FROM ROUGH PATH ESTIMATES TO MULTILEVEL MONTE CARLO

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

Abstract. Discrete approximations to solutions of stochastic differential equations are well-known to converge with "strong" rate 1/2. Such rates have played a key-role in Giles' multilevel Monte Carlo method [Giles, Oper. Res. 2008] which gives a substantial reduction of the computational effort necessary for the evaluation of diffusion functionals. In the present article similar results are established for large classes of rough differential equations driven by Gaussian processes (including fractional Brownian motion with $H > 1/4$ as special case).

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

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

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]. This requires 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

$$\mathbf{X}_{\cdot}(\omega) = 1 + \sum_{i=1}^{N} \int_{0<s_1<\cdots<s_i<\cdot} dX_{s_1}(\omega) \otimes \cdots \otimes dX_{s_i}(\omega),$$

a situation fairly well understood, cf. [FV10b, Ch. 15] and the references therein. 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

$$dY = V(Y)\,d\mathbf{X}. \tag{2}$$

Of course, it would be easy to include equations of the form (1) into the framework (2), e.g., by including time t as an additional (smooth) component of the noise X. This setting includes, for instance, fractional Brownian motion [CQ02] with Hurst parameter $H > 1/4$. 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 $\mathbb{B}_{s,t} = \int_s^t B_{s,\cdot} \otimes dB$. 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

Acknowledgements: P.F. has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement nr. 258237. S.R. was supported by a scholarship from the Berlin Mathematical School (BMS).

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

$$Y_t - Y_s \equiv Y_{s,t} = V_i(Y_s) B^i_{s,t} + V^k_i(Y_s) \partial_k V_j(Y_s) \mathbb{B}^{i,j}_{s,t} + o(|t-s|) \text{ as } t \downarrow s, \tag{3}$$

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

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

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 (3). The practical trouble is that Lévy's area, the antisymmetric 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 (3), 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, [GS12] 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 ever: 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 [CLL, FR13], initially developed to prove smoothness of laws for (non-Markovian) SDEs driven by Gaussian signals under a Hörmander condition, [CF10, HP].

Having established L^r-convergence (any $r < \infty$, with rates) for implementable "simplified" Milstein schemes we move to our second contribution: a multilevel algorithm, in the sense of Giles [Gil08b], for stochastic differential equations driven by large classes of Gaussian signals. 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" then Brownian motion. Luckily this can be done without too much trouble. Moreover, we more carefully keep track of the relevant constants in front of the asymptotic terms, a necessity in such an irregular regime.

More precisely, we consider the following scheme for approximating Y, see [DNT12, FV10b]. Given a 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

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

$X_{t_k,t_{k+1}}$ for the corresponding increments. We then define $\overline{Y}_0 \equiv Y_0$ and

$$\overline{Y}_{t_{k+1}} = \overline{Y}_{t_k} + \sum_{l=1}^{3} \frac{1}{l!} V_{i_1} \cdots V_{i_l} I\left(\overline{Y}_{t_k}\right) X_{t_k,t_{k+1}}^{i_1} \cdots X_{t_k,t_{k+1}}^{i_l}, \tag{4}$$

where $I(y) = y$ is the identity function and the vector fields $V_1, \ldots, V_d$ are viewed as linear first order operators, acting on functions by $V_i g(y) = \nabla g(y) \cdot V_i(y)$. Whenever convenient we extend $\overline{Y}$ 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 2.3. We are now able to state our main results; cf. Corollary 17:

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

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

where σ^2 is concave and $\sigma^2(\tau) = \mathcal{O}\left(\tau^{1/\rho}\right)$ as $\tau \to 0$ for some $\rho \in [1,2)$.
Let Y be the solution to the rough differential equation (2) driven by (the rough path lift) of X and $\overline{Y} = \overline{Y}^h$ be the approximate solution based on (4). Then we have strong convergence of (almost) rate $1/\rho - 1/2$. More precisely, for any $1 \le r < \infty$ and $\delta > 0$, there exists a constant C such that

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

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 $H \in (1/4, 1/2)$. Using Giles' multi-level Monte Carlo methodology, we can greatly improve the complexity bounds for the discretization algorithm (4), see Theorem 22.

Theorem 2 (Multilevel complexity estimate)**.** *Let X and Y be as in the previous theorem and $f : C([0,T], \mathbb{R}^m) \to \mathbb{R}^n$ a Lipschitz continuous functional. Then the Monte Carlo evaluation of a path-dependent functional of the form*

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

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

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

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 $\mathcal{O}\left(\varepsilon^{-2}\left(\log\varepsilon\right)^2\right)$ [Gil08a, Gil08b], recently sharpened to $\mathcal{O}\left(\varepsilon^{-2}\right)$ [GS12] 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 (4) 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 convergence[2], 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.

1. Rough path estimates revisited

In this section, we revisit and improve 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] or [FV10b].

1.1. Improved bounds for the Lipschitz constant of the Itô-Lyons map.

It is well-known that the Lipschitz constant of the Itô-Lyons solution map for an RDE driven by a rough path $\mathbf{x}$ is of the order $\mathcal{O}(\exp(C\|\mathbf{x}\|^p_{p-\mathrm{var}}))$. Considering Gaussian driving signals, this (random) constant fails to have finite q-th moments for any q. In this section, we improve the deterministic estimates for the Lipschitz constant slightly which will allow us to derive the desired estimates.

Recall the following definition, taken from [CLL]:

Definition 3. *Let ω be a control function. For $\alpha > 0$ and $[s,t] \subset [0,T]$ we set*

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

and define

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

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

It is easy to see that $N_{\alpha,[0,T]}(\mathbf{x}) \lesssim \|\mathbf{x}\|^p_{p-\mathrm{var};[0,T]}$, but the tail estimates for $N_{\alpha,[0,T]}(\mathbf{X})$ are significantly better than for $\|\mathbf{X}\|^p_{p-\mathrm{var};[0,T]}$ when we consider Gaussian lifts $\mathbf{X}$, cf. [CLL] and [FR13]. Next, we give the main result from this section.

Theorem 4. *Consider the RDEs*

$$dy^i_t = V^i(y^i_t)\, d\mathbf{x}^i_t; \quad y^i_0 \in \mathbb{R}^e$$

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

$$\begin{aligned} \left|y^1 - y^2\right|_{\infty;[0,T]} \leq\ & C\left[|y^1_0 - y^2_0| + \left|V^1 - V^2\right|_{Lip^{\gamma-1}} + \rho_{p-var;[0,T]}(\mathbf{x}^1, \mathbf{x}^2)\right] \\ & \times \exp\left\{C\left(N_{\alpha,[0,T]}(\mathbf{x}^1) + N_{\alpha,[0,T]}(\mathbf{x}^2) + 1\right)\right\} \end{aligned}$$

holds.

[2]By 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 4, where the weak order is equal to the strong order even in a standard Brownian motion setting.

Remark 5. *Comparing the result of Theorem 4 with* [FV10b, Theorem 10.26]*, one sees that we obtain a slightly weaker result; namely, the distance between y^1 and y^2 is measured here in uniform topology instead of p-variation topology. However, with little more effort, one can show that the same estimate holds for $\rho_{p\text{-}var;\,[0,T]}\left(y^1,y^2\right)$ instead of $\left|y^1-y^2\right|_{\infty;[0,T]}$.*

The proof of Theorem 4 will be given at the end of this section. We first prove some preparational Lemmata.

Recall that if ω^1 and ω^2 are controls, also $\omega^1+\omega^2$ is a control.

Lemma 6. *Let ω^1 and ω^2 be two controls. Then*

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

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

Proof. If ω is any control, set

$$\omega_\alpha(s,t) := \sup_{\substack{(t_i)=D\subset[s,t]\\ \omega(t_i,t_{i+1})\le\alpha}} \sum_{t_i}\omega(t_i,t_{i+1}).$$

If $\bar\omega := \omega^1+\omega^2$, $\bar\omega(t_i,t_{i+1})\le\alpha$ implies $\omega^i(t_i,t_{i+1})\le\alpha$ for $i=1,2$ and therefore $\bar\omega_\alpha(s,t)\le \omega^1_\alpha(s,t)+\omega^2_\alpha(s,t)$. From Proposition 4.6 in [CLL] we know that $\omega^i_\alpha(s,t)\le\alpha\left(2N_{\alpha,[s,t]}\left(\omega^i\right)+1\right)$ for $i=1,2$. (Strictly speaking, Proposition 4.6 is formulated for a particular control ω, namely the control induced by the p-variation of a rough path. However, the proof only uses general properties of control functions and the conclusion remains valid.) We conclude

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

□

Lemma 7. *Let ω^1 and ω^2 be two controls and assume that $\omega^2(s,t)\le K$. Then*

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

for every $\alpha>K$.

