt + \sum_{i=1}^d V_i(Y_t) dX_t^i(\omega)$$

driven by multidimensional Gaussian signals, say $X = X_t(\omega) \in \mathbb{R}^d$. The interpretation of these equations is in Lyons’ *rough path* sense [LQ02, LCL07, FV10b, FH14]. In essence, one deals with generalizations of classical Stratonovich SDE meaning for such equations. One requires some smoothness/boundedness conditions on the vector fields V_0 and $V \equiv (V_1, \dots, V_d)$; for the sake of this introduction, the reader may assume bounded vector fields with bounded derivatives of all order (but we will be more specific later). This also requires a “natural” lift of $X(\cdot, \omega)$ to a (random) rough path

$$(2) \quad \mathbf{X}_t(\omega) = \sum_{i=1}^N \int_{0 < s_1 < \dots < s_i < t} dX_{s_1}(\omega) \otimes \dots \otimes dX_{s_i}(\omega),$$

see, e.g., [LCL07, Ch. 3]. In (2), N is related to the roughness of X . For instance, in the case of Brownian motion, we need $N = 2$, but rougher processes X require $N > 2$. The reader not familiar with rough path theory may think of Y as “Stratonovich” solution to (1). In fact, Y is known to be the Wong-Zakai limit, obtained by replacing X in (1) by piecewise-linear approximation followed by taking the mesh-to-zero limit.

We shall simplify the discussion by choosing $V_0 \equiv 0$ and using the short-hand notation

$$(3) \quad dY_t = V(Y_t) d\mathbf{X}_t,$$

indicating that the differential equation is driven by the rough path \mathbf{X} given in (2). Of course, it would be easy to include equations of the form (1) into the framework (3), e.g., by including time t as an additional (smooth) component of the noise X . This setting includes, for instance, fractional

2010 *Mathematics Subject Classification.* 60G15, 60H10, 60H35, 65C05, 65C30.

Key words and phrases. Gaussian processes, stochastic differential equations, numerical methods for stochastic equations, Monte Carlo methods.

Brownian motion (fBm) with Hurst parameter $H > 1/4$, see [CQ02]. It may help the reader to recall that, in the case when $X = B$, a multidimensional Brownian motion, all this amounts to enhance B with *Lévy's stochastic area* or, equivalently, with all iterated stochastic integrals of B against itself, say $\mathbb{B}_{s,t} = \int_s^t B_{s,u} \otimes dB_u$. The (rough-)pathwise solution concept then agrees with the usual notion of an SDE solution (in Itô- or Stratonovich sense, depending on which integration was used in defining \mathbb{B}). As is well-known this provides a robust extension of the usual Itô framework of stochastic differential equations with an exploding number of new applications (including non-linear SPDE theory, robustness of the filtering problem, non-Markovian Hörmander theory).

In a sense, the rough path interpretation of a differential equation is closely related to strong, pathwise error estimates of Euler- resp. Milstein-approximation to stochastic differential equations. For instance, Davie's definition [Dav07] of a (rough)pathwise SDE solution is

$$(4) \quad Y_t - Y_s \equiv Y_{s,t} = V_i(Y_s) B_{s,t}^i + V_i^k(Y_s) \partial_k V_j(Y_s) \mathbb{B}_{s,t}^{i,j} + o(|t-s|) \text{ as } t-s \downarrow 0,$$

where we employ Einstein's convention. In fact, this becomes an entirely *deterministic* definition, only assuming

$$\exists \alpha \in (1/3, 1/2) : |B_{s,t}| \leq C |t-s|^\alpha, |\mathbb{B}_{s,t}| \leq C |t-s|^{2\alpha},$$

something which is known to hold true almost surely (i.e. for $C = C(\omega) < \infty$ a.s.), and something which is not at all restricted to Brownian motion. As the reader may suspect this approach leads to almost-sure convergence (with rates) of schemes which are based on the iteration of the approximation seen in the right-hand-side of (4). The practical trouble is that Lévy's area, the anti-symmetric part of \mathbb{B} , is notoriously difficult to simulate; leave alone the simulation of Lévy's area for other Gaussian processes. It has been understood for a while, at least in the Brownian setting, that the truncated (or: simplified) Milstein scheme, in which Lévy's area is omitted, i.e. replace $\mathbb{B}_{s,t}$ by $Sym(\mathbb{B}_{s,t})$ in (4), still offers benefits: For instance, Talay [Tal86] replaces Lévy area by suitable Bernoulli r.v. such as to obtain weak order 1 (see also Kloeden–Platen [KP92] and the references therein).¹ In the multilevel context, [GS14] use this truncated Milstein scheme together with a sophisticated antithetic (variance reduction) method. Finally, in the rough path context this scheme was used in [DNT12]: the convergence of the scheme can be traced down to an underlying Wong-Zakai type approximation for the driving random rough path – a (probabilistic!) result which is known to hold in great generality for stochastic processes, starting with [CQ02] in the context of fractional Brownian motion, see [FV10b, Ch. 15] and the references therein.

A rather difficult problem is to go from almost-sure convergence (with rates) to L^1 (or even: L^r any $r < \infty$) convergence. Indeed, as pointed out in [DNT12, Remark 1.2]: *"Note that the almost sure estimate [for the simplified Milstein scheme] cannot be turned into an L^1 -estimate [...]. This is a consequence of the use of the rough path method, which exhibits non-integrable (random) constants."* The resolution of this problem forms the first contribution of this paper. It is based on some recent progress [CLL13], see also [FR13], initially developed to prove smoothness of laws for (non-Markovian) SDEs driven by Gaussian signals under a Hörmander condition, [CF10, HP13].

Having established L^r -convergence (any $r < \infty$, with rates) for implementable "simplified" Milstein schemes we move to the second contribution of the paper: a multilevel algorithm, in the sense of Giles [Gil08b], for stochastic differential equations driven by large classes of Gaussian signals.

The savings here are rather dramatic. In absence of Markovian structure, the strong rate must proxy for the weak rate, which leaves one with complexity $\mathcal{O}(\epsilon^{-\theta-2})$, any $\theta > 2\rho/(2-\rho)$ where

¹A well-known counter-example by Clark and Cameron [CC80] shows that it is impossible to get strong order 1 if only using Brownian increments.

the parameter ρ quantifies the roughness of the noise. With multilevel, we reduce this to $\mathcal{O}(\epsilon^{-\theta})$. For instance, in the case of fBm with $H = 0.4$, one has $\rho = 1.25$ and thus $\theta \sim 3.33$, which is only marginally worse than single level Monte Carlo in the Brownian motion context (where one has $\theta = 3$). On the other hand, the computational cost of single-level Monte Carlo would be proportional to $\epsilon^{-\theta-2} \sim \epsilon^{-5.33}$ which corresponds to a numerically non-feasible situation.

A *strong*, L^2 error estimate (“rate $\beta/2$ ”) is the key assumption in Giles’ complexity theorem, and this is precisely what we have established in the first part. Some other extension of the Giles theorem are necessary; indeed it is crucial to allow for a weak rate of convergence $\alpha < 1/2$ (ruled out explicitly in [Gil08b]) whenever we deal with driving signals with sample path regularity “worse” than Brownian motion. We are able to do all this, moreover we carefully keep track of the relevant constants in front of the asymptotic terms, a necessity in such an irregular regime.

Let us now discuss the algorithm in more detail. We consider the following scheme for approximating Y , see [DNT12, FV10b]. Given an equi-distant dissection $D = (t_k)$ of $[0, T]$ with mesh h , so that $t_{k+1} - t_k \equiv h$ for all k , write $X_{t_k, t_{k+1}}$ for the corresponding increments. We then define $\bar{Y}_0 \equiv Y_0$ and the “simplified” step-3 Euler scheme

$$(5) \quad \bar{Y}_{t_{k+1}} = \bar{Y}_{t_k} + \sum_{l=1}^3 \frac{1}{l!} V_{i_1} \cdots V_{i_l} I(\bar{Y}_{t_k}) X_{t_k, t_{k+1}}^{i_1} \cdots X_{t_k, t_{k+1}}^{i_l},$$

where $I(y) = y$ is the identity function and the vector fields V_1, \dots, V_d , unless otherwise stated assumed bounded with bounded derivatives of all orders, are viewed as linear first order operators. Whenever convenient we extend \bar{Y} to $[0, T]$ by linear interpolation. Moreover, the Einstein summation convention is in force. For a more detailed description of the algorithm we refer to Section 3.3. We are now able to state (a simple version of) our main results; cf. Corollary 17:

Theorem 1 (Strong rates). *Let $X = (X^1, \dots, X^d)$ be a continuous, zero-mean Gaussian process with independent components. Assume furthermore that each component has stationary increments and that*

$$\sigma^2(t-s) := E |X_t^i - X_s^i|^2$$

where σ^2 is concave and $\sigma^2(\tau) = \mathcal{O}(\tau^{1/\rho})$ as $\tau \rightarrow 0$ for some $\rho \in [1, 2)$.

Let Y be the solution to the rough differential equation (3) driven by (the rough path lift) of X and $\bar{Y} = \bar{Y}^h$ be the approximate solution based on (5). Then we have strong convergence of (almost) rate $1/\rho - 1/2$. More precisely, for any $1 \leq r < \infty$ and $\delta > 0$, there exists a constant C such that

$$\left| E \left(\sup_{t \in [0, T]} |Y_t - \bar{Y}_t^h|^r \right) \right|^{\frac{1}{r}} \leq Ch^{1/\rho - 1/2 - \delta}.$$

The reader should notice that the assumption on X is met by multidimensional Brownian motion (with $\rho = 1$) in which case Y is nothing but a Stratonovich solution of the SDE (1), which of course may be rewritten as Itô equation. More interestingly, X may be a fractional Brownian motion (with $\rho = \frac{1}{2H} > 1$) in the (interesting) “rougher than Brownian” regime $H \in (1/4, 1/2)$. However, stationarity of increments of X plays very little role aside from allowing an easy-to-state formulation above. The precise technical requirement is given in Condition 10 below and is satisfied in many examples, see [FGGR15].

Using Giles’ multi-level Monte Carlo methodology, we can greatly improve the complexity bounds for the discretization algorithm (5), see Theorem 23. Note that the overall computational cost depends on the computational cost of simulating the increments of the process X as needed in

equation (5). For each trajectory of Y , we need to sample a vector of increments of X of length $N \equiv T/h$, i.e., we need to sample from an N -dimensional normal distribution with known non-diagonal covariance matrix. In general, this can be achieved at cost proportional to N^2 by multiplication of a standard normal vector with the lower triangular factor obtained by the Cholesky factorization of the covariance matrix. In many special cases, for instance for fBm, the cost can be reduced to $N \log N$, see [Die04], and we concentrate on this scenario below. The precise statement is given in Theorem 23.

Theorem 2 (Multilevel complexity estimate). *Let X and Y be as in the previous theorem and $f : C([0, T], \mathbb{R}^m) \rightarrow \mathbb{R}^n$ a Lipschitz continuous functional. Assume that the computational cost of generating a vector of (non-overlapping) increments of X of length N is proportional to $N \log N$. Then the Monte Carlo evaluation of a path-dependent functional of the form*

$$E(f(Y_t : 0 \leq t \leq T))$$

to within a MSE of ε^2 , can be achieved with computational work

$$\mathcal{O}(\varepsilon^{-\theta}), \quad \forall \theta > \frac{2\rho}{2-\rho}.$$

As a sanity check, let us compare this results with the corresponding, well-known results for classical stochastic differential equations (here: in Stratonovich sense) driven by d -dimensional Brownian motion B . The assumptions on X are clearly met with $\rho = 1$. As a consequence, we obtain strong convergence of (almost) rate $1/2$ in agreement with the well-known strong rate $1/2$ in the classical setting. Concerning our multilevel complexity estimate, we obtain (almost) order ε^{-2} which is arbitrarily "close" to known result $\mathcal{O}(\varepsilon^{-2} (\log \varepsilon)^2)$ [Gil08a, Gil08b], recently sharpened to $\mathcal{O}(\varepsilon^{-2})$ [GS14] with the aid of a suitable antithetic multilevel correction estimator.

Let us summarize the (computational) benefits of the multilevel approach in the present ("rougher than Brownian") setting. A direct Monte Carlo implementation of the scheme (5) would require a complexity of $\mathcal{O}(\varepsilon^{-(2+1/\alpha)})$ in order to attain an MSE of no more than ε^2 . Here, α is the *weak* rate of convergence of the scheme. On the other hand, we show in Theorem 18 that the complexity is only $\mathcal{O}(\varepsilon^{-(1+2\alpha-\beta)/\alpha})$ for the multi-level Monte Carlo estimator, where β is two times the strong rate of convergence. Thus, when the weak rate of convergence is equal to the strong rate of convergence², then the complexity of the multi-level estimator is reduced by a factor ε^2 as compared by the complexity of the standard Monte Carlo estimator. When the weak rate is two times the strong rate, the speed up is still by a factor ε , see Table 1 and Table 2.

2. ROUGH PATH ESTIMATES REVISITED

In this section, we revisit some classical estimates used in rough paths theory. Definitions of the basic objects and all relevant notation may be found in the appendix. A more detailed account to the theory of rough paths may be found in the monographs [LQ02], [LCL07], [FV10b] or [FH14].

Versions of the results we are interested in (cf. the forthcoming theorems 4 and 8) are already stated in the above mentioned references, and the given estimates are (essentially) sharp when the oscillations of the driving rough path (i.e. its p -variation) become small. However, it turns out that they are less useful when its oscillations get large. In this case, we will show that the estimates can

²By lack of the Markov property, the standard techniques of deriving weak error estimates fail in the setting of an RDE driven by a general Gaussian process such as a fBm. Thus, computing the weak rate of convergence for the simplified Euler scheme would be a non-trivial task. On the other hand, we present a numerical example in Section 5, where the weak order is equal to the strong order even in a standard Brownian motion setting.

be improved by replacing the occurring p -variation norm of the rough path \mathbf{x} in the inequality by another quantity $N(\mathbf{x})$ which was first introduced by Cass-Litterer-Lyons in [CLL13] (and which we recall in Definition 3 below). This becomes crucial when substituting the deterministic rough path \mathbf{x} by the lift of a Gaussian process \mathbf{X} : A key result in [CLL13] states that the quantity $N(\mathbf{X})$ enjoys significantly better integrability properties than the p -variation of \mathbf{X} .

The aim of this section is to show that one can indeed improve the bound for the Lipschitz constant of the Itô-Lyons map (Theorem 4) and the estimate for the distance of the solution of a rough differential equation (RDE) and its Euler (or Milstein) approximation (Theorem 8). This will allow us to deduce the desired probabilistic estimates in the forthcoming section 3.

2.1. Improved bounds for the Lipschitz constant of the Itô-Lyons map. Recall the following definition, taken from [CLL13]:

Definition 3. Let ω be a control function, that is, a continuous function $\omega: \{(s, t) : 0 \leq s \leq t \leq T\} \rightarrow [0, \infty)$ for which $\omega(s, t) + \omega(t, u) \leq \omega(s, u)$ holds for every $s \leq t \leq u$. For $\alpha > 0$ and $[s, t] \subset [0, T]$, we set

$$\begin{aligned} \tau_0(\alpha) &= s \\ \tau_{i+1}(\alpha) &= \inf \{u : \omega(\tau_i(\alpha), u) \geq \alpha, \tau_i(\alpha) < u \leq t\} \wedge t \end{aligned}$$

and define

$$N_{\alpha, [s, t]}(\omega) = \sup \{n \in \mathbb{N} \cup \{0\} : \tau_n(\alpha) < t\}.$$

When ω arises from the (homogenous) p -variation norm $\|\cdot\|_{p\text{-var}}$ of a (p -rough) path, \mathbf{x} , i.e. $\omega_{\mathbf{x}} = \|\mathbf{x}\|_{p\text{-var}; [\cdot, \cdot]}^p$ with $p \geq 1$, we shall also write $N_{\alpha, [s, t]}(\mathbf{x}) := N_{\alpha, [s, t]}(\omega_{\mathbf{x}})$.

It is easy to see that $\alpha N_{\alpha, [0, T]}(\mathbf{x}) \leq \|\mathbf{x}\|_{p\text{-var}; [0, T]}^p$ ([CLL13, Lemma 4.9]), and this is sharp (as one can see choosing $\alpha \nearrow \|\mathbf{x}\|_{p\text{-var}; [0, T]}^p$). However, for fixed α , the tail estimates for $N_{\alpha, [0, T]}(\mathbf{X})$ are significantly better than for $\|\mathbf{X}\|_{p\text{-var}; [0, T]}^p$ when we consider Gaussian lifts \mathbf{X} , cf. [CLL13] and [FR13].