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

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

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 $\bar\tau_i\ge\tau_i$ for $i=0,\dots,N_{\alpha,[s,t]}(\bar\omega)$. We do this by induction. For $i=0$, this is clear. If $\bar\tau_i\ge\tau_i$ for some $i\le N_{\alpha,[s,t]}(\bar\omega)-1$, superadditivity of control functions gives

$$\alpha=\bar\omega(\bar\tau_i,\bar\tau_{i+1})\le\omega^1(\tau_i,\bar\tau_{i+1})+K$$

which implies $\tau_{i+1}\le\bar\tau_{i+1}$. □

Lemma 8. *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*

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

for every $\alpha > 0$.

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

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

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

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

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

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

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

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

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

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

therefore

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

and finally

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

□

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 8 we know that there is a constant $C = C(\gamma, p, \nu, \alpha)$ such that

$$\begin{aligned}\big|y^1 - y^2\big|_{\infty;[0,T]} \leq \Big[|y^1_0 - y^2_0| + \big|V^1 - V^2\big|_{\mathrm{Lip}^{\gamma-1}} + \rho_{p-\omega;[s,t]}(\mathbf{x}^1, \mathbf{x}^2)\Big] \\ \times \exp\left\{C(N_{\alpha,[s,t]}(\omega) + 1)\right\}.\end{aligned}$$

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

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

It is easy to check that

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

Finally, if $\alpha > \lfloor p \rfloor$ we can use Lemma 7 and Lemma 6 to see that

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

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

1.2. **Improved bounds for Euler approximations based on entire rough path.** 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 $\mathbb{R}^e$, $g \in T^N\left(\mathbb{R}^d\right)$ 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 $\mathbf{x} \in C_0^{p-var}\left([0,T]; G^{\lfloor p\rfloor}\left(\mathbb{R}^d\right)\right)$ we define the (step-N) Euler approximation to the RDE solution y of

$$dy = V(y)\, d\mathbf{x} \tag{5}$$

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

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

Theorem 9. *Let $\mathbf{x} \in C_0^{p-var}\left([0,T]; G^{\lfloor p\rfloor}\left(\mathbb{R}^d\right)\right)$ and set $\omega(s,t) = \|\mathbf{x}\|_{p-var;[s,t]}^p$. Assume that $V \in Lip^\theta$ for some $\theta > p$ and let $\nu \ge |V|_{Lip^\theta}$. Choose $N \in \mathbb{N}$ such that $\lfloor p \rfloor \le N \le \theta$. Fix a dissection $D = \{0 = t_0 < \dots < t_n = T\}$ of $[0,T]$ and let $y_T^{Euler;D}$ 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*

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

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

$$\left|y_T - y_T^{Euler;D}\right| \le CT\, \|\mathbf{x}\|_{1/p\text{-Höl};[0,T]}^{\zeta p} \exp\left\{C\left(N_{\alpha,[0,T]}(\mathbf{x})+1\right)\right\} |D|^{\zeta-1} \tag{6}$$

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 (5) with starting point y_s at time s. Set

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

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

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

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 $\alpha > 0$

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

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

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

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

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

which gives the claim. □

2. Probabilistic convergence results for RDEs

Recall that our basic object is a multidimensional, zero-mean Gaussian process $X = (X^1, \dots, X^d)$ with independent components and continuous sample paths. The covariance function will be denoted by R_X. The existence of a lift $\mathbf{X}$ of X to a process with sample paths in a rough paths space follows if R_X is smooth enough in terms of 2-dimensional ρ-variation; more precisely, if $V_\rho(R_X; [0, T]^2) < \infty$ for some $\rho < 2$, $\mathbf{X}$ exists in a natural way (cf. [FV10a] or Theorem 25 in the appendix). The associated Cameron-Martin space of X will be denoted by $\mathcal{H}$. Recall that every path $\phi \in \mathcal{H}$ has the form $\phi_t = E(X_t Z)$ where Z is a random variable lying in the L^2-closure of $\text{span}\{X_t : t \in [0, T]\}$ and $\langle \phi, \tilde{\phi} \rangle_{\mathcal{H}} = E(Z\tilde{Z})$ if $\phi_t = E(ZX_t)$, $\tilde{\phi}_t = E(\tilde{Z}X_t)$. In the following, smoothness of the Cameron-Martin paths in terms of its p-variation index will be crucial. More precisely, we will say that *complementary Young-regularity* holds for the trajectories of X and the paths in the Cameron-Martin space if the trajectories have finite p-variation almost surely and if there is a continuous embedding

$$\iota \colon \mathcal{H} \hookrightarrow C^{q-\text{var}} \quad \text{for some } q \le p \text{ such that } \frac{1}{p} + \frac{1}{q} > 1.$$

Complementary Young-regularity will be a fundamental assumption. We recall some sufficient conditions:

(1) If R_X has finite ρ-variation, the Cameron-Martin paths have finite ρ-variation and complementary Young-regularity holds for $\rho \in [1, 3/2)$. In this case,

$$\|\iota\|_{\mathcal{H}\to C^{q-\mathrm{var}}} \le \sqrt{V_\rho(R_X; [0,T]^2)}$$

(cf. [FV10a, Proposition 17]).

(2) If R_X has finite mixed $(1,\rho)$-variation, the Cameron-Martin paths have finite q-variation for

$$q = \frac{1}{\frac{1}{2\rho} + \frac{1}{2}}$$

and complementary Young-regularity holds for $\rho \in [1, 2)$. In this case,

$$\|\iota\|_{\mathcal{H}\to C^{q-\mathrm{var}}} \le \sqrt{V_{1,\rho}(R_X; [0,T]^2)}$$

(cf. [FGR13, Theorem 1]).

(3) If X has stationary increments and $\sigma^2(u) := E(|X_u - X_0|^2)$ is concave with $\sigma^2(u) \le C_\sigma |u|^{\frac{1}{\rho}}$, there is a constant $l > 0$ such that R_X has finite mixed $(1,\rho)$-variation on squares $[s,t]^2$ with $|t-s| \le l$ and

$$V_{1,\rho}(R_X; [s,t]^2) \le 5 C_\sigma |t-s|^{\frac{1}{\rho}}$$

holds for all $|t-s| \le l$ (cf. [FGR13, Theorem 6]). Hence we have complementary Young-regularity for $\rho \in [1,2)$ and $\|\iota\|_{\mathcal{H}\to C^{q-\mathrm{var}}} \le C_{K,l,\rho,T}$ with $K \ge C_\sigma$. This covers fractional Brownian motion with Hurst parameter $H \in (1/4, 1/2]$.

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. Although these schemes are hard to implement (the distributions of the iterated integrals are in general not known) it will serve as a stepping stone to more simple schemes. 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 [FR] also hold in L^p.

2.1. 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 10. *Assume that $V_\rho\left(R_X; [s,t]^2\right) \le K\,|t-s|^{1/\rho}$ holds for all $s < t$, some $\rho \in [1,2)$ and a constant K. Assume that*

$$\iota \colon \mathcal{H} \hookrightarrow C^{q-var}$$

for some $q < 2$, let $M \ge \|\iota\|_{\mathcal{H}\to C^{q-var}}$ and assume that complementary Young-regularity holds. Choose $p > 2\rho$, assume that $V \in Lip^\theta$ for some $\theta > p$ and let $\nu \ge |V|_{Lip^\theta}$. Set

$$D = \{0 < h < 2h < \ldots < (\lfloor T/h \rfloor - 1)h < T\}$$

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 $\lfloor p \rfloor \le N \le \theta$.

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

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

holds for all $h > 0$.

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

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

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

Proof of Theorem 10. Lemma 5 together with Corollary 2 and Remark 1 in [FR13] show that there is a $\alpha = \alpha(p, \rho, K) > 0$ and a positive constant $c_1 = c_1(p, q, \rho, K)$ such that we have the tail estimate

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

for all $u > 0$. Now we use the pathwise estimate (6) and take the L^r norm on both sides. The Hölder inequality shows that

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

holds for some (possibly large) $r'' > r$. Our tail estimate for $N_{\alpha,[0,T]}(\mathbf{X})$ shows that the right hand side is finite which gives the claim. □

2.2. **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

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

$$\sup_{h \in (0,1]} V_\rho\left(R_{(X^h, X)}; [0,T]^2\right) =: K < \infty \tag{7}$$

for some $\rho \in [1, 2)$.

(2) Uniform convergence of the second moments:

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

Example 12. *A typical example of such approximations are the piecewise linear approximations of $X(\omega)$ at the time points $\{0 < h < 2h < \ldots < (\lfloor T/h \rfloor - 1)h < T\}$ (see* [FV10b, Chapter 15.5]*). In the case $V_\rho\left(R_X; [s,t]^2\right) \lesssim |t-s|^{1/\rho}$ (i.e. if we deal with Hölder rough paths), one can show that $\delta(h) \lesssim h$.*

Theorem 13. *Let $X : [0,T] \to \mathbb{R}^d$ be a centered Gaussian process with continuous sample paths, independent components and covariance of finite ρ-variation for some $\rho \in [1, 2)$. Let $(X^h)_{h>0}$ be a family of suitable approximations as seen in the beginning of this section. Let $\mathcal{H}^h$ and $\mathcal{H}^0$ denote the Cameron-Martin spaces of the processes X^h resp. X. Assume that*

$$\iota^h : \mathcal{H}^h \hookrightarrow C^{q-var}$$

for some $q < 2$, let $M \geq \|\iota^h\|_{\mathcal{H} \to C^{q-var}}$ for all $h > 0$ and assume that complementary Young-regularity holds. 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 $|V|_{Lip^\theta} \le \nu < \infty$ for some $\theta > \frac{2\rho}{1-2\rho\eta}$. Let $Y, Y^h \colon [0,T] \to \mathbb{R}^e$ denote the pathwise solutions to the equations

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

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

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

holds for all $h > 0$.

Remark 14. *We give some sufficient conditions under which the assumptions on the Cameron-Martin paths in Theorem 13 are fulfilled:*

1. *If the Cameron-Martin paths associated to the process X have finite q-variation, if complementary Young-regularity holds for the trajectories of X and the Cameron-Martin paths and if the operators $\Lambda_h \colon \omega \mapsto \omega^h$ are uniformly bounded in the sense that*

$$\sup_{h>0} \|\Lambda_h\|_{C^{q-var} \to C^{q-var}} < \infty,$$

   we have $\sup_{h>0} \|\iota^h\|_{\mathcal{H} \to C^{q-var}} < \infty$. This is the case, for instance, when dealing with piecewise-linear or mollifier approximations.
2. *In the case $\rho \in [1, 3/2)$, the assumptions are always fulfilled using (7) and the Cameron-Martin embedding from* [FV10a, Proposition 17] *and we can set $M = \sqrt{K}$.*
3. *For $\rho \in [1, 2)$, another sufficient condition is uniform mixed $(1, \rho)$-variation:*

$$\sup_{h \in (0,1]} V_{1,\rho}(R_{(X^h, X)}; [0,T]^2) =: K' < \infty.$$

   In this case, we can choose $K = K'$ and $M = \sqrt{K'}$ (see [FGR13, Theorem 1]*). This holds, for instance, for fractional Brownian motion with Hurst parameter $H > 1/4$.*

Proof of Theorem 13. By assumption, we know that

$$|\phi|_{q-\mathrm{var}} \le M\, |\phi|_{\mathcal{H}^h}$$

holds for every $\phi \in \mathcal{H}^h$ and $h \ge 0$. As in the proof of Theorem 10, we can find a $\alpha = \alpha(p, \rho, K) > 0$ and a positive constant $c_1 = c_1\,(p, q, \rho, M)$ such that we have the uniform tail estimate

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

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

$$P(N_{\alpha,[0,T]}(\hat{\mathbf{X}}^h) > u) \le \exp\left\{-c_1 \alpha^{2/p} u^{2/q}\right\} \tag{8}$$

hold for all $u > 0$ and $h \ge 0$ for a possibly smaller $\alpha > 0$. Now we use Corollary 4 and the Cauchy-Schwarz inequality to see that

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

for a constant c_2. The uniform tail estimates (8) show that

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

Using [FR, Theorem 6] gives

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

for a constant c_4 which finishes the proof. □

2.3. **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

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

for $s<t$ and

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

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

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

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

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

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. *Let $X\colon[0,T]\to\mathbb{R}^d$ be as in Theorem 10 with covariance of finite ρ-variation, $\rho<2$. Choose $N\geq\lfloor 2\rho\rfloor$ and*

$$\eta_1<\frac{1}{\rho}-\frac{1}{2},\quad \eta_2<\frac{1}{2\rho}\quad\text{and}\quad \eta_3<\frac{N+1}{2\rho}-1.$$

Assume that $|V|_{Lip^\theta}\leq\nu<\infty$ for some $\theta\in(0,\infty]$ chosen such that $\theta>\frac{2\rho}{1-2\rho\eta_1}$ and $\theta\geq N$. Set

$$D=\{0<h<2h<\ldots<(\lfloor T/h\rfloor-1)h<T\}$$

for $h>0$. Then for any $r\geq 1$,

$$\left|\left|Y-Y^{simple\ Euler;D}\right|_{\infty}\right|_{L^r}\lesssim h^{\eta_1}+h^{\eta_2}+h^{\eta_3}$$

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, the rate η_2 comes from the (almost) $\frac{1}{2\rho}$-Hölder-regulartiy of the sample paths of Y and η_3 comes from the step-N Euler approximation. Since $\rho \geq 1$, the Wong-Zakai error always dominates η_2. In particular, for $\rho = 1$ we can choose $N = 2$ to obtain a rate arbitrary close to $\frac{1}{2}$. For $\rho > 1$, the choice $N = 3$ gives a rate of almost $\frac{1}{\rho} - \frac{1}{2}$. In both cases the rate does not increase for larger choices of N.*

From this remark, we immediately obtain

Corollary 17. *Assume that the vector fields $V = (V_1, \dots, V_d)$ are bounded, C^∞ with bounded derivatives and that the covariance of X has finite mixed $(1, \rho)$-variation for some $\rho < 2$. Then the simlified step-3 Euler scheme (step-2 in the case $\rho = 1$) converges in L^r, for any $r \geq 1$, and rate $\frac{1}{\rho} - \frac{1}{2} - \delta$, for any $\delta > 0$, to the solution of the corresponding rough differential equation.*

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 $Y^h \colon [0,T] \to \mathbb{R}^e$ 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

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

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

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

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 $b = \frac{t_{k+1} - t}{t_{k+1} - t_k}$. Then $a + b = 1$ and by the triangle in equality,

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

for any $p > 2\rho$. Since the right hand side does not depend on t, we can pass to the sup-norm on the left hand side. We now take the L^r-norm on both sides and check that the conditions of Theorem 13 and 10 are fulfilled and that the constants can be chosen independently from h. Since we are dealing with piecewise linear approximations, we have

$$\sup_{h>0} V_\rho \left(R_{(X^h, X)}; [0,T]^2 \right) < \infty \quad \text{and} \quad \sup_{t \in [0,T]} E\left[\left| X_t^h - X_t \right|^2 \right] \lesssim h^{1/\rho}$$

(cf. [FV10b, Chapter 15]). Furthermore, for every $\omega \in \Omega$ one has $\left|\omega^h\right|_{p-\text{var}} \leq 3^{1-1/p} \left|\omega\right|_{p-\text{var}}$ and $\left|\omega^h\right|_{1/p-\text{Höl}} \leq 3^{1-1/p} \left|\omega\right|_{1/p-\text{Höl}}$ for every $p \geq 1$ and $h > 0$ (this follows, for instance, from [FV10b,

Theorem 5.23]). This shows that we can apply Theorem 13 to see that for any $\delta > 0$,