Next, we give the main result of this section. The following theorem is a variant of [FV10b, Theorem 10.38]. The main difference is that in [FV10b, Theorem 10.38], the Lipschitz constant is (essentially) given by $C \exp \left\{ C \left(\|\mathbf{x}^1\|_{p\text{-var}; [0, T]}^p + \|\mathbf{x}^2\|_{p\text{-var}; [0, T]}^p \right) \right\}$, whereas in the following theorem, $\|\mathbf{x}^i\|_{p\text{-var}; [0, T]}^p$ is replaced by $N_{\alpha, [0, T]}(\mathbf{x}^i)$, $i = 1, 2$.

Theorem 4. Consider the RDEs

$$dy_t^i = V^i(y_t^i) d\mathbf{x}_t^i; \quad y_0^i \in \mathbb{R}^e$$

for $i = 1, 2$ on $[0, T]$ where V^1 and V^2 are two families of vector fields, $\gamma > p$ and ν is a bound on $|V^1|_{Lip^\gamma}$ and $|V^2|_{Lip^\gamma}$. Then for every $\alpha > 0$ there is a constant $C = C(\gamma, p, \nu, \alpha)$ such that

$$\begin{aligned} |y^1 - y^2|_{\infty; [0, T]} &\leq C \left[|y_0^1 - y_0^2| + |V^1 - V^2|_{Lip^{\gamma-1}} + \rho_{p\text{-var}; [0, T]}(\mathbf{x}^1, \mathbf{x}^2) \right] \\ &\quad \times \exp \left\{ C \left(N_{\alpha, [0, T]}(\mathbf{x}^1) + N_{\alpha, [0, T]}(\mathbf{x}^2) \right) \right\} \end{aligned}$$

holds.

The proof of Theorem 4 will be given at the end of this section. We first prove some preparatory lemmata. Recall that if ω^1 and ω^2 are controls, also $\omega^1 + \omega^2$ is a control.

Lemma 5. *Let ω^1 and ω^2 be two controls. Then*

$$N_{\alpha,[s,t]}(\omega^1 + \omega^2) \leq 2N_{\alpha,[s,t]}(\omega^1) + 2N_{\alpha,[s,t]}(\omega^2) + 2$$

for every $s < t$ and $\alpha > 0$.

Proof. If ω is any control, set

$$\omega_\alpha(s, t) := \sup \left\{ \sum_{i=0}^{M-1} \omega(t_i, t_{i+1}) : s = t_0 < t_1 < \dots < t_M = t, \omega(t_i, t_{i+1}) \leq \alpha, n \in \mathbb{N} \right\}.$$

If $\bar{\omega} := \omega^1 + \omega^2$, $\bar{\omega}(t_i, t_{i+1}) \leq \alpha$ implies $\omega^i(t_i, t_{i+1}) \leq \alpha$ for $i = 1, 2$ and therefore $\bar{\omega}_\alpha(s, t) \leq \omega_\alpha^1(s, t) + \omega_\alpha^2(s, t)$. From Proposition 4.6 in [CLL13] we know that $\omega_\alpha^i(s, t) \leq \alpha(2N_{\alpha,[s,t]}(\omega^i) + 1)$ for $i = 1, 2$. We conclude

$$\begin{aligned} \alpha N_{\alpha,[s,t]}(\bar{\omega}) &= \sum_{i=0}^{N_{\alpha,[s,t]}(\bar{\omega})-1} \bar{\omega}(\tau_i(\alpha), \tau_{i+1}(\alpha)) \leq \bar{\omega}_\alpha(s, t) \leq \omega_\alpha^1(s, t) + \omega_\alpha^2(s, t) \\ &\leq \alpha(2N_{\alpha,[s,t]}(\omega^1) + 2N_{\alpha,[s,t]}(\omega^2) + 2). \end{aligned}$$

□

Lemma 6. *Let ω^1 and ω^2 be two controls and assume that $\omega^2(s, t) \leq K$. Then*

$$N_{\alpha,[s,t]}(\omega^1 + \omega^2) \leq N_{\alpha-K,[s,t]}(\omega^1)$$

for every $\alpha > K$.

Proof. Set $\bar{\omega} := \omega^1 + \omega^2$ and

$$\begin{aligned} \bar{\tau}_0(\alpha) &= s \\ \bar{\tau}_{i+1}(\alpha) &= \inf \{u : \bar{\omega}(\bar{\tau}_i(\alpha), u) \geq \alpha, \bar{\tau}_i(\alpha) < u \leq t\} \wedge t. \end{aligned}$$

Similarly, we define $(\tau_i)_{i \in \mathbb{N}} = (\tau_i(\alpha - K))_{i \in \mathbb{N}}$ for ω^1 . It suffices to show that $\bar{\tau}_i \geq \tau_i$ for $i = 0, \dots, N_{\alpha,[s,t]}(\bar{\omega})$. We do this by induction. For $i = 0$, this is clear. If $\bar{\tau}_i \geq \tau_i$ for some $i \leq N_{\alpha,[s,t]}(\bar{\omega}) - 1$, superadditivity of control functions gives

$$\alpha = \bar{\omega}(\bar{\tau}_i, \bar{\tau}_{i+1}) \leq \omega^1(\tau_i, \bar{\tau}_{i+1}) + K$$

which implies $\tau_{i+1} \leq \bar{\tau}_{i+1}$. □

For the next Lemma, recall the definition of the homogenous p - ω distance and -norm given in the appendix and in [FV10b, Definition 8.2].

Lemma 7. *Let $s < t \in [0, T]$ and assume that $\|\mathbf{x}^i\|_{p-\omega;[s,t]} \leq 1$ for $i = 1, 2$. Then there is a constant $C = C(\gamma, p)$ such that*

$$\begin{aligned} \nu|y^1 - y^2|_{\infty;[s,t]} &\leq \left[\nu|y_s^1 - y_s^2| + |V^1 - V^2|_{Lip^{\gamma-1}} + \nu\rho_{p-\omega;[s,t]}(\mathbf{x}^1, \mathbf{x}^2) \right] \\ &\quad \times (N_{\alpha,[s,t]}(\omega) + 1) \exp \{C\nu^p \alpha(N_{\alpha,[s,t]}(\omega) + 1)\} \end{aligned}$$

for every $\alpha > 0$.

Proof. Set $\bar{y} = y^1 - y^2$ and

$$\kappa = \frac{|V^1 - V^2|_{\text{Lip}^{\gamma-1}}}{\nu} + \rho_{p-\omega;[s,t]}(\mathbf{x}^1, \mathbf{x}^2).$$

From [FV10b, Theorem 10.26] we can deduce that there is a constant $C = C(\gamma, p)$ such that

$$|\bar{y}_{u,v}| \leq C\nu\omega(u, v)^{1/p} [|\bar{y}_u| + \kappa] \exp\{C\nu^p\omega(u, v)\}$$

for every $u < v \in [s, t]$. From $|\bar{y}_{u,v}| \geq |\bar{y}_{s,v}| - |\bar{y}_{s,u}|$ we obtain

$$\begin{aligned} |\bar{y}_{s,v}| &\leq C\nu\omega(u, v)^{1/p} [|\bar{y}_u| + \kappa] \exp\{C\nu^p\omega(u, v)\} + |\bar{y}_{s,u}| \\ &\leq [|\bar{y}_s| + |\bar{y}_{s,u}| + \kappa] \exp\{C\nu^p\omega(u, v)\} \end{aligned}$$

for $s \leq u < v \leq t$. Now let $s = \tau_0 < \tau_1 < \dots < \tau_M < \tau_{M+1} = v \leq t$ for $M \geq 0$. By induction, one sees that

$$\begin{aligned} |\bar{y}_{s,v}| &\leq (M+1)(|\bar{y}_s| + \kappa) \exp\left\{C\nu^p \sum_{i=0}^M \omega(\tau_i, \tau_{i+1})\right\} \\ &\leq C^{M+1} [|\bar{y}_s| + \kappa] \exp\left\{C\nu^p \sum_{i=0}^M \omega(\tau_i, \tau_{i+1})\right\}. \end{aligned}$$

It follows that for every $v \in [s, t]$,

$$|\bar{y}_{s,v}| \leq [|\bar{y}_s| + \kappa] (N_{\alpha,[s,t]}(\omega) + 1) \exp\{C\nu^p\alpha(N_{\alpha,[s,t]}(\omega) + 1)\},$$

therefore

$$|\bar{y}_v| \leq [|\bar{y}_s| + \kappa] (N_{\alpha,[s,t]}(\omega) + 1) \exp\{C\nu^p\alpha(N_{\alpha,[s,t]}(\omega) + 1)\} + |\bar{y}_s|$$

and finally

$$|\bar{y}|_{\infty;[s,t]} \leq [|\bar{y}_s| + \kappa] (N_{\alpha,[s,t]}(\omega) + 1) \exp\{C\nu^p\alpha(N_{\alpha,[s,t]}(\omega) + 1)\}.$$

□

Proof of Theorem 4. Let ω be a control such that $\|\mathbf{x}^i\|_{p-\omega;[0,T]} \leq 1$ for $i = 1, 2$ (the precise choice of ω will be made later). From Lemma 7 we know that there is a constant $C = C(\gamma, p, \nu, \alpha)$ such that

$$\begin{aligned} |y^1 - y^2|_{\infty;[0,T]} &\leq \left[|y_0^1 - y_0^2| + |V^1 - V^2|_{\text{Lip}^{\gamma-1}} + \rho_{p-\omega;[s,t]}(\mathbf{x}^1, \mathbf{x}^2)\right] \\ &\quad \times \exp\{C(N_{\alpha,[s,t]}(\omega) + 1)\}. \end{aligned}$$

Now we set $\omega = \omega_{\mathbf{x}^1, \mathbf{x}^2}$ where

$$\omega_{\mathbf{x}^1, \mathbf{x}^2}(s, t) = \|\mathbf{x}^1\|_{p-\text{var};[s,t]}^p + \|\mathbf{x}^2\|_{p-\text{var};[s,t]}^p + \sum_{k=1}^{\lfloor p \rfloor} \frac{\left(\rho_{p-\text{var};[s,t]}^{(k)}(\mathbf{x}^1, \mathbf{x}^2)\right)^{p/k}}{\left(\rho_{p-\text{var};[0,T]}^{(k)}(\mathbf{x}^1, \mathbf{x}^2)\right)^{p/k}}$$

(the definition of $\rho_{p-\text{var}}^{(k)}(\cdot, \cdot)$ may be found in the appendix). It is easy to check that

$$\begin{aligned} \|\mathbf{x}^1\|_{p-\omega_{\mathbf{x}^1, \mathbf{x}^2};[0,T]} &\leq 1, \quad \|\mathbf{x}^2\|_{p-\omega_{\mathbf{x}^1, \mathbf{x}^2};[0,T]} \leq 1 \quad \text{and} \\ \rho_{p-\omega_{\mathbf{x}^1, \mathbf{x}^2};[0,T]}(\mathbf{x}^1, \mathbf{x}^2) &\leq \rho_{p-\text{var};[0,T]}(\mathbf{x}^1, \mathbf{x}^2). \end{aligned}$$

Finally, if $\alpha > \lfloor p \rfloor$ we can use Lemma 6 and Lemma 5 to see that

$$\begin{aligned} N_{\alpha, [0, T]}(\omega_{\mathbf{x}^1, \mathbf{x}^2}) + 1 &\leq N_{\alpha - \lfloor p \rfloor, [0, T]}(\omega_{\mathbf{x}^1} + \omega_{\mathbf{x}^2}) + 1 \\ &\leq 3 (N_{\alpha - \lfloor p \rfloor, [0, T]}(\mathbf{x}^1) + N_{\alpha - \lfloor p \rfloor, [0, T]}(\mathbf{x}^2) + 1). \end{aligned}$$

Substituting $\alpha \mapsto \alpha + \lfloor p \rfloor$ gives the claimed estimate. \square

2.2. Improved bounds for the higher order Euler approximations. We are now interested in proving a similar estimate for the distance between Euler-/Milstein approximations for rough paths and the actual solution (for the purpose of unified terminology, in the sequel we will only speak of Euler-schemes). Recall the notation from [FV10b]: If $V = (V_1, \dots, V_d)$ is a collection of sufficiently smooth vector fields on \mathbb{R}^e , $g \in T^N(\mathbb{R}^d)$ and $y \in \mathbb{R}^e$, we define an increment of the step- N Euler scheme by

$$\mathcal{E}_{(V)}(y, g) := \sum_{k=1}^N V_{i_1} \dots V_{i_k} I(y) g^{k, i_1, \dots, i_k}$$

where $g^{k, i_1, \dots, i_k} = \pi_k(g)^{i_1, \dots, i_k} \in \mathbb{R}$, I is the identity on \mathbb{R}^e and every V_j is identified with the first-order differential operator $V_j^k(y) \frac{\partial}{\partial y^k}$ (throughout, we use the Einstein summation convention). Furthermore, we set

$$\mathfrak{E}^g y := y + \mathcal{E}_{(V)}(y, g).$$

Given $D = \{0 = t_0 < \dots < t_n = T\}$ and a path $\mathbf{x} \in C_0^{p-\text{var}}([0, T]; G^{\lfloor p \rfloor}(\mathbb{R}^d))$ we define the (step- N) Euler approximation to the RDE solution y of

$$(6) \quad dy = V(y) d\mathbf{x}$$

with starting point $y_0 \in \mathbb{R}^e$ at time $t_k \in D$ by

$$y_{t_k}^{\text{Euler}; D} := \mathfrak{E}^{t_k \leftarrow t_0} y_0 := \mathfrak{E}^{S_N(\mathbf{x})_{t_{k-1}, t_k}} \circ \dots \circ \mathfrak{E}^{S_N(\mathbf{x})_{t_0, t_1}} y_0$$

where $S_N(\mathbf{x})$ denotes the Lyons lift of the rough path \mathbf{x} , see [FV10b, Section 9.1].

The following theorem is a version of [FV10b, Theorem 10.30] where, as in Theorem 4, the estimate is improved by replacing the quantity $\|\cdot\|_{p-\text{var}; [0, T]}^p$ by $N_{\alpha, [0, T]}(\cdot)$.