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

holds for all $h > 0$. Furthermore, can choose $p' > 2\rho$ such that $\frac{N+1}{p'} - 1 = \eta_3$ and then apply Theorem 10. Since $\left|d_{1/p'\text{-Höl}}(\mathbf{X}^h, \mathbf{X})\right|_{L^r} \to 0$ for $h \to 0$, clearly $\sup_{h>0} \left|\left\|\mathbf{X}^h\right\|_{1/p'\text{-Höl};[0,T]}\right|_{L^r} < \infty$ and the constants on the right hand side can be chosen independently of h. Choosing p such that $\frac{1}{p} = \eta_2$ gives the claim. □

3. 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 2008 in certain respects. On the one hand we relax the requirement $\alpha \geq 1/2$ in Giles 2008 concerning the bias rate, and on the other we keep track of various proportionality constants more carefully. Naturally, in case one is only interested in achieving asymptotically the desired multilevel complexity rate, one may fix the trade-off between the bias and variance of the estimator, and also one doesn't need to distinguish between proportionality constants due to the zero-level and the other ones (cf. Giles 2008). However, it so may happen that the proportionality coefficient of the estimated asymptotic complexity is so large that the "critical accuracy" where the multilevel estimator starts outperforming the standard one is too close to zero. For example, for certain RDEs the multilevel complexity rate still contains a relatively high negative power of ϵ. A too large proportionality constant would then lead to a critical accuracy that cannot be reached within a reasonable computation time by the multilevel estimator and by the standard estimator even more. (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].)

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

$$P = f(Y_\cdot)$$

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 (4), based on a meshsize $h_l = T/(M_0 M^l)$. Recall the basic idea

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

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

$$\widehat{Y}_l = \frac{1}{N_l} \sum_{i=1}^{N_l} \left(\widehat{P}_l^{(i)} - \widehat{P}_{l-1}^{(i)}\right) \tag{9}$$

based on $i = 1, \ldots, N_l$ 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.

3.1. **Giles' complexity theorem revisited.** The following abstract theorem is an extension of [Gil08b] to the case $\alpha < 1/2$, with special consideration of the involved constants.

Theorem 18. *Let $0 < \alpha < 1/2$ and $0 < \beta \le 2\alpha$. In the spirit of Giles, we assume that there are constants c_1, c_2', c_2 and c_3 such that*

(i) $E\left[\widehat{P}_l - P\right] \le c_1 h_l^\alpha$,

(ii) $E\left[\widehat{Y}_0\right] = E\left[\widehat{P}_0\right]$ *and* $E\left[\widehat{Y}_l\right] = E\left[\widehat{P}_l - \widehat{P}_{l-1}\right]$, $l > 0$,

(iii) $\operatorname{var}\left[\widehat{Y}_0\right] \le c_2' N_0^{-1}$ *and* $\operatorname{var}\left[\widehat{Y}_l\right] \le c_2 N_l^{-1} h_l^\beta$ *for* $l \in \mathbb{N}$,[3]

(iv) *the complexity C_l of computing $\widehat{Y}_l$ is bounded by $C_0 \le c_3 N_0 h_0^{-1}$ for $l = 0$ and $C_l \le c_3 N_l (h_l^{-1} + h_{l-1}^{-1})$ for $l \ge 1$.*[4]

Then for every $\varepsilon > 0$, there are choices L and N_l, $0 \le l \le L$, to be given below in (13) *and* (14)*, respectively, and constants c_4 and c_5 given in* (15) *together with* (16) *such that the multilevel estimator $\widehat{Y} = \sum_{l=0}^{L} \widehat{Y}_l$ satisfies the mean square error bound*

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

with complexity bound

$$C \le \mathrm{const}\, \varepsilon^{-\frac{1+2\alpha-\beta}{\alpha}} + o\left(\varepsilon^{-\frac{1+2\alpha-\beta}{\alpha}}\right),$$

where $\mathrm{const} = c_4$ *for* $\beta < 2\alpha$ *and* $\mathrm{const} = c_4 + c_5$ *for* $\beta = 2\alpha$.

Proof. We first ignore the basic requirement of L and N_l being integer values to obtain (almost) optimal real-valued choices for L and N_l. Then we are going to verify the above given bounds for the MSE and the complexity using the smallest integers dominating our real-valued choices. In this proof, we abuse notation by setting $T = T/M_0$, noting that both complexity and MSE only depend on T and M_0 by T/M_0.

The mean-square-error satisfies

$$\begin{aligned}
\mathrm{MSE} &= E\left[\left(\widehat{Y} - E[P]\right)^2\right] \\
&= E\left[\left(\widehat{Y} - E\left[\widehat{Y}\right]\right)^2 + \left(E\left[\widehat{Y}\right] - E[P]\right)^2\right] \\
&= \operatorname{var}\left[\widehat{Y}\right] + \left(E\left[\widehat{P}_L\right] - E[P]\right)^2 \\
&\le \sum_{l=0}^{L} \operatorname{var}\left[\widehat{Y}_l\right] + c_1^2 h_L^{2\alpha} \\
&\le c_2' N_0^{-1} + c_2 T^\beta \sum_{l=1}^{L} N_l^{-1} M^{-l\beta} + c_1^2 h_L^{2\alpha}.
\end{aligned}$$

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

[4]Note that the complexity at the 0-level is proportional to the number of timesteps h_0^{-1}, whereas at higher levels, we need to apply the numerical scheme twice, once for the finer and once for the coarser grid.

Now we need to minimize the total computational work

$$\begin{aligned} C &\leq c_3 N_0 h_0^{-1} + c_3 \sum_{l=1}^{L} N_l \left(h_l^{-1} + h_{l-1}^{-1} \right) \\ &= c_3 T^{-1} \left[N_0 + \frac{M+1}{M} \sum_{l=1}^{L} N_l M^l \right] \end{aligned}$$

under the constraint 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

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

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

$$\begin{aligned} \frac{\partial f}{\partial N_0} &= c_3 T^{-1} - \lambda c_2' N_0^{-2} = 0, \\ \frac{\partial f}{\partial N_l} &= c_3 T^{-1} M^l \frac{M+1}{M} - \lambda c_2 T^{\beta} M^{-l\beta} N_l^{-2} = 0, \end{aligned}$$

implying that

$$N_0 = \sqrt{\lambda} \sqrt{\frac{c_2'}{c_3}} T, \tag{10a}$$

$$N_l = \sqrt{\lambda} \sqrt{\frac{c_2}{c_3}} T^{(1+\beta)/2} \sqrt{\frac{M}{M+1}} M^{-l(1+\beta)/2}, \ 1 \leq l \leq L, \tag{10b}$$

which we insert into the bound for the MSE to obtain

$$\sqrt{\lambda} = \left[\sqrt{c_2' c_3 T^{-\beta}} + \sqrt{c_2 c_3} \sqrt{\frac{M+1}{M}} M^{(1-\beta)/2} \frac{M^{L(1-\beta)/2} - 1}{M^{(1-\beta)/2} - 1} \right] \frac{T^{-(1-\beta)/2}}{\varepsilon^2 - c_1^2 T^{2\alpha} M^{-2\alpha L}}. \tag{11}$$

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

$$\begin{aligned} C(L) &\equiv c_3 T^{-1} \left[\sqrt{\lambda} \sqrt{\frac{c_2'}{c_3}} T + \frac{M+1}{M} \sqrt{c_2 c_3} \sum_{l=1}^{L} M^l \sqrt{\lambda} \sqrt{\frac{c_2}{c_3}} T^{(1+\beta)/2} \sqrt{\frac{M}{M+1}} M^{-l(1+\beta)/2} \right] \\ &= \sqrt{\lambda} T^{-(1-\beta)/2} \left[\sqrt{c_2' c_3 T^{-\beta}} + \sqrt{c_2 c_3} \sqrt{\frac{M+1}{M}} M^{(1-\beta)/2} \frac{M^{L(1-\beta)/2} - 1}{M^{(1-\beta)/2} - 1} \right] \qquad (12) \\ &= \left[\sqrt{c_2' c_3 T^{-\beta}} + \sqrt{c_2 c_3} \sqrt{\frac{M+1}{M}} M^{(1-\beta)/2} \frac{M^{L(1-\beta)/2} - 1}{M^{(1-\beta)/2} - 1} \right]^2 \frac{T^{-(1-\beta)}}{\varepsilon^2 - c_1^2 T^{2\alpha} M^{-2\alpha L}}. \end{aligned}$$

In general, the optimal (but real-valued) choice of L would now be the arg-min of the above function, which we could not determine explicitly in an arbitrary regime. We parametrize the optimal choice

of L by d_1 in

$$L = \left\lceil \frac{\log\left(d_1 c_1 T^\alpha \varepsilon^{-1}\right)}{\alpha \log(M)} \right\rceil. \tag{13}$$

There are three different approaches to the choice of L: Giles chooses $d_1 = 1/2$, which is probably motivated by the considerations in Remark 20 below. If all the constants involved have already been estimated, then one could choose L by numerical minimization of the complexity, or one could provide an asymptotic optimizer L (for $\varepsilon \to 0$). The latter approach has been carried out for the special case $\beta = 1$ in Theorem 21 below, and in this case the optimal L is indeed (almost) of the form (13) with d_1 weakly depending on ε. Moreover, we choose with $\kappa = \frac{1-\beta}{2\alpha}$

(14a)

$$N_0 = \left\lceil \frac{\sqrt{c_2' T^\beta}}{\varepsilon^2 (1 - d_1^{-2})} \left(\sqrt{c_2' T^{-\beta}} + \sqrt{c_2} \sqrt{\frac{M+1}{M}} M^{(1-\beta)/2} \frac{d_1^\kappa c_2^\kappa T^{(1-\beta)/2} \varepsilon^{-\kappa} - 1}{M^{(1-\beta)/2} - 1} \right) \right\rceil,$$

(14b)

$$N_l = \left\lceil \frac{\sqrt{c_2}\sqrt{\frac{M}{M+1}} T^\beta}{\varepsilon^2 (1 - d_1^{-2})} \left(\sqrt{c_2' T^{-\beta}} + \sqrt{c_2} \sqrt{\frac{M+1}{M}} M^{(1-\beta)/2} \frac{d_1^\kappa c_2^\kappa T^{(1-\beta)/2} \varepsilon^{-\kappa} - 1}{M^{(1-\beta)/2} - 1} \right) M^{-l(1+\beta)/2} \right\rceil,$$

$1 \le l \le L$.

By construction, the MSE will be bounded by ε^2 using the choices (13) and (14). As $x \le \lceil x \rceil \le x + 1$ and using the inequalities

$$M^L \le d_1^{1/\alpha} c_1^{1/\alpha} T M \varepsilon^{-1/\alpha},$$
$$M^{L(1-\beta)/2} \le d_1^\kappa c_1^\kappa T^{(1-\beta)/2} M^{(1-\beta)/2} \varepsilon^{-\kappa},$$

together with the shorthand-notations

$$e_1 = \sqrt{c_2' T^{-\beta}} - \sqrt{c_2} \sqrt{\frac{M+1}{M}} \frac{M^{(1-\beta)/2}}{M^{(1-\beta)/2} - 1},$$
$$e_2 = d_1^\kappa c_2^{\kappa + 1/2} \sqrt{\frac{M+1}{M}} \frac{M^{(1-\beta)/2}}{M^{(1-\beta)/2} - 1} T^{(1-\beta)/2}$$

motivated from our choice (14), we arrive after a tedious calculation at

$$\begin{aligned} C &\le c_3 T^{-1} \left[N_0 + \frac{M+1}{M} \sum_{l=1}^{L} N_l M^l \right] \\ &\le c_3 T^{-1} \Bigg[1 + \sqrt{c_2' T^\beta} (1 - d_1^{-2})^{-1} (e_1 + e_2 \varepsilon^{-\kappa}) \varepsilon^{-2} + \\ &\quad + \sqrt{c_2} T^\beta (1 - d_1^{-2})^{-1} \sqrt{\frac{M+1}{M}} \frac{M^{(1-\beta)/2}}{M^{(1-\beta)/2} - 1} \left(d_1^\kappa c_1^\kappa T^{(1-\beta)/2} M^{(1-\beta)/2} \varepsilon^{-\kappa} - 1 \right) (e_1 + e_2 \varepsilon^{-\kappa}) \varepsilon^{-2} + \\ &\quad + \frac{M+1}{M-1} \left(d_1^{1/\alpha} c_1^{1/\alpha} T M \varepsilon^{-1/\alpha} - 1 \right) \Bigg]. \end{aligned}$$

Arranging the terms according to powers of ε and recalling $\kappa = \frac{1-\beta}{2\alpha}$, we get

$$C \leq c_4 \varepsilon^{-2(1+\kappa)} + c_5 \varepsilon^{-1/\alpha} + c_6 \varepsilon^{-(2+\kappa)} + c_7 \varepsilon^{-2} + c_8, \tag{15}$$

where

$$c_4 = c_1^\kappa c_2^{1+\kappa} c_3 \frac{d_1^{2\kappa}}{1 - d_1^{-2}} \frac{M+1}{M} \frac{M^{3(1-\beta)/2}}{\left(M^{(1-\beta)/2} - 1\right)^2}, \tag{16a}$$

$$c_5 = c_1^{1/\alpha} c_3 d_1^{1/\alpha} \frac{M(M+1)}{M-1}, \tag{16b}$$

$$c_6 = (c_1^\kappa + c_2^\kappa)\sqrt{c_2} c_3 \frac{d_1^{2\kappa}}{1 - d_1^{-2}} T^{-(1-\beta)/2} \sqrt{\frac{M+1}{M}} \frac{M^{(1-\beta)/2}}{M^{(1-\beta)/2} - 1} \tag{16c}$$

$$\times \left(\sqrt{c_2' T^{-\beta}} - \sqrt{c_2} \sqrt{\frac{M+1}{M}} \frac{M^{(1-\beta)/2}}{M^{(1-\beta)/2} - 1} \right), \tag{16d}$$

$$c_7 = \frac{c_3 T^{-(1-\beta)}}{1 - d_1^{-2}} e_1^2, \tag{16e}$$

$$c_8 = -2 \frac{c_3 T^{-1}}{M-1}. \tag{16f}$$

We remark that, under the condition that $\beta \leq 2\alpha$, we have $2(1+\kappa) \geq 1/\alpha$ with equality iff $\beta = 2\alpha$. Consequently, $\varepsilon^{-2(1+\kappa)}$ is the dominating term in the complexity-expansion. We further note that the second term in the expression can be either $\varepsilon^{-1/\alpha}$ or $\varepsilon^{-(2+\kappa)}$. □

The leading order coefficients c_4 and c_5 are positive, whereas the sign of c_6 is not clear. In particular, if we do not distinguish between the variance of $\widehat{Y}_0$ (controlled by c_2') and the variances of the differences $\widehat{Y}_l$, $l = 1, \ldots, L$, controlled by c_2, then c_6 will be negative. c_7 is again positive (but often small) and c_8 is negative. Clearly, we could simplify the complexity bound by omitting all terms with negative coefficients in (15). We further note that the leading order terms of the complexity do not depend on T or M_0.

We do not know the rate of weak convergence α of our simplified Milstein scheme, but for Lipschitz functions f, we clearly have

$$|E[f(X_\cdot) - f(Y_\cdot)]| \leq |f|_{\mathrm{Lip}} E\left[|X - Y|_\infty\right],$$

so that the weak rate of convergence is at least as good as the strong rate of convergence, i.e., $\alpha \geq \beta/2$ in the above notation. In fact, if we only impose minimal regularity conditions on f, then it is highly unlikely that we can get anything better than $\alpha = \beta/2$.