Theorem 8. *Let $\mathbf{x} \in C_0^{p-\text{var}}([0, T]; G^{\lfloor p \rfloor}(\mathbb{R}^d))$ and set $\omega(s, t) = \|\mathbf{x}\|_{p-\text{var}; [s, t]}^p$. Assume that $V \in \text{Lip}^\theta$ for some $\theta > p$ and let $\nu \geq |V|_{\text{Lip}^\theta}$. Choose $N \in \mathbb{N}$ such that $\lfloor p \rfloor \leq N \leq \theta$. Fix a dissection $D = \{0 = t_0 < \dots < t_n = T\}$ of $[0, T]$ and let $y_T^{\text{Euler}; D}$ denote the step- N Euler approximation of y . Then for every $\zeta \in \left[\frac{N}{p}, \frac{N+1}{p}\right)$ and $\alpha > 0$ there is a constant $C = C(p, \theta, \zeta, N, \nu, \alpha)$ such that*

$$\left| y_T - y_T^{\text{Euler}; D} \right| \leq C \exp \left\{ C (N_{\alpha, [0, T]}(\mathbf{x}) + 1) \right\} \sum_{k=1}^n \omega(t_{k-1}, t_k)^\zeta.$$

In particular, if \mathbf{x} is a Hölder rough path and $|t_{k+1} - t_k| \leq |D|$ for all k we obtain

$$(7) \quad \left| y_T - y_T^{\text{Euler}; D} \right| \leq CT \|\mathbf{x}\|_{1/p\text{-Hö}; [0, T]}^p \exp \left\{ C (N_{\alpha, [0, T]}(\mathbf{x}) + 1) \right\} |D|^{\zeta-1}$$

Proof. We basically repeat the proof of [FV10b, Theorem 10.30]. Recall the notation $\pi_{(V)}(s, y_s; \mathbf{x})$ for the (unique) solution of (6) with starting point y_s at time s . Set

$$z^k = \pi_{(V)}(t_k, \mathfrak{E}^{t_k \leftarrow t_0} y_0; \mathbf{x}).$$

Then $z_t^0 = y_t$, $z_{t_k}^k = \mathfrak{E}^{t_k \leftarrow t_0} y_0$ for every $k = 1, \dots, n$ and $z_T^n = y_T^{\text{Euler};D}$, hence

$$\left| y_T - y_T^{\text{Euler};D} \right| \leq \sum_{k=1}^n \left| z_T^k - z_T^{k-1} \right|.$$

One can easily see that

$$z_T^{k-1} = \pi_{(V)} \left(t_{k-1}, z_{t_{k-1}}^{k-1}; \mathbf{x} \right) = \pi_{(V)} \left(t_k, z_{t_k}^{k-1}; \mathbf{x} \right)$$

for all $k = 1, \dots, n$. Applying Theorem 4 (in particular the Lipschitzness in the starting point) we obtain for any $\alpha > 0$

$$\left| z_T^k - z_T^{k-1} \right| \leq c_1 \left| z_{t_k}^k - z_{t_k}^{k-1} \right| \exp \left\{ c_1 \left(N_{\alpha, [0, T]}(\mathbf{x}) + 1 \right) \right\}.$$

Moreover (cf. [FV10b, Theorem 10.30]),

$$\left| z_{t_k}^k - z_{t_k}^{k-1} \right| \leq \left| \pi_{(V)} \left(t_{k-1}, \cdot, \mathbf{x} \right)_{t_{k-1}, t_k} - \mathcal{E}_{(V)} \left(\cdot, S_N(\mathbf{x})_{t_{k-1}, t_k} \right) \right|_{\infty}.$$

Let $\delta \in [0, 1)$ such that $\zeta = \frac{N+\delta}{p}$. Since $(N + \delta) - 1 < N \leq \gamma$ we have $V \in \text{Lip}^{(N+\delta)-1}$. Thus we can apply [FV10b, Corollary 10.15] to see that

$$\begin{aligned} \left| \pi_{(V)} \left(t_{k-1}, \cdot, \mathbf{x} \right)_{t_{k-1}, t_k} - \mathcal{E}_{(V)} \left(\cdot, S_N(\mathbf{x})_{t_{k-1}, t_k} \right) \right|_{\infty} &\leq c_2 \left(|V|_{\text{Lip}^{(N+\delta)-1}} \|\mathbf{x}\|_{p\text{-var}; [t_{k-1}, t_k]} \right)^{N+\delta} \\ &\leq c_2 |V|_{\text{Lip}^{\zeta}}^{p\zeta} \omega(t_{k-1}, t_k)^{\zeta} \end{aligned}$$

which gives the claim. \square

3. PROBABILISTIC CONVERGENCE RESULTS FOR RDEs

Let $X : [0, T] \rightarrow \mathbb{R}$ be a real valued, centered, continuous Gaussian process with covariance

$$R_X(s, t) = EX_s X_t.$$

We recall the definition of mixed right (γ, ρ) -variation: For $\gamma, \rho \geq 1$ let

$$(8) \quad V_{\gamma, \rho}(R_X; [s, t] \times [u, v]) := \sup_{\substack{(t_i) \in \mathcal{D}([s, t]) \\ (t'_j) \in \mathcal{D}([u, v])}} \left(\sum_{t'_j} \left(\sum_{t_i} \left| R_X \left(\begin{array}{c} t_i, t_{i+1} \\ t'_j, t'_{j+1} \end{array} \right) \right|^{\gamma} \right)^{\frac{\rho}{\gamma}} \right)^{\frac{1}{\rho}},$$

where $\mathcal{D}([s, t])$ denotes the set of all dissections of $[s, t]$ and

$$R_X \left(\begin{array}{c} t_i, t_{i+1} \\ t'_j, t'_{j+1} \end{array} \right) = EX_{t_i, t_{i+1}} X_{t'_j, t'_{j+1}}.$$

We note that $V_{\rho} \equiv V_{\rho, \rho}$ regularity plays a key role in Gaussian rough path theory [FV10b, FV10a, FH14] and in particular yields a stochastic integration theory for large classes of multidimensional Gaussian processes. The importance of finite mixed $(1, \rho)$ -variation was understood in [FGGR15], where it is shown to allow for concentration of measure results (via Cass-Litterer-Lyons [CLL13]) which are pivotal for our result. In effect, one then has a substitute for good moment bounds in Itô theory which are no more available in our general Gaussian setting. (Recall that fBm, other than Brownian motion, is not a semimartingale.)

The following condition will be in place throughout the paper.

Condition 9. Let $X = (X^1, \dots, X^d): [0, T] \rightarrow \mathbb{R}^d$ be a centered, continuous Gaussian process with independent components. Assume that the covariance of every component has finite mixed $(1, \rho)$ -variation for some $\rho \in [1, 2)$ on $[0, T]^2$, that is, for $k = 1, \dots, d$,

$$\sup_{(t_i), (t'_j) \in \mathcal{D}([0, T])} \left(\sum_{t'_j} \left(\sum_{t_i} \left| EX_{t_i, t_{i+1}}^k X_{t'_j, t'_{j+1}}^k \right| \right)^\rho \right)^{\frac{1}{\rho}} < \infty.$$

We note that Condition 9 is satisfied by large number of Gaussian examples [FGGR15], in particular classical fBm, with Hurst parameter $H > 1/4$, and many variants thereof. (The intuition behind this condition is that ρ measures the roughness of the covariance close to the diagonal, whereas 1 deals with the off-diagonal part and somehow expresses a “sign” in the correlation structure (negative in case of fBm with $H < 1/2$), while still allowing for sufficiently regular perturbations.) There are two consequence of Condition 9 that will be most important to us. First, it guarantees the existence of a “canonical” Gaussian rough path associated to X , denoted by \mathbf{X} (or $\mathbf{X}(\omega)$ if we want to stress its random nature). Secondly, it provides regularity of the Cameron–Martin space \mathcal{H} associated to X ,

$$(9) \quad \iota: \mathcal{H} \hookrightarrow C^{q-\text{var}} \text{ with } q = \frac{1}{\frac{1}{2\rho} + \frac{1}{2}} < 2$$

which - cutting a long story ([CLL13] or [FH14, Ch.12]) short - leads to probabilistic estimates akin to those available within Itô theory.

We will sometimes need a slightly stronger version of Condition 9 which will ensure that our rough paths live in Hölder-type spaces:

Condition 10. Let $X = (X^1, \dots, X^d): [0, T] \rightarrow \mathbb{R}^d$ be a centered, continuous Gaussian process with independent components. Assume that the covariance of every component has Hölder dominated finite mixed $(1, \rho)$ -variation for some $\rho \in [1, 2)$ on $[0, T]^2$, that is, there exists $K < \infty$ such that, for $k = 1, \dots, d$ and uniformly over $s < t$ in $[0, T]$,

$$\sup_{(t_i), (t'_j) \in \mathcal{D}([s, t])} \left(\sum_{t'_j} \left(\sum_{t_i} \left| EX_{t_i, t_{i+1}}^k X_{t'_j, t'_{j+1}}^k \right| \right)^\rho \right)^{\frac{1}{\rho}} \leq K (|t - s|^{1/\rho}).$$

Note that fBm (with $\rho = \frac{1}{2H}$) and then any centered Gaussian process with stationary increments s.t. $\sigma^2(t - s) := E |X_t^i - X_s^i|^2$ is concave and $\sigma^2(\tau) = \mathcal{O}(\tau^{1/\rho})$ satisfies Condition 10, provided $\rho \in [1, 2)$. This, and more examples, are discussed in [FGGR15].

In the following subsection, we will establish L^p -convergence rates for step- N Euler approximations based on the entire Gaussian rough paths, i.e. schemes involving iterated (random) integrals up to order N . We continue by giving L^p -rates for the Wong-Zakai theorem in the Gaussian case. Putting together both results, we can give L^p convergence rates for an (easy-to-implement) simplified Euler scheme presented first in [DNT12]. We will see that the (sharp) almost sure convergence rates obtained in [FR14] also hold in L^p .

3.1. L^r -rates for step- N Euler approximation (based on entire rough path). For simplicity, the following Theorem is formulated only in the Hölder case.

Theorem 11. Assume the driving Gaussian noise X satisfies Condition 10. Choose $p > 2\rho$, assume that $V \in Lip^\theta$ for some $\theta > p$ and let $\nu \geq |V|_{Lip^\theta}$. Let D be a dissection of $[0, T]$ with mesh

size at most $h > 0$ and let $Y_T^{Euler;D}$ denote the step- N Euler approximation of Y , the (pathwise) solution of

$$dY = V(Y) d\mathbf{X} ; \quad Y_0 \in \mathbb{R}^e$$

where N is chosen such that $\lfloor p \rfloor \leq N \leq \theta$.

Then for every $r \geq 1$, $r' > r$ and $\zeta \in \left[\frac{N}{p}, \frac{N+1}{p} \right)$ there is a constant $C = C(\rho, p, \theta, \nu, K, r, r', N, \zeta)$ such that

$$\left| Y_T - Y_T^{Euler;D} \right|_{L^r} \leq CT \left\| \|\mathbf{X}\|_{1/p\text{-Hö};[0,T]}^{\zeta p} \right|_{L^{r'}} h^{\zeta-1}$$

holds for all $h > 0$.

Remark 12. By choosing $\hat{p} \in (2\rho, p)$ one has $\frac{N+1}{\hat{p}} < \frac{N+1}{p}$ and applying Theorem 11 with \hat{p} instead of p shows that

$$\left| Y_T - Y_T^{Euler;D} \right|_{L^r} \leq Ch^{\frac{N+1}{\hat{p}}-1}$$

holds for every $p > 2\rho$ if $h \rightarrow 0$.

Proof of Theorem 11. The embedding (9) provides the so-called *complementary Young regularity* (see e.g. [FH14, Sec. 11.1]) meaning that integrals of the form $\int hdX$, $\int Xdh$ are well-defined Young integrals. With this condition in place we can use (a variation of the theme of) [CLL13] to conclude sharp tail estimates for $N_{\alpha,[0,T]}(\mathbf{X})$. More precisely, choosing q as in (9), [FR13, Lemma 5 and Corollary 2] show that there is an $\alpha = \alpha(p, \rho, K) > 0$ and a positive constant $c_1 = c_1(p, \rho, K)$ such that

$$P(N_{\alpha,[0,T]}(\mathbf{X}) > u) \leq \exp \left\{ -c_1 \alpha^{2/p} u^{2/q} \right\}$$

holds for all $u > 0$. Now we use the pathwise estimate (7) and take the L^r norm on both sides of the inequality. The Hölder inequality shows that

$$\left| Y_T - Y_T^{Euler;D} \right|_{L^r} \leq c_1 T \left\| \|\mathbf{X}\|_{1/p\text{-Hö};[0,T]}^{\zeta p} \right|_{L^{r'}} \left| \exp \left\{ C(N_{\alpha,[0,T]}(\mathbf{X}) + 1) \right\} \right|_{L^{r''}} |D|^{\zeta-1}$$

holds for some (possibly large) $r'' > r$. Our tail estimate for $N_{\alpha,[0,T]}(\mathbf{X})$ shows that the $L^{r''}$ norm of the exponential term above is finite which yields the claim. \square

3.2. L^r -rates for Wong-Zakai approximations. We aim to formulate a version of the Wong-Zakai Theorem which contains convergence rates in L^r , any $r \geq 1$ for a class of suitable approximations X^h of X . By this, we mean that X^h is a centered, continuous Gaussian process with independent components for every $h \in (0, 1]$ and that

- (i) $(X^h, X) : [0, T] \rightarrow \mathbb{R}^{d+d}$ is jointly Gaussian, $(X^{h;i}, X^i)$ and $(X^{h;j}, X^j)$ are independent for $i \neq j$ and

$$(10) \quad \sup_{h \in (0,1]} V_{1,\rho}(R_{(X^h, X)}; [0, T]^2) =: K < \infty$$

for $\rho \in [1, 2)$ as in Condition 9.

(Note that this implies that Condition 9 also holds for every X^h , $h \in (0, 1]$.)

- (ii) Uniform convergence of the second moments:

$$\sup_{t \in [0, T]} E \left[|X_t^h - X_t|^2 \right] =: \delta(h)^{1/\rho} \rightarrow 0 \quad \text{for } h \rightarrow 0.$$

Theorem 13. *Assume the driving Gaussian noise X satisfies Condition 9 and let $(X^h)_{h>0}$ be a family of suitable approximations as above. Let \mathbf{X} and \mathbf{X}^h denote the lift of X resp. X^h to a process with p -rough sample paths for some $p > 2\rho$. Let $V = (V_1, \dots, V_d)$ be a collection of vector fields in \mathbb{R}^e . Choose $\eta < \frac{1}{\rho} - \frac{1}{2}$ and assume that $|V|_{Lip^\theta} \leq \nu < \infty$ for some $\theta > \frac{2\rho}{1-2\rho\eta}$. Let $Y, Y^h : [0, T] \rightarrow \mathbb{R}^e$ denote the pathwise solutions to the equations*

$$\begin{aligned} dY_t &= V(Y_t) d\mathbf{X}_t; & Y_0 &\in \mathbb{R}^e \\ dY_t^h &= V(Y_t^h) d\mathbf{X}_t^h; & Y_0^h &= Y_0 \in \mathbb{R}^e. \end{aligned}$$

Then, for any $r \geq 1$ there is a constant $C = C(\rho, p, \theta, \nu, K, \eta, r)$ such that

$$\left\| |Y^h - Y|_{\infty; [0, T]} \right\|_{L^r} \leq C\delta(h)^\eta$$

holds for all $h > 0$.

A typical example of such approximations are the *piecewise linear approximations*:

Corollary 14. *Assume the driving Gaussian noise X satisfies Condition 10 and that X^h is a piecewise linear approximation of X with mesh-size at most h . Then*

$$\left\| |Y^h - Y|_{\infty; [0, T]} \right\|_{L^r} \leq Ch^\eta$$

for any $\eta < \frac{1}{\rho} - \frac{1}{2}$

Proof of Corollary 14. We need to check that piecewise linear approximations are “suitable” in the sense of the beginning of this section. Secondly, we need to see that $E|X_t^h - X_t|^2 \leq Ch^{1/\rho}$ uniformly in $t \in [0, T]$. To simplify notation, we will assume $d = 1$. Let $D = \{0 < s_1 < \dots < s_M = T\}$ be a dissection of $[0, T]$ such that $|s_{k+1} - s_k| \leq h$, $k = 0, \dots, M-1$, and let X^h denote the piecewise linear approximation of X at the time points given by D . Concerning the first point, we have to show that $V_{1,\rho}(R_{(X^h, X)}; [0, T]^2)$ is uniformly bounded in h . Let D_1 and D_2 be two arbitrary dissections of $[0, T]$. Set $\bar{D}_1 := D_1 \cup D$. By the triangle inequality,

$$\begin{aligned} \sum_{t'_j \in D_2} \left(\sum_{t_i \in D_1} |EX_{t_i, t_{i+1}}^h X_{t'_j, t'_{j+1}}| \right)^\rho &\leq \sum_{t'_j \in D_2} \left(\sum_{\bar{t}_i \in \bar{D}_1} |EX_{\bar{t}_i, \bar{t}_{i+1}}^h X_{t'_j, t'_{j+1}}| \right)^\rho \\ &= \sum_{t'_j \in D_2} \left(\sum_{s_k \in D} |EX_{s_k, s_{k+1}} X_{t'_j, t'_{j+1}}| \right)^\rho \\ &\leq V_{1,\rho}(R_X; [0, T]^2)^\rho. \end{aligned}$$

Using the basic estimate $(a + b)^\rho \leq 2^{\rho-1}(a^\rho + b^\rho)$ instead of the triangle inequality, we similarly obtain

$$\sum_{t'_j \in D_2} \left(\sum_{t_i \in D_1} |EX_{t_i, t_{i+1}} X_{t'_j, t'_{j+1}}^h| \right)^\rho \leq 2^{\rho-1} V_{1,\rho}(R_X; [0, T]^2)^\rho$$

and

$$\sum_{t'_j \in D_2} \left(\sum_{t_i \in D_1} |EX_{t_i, t_{i+1}}^h X_{t'_j, t'_{j+1}}^h| \right)^\rho \leq 2^{\rho-1} V_{1,\rho}(R_X; [0, T]^2)^\rho.$$

Taking the supremum over all dissections, these estimates imply that $V_{1,\rho}(R_{(X^h,X)}; [0, T]^2)$ is bounded from above by $V_{1,\rho}(R_X; [0, T]^2)$ times a constant which only depends on ρ . Concerning the second point, note that for $t \in [s_k, s_{k+1}]$,

$$|X_t^h - X_t| \leq |X_{s_{k+1}} - X_{s_k}| + |X_t - X_{s_k}|$$

and therefore, using Condition 10,

$$\begin{aligned} E|X_t^h - X_t|^2 &\leq 2E|X_{s_{k+1}} - X_{s_k}|^2 + 2E|X_t - X_{s_k}|^2 \leq 4V_{1,\rho}(R_X; [s_k, s_{k+1}]^2) \leq 4K|s_{k+1} - s_k|^{1/\rho} \\ &\leq 4Kh^{1/\rho} \end{aligned}$$

which implies the second point. \square

Proof of Theorem 13. Set $X^0 := X$ and $\mathbf{X}^0 := \mathbf{X}$. Let \mathcal{H}^h denote the Cameron–Martin space associated to X^h , $h \geq 0$. Using the uniform bound (10), [FGGR15, Theorem 1] implies that

$$|\phi|_{q\text{-var}} \leq \sqrt{K} |\phi|_{\mathcal{H}^h}$$

holds for every $\phi \in \mathcal{H}^h$ and $h \geq 0$ with q as in (9). As in the proof of Theorem 11, we can find an $\alpha = \alpha(p, \rho, K) > 0$ and a positive constant $c_1 = c_1(p, \rho, K)$ such that the uniform tail estimate

$$P(N_{\alpha,[0,T]}(\mathbf{X}^h) > u) \leq \exp\left\{-c_1 \alpha^{2/p} u^{2/q}\right\} \quad \text{for all } u > 0$$

holds for all $h \geq 0$. Choose $\hat{p} \in \left(\frac{2\rho}{1-2\rho\eta}, \theta\right)$ and set $\hat{\mathbf{X}}^h = S_{[\hat{p}]}(\mathbf{X}^h)$ for $h \geq 0$. Lipschitzness of the map $S_{[\hat{p}]}$ and [FR13, Lemma 2] show that also

$$(11) \quad P(N_{\alpha,[0,T]}(\hat{\mathbf{X}}^h) > u) \leq \exp\left\{-c_1 \alpha^{2/p} u^{2/q}\right\} \quad \text{for all } u > 0$$

holds for all $h \geq 0$ for a possibly smaller $\alpha > 0$ (depending on \hat{p}). Now we use Theorem 4 and the Cauchy-Schwarz inequality to see that

$$\left\| |Y^h - Y|_{\infty,[0,T]} \right\|_{L^r} \leq c_2 \left| \rho_{\hat{p}\text{-var};[0,T]}(\hat{\mathbf{X}}^h, \hat{\mathbf{X}}) \right\|_{L^{2r}} \left| \exp\left\{c_2 \left(N_{\alpha,[0,T]}(\hat{\mathbf{X}}^h) + N_{\alpha,[0,T]}(\hat{\mathbf{X}}) + 1\right)\right\} \right\|_{L^{2r}}$$

holds for a constant $c_2 > 0$. The uniform tail estimates (11) show that

$$\sup_{h \geq 0} \left| \exp\left\{c_2 \left(N_{\alpha,[0,T]}(\hat{\mathbf{X}}^h) + N_{\alpha,[0,T]}(\hat{\mathbf{X}}) + 1\right)\right\} \right\|_{L^{2r}} \leq c_3 < \infty.$$

Applying [FR14, Theorem 5] with $\gamma = \frac{\rho}{1-2\rho\eta}$ shows that

$$\left| \rho_{\hat{p}\text{-var};[0,T]}(\hat{\mathbf{X}}^h, \hat{\mathbf{X}}) \right\|_{L^{2r}} \leq c_4 \sup_{t \in [0,T]} |X_t^h - X_t|_{L^2}^{1-\frac{\rho}{\gamma}} = c_4 \delta(h)^\eta$$

for a constant c_4 which yields the claim. \square

3.3. L^r -rates for the simplified Euler schemes. For $N \geq 2$, step- N Euler schemes contain iterated integrals whose distributions are not easy to simulate when dealing with Gaussian processes. In contrast, the simplified step- N Euler schemes avoid this difficulty by substituting the iterated integrals by a product of increments. In the context of fractional Brownian motion, it was introduced in [DNT12]. We make the following definition: If $V = (V_1, \dots, V_d)$ is sufficiently smooth, \mathbf{x} is a p -rough path, $y \in \mathbb{R}^e$ and $N \geq \lfloor p \rfloor$, we set

$$\mathcal{E}_{(V)}^{\text{simple}}(y, S_N(\mathbf{x})_{s,t}) := \sum_{k=1}^N \frac{1}{k!} V_{i_1} \dots V_{i_k} I(y) x_{s,t}^{i_1} \dots x_{s,t}^{i_k}$$

for $s < t$ and

$$\mathfrak{E}_{\text{simple}}^{S_N(\mathbf{x})_{s,t}} y := y + \mathcal{E}_{(V)}^{\text{simple}} \left(y, S_N(\mathbf{x})_{s,t} \right).$$

Given $D = \{0 = t_0 < \dots < t_n = T\}$ and a path $\mathbf{x} \in C_0^{p-\text{var}}([0, T]; G^{[p]}(\mathbb{R}^d))$ we define the *simplified* (step- N) Euler approximation to the RDE solution y of

$$dy = V(y) d\mathbf{x}$$

with starting point $y_0 \in \mathbb{R}^e$ at time $t_k \in D$ by

$$y_{t_k}^{\text{simple Euler}; D} := \mathfrak{E}_{\text{simple}}^{t_k \leftarrow t_0} y_0 := \mathfrak{E}_{\text{simple}}^{S_N(\mathbf{x})_{t_{k-1}, t_k}} \circ \dots \circ \mathfrak{E}_{\text{simple}}^{S_N(\mathbf{x})_{t_0, t_1}} y_0$$

and at time $t \in (t_k, t_{k+1})$ by

$$y_t^{\text{simple Euler}; D} := \left(\frac{t - t_k}{t_{k+1} - t_k} \right) \left(y_{t_{k+1}}^{\text{simple Euler}; D} - y_{t_k}^{\text{simple Euler}; D} \right) + y_{t_k}^{\text{simple Euler}; D}.$$

Theorem 15. *Assume the driving Gaussian noise X satisfies Condition 10. Choose $N \in \{2, 3\}$, $\eta_1 > 0$ and $\eta_2 > 0$ such that*

$$N > 2\rho - 1, \quad \eta_1 < \frac{1}{\rho} - \frac{1}{2} \quad \text{and} \quad \eta_2 < \frac{N+1}{2\rho} - 1.$$

Assume that $|V|_{Lip^\theta} \leq \nu < \infty$ for some $\theta \in (2, \infty]$ which satisfies $\theta > \frac{2\rho}{1-2\rho\eta_1}$ and $\theta \geq N$. Let D be a dissection of $[0, T]$ with mesh size at most $h > 0$.

Then for any $r \geq 1$ there is a constant $C = C(\rho, K, N, \eta_1, \eta_2, \theta, \nu, r)$ such that

$$\|Y - Y^{\text{simple Euler}; D}\|_{\infty} |_{L^r} \leq C(h^{\eta_1} + h^{\eta_2})$$

for all $h > 0$.

Remark 16. *In the proof, we will see that the rate η_1 is the rate for the Wong-Zakai approximation and η_2 comes from the rate of the step- N Euler approximation. In particular, for $\rho = 1$, we can choose $N = 2$ to obtain a rate arbitrary close to $\frac{1}{2}$ and the rate does not increase even if we choose $N = 3$. For $\rho > 1$, the choice $N = 3$ gives a rate of almost $\frac{1}{\rho} - \frac{1}{2}$.*

From this remark, we immediately obtain

Corollary 17. *Assume the driving Gaussian noise X satisfies Condition 10. Assume that the vector fields $V = (V_1, \dots, V_d)$ are bounded, C^∞ with bounded derivatives.*

(i) Case of $\rho < 3/2$. The simplified step-2 Euler scheme converges in L^r , for any $r \geq 1$, and rate $\frac{3}{2\rho} - 1 - \delta$, for any $\delta > 0$.

(ii) In the general case of $\rho < 2$, the simplified step-3 Euler scheme converges in L^r , for any $r \geq 1$, and rate $\frac{1}{\rho} - \frac{1}{2} - \delta$, for any $\delta > 0$,

Proof of Theorem 15. Let X^h denote the Gaussian process whose sample paths are piecewise linear approximated at the time points given by D and let $Y^h: [0, T] \rightarrow \mathbb{R}^e$ denote the pathwise solution to the equation

$$dY^h = V(Y^h) dX^h; \quad Y_0^h = Y_0 \in \mathbb{R}^e.$$

Then for any $t_k, t_{k+1} \in D$ we have

$$\mathbf{X}_{t_k, t_{k+1}}^{h; k; i_1, \dots, i_k} = \frac{1}{k!} X_{t_k, t_{k+1}}^{i_1} \cdots X_{t_k, t_{k+1}}^{i_k},$$

hence $Y_t^{\text{simple Euler};D} = Y_t^{h; \text{Euler};D}$ for any $t \in D$ and thus

$$\left| Y_t - Y_t^{\text{simple Euler};D} \right| \leq |Y - Y^h|_\infty + \max_{t_k \in D} \left| Y_{t_k}^h - Y_{t_k}^{h; \text{Euler};D} \right|$$

if $t \in D$. For $t \notin D$, choose $t_k \in D$ such that $t_k < t < t_{k+1}$. Set $a = \frac{t-t_k}{t_{k+1}-t_k}$ and $b = \frac{t_{k+1}-t}{t_{k+1}-t_k}$, i.e. $a + b = 1$. In the following, the relation $A \lesssim B$ means $A \leq \text{const.} B$ where the constant does not depend on h or t . By the triangle inequality,

$$\begin{aligned} \left| Y_t - Y_t^{\text{simple Euler};D} \right| &\leq a |Y_t - Y_{t_{k+1}}| + b |Y_t - Y_{t_k}| + a \left| Y_{t_{k+1}} - Y_{t_{k+1}}^{\text{simple Euler};D} \right| \\ &\quad + b \left| Y_{t_k} - Y_{t_k}^{\text{simple Euler};D} \right| \\ &\lesssim h^{1/p} \|Y\|_{1/p\text{-Hö};[0,T]} + \max_{t_k \in D} \left| Y_{t_k} - Y_{t_k}^{\text{simple Euler};D} \right| \\ &\lesssim h^{1/p} \left(\|\mathbf{X}\|_{1/p\text{-Hö};[0,T]} \vee \|\mathbf{X}\|_{1/p\text{-Hö};[0,T]}^p \right) + |Y - Y^h|_\infty \\ &\quad + \max_{t_k \in D} \left| Y_{t_k}^h - Y_{t_k}^{h; \text{Euler};D} \right| \end{aligned}$$

for $p > 2\rho$ sufficiently small, where we used [FV10b, Theorem 10.14] in the last inequality. Since the estimate holds uniformly over t , we can pass to the sup-norm on the left hand side of the inequality. We now take the L^r -norm and use the triangle inequality on the right hand side.

Since \mathbf{X} is the lift of a Gaussian process, $\|\mathbf{X}\|_{1/p\text{-Hö};[0,T]}$ has Gaussian tails ([FV10b, Theorem 15.33]). Therefore, all its moments are finite, and we can choose p such that $1/\eta_1 \geq p > 2\rho$ to obtain

$$\left| h^{1/p} \left(\|\mathbf{X}\|_{1/p\text{-Hö};[0,T]} \vee \|\mathbf{X}\|_{1/p\text{-Hö};[0,T]}^p \right) \right|_{L^r} \lesssim h^{\eta_1}.$$

Corollary 14 implies that

$$\left| |Y - Y^h|_\infty \right|_{L^r} \lesssim h^{\eta_1}$$

holds for all $h > 0$. Now we choose $p' > 2\rho$ such that $\frac{N+1}{p'} - 1 = \eta_2$ and apply Theorem 11 to estimate the last term. Since $|d_{1/p'\text{-Hö}}(\mathbf{X}^h, \mathbf{X})|_{L^r} \rightarrow 0$ for $h \rightarrow 0$, clearly $\sup_{h>0} \left\| \|\mathbf{X}^h\|_{1/p'\text{-Hö};[0,T]} \right\|_{L^r} < \infty$ and we obtain a uniform estimate of the form

$$\left| \max_{t_k \in D} \left| Y_{t_k}^h - Y_{t_k}^{h; \text{Euler};D} \right| \right|_{L^r} \lesssim h^{\eta_2}$$

which yields the claim. \square

4. MULTILEVEL SIMULATION OF RDES

In the spirit of Giles [Gil08b] we consider a multilevel Monte Carlo procedure in connection with the developed schemes for RDEs. In this context we reconsider and refine the complexity analysis by Giles [Gil08b] in certain respects. On the one hand we relax the requirement $\alpha \geq 1/2$ in Giles [Gil08b] concerning the bias rate, and on the other we keep track of various proportionality constants more carefully. Müller-Gronbach and Ritter [MGR09, Theorem 1] give a very general abstract multilevel Monte Carlo complexity result, which includes Theorem 18 (but not Theorem 20) as a special case as far as rates are concerned. However, we also feel that the balance between the various constants of proportionality involved can make a big difference for the performance of a multilevel algorithm in practice.

(Cf. the importance of various proportionality constants in the multilevel Andersen-Broadie algorithm for simulating dual prices of American options due to multilevel sub-simulation in [BSD13]. See also Collier et al. [CHAN⁺15] for an empirical approach to constructing optimal multilevel Monte Carlo algorithms.) Furthermore, we will also give a discussion about the optimal balance between bias and variance in the multilevel Monte Carlo algorithm in Section 4.2 below.

We adapt the main theorem of [Gil08b] to our needs. Below one should think

$$P = f(Y)$$

for a Lipschitz function f and Y the solution to the Gaussian RDE $dY = V(Y)dX$. Let \widehat{P}_l denote some (modified) Milstein approximation à la [DNT12], for instance (5), based on a meshsize $h_l = T/(M_0 M^l)$. Recall the basic idea

$$\begin{aligned} E[P] &\approx E[\widehat{P}_L] \text{ for } L \text{ large} \\ &= E[\widehat{P}_0] + \sum_{l=1}^L E[\widehat{P}_l - \widehat{P}_{l-1}] \end{aligned}$$

set $\widehat{P}_{-1} \equiv 0$ and define the (unbiased) estimator \widehat{Y}_l of $E[\widehat{P}_l - \widehat{P}_{l-1}]$, say

$$(12) \quad \widehat{Y}_l = \frac{1}{N_l} \sum_{i=1}^{N_l} (\widehat{P}_l^{(i)} - \widehat{P}_{l-1}^{(i)})$$

based on $i = 1, \dots, N_l$ independent samples. Note that $\widehat{P}_l^{(i)} - \widehat{P}_{l-1}^{(i)}$ comes from approximations with different mesh but the same realization of the driving noise.

Any implementation of our proposed algorithm relies on samples of the increments of the underlying Gaussian process X , say on a grid with size h_l^{-1} . In the following, we assume that we know the covariances of those increments in closed form. For concreteness, let Σ be the covariance matrix of the vector ΔX (of size h_l^{-1}) of increments of X . Clearly, we can always obtain samples from ΔX by the Cholesky factorization of Σ , at cost proportional to h_l^{-2} —disregarding the one-off cost of computing the Cholesky factorization itself. In this case, the cost of simulating one trajectory of the approximate solution at level l is, thus, proportional to h_l^{-2} .

On the other hand, when the increments are stationary, Σ can be embedded into a circulant matrix, and FFT-methods can be employed to sample ΔX at cost proportional to $h_l^{-1} \log(h_l^{-1})$. (This case includes the fractional Brownian motion.) We refer to Dieker [Die04] for a description of this and other related methods.

Finally, in case of a standard Brownian motion, it is of course possible to simulate ΔX at cost proportional to h_l^{-1} . For the multilevel analysis below, all three cases are going to be addressed.

4.1. Giles' complexity theorem revisited.

Theorem 18. *In the spirit of Giles, we assume that there are constants c_1, c'_2, c_2, c_3 and a rate γ such that*

- (i) $E[\widehat{P}_l - P] \leq c_1 h_l^\alpha,$
- (ii) $E[\widehat{Y}_0] = E[\widehat{P}_0]$ and $E[\widehat{Y}_l] = E[\widehat{P}_l - \widehat{P}_{l-1}], l > 0,$

(iii) $\text{var} [\widehat{Y}_0] \leq c'_2 N_0^{-1}$ and $\text{var} [\widehat{Y}_l] \leq c_2 N_l^{-1} h_l^\beta$ for $l \in \mathbb{N}$,³

(iv) $C_l \leq c_3 N_l h_l^{-\gamma}$, $l \geq 0$,

where C_l denotes the computational cost at level l . We further need to assume that $0 < \beta < \gamma$, $0 < \alpha$.