Corollary 19. *Under the assumptions of Theorem 18, let us additionally assume that $\alpha = \beta/2$. Then the complexity of the above multi-level algorithm is bounded by*

$$C \leq c_4' \varepsilon^{-1/\alpha} + o\left(\varepsilon^{-1/\alpha}\right),$$

where

$$c_4' = c_3 d_1^{1/\alpha} (M+1) \left[\frac{c_1^{(1-\beta)/\beta} c_2^{1/\beta}}{d_1^2 - 1} \frac{M^{3(1-\beta)/2}}{M\left(M^{(1-\beta)/2} - 1\right)^2} + c_1^{2/\beta} \frac{M}{M-1} \right].$$

The optimal choice of d_1 minimizing c_4' is obtained by

$$d_1 = \sqrt{1 - \frac{(1-\beta)f_1}{2f_2} + \frac{\sqrt{(1-\beta)^2 f_1^2 + 4\beta f_1 f_2}}{2f_2}},$$

with

$$f_1 = c_1^{(1-\beta)/\beta} c_2^{1/\beta} c_3 \frac{M+1}{M} \frac{M^{3(1-\beta)/2}}{\left(M^{(1-\beta)/2}-1\right)^2},$$
$$f_2 = c_1^{2/\beta} c_3 \frac{M(M+1)}{M-1}.$$

Proof. We use $c_4' = c_4 + c_5$ in order to obtain the formula for the constant. Then we consider c_4' as a function of d_1 and get the minimizer as the unique zero of the derivative in $]1, \infty[$, noting that c_4' approaches ∞ on both boundaries of the domain. □

Remark 20. *In the now classical works of Giles on multilevel Monte Carlo, he usually chooses $d_1 = \sqrt{2}$, 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. In many situations, this choice 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 discretization error is $\varepsilon/3$ and the statistical error is $2\varepsilon/3$.

In the above, the choice of d_1 corresponds to the distribution of the total MSE ε^2 between the statistical and the discretization error according to

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

So, depending on the parameters, Corollary 19 shows that the canonical error distribution is not optimal.

As the leading order coefficients c_4' depends only mildly on M, we do not try to find an optimal choice of the parameter M.

The above analysis has also given us new insight into the classical multi-level Monte Carlo algorithm corresponding to the choice $\beta = 1$. Indeed, even in this case an equal distribution of the error tolerance ε among the bias and the statistical error is far from optimal. Indeed, we have

Theorem 21. *For $\beta = 1$ and $\alpha \geq 1/2$, the optimal choice of L is (apart from rounding up) given by*

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

which is of the form (13) *with*

$$d_1 \approx \left(1 + \sqrt{\frac{c_2'}{c_2 T}\frac{M}{M+1}}\alpha \log M + \log \varepsilon^{-1}\right)^{1/2}.$$

Proof. Let us investigate the behaviour of (12) for $\beta \uparrow 1$. For $\beta \uparrow 1$ we obtain

$$C^{\beta=1}(L) = \left[\sqrt{c_2' c_3 T^{-1}} + \sqrt{c_2 c_3}\sqrt{\frac{M+1}{M}}L\right]^2 \frac{1}{\varepsilon^2 - c_1^2 T^{2\alpha} M^{-2\alpha L}}, \tag{17}$$

and we want to minimize this object for L. Let us abbreviate to

$$D(L) := \frac{C(L)}{c_2' c_3 T^{-1}} = (1+aL)^2 \frac{1}{\varepsilon^2 - bM^{-cL}}$$

with obvious definitions for a, b, c. Setting the derivative to zero yields

$$D'(L) = 2a\,(1+aL)\,\frac{1}{\varepsilon^2 - bM^{-cL}} - (1+aL)^2\,\frac{bcM^{-cL}\log M}{\left(\varepsilon^2 - bM^{-cL}\right)^2} = 0$$

i.e.

$$2a\left(\varepsilon^2 - bM^{-cL}\right) = (1+aL)\,bcM^{-cL}\log M$$

$$2a\varepsilon^2 M^{cL} = 2ab + bc\log M + abcL\log M$$

abbreviate again

$$\varepsilon^2 M^{cL} = b + \frac{bc}{2a}\log M + \frac{bc}{2}L\log M =: p + qL$$

and write the latter as

$$\frac{M^{cL}}{p+qL} = \frac{1}{\varepsilon^2} \tag{18}$$

We now derive an asymptotic expansion for the solution $L(\varepsilon)$ to (18) for $\varepsilon \downarrow 0$. For this we take the logarithm of (18) to obtain with $y := \log \varepsilon^{-2}$ (with $y \to \infty$ as $\varepsilon \downarrow 0$).

$$L = \frac{y}{c\log M} + \frac{\log(p+qL)}{c\log M}. \tag{19}$$

By reformulating (19) as

$$\begin{aligned} L &= \frac{y}{c\log M} + \frac{1}{c\log M}\left(\log L + \log q + \log\left(\frac{p}{qL}+1\right)\right) \\ &= \frac{y}{c\log M} + \frac{\log q}{c\log M} + \frac{\log L}{c\log M} + O(L^{-1}), \quad y\to\infty, \end{aligned} \tag{20}$$

and then writing (20) as

$$L = \frac{\frac{y}{c\log M} + \frac{\log q}{c\log M} + O(L^{-1})}{1 - \frac{\log L}{Lc\log M}}$$

we easily observe that $L = O(y)$ as $y \to \infty$. We thus get by iterating (19),

$$L = \frac{y}{c\log M} + O(\log y),$$

and iterating once again,

$$\begin{aligned} L &= \frac{y}{c\log M} + \frac{\log\left(p + \frac{qy}{c\log M} + O(\log y)\right)}{c\log M} \\ &= \frac{y}{c\log M} + \frac{\log\left(p + \frac{qy}{c\log M}\right)}{c\log M} + O(\frac{\log y}{y}). \end{aligned}$$

The next few iterations yield

$$L = \frac{y}{c\log M} + \frac{\log\left(p + \frac{qy}{c\log M} + \frac{q\log\left(p+\frac{qy}{c\log M}\right)}{c\log M}\right)}{c\log M} + O(\frac{\log y}{y^2}), \quad \text{and}$$

$$L = \frac{y}{c\log M} + \frac{\log\left(p + \frac{qy}{c\log M} + \frac{q\log\left(p+\frac{qy}{c\log M} + \frac{q\log\left(p+\frac{qy}{c\log M}\right)}{c\log M}\right)}{c\log M}\right)}{c\log M} + O(\frac{\log y}{y^3}),$$

etc. After re-expressing the asymptotic solution in the origonal terms via

$$\begin{aligned} a &= \sqrt{\frac{c_2}{c_2'}\frac{M+1}{M}T} \\ b &= c_1^2 T^{2\alpha} \end{aligned}$$

$$\begin{aligned} p &= c_1^2 T^{2\alpha}\left(1 + \sqrt{\frac{c_2'}{c_2 T}\frac{M}{M+1}}\alpha\log M\right) \\ q &= c_1^2\alpha T^{2\alpha}\log M \end{aligned}$$

we gather, respectively,

$$\begin{aligned} L(\varepsilon) &= \frac{\log\varepsilon^{-1}}{\alpha\log M} + O(\log\log\varepsilon^{-1}), \\ L(\varepsilon) &= \frac{1}{\alpha\log M}\log\varepsilon^{-1} \\ &\quad + \frac{1}{2\alpha\log M}\log\left(c_1^2 T^{2\alpha}\left(1 + \sqrt{\frac{c_2'}{c_2 T}\frac{M}{M+1}}\alpha\log M\right) + c_1^2 T^{2\alpha}\log\varepsilon^{-1}\right) \\ &\quad + O(\frac{\log\log\varepsilon^{-1}}{\log\varepsilon^{-1}}) \end{aligned} \tag{21}$$

etc. The resulting complexity (12) will be obviously

$$C^{\beta=1}(L) = O(\frac{\log^2\varepsilon^{-1}}{\varepsilon^2}),$$

where sharper expressions can be obtained by inserting one of the above asymptotic expansions for $L(\varepsilon)$. Note that (21) may be written as

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

which suggest that in (13),

$$d_1 \approx \left(1 + \sqrt{\frac{c_2'}{c_2 T} \frac{M}{M+1}} \alpha \log M + \log \varepsilon^{-1} \right)^{1/2}$$

for the case $\beta = 1$. □

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

TABLE 1. Comparison of asymptotic complexity for multilevel and standard Monte Carlo in the framework of Theorem 18. α denotes the weak order of convergence, $\beta/2$ the strong order. We distinguish the cases $\alpha = \beta/2$ and $\alpha = \beta$.

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.