Then for every $\varepsilon > 0$, there are choices L and N_l , $0 \leq l \leq L$, to be given below in (16) and (17), respectively, and constants c_4 and c_5 given in (18) together with (19) such that the multilevel estimator $\widehat{Y} = \sum_{l=0}^L \widehat{Y}_l$ satisfies the mean square error bound

$$\text{MSE} \equiv E \left[\left(\widehat{Y} - E[P] \right)^2 \right] \leq \varepsilon^2,$$

with complexity bound

$$C \leq \begin{cases} c_4 \varepsilon^{-\frac{\gamma+2\alpha-\beta}{\alpha}} + o \left(\varepsilon^{-\frac{\gamma+2\alpha-\beta}{\alpha}} \right), & 2\alpha > \beta, \\ (c_4 + c_5) \varepsilon^{-\frac{\gamma+2\alpha-\beta}{\alpha}} + o \left(\varepsilon^{-\frac{\gamma+2\alpha-\beta}{\alpha}} \right), & 2\alpha = \beta, \\ c_5 \varepsilon^{-\gamma/\alpha} + o \left(\varepsilon^{-\gamma/\alpha} \right), & 2\alpha < \beta. \end{cases}$$

Proof. The basic structure of the proof is closely based on the corresponding proof of Giles [Gil08b]. Hence, we will not give all the details. Note that the parameters T and M_0 only enter into the picture in the form T/M_0 . Without loss of generality, we may therefore set $M_0 = 1$.

As typical, the first step consists in a standard Lagrangian optimization procedure (minimizing the complexity constraint by the MSE), where one ignores the requirement of L and N_l being integers. In the second step one then chooses integer valued parameters that are close to the optimal real-valued ones.

The mean-square-error satisfies

$$\text{MSE} = E \left[\left(\widehat{Y} - E[P] \right)^2 \right] = \text{var} [\widehat{Y}] + \left(E [\widehat{P}_L] - E[P] \right)^2 \leq c'_2 N_0^{-1} + c_2 T^\beta \sum_{l=1}^L N_l^{-1} M^{-l\beta} + c_1^2 h_L^{2\alpha}.$$

Now we need to minimize the total computational work

$$C \leq c_3 N_0 h_0^{-\gamma} + c_3 \sum_{l=1}^L N_l h_l^{-\gamma} = c_3 T^{-\gamma} \left[N_0 + \sum_{l=1}^L N_l M^{\gamma l} \right]$$

under the constraint $\text{MSE} \leq \varepsilon^2$. We first assume L to be given and minimize over N_0, \dots, N_L , and then we try to find an optimal L . We consider the Lagrange function

$$\begin{aligned} f(N_0, \dots, N_L, \lambda) \equiv & c_3 T^{-\gamma} \left[N_0 + \sum_{l=1}^L N_l M^{\gamma l} \right] + \\ & + \lambda \left(c'_2 N_0^{-1} + c_2 T^\beta \sum_{l=1}^L N_l^{-1} M^{-l\beta} + c_1^2 h_L^{2\alpha} - \varepsilon^2 \right). \end{aligned}$$

³We distinguish between c'_2 and c_2 , since the former controls the variance $\text{var} [\widehat{Y}_0]$, which is often already proportional to the variance of $f(Y)$, whereas the latter controls the variance of the *difference* \widehat{Y}_l , which is often much smaller in size.

Taking derivatives with respect to N_l , $0 \leq l \leq L$, we arrive at

$$\begin{aligned}\frac{\partial f}{\partial N_0} &= c_3 T^{-\gamma} - \lambda c_2' N_0^{-2} = 0, \\ \frac{\partial f}{\partial N_l} &= c_3 T^{-\gamma} M^{\gamma l} - \lambda c_2 T^{\beta} M^{-l\beta} N_l^{-2} = 0,\end{aligned}$$

implying that

$$(13a) \quad N_0 = \sqrt{\lambda} \sqrt{\frac{c_2'}{c_3}} T^{\gamma/2},$$

$$(13b) \quad N_l = \sqrt{\lambda} \sqrt{\frac{c_2}{c_3}} T^{(\gamma+\beta)/2} M^{-l(\gamma+\beta)/2}, \quad 1 \leq l \leq L,$$

which we insert into the bound for the MSE to obtain the Lagrange multiplier

$$(14) \quad \sqrt{\lambda} = \left[\sqrt{c_2' c_3} T^{-\gamma/2} + \sqrt{c_2 c_3} T^{-(\gamma-\beta)/2} \frac{M^{L(\gamma-\beta)/2} - 1}{M^{(\gamma-\beta)/2} - 1} \right] \cdot [\varepsilon^2 - c_1^2 T^{2\alpha} M^{-2\alpha L}]^{-1}.$$

By construction, we see that for any such choice of N_0, \dots, N_L , the MSE is, indeed, bounded by ε^2 . For fixed L , the total complexity is now given by

$$(15) \quad C(L) = \left[\sqrt{c_2' c_3} T^{-\gamma/2} + \sqrt{c_2 c_3} T^{-(\gamma-\beta)/2} M^{(\gamma-\beta)/2} \frac{M^{L(\gamma-\beta)/2} - 1}{M^{(\gamma-\beta)/2} - 1} \right]^2 \frac{1}{\varepsilon^2 - c_1^2 T^{2\alpha} M^{-2\alpha L}}.$$

In general, the optimal (but real-valued) choice of L would now be the arg-min of complexity estimate corresponding to the above choices of N_l , which we could not determine explicitly in an arbitrary regime.

We now turn to the second step, i.e., to integer-valued parameter choices. We parametrize the optimal choice of L by d_1 in

$$(16) \quad L = \left\lceil \frac{\log(d_1 c_1 T^\alpha \varepsilon^{-1})}{\alpha \log(M)} \right\rceil.$$

The proper choice of the parameter d_1 is discussed in detail in the subsequent subsection 4.2. Moreover, we choose with $\kappa = \frac{\gamma-\beta}{2\alpha}$

$$(17a) \quad N_0 = \left[\sqrt{c_2'} \frac{d_1^2}{d_1^2 - 1} \left(\sqrt{c_2'} + \sqrt{c_2} T^{\beta/2} M^{\alpha\kappa} \frac{d_1^\kappa c_1^\kappa T^{\alpha\kappa} \varepsilon^{-\kappa} - 1}{M^{\alpha\kappa} - 1} \right) \varepsilon^{-2} \right],$$

$$(17b) \quad N_l = \left[\sqrt{c_2'} \frac{d_1^2}{d_1^2 - 1} \left(\sqrt{c_2'} T^{\beta/2} + \sqrt{c_2} T^\beta \frac{d_1^\kappa c_1^\kappa T^{\alpha\kappa} \varepsilon^{-\kappa} - 1}{M^{\alpha\kappa} - 1} \varepsilon^{-2} M^{-l(\beta+\gamma)/2} \right) \right],$$

$1 \leq l \leq L$.

By construction, the MSE will be bounded by ε^2 using the choices (16) and (17). In the next step, we insert these definitions into the complexity bound. In order to obtain suitable simplifications, we use that $\lceil x \rceil \leq x + 1$ for real x .

After a tedious calculation, we finally arrive at the expression

$$(18) \quad C \leq c_4 \varepsilon^{-2(1+\kappa)} + c_5 \varepsilon^{-\gamma/\alpha} + c_6 \varepsilon^{-(2+\kappa)} + c_7 \varepsilon^{-2} + c_8,$$

where (once more including the dependence on M_0)

$$(19a) \quad c_4 = c_1^{2\kappa} c_2 c_3 \frac{d_1^{2+2\kappa}}{d_1^2 - 1} \frac{M^{3\alpha\kappa}}{(M^{\alpha\kappa} - 1)^2},$$

$$(19b) \quad c_5 = c_1^{\gamma/\alpha} c_3 d_1^{\gamma/\alpha} \frac{M^\gamma}{M^\gamma - 1},$$

$$(19c) \quad c_6 = c_1^\kappa c_3 \frac{d_1^{2+\kappa}}{d_1^2 - 1} \left[\sqrt{c_2 c_2'} T^{-\gamma/2} M_0^{\gamma/2} - 2c_2 T^{-\kappa\alpha} M_0^{\kappa\alpha} \frac{M^{\alpha\kappa}}{M^{\alpha\kappa} - 1} \right] \frac{M^{\alpha\kappa}(1 + M^{\alpha\kappa})}{M^{\alpha\kappa} - 1},$$

$$(19d) \quad c_7 = c_3 \frac{d_1^2}{d_1^2 - 1} \left[c_2' T^{-\gamma} M_0^\gamma - \sqrt{c_2'} (1 + \sqrt{c_2'}) T^{-(\gamma-\beta/2)} M_0^{\gamma-\beta/2} \frac{M^{\alpha\kappa}}{M^{\alpha\kappa} - 1} \right. \\ \left. + c_2 T^{-2\alpha\kappa} M_0^{2\alpha\kappa} \frac{M^{2\alpha\kappa}}{(M^{\alpha\kappa} - 1)^2} \right],$$

$$(19e) \quad c_8 = c_3 T^{-\gamma} M_0^\gamma.$$

If $2\alpha > \beta$, then $\varepsilon^{-2(1+\kappa)}$ is the leading order term, and we obtain the first alternative in the complexity bound of the theorem statement. If $2\alpha < \beta$, then $\varepsilon^{-\gamma/\kappa}$ is the leading order term—i.e., the total work is dominated by the work needed for simulating one trajectory at the finest level—, and we obtain the third alternative in the complexity bound. Finally, when $2\alpha = \beta$, then the two terms are of the same order. \square

Remark 19. Note that the complexity bound (18) needs to be considered with care when β is close to γ . Indeed, the coefficients c_4 and c_6 explode as $\beta \rightarrow \gamma$. However, please note that for ε fixed and $\beta \rightarrow \gamma$ we have that the corresponding powers $\varepsilon^{-2(1+\kappa)}$ and $\varepsilon^{-(2+\kappa)}$ converge to ε^{-2} as $\kappa \rightarrow 0$. Moreover, a closer look at (19) verifies that $c_4 \varepsilon^{-2(1+\kappa)} + c_6 \varepsilon^{-(2+\kappa)}$ remains bounded as $\beta \rightarrow \gamma$. Hence, there is no inconsistency in the results reported in Theorem 18 with the results of [Gil08b].

Next, we consider the situation when the actual complexity bound for computing one trajectory has logarithmic terms in h_l^{-1} , for instance in the case of fractional Brownian motion.

Theorem 20. In the setting of Theorem 18, we replace the complexity bound (iv) by the new condition

$$(iv') \quad C_l \leq c_3 N_l h_l^{-\gamma} \log(h_l^{-1}), \quad 0 \leq l.$$

Then the choices of L and N_l , $0 \leq l \leq L$, given in (16) and (17), respectively, lead to a mean squared error bound $\text{MSE} \leq \varepsilon^2$ with a complexity bound

$$C \leq \begin{cases} c_4' \varepsilon^{-\frac{\gamma+2\alpha-\beta}{\alpha}} \log(\varepsilon^{-1}) + o\left(\varepsilon^{-\frac{\gamma+2\alpha-\beta}{\alpha}} \log(\varepsilon^{-1})\right), & 2\alpha > \beta, \\ (c_4' + c_5') \varepsilon^{-\frac{\gamma+2\alpha-\beta}{\alpha}} \log(\varepsilon^{-1}) + o\left(\varepsilon^{-\frac{\gamma+2\alpha-\beta}{\alpha}} \log(\varepsilon^{-1})\right), & 2\alpha = \beta, \\ c_5' \varepsilon^{-\gamma/\alpha} \log(\varepsilon^{-1}) + o\left(\varepsilon^{-\gamma/\alpha} \log(\varepsilon^{-1})\right), & 2\alpha < \beta, \end{cases}$$

where $c_i' = c_i/\alpha$, $i = 4, 5$.

Proof. We choose L and N_0, \dots, N_L as in Theorem 18. As the mean squared error is not effected by the changed complexity bound, we obtain that the mean squared error is, once more, bounded by ε^2 . On the other hand, for the complexity bound, we note that

$$C \leq c_3 \sum_{l=0}^L N_l h_l^{-\gamma} \log(h_l^{-1}) \leq c_3 \log(h_L^{-1}) \sum_{l=0}^L N_l h_l^{-\gamma}.$$

The last sum gives the upper bound given in (18). On the other hand, by (16) (using once more $\lceil x \rceil \leq x + 1$ for $x \in \mathbb{R}$), we have

$$\log(h_L^{-1}) \leq \log\left(M_0 M d_1^{1/\alpha} c_1^{1/\alpha}\right) + \frac{\log(\varepsilon^{-1})}{\alpha}. \quad \square$$

	MLMC	classical MC	speed up of MLMC
Generic	$\varepsilon^{-(\gamma+2\alpha-\beta)/\alpha}$	$\varepsilon^{-(2+\gamma/\alpha)}$	$\varepsilon^{-\beta/\alpha}$
$\alpha = \beta/2$	$\varepsilon^{-\gamma/\alpha}$	$\varepsilon^{-(2+\gamma/\alpha)}$	ε^{-2}
$\alpha = \beta$	$\varepsilon^{-(\gamma/\alpha+1)}$	$\varepsilon^{-(2+\gamma/\alpha)}$	ε^{-1}

TABLE 1. Comparison of asymptotic complexity for multilevel and standard Monte Carlo in the framework of Theorem 18—i.e., ignoring log-terms. α denotes the weak order of convergence, $\beta/2$ the strong order. We distinguish the cases $\alpha = \beta/2$ and $\alpha = \beta$. We choose the work rate to be $\gamma = 1$, which is the most relevant case (up to logarithmic terms).

Under the assumptions of Theorem 18, we can summarize the complexity requirements for the multi-level and the classical Monte Carlo methods, respectively, to obtain an MSE of order ε^2 , see Table 1. In particular, note that the complexity of classical Monte Carlo is asymptotically worse by a factor ε^{-2} in the “non-regular” case, when the weak rate is equal to the strong rate, but still worse by a factor ε^{-1} when the weak rate is actually twice as good as the strong rate. With Theorem 20 we get the same speed ups, as the logarithmic terms appear for both multilevel and single level Monte Carlo.

4.2. Balancing bias and variance in the multilevel algorithm. In the now classical works of Giles on multilevel Monte Carlo, the choice $d_1 = \sqrt{2}$ is advocated, see for instance [Gil08b]. This means that we reserve the same error tolerance $\varepsilon/2$ both for the bias or discretization error and for the statistical or Monte Carlo error. Indeed, the choice of d_1 corresponds to the distribution of the total MSE ε^2 between the statistical and the discretization error according to

$$\varepsilon^2 = \underbrace{\frac{\varepsilon^2}{d_1^2}}_{\text{disc. error}} + \underbrace{\left(1 - \frac{1}{d_1^2}\right)}_{\text{stat. error}} \varepsilon^2.$$

In many situations, the choice $d_1 = \sqrt{2}$ is not optimal. In fact, even in an ordinary Monte Carlo framework, one should not blindly follow this rule. For instance, for an SDE driven by a Brownian motion, the Euler scheme usually (i.e., under suitable regularity conditions) exhibits weak convergence with rate 1. Assuming the same constants for the weak error and the statistical error, a straightforward optimization will show that it is optimal to choose the number of timesteps and the number of Monte Carlo samples such that the (squared) discretization error is $\varepsilon^2/3$ and the (squared) statistical error is $2\varepsilon^2/3$.