3.2. **Multilevel Monte Carlo for RDEs.** Let $X\colon [0,T] \to \mathbb{R}^d$ be Gaussian with the same assumptions as in Theorem 10 for some $\rho \in [1,2)$. Consider the solution $Y\colon [0,T] \to \mathbb{R}^m$ 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 $|V|_{\mathrm{Lip}^\gamma} < \infty$ for some $\gamma \geq \frac{2\rho}{\rho-1}$. 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 $f\colon C([0,T], \mathbb{R}^m) \to \mathbb{R}^n$ be a Lipschitz continuous functional and set $P := f(S)$, $\widehat{P}_l := f(S^{(h_l)})$. We want to calculate the quantities needed in Theorem 18. It suffices to apply the modified complexity theorem with $\alpha = \beta/2$. To wit,

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

and

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

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

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

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

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

Condition (ii), "unbiasedness" is obvious for the estimator (9). Theorem 15 and 19 then imply

Theorem 22. *The Monte Carlo evaluation of a functional of an RDE driven by Gaussian signal, to within a MSE of ε^2, can be achieved with computational complexity*

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

Remark 23. *In the context of an RDE driven by fractional Brownian motion, the condition (iv) of Theorem 18 is not precisely satisfied, as there seems no exact simulation algorithm for increments of a fractional Brownian motion with linear complexity, see* [Die04]. *There are inexact simulation algorithms satisfying (iv), but those would introduce additional error terms in the other conditions. On the other hand, there are exact simulation algorithms with "almost linear" complexity, i.e., linear complexity up to logarithmic terms. Obviously, these logarithmic terms would propagate into the complexity statement of Theorem 18. However, Theorem 22 above remains unchanged, due to the "$\delta > 0$"-term inherited from Corollary 17.*

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 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 the complexity of both standard and multi-level Monte Carlo are probably too bad for many situations.

4. Numerical experiments

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

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

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. We implement the simplified Euler scheme (4), where the increments

	$H = 2/5$		$H = 1/3$	
	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 $\beta/2 = 1/6$.

of the fractional Brownian motion were simulated by Hosking's method, see [Die04].[5] Hosking's method is an exact simulation method, i.e., if fed with truly Gaussian random numbers, it will produce samples from the true 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. There are non-exact methods with linear complexity, and even exact methods with linear complexity up to logarithmic terms, but we prefer Hosking's algorithm due to its simplicity.

Starting at $Y_0 = (1, 0, 0)$, Figure 1 shows the strong and weak convergence of the scheme for $H = 0.4$. More precisely, let $\overline{Y}_1^N$ denote the result of the scheme based on a uniform grid on $[0, 1]$ based on N time-step. Then consider $\overline{Y}_1^{2N}$ based on the increments of the *same fBM*.[6] Then, the lower part of Figure 1 shows the Monte Carlo estimator of $E\left[\left|\overline{Y}_1^N - \overline{Y}_1^{2N}\right|\right]$ 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 1, 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 2 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

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

[6]In 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.

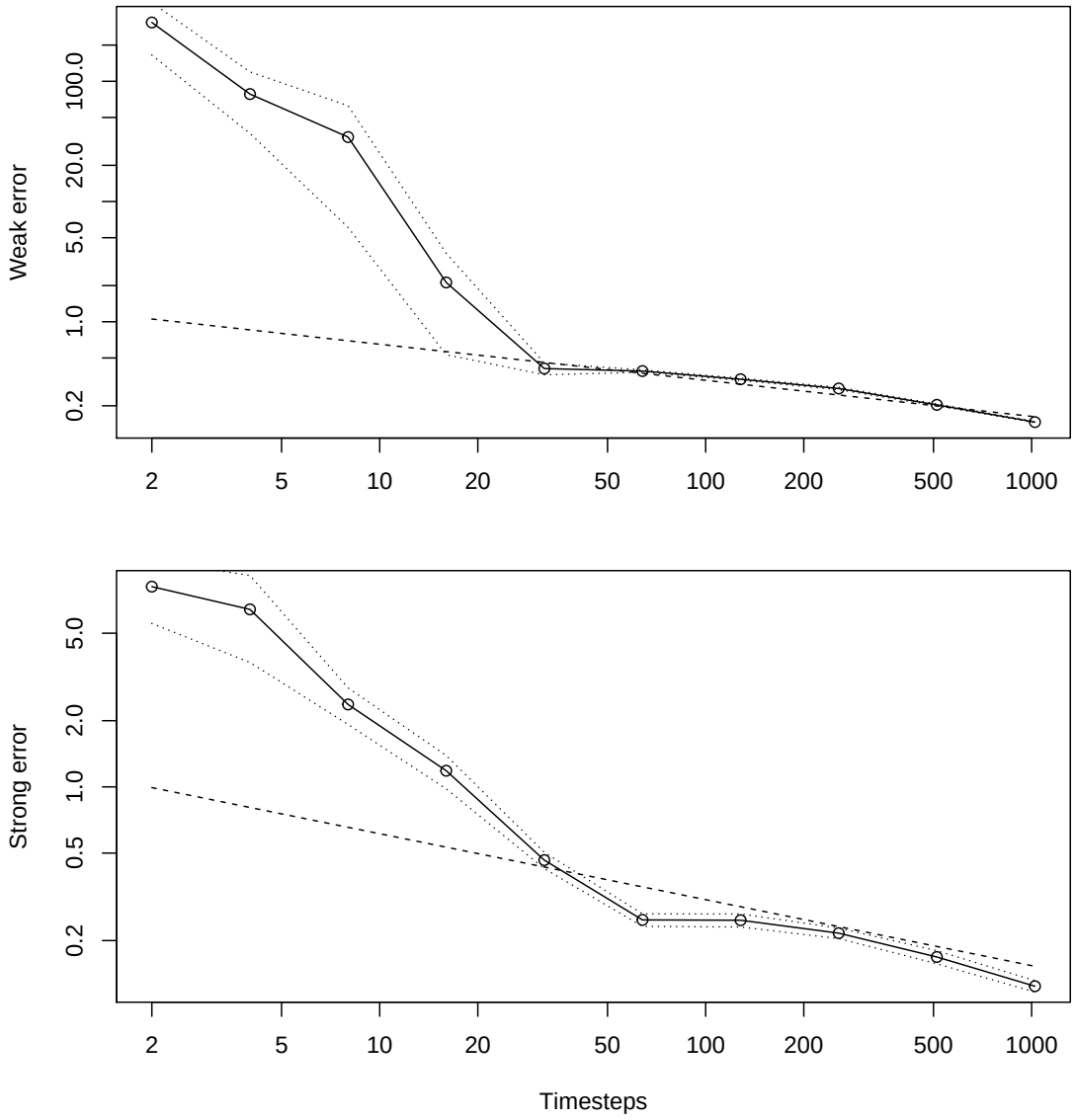

FIGURE 1. Strong and weak error for a fBM with Hurst index $H = 0.4$. Dashed line corresponds to the theoretical strong rate of convergence 0.3, dotted lines show confidence intervals around the error due to the integration error. Weak error corresponds to the functional $f(y) := (|y(1)| - 1)^+$.

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.