Let us now study in detail the asymptotic behavior for $\varepsilon \downarrow 0$ of the optimal choice for d_1 for a given fixed M in the context of Theorem 20. Once again assuming $M_0 = 1$ by absorbing it into T , we obtain the complexity

$$D(L) = \left[\sqrt{c_2' c_3} T^{-\gamma/2} + \sqrt{c_2 c_3} T^{(\beta-\gamma)/2} M^{(\gamma-\beta)/2} \frac{M^{L(\gamma-\beta)/2} - 1}{M^{(\gamma-\beta)/2} - 1} \right]^2 \frac{L \log M - \log T}{\varepsilon^2 - c_1^2 T^{2\alpha} M^{-2\alpha L}},$$

i.e., (15) with an extra factor stemming from $\log h_L^{-1} = L \log M - \log T$. (We note that we disregard integer constraints on L and N_l for this asymptotic analysis.) We assume in advance that L is such that $L \log M - \log T > 0$. Setting $\frac{d}{dL} \log(D(L)) = 0$ yields

$$\underbrace{\frac{2\sqrt{c_2}T^{\beta/2}M^{(\gamma-\beta)/2}}{\sqrt{c_2} + \sqrt{c_2}T^{\beta/2}M^{(\gamma-\beta)/2} \frac{M^{L(\gamma-\beta)/2-1}}{M^{(\gamma-\beta)/2-1}}} \frac{M^{L(\gamma-\beta)/2}}{M^{(\gamma-\beta)/2} - 1} \frac{\gamma - \beta}{2} + \frac{1}{L \log M - \log T}}_{\equiv h_M(L, \alpha, \beta, \gamma)} = \frac{2\alpha c_1^2 T^{2\alpha}}{\varepsilon^2 M^{2\alpha L} - c_1^2 T^{2\alpha}}.$$

Noting that $h_M > 0$,

$$\lim_{L \rightarrow \infty} h_M(L, \alpha, \beta, \gamma) = \gamma - \beta > 0, \quad h_M(0, \alpha, \beta, \gamma) = \frac{2\sqrt{c_2'c_3}T^{-(\gamma-\beta)/2}M^{(\gamma-\beta)/2}}{\sqrt{c_2'c_3}T^{-\gamma/2} \frac{2}{\gamma-\beta}(M^{(\gamma-\beta)/2} - 1)},$$

we may rewrite the above equality as

$$(20) \quad \frac{2\alpha c_1^2 T^{2\alpha}}{h_M(L, \alpha, \beta, \gamma) + \frac{1}{L \log M - \log T}} + c_1^2 T^{2\alpha} = \varepsilon^2 M^{2\alpha L}.$$

Clearly, the l.h.s. of (20) is bounded from below by $c_1^2 T^{2\alpha}$ if $L > \log(T/M)$. That is, if $\varepsilon \downarrow 0$ we must have $M^{2\alpha L} \rightarrow \infty$, hence $L \rightarrow \infty$. Then by taking logarithms of (20) and solving for L we get for $\varepsilon \downarrow 0$,

$$L = \frac{\log \varepsilon^{-1}}{\alpha \log M} + \frac{1}{2\alpha \log M} \log \left(\frac{2\alpha c_1^2 T^{2\alpha}}{h_M(L, \alpha, \beta, \gamma) + \frac{1}{L \log M - \log T}} + c_1^2 T^{2\alpha} \right).$$

Then, for $\varepsilon \downarrow 0$, i.e. $L \rightarrow \infty$, it follows that

$$(21) \quad L^* = \frac{1}{\alpha \log M} \log \left(\varepsilon^{-1} \left(\frac{2\alpha c_1^2 T^{2\alpha}}{\gamma - \beta} + c_1^2 T^{2\alpha} \right)^{1/2} \right) + O\left(\frac{1}{\log \varepsilon^{-1}}\right).$$

Now, with d_1 implicitly defined by $L^* = \frac{\log(d_1 c_1 T^\alpha \varepsilon^{-1})}{\alpha \log M}$, cf. (16), we obtain

$$(22) \quad d_1 = \left(\frac{2\alpha}{\gamma - \beta} + 1 \right)^{1/2} \left(1 + O\left(\frac{1}{\log \varepsilon^{-1}}\right) \right).$$

Remark 21. Notice that the dominant term in the asymptotically optimal choice for d_1 only depends on the rates α, β, γ , but not on the constants.

Plugging the optimal choice $L = L^*$ into the (approximate) computational cost D , we obtain

$$D(L^*) = \frac{2\alpha + \gamma - \beta}{2\alpha^2} \frac{4c_2c_3T^{-\gamma}M^{(\gamma-\beta)}c_1^{\frac{\gamma-\beta}{\alpha}}}{2(M^{(\gamma-\beta)/2} - 1)^2} f\left(\frac{\gamma - \beta}{2\alpha}\right) \varepsilon^{-\frac{2\alpha + \gamma - \beta}{\alpha}} \log \varepsilon^{-1} \left(1 + O\left(\frac{1}{\log \varepsilon^{-1}}\right) \right)$$

with $f(x) := (1 + \frac{1}{x})^x$ being increasing with $f(0+) = 1$ and $f(x) \rightarrow e$ when $x \rightarrow \infty$. Comparing with Giles' choice $d_1 = \sqrt{2}$, i.e.,

$$L^{\text{Giles}} = \frac{\log(\sqrt{2}c_1T^\alpha\varepsilon^{-1})}{\alpha \log M}$$

we get

$$\frac{D(L^{\text{Giles}})}{D(L^*)} = \frac{2^{1 + \frac{\gamma - \beta}{2\alpha}}}{\left(1 + \frac{\gamma - \beta}{2\alpha}\right) f\left(\frac{\gamma - \beta}{2\alpha}\right)} \left(1 + O\left(\frac{1}{\log \varepsilon^{-1}}\right) \right).$$

Note that the above fraction—plotted in Figure 1—(asymptotically) takes its minimum value of 1 when $\frac{\gamma-\beta}{2\alpha} = 1$. In this case, our proposed asymptotically optimal choice of d_1 takes the value $\sqrt{2}$ (up to higher order terms) and, thus, coincides with Giles' choice.

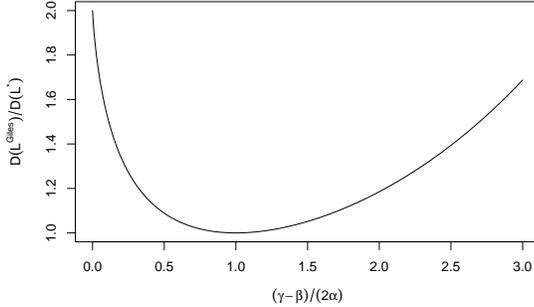


FIGURE 1. Dominant term of $\frac{D(L^{\text{Giles}})}{D(L^*)}$ as a function of $\frac{\gamma-\beta}{2\alpha}$

Remark 22. *The special case $\gamma = \beta$ requires new calculations (due to exploding terms). In this case, one can see that*

$$L^* = \frac{\log \varepsilon^{-2}}{2\alpha \log M} + \frac{\log \log \varepsilon^{-2}}{2\alpha \log M} - \frac{\log \log M^{2\alpha}}{2\alpha \log M} + \frac{1}{2} c_1^2 T^{2\alpha} + O\left(\frac{\log \log \varepsilon^{-2}}{\log \varepsilon^{-2}}\right)$$

corresponding to

$$d_1 = \frac{M c_1^2 \alpha T^{2\alpha/2}}{c_1 T^\alpha \sqrt{\alpha M}} \sqrt{\log \varepsilon^{-1}} \left(1 + O\left(\frac{\log \log \varepsilon^{-2}}{\log \varepsilon^{-2}}\right)\right).$$

We see that $d_1 \rightarrow \infty$ as $\varepsilon \downarrow 0$. Nonetheless, the relative computational costs compared to $d_1 = \sqrt{2}$ stays bounded as

$$\frac{D(L^{\text{Giles}})}{D(L^*)} = 2 \left(1 + O\left(\frac{1}{\log \varepsilon^{-1}}\right)\right).$$

Let us note an important difference between the multilevel algorithm for the irregular case $\beta < 1$ explored here and the classical multilevel algorithm of Giles [Gil08b] regarding the distribution of the work-load. For the case of a classical SDE, the work is going to be essentially equi-distributed among the levels. For the rough SDE case considered here, we see from the proof of Theorem 18 that most of the computational budget is actually spent on the fine grids, i.e., on the levels with high index l . This is schematically represented in Figure 2, where we used the theoretical complexity estimates from the proof of Theorem 18 with $\beta = 0.6$, $\alpha = 0.3$ and the remaining constants set to arbitrary values.

4.3. Multilevel Monte Carlo for RDEs. We shall now combine the results of Section 3 and Section 4. For convenience, we recall the regularity assumptions of the convergence analysis (cf. Condition 10): Let $X = (X^1, \dots, X^d): [0, T] \rightarrow \mathbb{R}^d$ be a centered, continuous Gaussian process with

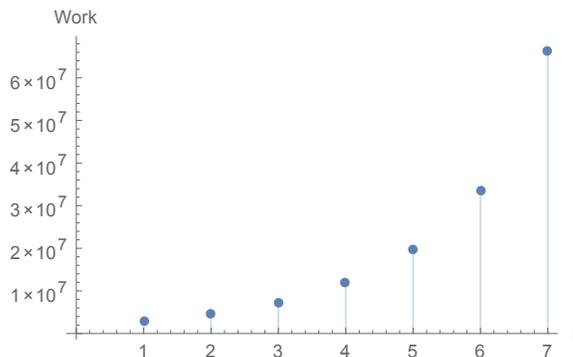


FIGURE 2. Work distribution among levels. Schematic presentation based on a case of $\beta = 0.6$.

independent components. Assume that the covariance of every component has finite mixed $(1, \rho)$ -variation for some $\rho \in [1, 2)$ on $[0, T]^2$, that is, for $k = 1, \dots, d$,

$$\sup_{(t_i), (t'_j) \in \mathcal{D}([0, T])} \left(\sum_{t'_j} \left(\sum_{t_i} \left| E \left[X_{t_i, t_{i+1}}^k X_{t'_j, t'_{j+1}}^k \right] \right| \right) \right)^\rho < \infty.$$

Consider the solution $Y: [0, T] \rightarrow \mathbb{R}^m$ of the RDE

$$dY_t = V(Y_t) d\mathbf{X}_t; \quad Y_0 \in \mathbb{R}^m$$

where $V = (V_1, \dots, V_d)$ is a collection of vector fields in \mathbb{R}^m with $|V|_{\text{Lip}^\eta} < \infty$ for some $\eta \geq \frac{2\rho}{\rho-1}$. Set $S := Y$ and let $S^{(h_l)}$ be the simplified step-3 Euler approximation of Y with mesh-size h_l (in the case $\rho = 1$, it suffices to consider a step-2 approximation). Let $f: C([0, T], \mathbb{R}^m) \rightarrow \mathbb{R}^n$ be a Lipschitz continuous functional and set $P := f(S)$, $\hat{P}_l := f(S^{(h_l)})$.

Theorem 23. *Consider a functional of an RDE driven by Gaussian signal satisfying the above assumptions, which is evaluated to within a MSE of ε^2 .*

(a) *If we assume that the cost of sampling a vector of increments of X of length N is proportional to $N \log(N)$, then an upper bound for the complexity is given by*

$$\mathcal{O}(\varepsilon^{-\theta}), \quad \forall \theta > \frac{2\rho}{2-\rho}.$$

(b) *On the other hand, if we assume that the cost of obtaining such a sample is proportional to N^2 , then an upper bound of the total complexity is*

$$\mathcal{O}(\varepsilon^{-\theta}), \quad \forall \theta > \frac{4\rho}{2-\rho}.$$

Proof. We want to calculate the quantities needed in Theorem 18. As f is assumed to be Lipschitz, the weak rate of convergence is (at least) the strong rate of convergence, i.e., $\alpha = \beta/2$ in the notation of Theorem 18. By Corollary 17, the strong rate is $\beta/2 = \frac{2}{\rho} - \frac{1}{2} - \delta$ for any $\delta > 0$. Observe that

$$\text{var} [\hat{P}_l - P] \leq E \left[\left(\hat{P}_l - P \right)^2 \right] \leq |f|_{\text{Lip}}^2 E \left[\left| S^{(h_l)} - S \right|^2 \right] = \mathcal{O} \left(h_l^\beta \right)$$

and

$$\text{var} \left[\widehat{P}_l - \widehat{P}_{l-1} \right] \leq \left(\text{var} \left[\widehat{P}_l - P \right]^{1/2} + \text{var} \left[\widehat{P}_{l-1} - P \right]^{1/2} \right)^2 = O \left(h_l^\beta \right)$$

for all $\beta < \frac{2}{\rho} - 1$. Of course the variance of the average of N_l IID samples becomes

$$\text{var} \left[\widehat{Y}_l \right] = \frac{1}{N_l} \text{var} \left[\widehat{P}_l - \widehat{P}_{l-1} \right] = O \left(h_l^\beta / N_l \right).$$

This shows condition (iii) in Theorem 18. Trivially, a strong rate is also a weak rate, in the sense that

$$E \left(\widehat{P}_l - P \right) \leq E \left[\left(\widehat{P}_l - P \right)^2 \right]^{1/2} = O \left(h_l^{\beta/2} \right),$$

which gives condition (i) Condition (ii), “unbiasedness” is obvious for the estimator (12).

In case (a), condition (iv’) of Theorem 20 holds, and the theorem implies that an MSE ε^2 can be achieved at cost proportional to

$$\varepsilon^{-\frac{1}{\alpha}} \log(\varepsilon^{-1}) = \varepsilon^{-\left(\frac{1}{\rho} - \frac{1}{2} + \delta\right)} \log(\varepsilon^{-1}) = \varepsilon^{-\left(\frac{2\rho}{2-\rho} + \delta\right)} \log(\varepsilon^{-1})$$

for any $\delta > 0$. By choosing δ slightly larger, we may get rid of the logarithmic term. In the end, we get $\mathcal{O}(\varepsilon^{-\theta})$ for any $\theta > \frac{2\rho}{2-\rho}$.

On the other hand, in the general case (b), we rely on Theorem 18 with $\gamma = 2$. By similar calculations as above (replacing $1/\alpha$ by $2/\alpha$), we arrive at the result for this case. \square

Remark 24. *If X is a fractional Brownian motion with Hurst parameter $H > 1/4$ (with $\rho = 1/(2H) < 2$), then we can generate samples of increments at cost proportional to $N \log(N)$ (e.g., by circulant embedding methods, see [Die04]), and we are, hence, in the situation of Theorem 23 (a).*

In Table 2 we compare typical asymptotic complexities for RDEs driven by fractional Brownian motion for both the multi-level and the classical Monte Carlo estimators. We distinguish between the “non-regular” regime when $\alpha = \beta/2$ and the more favorable regime when $\alpha = \beta$. Moreover, we have simplified the presentation in Table 2 by neglecting the higher order and logarithmic terms. I.e., any complexity ε^{-a} in Table 2 should actually be understood as $\varepsilon^{-a-\delta}$ for any $\delta > 0$. Thus, when the Hurst parameter is not too small, multi-level can make the difference between a feasible simulation and a quite impossible one. E.g., when $H = 2/5$ and the payoff function f is so irregular that the weak rate of convergence is not better than the strong rate of convergence, the complexity for a standard Monte Carlo estimator would be roughly of order $\varepsilon^{-5.33}$, whereas the multi-level version would have complexity roughly of order $\varepsilon^{-3.33}$, which is not much worse than the complexity of a standard Monte Carlo estimator of the usual Brownian motion regime. Admittedly, when $H = 1/3$ and one has an irregular payoff, then both standard and multi-level Monte Carlo are very costly computation wise.

5. NUMERICAL EXPERIMENTS

5.1. A linear, non-smoothing example. We consider a linear RDE in \mathbb{R}^3 driven by a two-dimensional fractional Brownian motion with Hurst index H . In fact, we consider vector fields $V_i(y) = A_i y$, $y \in \mathbb{R}^3$, $i = 1, 2$, with

$$A_1 = \begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1/2 \\ -2 & -1/2 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0.7 & 0.9 \\ -0.7 & 0 & 1 \\ -0.9 & -1 & 0 \end{pmatrix}.$$

	$H = 2/5$		$H = 1/3$	
	MLMC	classical MC	MLMC	classical MC
$\alpha = \beta/2$	$\varepsilon^{-10/3} \approx \varepsilon^{-3.33}$	$\varepsilon^{-16/3} \approx \varepsilon^{-5.33}$	ε^{-6}	ε^{-8}
$\alpha = \beta$	$\varepsilon^{-8/3} \approx \varepsilon^{-2.67}$	$\varepsilon^{-11/3} \approx \varepsilon^{-3.67}$	ε^{-4}	ε^{-5}

TABLE 2. Comparison of asymptotic complexities for multi-level and classical Monte Carlo for RDEs driven by fractional Brownian motion with Hurst index $H = 2/5$ and $H = 1/3$. We distinguish between the cases $\alpha = \beta/2$ and $\alpha = \beta$. For this summary, we neglect the “higher order terms”, i.e., we neglect the δ in $\beta/2 = 1/\rho - 1/2 - \delta$. For $H = 2/5$, we set $\beta/2 = 3/10$, and for $H = 1/3$ we set $\beta/2 = 1/6$.

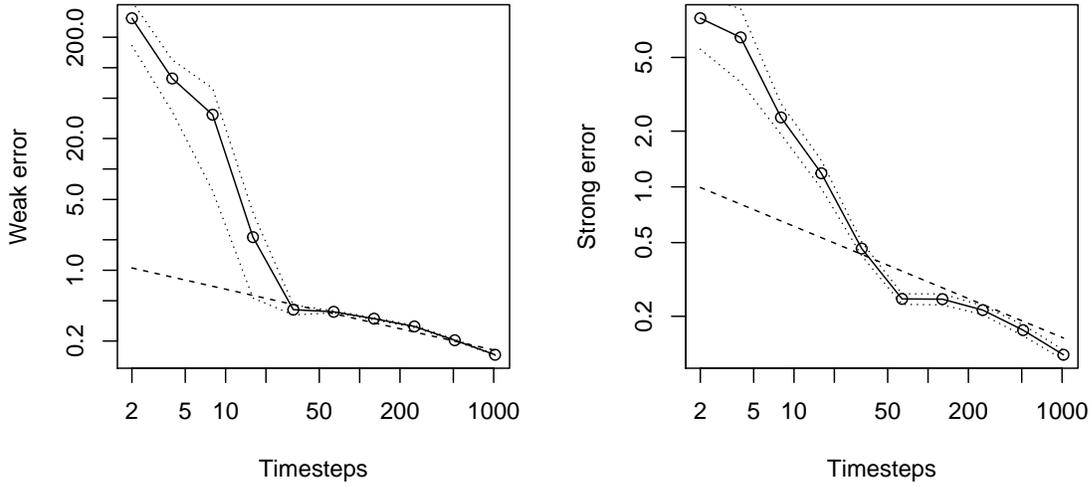


FIGURE 3. Strong and weak error for a fBm with Hurst index $H = 0.4$. Dashed line corresponds to the theoretical strong rate of convergence 0.3, dotted lines show confidence intervals around the error due to the integration error. Weak error corresponds to the functional $f(y) := (|y(1)| - 1)^+$.

Note that the matrices A_1 and A_2 are anti-symmetric, implying that the sphere S^2 is invariant under the solution of the SDE. Note that the RDE driven by these vector fields is rather challenging from a numerical point of view, if we try to solve them using general, non-geometric schemes. In particular, in the case $H = 1/2$, the equation is non-hypoelliptic and provides a smooth example in which the standard Euler scheme only has weak convergence rate $1/2$. In the case $H \neq 1/2$, we cannot expect any smoothing properties of the solution, either. Formally, further note that the vector fields are unbounded, violating one of our theoretical assumptions.

We implement the simplified Euler scheme (5), where the increments of the fractional Brownian motion were simulated by Hosking’s method, see [Die04].⁴ Hosking’s method is an exact simulation method, i.e., if fed with truly Gaussian random numbers, it will produce samples from the true

⁴The underlying Gaussian random numbers are simulated using the Box-Müller method. The pseudo random numbers are generated by the Mersenne-Twister [MN98].

distribution of increments of the fractional Brownian motion. It is similar to the more obvious simulation method based on the Cholesky factorization of the covariance matrix of the increments, but preferable in terms of memory requirement, especially when grids of sizes of up to $2^{14} = 16384$ are considered. As Cholesky's method, the complexity of simulating the increments of the fractional Brownian motion on a grid with size M is essentially proportional to M^2 , so we are working in the context of Theorem 23 (b).

Starting at $Y_0 = (1, 0, 0)$, Figure 3 shows the strong and weak convergence of the scheme for $H = 0.4$. More precisely, let \bar{Y}_1^N denote the result of the scheme based on a uniform grid on $[0, 1]$ based on N time-step. Then consider \bar{Y}_1^{2N} based on the increments of the *same* fBm.⁵ Then, the lower part of Figure 3 shows the Monte Carlo estimator of $E \left[\left| \bar{Y}_1^N - \bar{Y}_1^{2N} \right| \right]$ plotted against N . We, indeed, observe the expected rate of strong convergence, which, due to Theorem 15 is $2H - 1/2 = 0.3$, but only after a prolonged pre-asymptotic phase.

In the upper panel of Figure 3, we plot the *weak* error for the calculation of $E[f(Y_1)]$ for the functional

$$f(y) := (|y| - 1)^+.$$

This implies that $E[f(Y_1)] = 0$, so that we do not need to carry out lengthy calculations in order to find an appropriately accurate reference value. The figure indicates that the rate of the weak error is again equal to the strong rate 0.3. Note that the same would be true even in the case $H = 1/2$, because the Markov semigroup associated to the solution (in the case $H = 1/2$) is not smoothing and, in addition, the functional f is non-smooth on S^2 , i.e., with probability 1. Again, the roughness of the driving signal leads to a remarkably strong pre-asymptotic regime. Indeed, when the grid is too coarse, then the weak approximation error can be huge. Visually, it seems that the asymptotic error analysis accurately describes the true error when the mesh of the grid is at least around 0.02 for the case $H = 0.4$.

Figure 4 shows strong and weak errors for the same differential equation and the same function f , but in the even rougher case $H = 0.33$. In this case, the size of the errors for very coarse grids are even larger than for $H = 0.4$, and, moreover, the pre-asymptotic phase seems even longer: here the mesh of the grid should probably be at least 0.01 in order to describe the true computational error by the asymptotic error bounds.

5.2. Multilevel Monte Carlo for the linear example. This long pre-asymptotic phase, in which the computational error is very large, needs to be taken into account when constructing a successful multi-level estimator: indeed, it is advisable to choose the coarsest grid used in the multi-level iteration already within the asymptotic regime. Thus, in the case $H = 0.4$, we would recommend to choose $h_0 \leq 0.02$ for this particular example. This is remarkably different from the standard SDE case, where often h_0 is chosen to be equal to T , i.e., the coarsest grid contains only the start and end points of the interval $[0, T]$. However, when employing this strategy for the fBm example here, the constants in the error bound for the multi-level estimator will completely overshadow the asymptotic convergence rate, to the extent that even for long computation time no “empirical” convergence is exhibited. Indeed, the coarsest levels then combine a large error with an even larger variance, and this combination, while harmless in the asymptotic limit $\varepsilon \rightarrow 0$, renders the standard multi-level construction useless.

⁵In practice, this means that we generated the increments of the fBm X on the finer grid $\frac{k}{2N}$, $k = 0, \dots, 2N$ and then obtained the increments on the coarser grid by adding the respective increments on the fine grid.

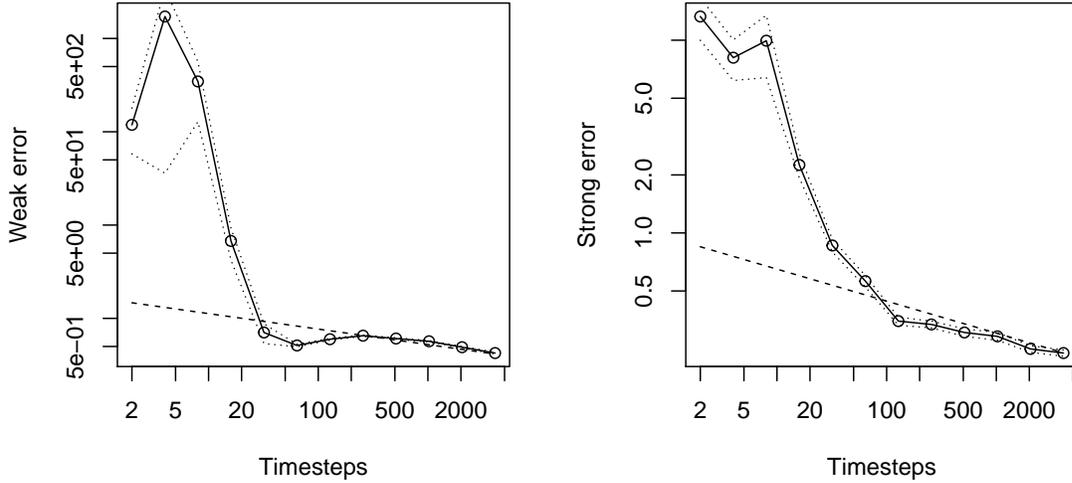


FIGURE 4. Strong and weak error for a fBm with Hurst index $H = 0.33$. Dashed line corresponds to the theoretical strong rate of convergence 0.16, dotted lines show confidence intervals around the error due to the integration error. Weak error corresponds to the functional $f(y) := (|y(1)| - 1)^+$.

Fortunately, the picture is completely different when the coarsest grid is chosen to be fine enough, in the current example for $H = 0.4$ this means $h_0 \leq 0.02$. Then the multi-level algorithm requires considerably less computational time for the same MSE tolerance than a classical MC estimator, even for quite moderate levels of the tolerance. For this demonstration, we choose a different function, namely

$$g(y) = |y| \mathbf{1}_{y^1 > 0}.$$

Indeed, the previously used function $f(y) = (|y| - 1)^+$ has the property that $f(Y_1) \equiv 0$, so that the variance of $f(\bar{Y}_1^N)$ goes to 0 when $N \rightarrow \infty$. This, however, makes the basic idea of the multi-level approach redundant, as the variance of estimators anyway decrease when the mesh is decreased, even without the telescoping procedure.

A direct comparison of the performance of the classical and the multi-level Monte-Carlo estimator is difficult in our situation, as it is very hard to obtain a reference value, i.e., a “true result”. Moreover, by the same reasoning the coefficients c_i in Theorem 18 are very difficult to estimate. Thus, we use the following procedure to test the respective performances:

- Fix L , the number of levels in the multi-level procedure, and h_0 , the coarsest grid. Here, we choose $h_0 = 1/64$ and $L = 7$. Thus, the finest grid in the multi-level Monte Carlo corresponds to $h_L = 0.00012 = 1/8192$. As the fixed L is probably sub-optimal, this choice is disadvantageous to the multi-level algorithm. We also choose the multiplication factor $M = 2$ here, and we parametrize the number of paths N_l for the level l by the number of paths N_0 at the coarsest level by some heuristic. In Table 3, we choose $N_0 = 100$. Again, these non-optimal choices favour the classical Monte Carlo estimator.
- Choose the mesh of the classical Monte Carlo estimator to be equal to h_L , the finest grid in the multi-level hierarchy. This guarantees that both estimators have the same bias – even though we cannot easily estimate this bias due to the absence of a reference value.

- Choose the number of paths in the classical Monte Carlo estimator and the number of paths in the coarsest grid for the multi-level estimator such that the complexity for the classical Monte Carlo estimator is equal to the complexity of the multi-level Monte Carlo estimator. We use an a-priori estimate for the complexity.
 - For the classical Monte Carlo method, the complexity is estimated by the number of trajectories multiplied by the size of the grid.
 - For the multi-level Monte Carlo method, the complexity at a level l is estimated by the product of the size of the finer grid and the number of trajectories for the level. The overall complexity is estimated by the sum of these complexity estimates for the individual levels.

Note that in practice, this complexity estimate is only given up to a constant of proportionality, which can be checked by comparing run-times on a computer.

- Compute the sample variance for both estimators. If the sample variance for the multi-level Monte Carlo estimator is (significantly) smaller than the sample variance for the classical Monte Carlo estimator, then we, indeed, have demonstrated that the multi-level estimator will have a smaller MSE than the classical Monte Carlo estimator given the same computational budget, i.e., the same complexity.

The nice aspect of this procedure is that it allows a reliable comparison of MSE given a certain complexity, even when the true MSE is not known because of the absence of a reference value. However, we stress again that the multi-level estimator constructed above will certainly not be optimal. In order to take care of the constant in the complexity bound, we also compare the actual run-times as empirical complexity estimates.

	Multilevel	Classical MC
Variance	1.47×10^{-2}	1.90×10^{-2}
Time	0.99 s	3.68 s

TABLE 3. Variance and run-times for the multi-level and the classical Monte Carlo algorithm for fixed complexity and bias. Calculations are normalized by $N_0 = 100$.

Table 3 finds that for comparable complexity the variance associated to the classical Monte Carlo estimator is considerably lower than the variance of the multi-level Monte Carlo estimator. It is interesting to note that the classical Monte Carlo estimator takes considerably longer computational time. The reason is that the multi-level algorithm uses the Euler scheme on coarser grids on average than the classical Monte Carlo algorithm. As the complexity for sampling the increments of the fractional Brownian motion increases quadratically in the size of the grid when Hosking’s method is applied, this explains why the computational time is almost four times larger for the classical Monte Carlo method. Note that there are other exact simulation methods with a complexity of order $\mathcal{O}(M \log(M))$ in the grid size M , and approximate simulation methods even with order $\mathcal{O}(M)$, see [Die04]. However, at least for the present, linear differential equation, the simulation of the increments of the fBm will always dominate the Euler steps, even when the complexity does only increase linearly. Thus, the conclusions of Table 3 should hold irrespective of the simulation method.⁶

⁶The following heuristic calculation also supports this conclusion: assuming that we replace Hosking’s algorithm by an algorithm with linear complexity and the same constant. Then we can easily predict the run-time of the classical Monte Carlo algorithm by dividing the run-time reported in Table 3 by the size of the (finest) grid, i.e., by

5.3. A fractional Heston model. As a third example, let us consider an application from finance. One of the most popular asset price models is the Heston model, a stochastic volatility model, meaning that the diffusion coefficient of the asset price is itself stochastic. Recently, it has emerged that the stochastic volatility component is rougher than a standard Brownian motion (or a diffusion process) and should be modelled by a fractional Brownian motion or a process driven by a fractional Brownian motion, respectively, see Gatheral, Jaisson and Rosenbaum [GJR14a] and Bayer, Friz and Gatheral [BFG15] and the references therein. Hence, the following *fractional Heston model*—corresponding to the classical Heston model in Stratonovich formulation for $h = 1/2$ —, see also Guennoun, Jacquier and Roome [GJR14b], could be considered:

$$(23a) \quad dS_t = -\frac{1}{2} \left(v_t + \frac{1}{2} \xi \varrho \right) S_t dt + \sqrt{v_t} S_t dW_t^1$$

$$(23b) \quad dv_t = \left(\kappa(\theta - v_t) - \frac{1}{4} \xi^2 \right) dt + \xi \sqrt{v_t} (\varrho dW_t^1 + \sqrt{1 - \varrho^2} dW_t^2),$$

for a standard Brownian motion W^1 and an independent fractional Brownian motion W^2 with Hurst index $1/4 < H < 1/2$. That is, in terms of the RDE (1), we choose the Gaussian process $X_t = (W_t^1, W_t^2)$, which satisfies our assumptions with $\rho = 1/(2H)$. Note that the driving noise of the price and the variance processes can be correlated using $-1 \leq \varrho \leq 1$.

Notice that the fractional Heston model does not satisfy our regularity assumptions in several ways:

- the vector fields are unbounded;
- the vector fields are not differentiable at $v = 0$ and not even defined for $v < 0$.

Hence, depending on the choice of parameters and initial values (S_0, v_0) , we may expect the rate to deteriorate. We also need to adjust the scheme in order to preserve positivity of (S, v) , in this case by simply taking the positive part after each time step.

On the other hand, one may expect the impact of the Brownian motion W^1 to be stronger than the impact of the fractional Brownian motion W^2 , especially when considering standard payoff functions depending on the asset price component S only. Thus, the actual error based on a step size h might look like $h^{-1/2}$ and h^{-1} in the strong and weak sense, respectively, when the parameters and initial values are “nice enough” and h is not too small.

We choose model parameters $\kappa = 1$, $\theta = 0.16$, $\xi = 0.2$, $\rho = -0.1$, $H = 0.4$. Notice that these parameters easily verify the Feller condition $2\kappa\theta > \xi^2$ (which would imply $v_t > 0$ for $H = 1/2$). We further choose $S_0 = 1$, $v_0 = \theta$ and $r = 0$ and consider a European call option with strike price $K = 1$.

As expected and shown in Figure 5, we empirically observe first order weak convergence and strong convergence with rate $1/2$ —actually, regression gives rates 1.02 and 0.42, respectively. This is significantly better than the theoretical (strong) rate 0.3. We expect the rate to deteriorate eventually, when the number of timesteps is increased even further.

8192, which gives a predicted run-time of 0.00045 seconds. For the multilevel Monte Carlo method, the corresponding factor would be (with $M_l = Th_l^{-1}$ and $N_l = N_0 2^{-l(1+\beta)/2} = N_0 2^{-0.8l}$)

$$\frac{M_0^2 N_0 + \dots + M_L^2 N_L}{M_0 N_0 + \dots + M_L N_L} = 2799,$$

giving a predicted run-time of 0.00035 seconds, which is still lower than the predicted run-time for the classical Monte Carlo algorithm.

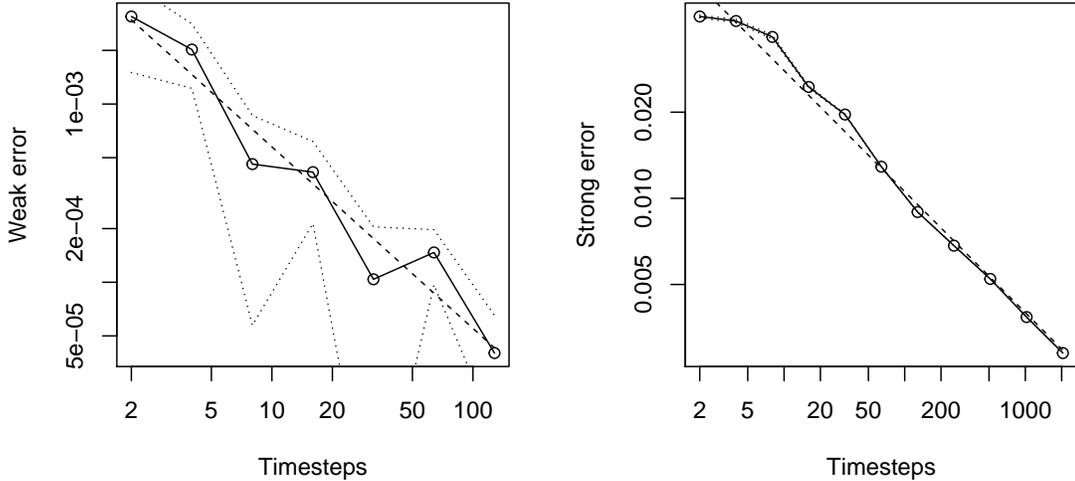


FIGURE 5. Weak and strong error for the fractional Heston model. Solid lines show the empirical errors, dashed lines show regression lines with rates 1.02 in the weak and 0.42 in the strong case. Dotted lines show confidence intervals around the solid lines.

We have also implemented the multilevel Monte Carlo algorithm for the fractional Heston model (using the same parameters as above). We normalize the workload to (roughly) 2 610 000 individual Euler steps—i.e., we set the cost of one Euler step to one. We choose $M_0 = M = 2$ and fix the bias by requiring $h_L = \frac{1}{16}$. Under these normalizations, the single level and multilevel algorithms produce the variances and run times reported in Table 4. As expected, the multilevel algorithm

	Multilevel	Classical MC
Variance	1.9×10^{-7}	7.4×10^{-7}
Time	0.88 s	0.79 s

TABLE 4. Variance and run-times for the multi-level and the classical Monte Carlo algorithm for fixed complexity and bias.

again clearly outperforms the single level algorithm in producing an estimate with essentially 4 times smaller variance. We note that the parameters of the multilevel algorithm were based on the *observed* weak and strong rates of convergence, not the theoretical ones.

APPENDIX A. ELEMENTS OF ROUGH PATH THEORY

We will now very briefly recall the elements of rough paths theory used in this paper. For more details we refer to [FV10b], [LCL07], [LQ02] or [FH14]. Our notation coincides with the one used in [FV10b].

Let $T^N(\mathbb{R}^d) = \mathbb{R} \oplus \mathbb{R}^d \oplus (\mathbb{R}^d \otimes \mathbb{R}^d) \oplus \dots \oplus (\mathbb{R}^d)^{\otimes N}$ be the truncated step- N tensor algebra. We are concerned with $T^N(\mathbb{R}^d)$ -valued paths, as naturally given by iterated integrations of \mathbb{R}^d -valued smooth paths (“lifted smooth paths”). Such a path \mathbf{x} has natural increments $\mathbf{x}_{s,t} \equiv \mathbf{x}_s^{-1} \otimes \mathbf{x}_t$. The projection of such a path \mathbf{x} on the first level is an \mathbb{R}^d -valued path and will be denoted by

$\pi_1(\mathbf{x})$, the projection to k th level is denoted by π_k . Lifted smooth paths actually take values in $G^N(\mathbb{R}^d) \subset T^N(\mathbb{R}^d)$, where $G^N(\mathbb{R}^d)$ denotes the free step- N nilpotent Lie group with d generators. The (left-invariant) Carnot-Caratheodory metric turns $(G^N(\mathbb{R}^d), d)$ into a metric space.

This already allows (see e.g. [FV10b]) to introduce the most commonly used (homogenous) rough path “norms”

$$\begin{aligned} \|\mathbf{x}\|_{p\text{-var};[0,T]} &= \sup_{(t_i) \subset [0,T]} \left(\sum_i d(\mathbf{x}_{t_i}, \mathbf{x}_{t_{i+1}})^p \right)^{1/p}, \\ \|\mathbf{x}\|_{1/p\text{-Höl};[0,T]} &= \sup_{0 \leq s < t \leq T} \frac{d(\mathbf{x}_s, \mathbf{x}_t)}{|t-s|^{1/p}}, \end{aligned}$$

and distances

$$\begin{aligned} d_{p\text{-var};[0,T]}(\mathbf{x}, \mathbf{y}) &= \left(\sup_{(t_i) \subset [0,T]} \sum_i d(\mathbf{x}_{t_i, t_{i+1}}, \mathbf{y}_{t_i, t_{i+1}})^p \right)^{1/p}, \\ d_{1/p\text{-Höl};[0,T]}(\mathbf{x}, \mathbf{y}) &= \sup_{0 \leq s < t \leq T} \frac{d(\mathbf{x}_{s,t}, \mathbf{y}_{s,t})}{|t-s|^{1/p}} \end{aligned}$$

where $p \in [1, \infty)$. Define also “inhomogenous” variation and Hölder distances as follows. For $k = 1, \dots, N$,

$$\rho_{p\text{-var};[0,T]}^{(k)}(\mathbf{x}, \mathbf{y}) = \sup_{(t_i) \subset [0,T]} \left(\sum_i |\pi_k(\mathbf{x}_{t_i, t_{i+1}} - \mathbf{y}_{t_i, t_{i+1}})|^{p/k} \right)^{k/p},$$

and

$$\rho_{p\text{-var};[0,T]}(\mathbf{x}, \mathbf{y}) = \max_{k=1, \dots, N} \rho_{p\text{-var};[0,T]}^{(k)}(\mathbf{x}, \mathbf{y}).$$

Similarly,

$$\rho_{1/p\text{-Höl};[0,T]}^{(k)}(\mathbf{x}, \mathbf{y}) = \sup_{0 \leq s < t \leq T} \frac{|\pi_k(\mathbf{x}_{s,t} - \mathbf{y}_{s,t})|}{|t-s|^{k/p}},$$

and

$$\rho_{1/p\text{-Höl};[0,T]}(\mathbf{x}, \mathbf{y}) = \max_{k=1, \dots, N} \rho_{1/p\text{-Höl};[0,T]}^{(k)}(\mathbf{x}, \mathbf{y}).$$

Recall that a control function ω is a continuous function from $\{0 \leq s \leq t \leq T\}$ to $[0, \infty)$, for which $\omega(s, t) + \omega(t, u) \leq \omega(s, u)$ holds for every $s \leq t \leq u$. Note that $\omega(s, t)^{1/p}$ is a natural generalization of the quantity $|t-s|^{1/p}$ which appeared in the definition of all “Hölder objects”

$$\|\mathbf{x}\|_{1/p\text{-Höl};[0,T]}, d_{1/p\text{-Höl};[0,T]}(\mathbf{x}, \mathbf{y}), \rho_{1/p\text{-Höl};[0,T]}^k(\mathbf{x}, \mathbf{y}), \rho_{1/p\text{-Höl};[0,T]}(\mathbf{x}, \mathbf{y}),$$

defined above. Replacing $|t-s|^{1/p}$ by $\omega(s, t)^{1/p}$ then gives rise to similar norms and distances, denoted by

$$\|\mathbf{x}\|_{p\text{-}\omega;[0,T]}, d_{p\text{-}\omega;[0,T]}(\mathbf{x}, \mathbf{y}), \rho_{p\text{-}\omega;[0,T]}^k(\mathbf{x}, \mathbf{y}), \rho_{p\text{-}\omega;[0,T]}(\mathbf{x}, \mathbf{y}).$$

By definition, a geometric $1/p$ -Hölder rough path \mathbf{x} is a path in $T^{\lfloor p \rfloor}(\mathbb{R}^d)$ which can be approximated by lifts of smooth paths in the $d_{1/p\text{-Höl}}$ (equivalently: $\rho_{1/p\text{-Höl}}$) metric; geometric p -rough

paths are defined similarly (with respect to the variation distance). Necessarily then, any such \mathbf{x} takes values in $G^{\lfloor p \rfloor}(\mathbb{R}^d) \subset T^{\lfloor p \rfloor}(\mathbb{R}^d)$. The resulting rough path spaces are known to be Polish and are denoted by

$$C_0^{0,1/p-\text{H\"{o}l}}([0, T], G^{\lfloor p \rfloor}(\mathbb{R}^d)) \quad \text{and} \quad C_0^{0,p-\text{var}}([0, T], G^{\lfloor p \rfloor}(\mathbb{R}^d)).$$

If $V = (V_i)_{i=1, \dots, d}$ is a collection of $\text{Lip}^\gamma(\mathbb{R}^e)$ vector fields (in the sense of Stein, cf. [FV10b]) for some $\gamma > p$ and \mathbf{x} is a geometric p -rough path, one can make sense of a unique solution $y: [0, T] \rightarrow \mathbb{R}^e$ of the equation

$$dy_t = V(y_t) d\mathbf{x}_t; \quad y_0 \in \mathbb{R}^e$$

and the solution depends (locally Lipschitz) continuously on the driving signal in the inhomogeneous rough paths metric.

Acknowledgements. P.F. has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) / ERC grant agreement nr. 258237. S.R. was supported by a scholarship from the Berlin Mathematical School (BMS). C. B. and P. F. acknowledge funding by the DFG grants BA5484/1 and FR2943/2). This paper contains results of S. Riedel's Ph.D. dissertation [Rie13, Ch. 6], which was written in collaboration with CB, PKF and JS.

REFERENCES

- [BFG15] Christian Bayer, Peter Friz, and Jim Gatheral, *Pricing under rough volatility*, Quantitative Finance (2015), to appear.
- [BSD13] Denis Belomestny, John Schoenmakers, and Fabian Dickmann, *Multilevel dual approach for pricing American style derivatives*, Finance Stoch. **17** (2013), no. 4, 717–742.
- [CC80] John M. C. Clark and R. J. Cameron, *The maximum rate of convergence of discrete approximations for stochastic differential equations*, Stochastic differential systems (Proc. IFIP-WG 7/1 Working Conf., Vilnius, 1978), Lecture Notes in Control and Information Sci., vol. 25, Springer, Berlin, 1980, pp. 162–171.
- [CF10] Thomas Cass and Peter K. Friz, *Densities for rough differential equations under Hörmander's condition*, Ann. of Math. (2) **171** (2010), no. 3, 2115–2141.
- [CHAN⁺15] Nathan Collier, Abdul-Lateef Haji-Ali, Fabio Nobile, Erik von Schwerin, and Ral Tempone, *A continuation multilevel monte carlo algorithm*, BIT Numerical Mathematics **55** (2015), no. 2, 399–432 (English).
- [CLL13] Thomas Cass, Christian Litterer, and Terry Lyons, *Integrability and tail estimates for Gaussian rough differential equations*, Ann. Probab. **41** (2013), no. 4, 3026–3050.
- [CQ02] Laure Coutin and Zhongmin Qian, *Stochastic analysis, rough path analysis and fractional Brownian motions*, Probab. Theory Related Fields **122** (2002), no. 1, 108–140.
- [Dav07] Alexander M. Davie, *Differential equations driven by rough paths: an approach via discrete approximation*, Appl. Math. Res. Express. AMRX (2007), no. 2, Art. ID abm009, 40.
- [Die04] Ton Dieker, *Simulation of fractional Brownian motion*, Master's thesis, University of Twente, 2004.
- [DNT12] Aurélien Deya, Andreas Neuenkirch, and Samy Tindel, *A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion*, Ann. Inst. Henri Poincaré Probab. Stat. **48** (2012), no. 2, 518–550.
- [FGGR15] Peter K. Friz, Benjamin Gess, Archil Gulisashvili, and Sebastian Riedel, *The Jain-Monrad criterion for rough paths and applications to random Fourier series and non-Markovian Hörmander theory*, Ann. Probab. (to appear; http://www.imstat.org/aop/future_papers.htm) (2015+).
- [FH14] Peter K. Friz and Martin Hairer, *A course on rough paths*, Universitext, Springer, Cham, 2014, With an introduction to regularity structures.
- [FR13] Peter K. Friz and Sebastian Riedel, *Integrability of (non-)linear rough differential equations and integrals*, Stoch. Anal. Appl. **31** (2013), no. 2, 336–358.

- [FR14] Peter Friz and Sebastian Riedel, *Convergence rates for the full Gaussian rough paths*, Ann. Inst. Henri Poincaré Probab. Stat. **50** (2014), no. 1, 154–194.
- [FV10a] Peter K. Friz and Nicolas B. Victoir, *Differential equations driven by Gaussian signals*, Ann. Inst. Henri Poincaré Probab. Stat. **46** (2010), no. 2, 369–413.
- [FV10b] ———, *Multidimensional stochastic processes as rough paths*, Cambridge Studies in Advanced Mathematics, vol. 120, Cambridge University Press, Cambridge, 2010, Theory and applications.
- [Gil08a] Michael B. Giles, *Improved multilevel Monte Carlo convergence using the Milstein scheme*, Monte Carlo and quasi-Monte Carlo methods 2006, Springer, Berlin, 2008, pp. 343–358.
- [Gil08b] ———, *Multilevel Monte Carlo path simulation*, Oper. Res. **56** (2008), no. 3, 607–617.
- [GJR14a] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum, *Volatility is rough*, preprint, 2014.
- [GJR14b] Hamza Guennoun, Antoine Jacquier, and Patrick Roome, *Asymptotic behaviour of the fractional heston model*, preprint, 2014.
- [GS14] Michael B. Giles and Lukasz Szpruch, *Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation*, Ann. Appl. Probab. **24** (2014), no. 4, 1585–1620. MR 3211005
- [HP13] Martin Hairer and Natesh S. Pillai, *Regularity of laws and ergodicity of hypoelliptic SDEs driven by rough paths*, Ann. Probab. **41** (2013), no. 4, 2544–2598.
- [KP92] Peter E. Kloeden and Eckhard Platen, *Numerical solution of stochastic differential equations*, Applications of Mathematics (New York), vol. 23, Springer-Verlag, Berlin, 1992.
- [LCL07] Terry J. Lyons, Michael Caruana, and Thierry Lévy, *Differential equations driven by rough paths*, Lecture Notes in Mathematics, vol. 1908, Springer, Berlin, 2007, Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, With an introduction concerning the Summer School by Jean Picard.
- [LQ02] Terry J. Lyons and Zhongmin Qian, *System control and rough paths*, Oxford Mathematical Monographs, Oxford University Press, Oxford, 2002, Oxford Science Publications.
- [MGR09] Thomas Müller-Gronbach and Klaus Ritter, *Variable subspace sampling and multi-level algorithms*, Monte Carlo and quasi-Monte Carlo methods 2008, Springer, Berlin, 2009, pp. 131–156. MR 2743892 (2012e:65012)
- [MN98] M. Matsumoto and T. Nishimura, *Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator*, ACM Trans. on Modeling and Computer Simulation **8** (1998), no. 1, 3–30.
- [Rie13] Sebastian Riedel, *Topics in Gaussian rough paths theory*, dissertation, Technische Universität Berlin, 2013.
- [Tal86] Denis Talay, *Discrétisation d'une équation différentielle stochastique et calcul approché d'espérances de fonctionnelles de la solution*, RAIRO Modél. Math. Anal. Numér. **20** (1986), no. 1, 141–179.

CHRISTIAN BAYER, WEIERSTRASS INSTITUTE BERLIN, GERMANY

E-mail address: `Christian.Bayer@wias-berlin.de`

PETER K. FRIZ, TECHNISCHE UNIVERSITÄT BERLIN AND WEIERSTRASS INSTITUTE BERLIN, GERMANY

E-mail address: `friz@math.tu-berlin.de`

SEBASTIAN RIEDEL, TECHNISCHE UNIVERSITÄT BERLIN, GERMANY

E-mail address: `riedel@math.tu-berlin.de`

JOHN SCHOENMAKERS, WEIERSTRASS INSTITUTE BERLIN, GERMANY

E-mail address: `schoenma@wias-berlin.de`