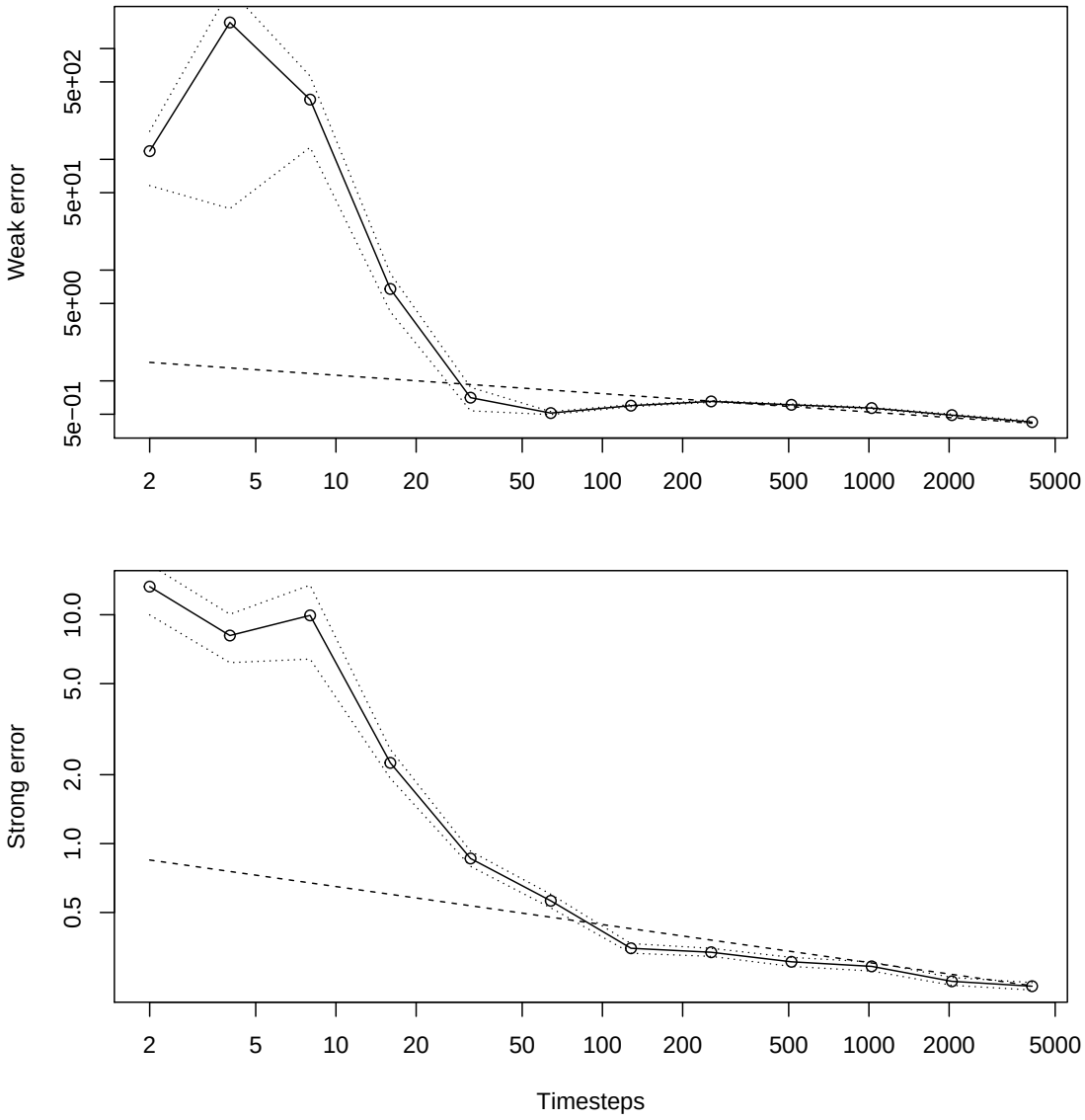

FIGURE 2. Strong and weak error for a fBM with Hurst index $H = 0.33$. Dashed line corresponds to the theoretical strong rate of convergence 0.16, dotted lines show confidence intervals around the error due to the integration error. Weak error corresponds to the functional $f(y) := (|y(1)| - 1)^+$.

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 $\varepsilon \to 0$, renders the standard multi-level construction useless.

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 $f(\overline{Y}_1^N)$ goes to 0 when $N \to \infty$. 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 classical 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.[7]

APPENDIX A. RECALLS ON RDES DRIVEN BY GAUSSIAN SIGNALS

In this section, we introduce the concepts and definitions from rough path theory that are necessary for our current application. For a detailed account of the theory, we refer readers to [FV10b], [LCL07] and [LQ02].

Fix the time interval $[0, T]$. For all $s < t \in [0, T]$, let $\Delta_{s,t}$ denote the simplex

$$\{(u_1, u_2) \mid s \leq u_1 \leq u_2 \leq t\},$$

[7]The 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 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 $N_l = N_0 2^{-l(1+\beta)/2} = N_0 2^{-0.8l}$)

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

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

and we simply write Δ for $\Delta_{0,T}$. In what follows, we will use x to denote an $\mathbb{R}^d$-valued path, and X to denote a stochastic process in $\mathbb{R}^d$, which is a Gaussian process in the current paper. Let (E,d) be a metric space and $x \in C([0,T],E)$. For $p \geq 1$, we define

$$\|x\|_{p-\mathrm{var};[s,t]} := \sup_{D \subset [s,t]} \left(\sum_{t_i, t_{i+1} \in D} |d(x_{t_i}, x_{t_{i+1}})|^p \right)^{\frac{1}{p}} \quad \text{and} \quad \|x\|_{\alpha-\mathrm{Höl};[s,t]} := \sup_{(u,v)\in\Delta_{s,t}} \frac{d(x_u, x_v)}{|v-u|^{\frac{1}{p}}}.$$

We will use the short hand notation $\|\cdot\|_{p-\mathrm{var}}$ and $\|\cdot\|_{\alpha-\mathrm{Höl}}$ for $\|\cdot\|_{p-\mathrm{var};[0,T]}$ resp. $\|\cdot\|_{\alpha-\mathrm{Höl};[0,T]}$ which are easily seen to be semi-norms. Given a positive integer N the truncated tensor algebra of degree N is given by the direct sum

$$T^N\left(\mathbb{R}^d\right) = \mathbb{R} \oplus \mathbb{R}^d \oplus ... \oplus \left(\mathbb{R}^d\right)^{\otimes N}.$$

With tensor product $\otimes$, vector addition and usual scalar multiplication, $T^N\left(\mathbb{R}^d\right) = \left(T^N\left(\mathbb{R}^d\right), \otimes, +, .\right)$ is an algebra. Let π_i denote the canonical projection from $T^N\left(\mathbb{R}^d\right)$ onto $\left(\mathbb{R}^d\right)^{\otimes i}$.

For a path $x : [0,1] \to \mathbb{R}^d$ of bounded variation, we define the canonical lift $\mathbf{x} \equiv S_N(x) : [0,T] \to T^N\left(\mathbb{R}^d\right)$ via iterated (Young) integration,

$$\mathbf{x}_t \equiv S_N(x)_t = 1 + \sum_{i=1}^{N} \int_{0<s_1<...<s_i<t} dx_{s_1} \otimes ... \otimes dx_{s_i}$$

noting that $\mathbf{x}_0 = 1 + 0 + ... + 0 =: e$ is the neutral element for $\otimes$, and that $\mathbf{x}_t$ really takes values in

$$G^N\left(\mathbb{R}^d\right) = \left\{ g \in T^N\left(\mathbb{R}^d\right) : \exists x \in C^{1\text{-var}}\left([0,T], \mathbb{R}^d\right) : \ g = S_N(X)_1 \ \right\},$$

a submanifold of $T^N\left(\mathbb{R}^d\right)$, called the free step-$N$ nilpotent Lie group with d generators. We will use the canonical notion of increments expressed by

$$\mathbf{x}_{s,t} := \mathbf{x}_s^{-1} \otimes \mathbf{x}_t.$$

The dilation operator $\delta : \mathbb{R} \times G^N\left(\mathbb{R}^d\right) \to G^N\left(\mathbb{R}^d\right)$ is defined by

$$\pi_i\left(\delta_\lambda(g)\right) = \lambda^i \pi_i(g), \ \ i = 0, ..., N.$$

The Carnot-Caratheodory norm, given by

$$\|g\| = \inf\left\{ \text{length}(s) \ : \ x \in C^{1\text{-var}}\left([0,1], \mathbb{R}^d\right), S_N(x)_1 = g \right\}$$

defines a continuous norm on $G^N\left(\mathbb{R}^d\right)$, homogeneous with respect to δ. This norm induces a (left-invariant) metric on $G^N\left(\mathbb{R}^d\right)$ known as Carnot-Caratheodory metric,

$$d(g,h) := \left\| g^{-1} \otimes h \right\|.$$

Let $x, y \in C_0\left([0,T], G^N\left(\mathbb{R}^d\right)\right)$, the space of continuous $G^N\left(\mathbb{R}^d\right)$-valued paths started at the neutral element. We define p-variation- and $\frac{1}{p}$-Hölder-distances by

$$d_{p-\mathrm{var}}(\mathbf{x},\mathbf{y}) := \sup_{D\subset[0,T]} \left(\sum_{t_i,t_{i+1}\in D} |d(\mathbf{x}_{t_i,t_{i+1}}, \mathbf{y}_{t_i,t_{i+1}})|^p \right)^{\frac{1}{p}} \quad \text{and}$$

$$d_{\frac{1}{p}-\mathrm{Höl}}(\mathbf{x},\mathbf{y}) := \sup_{(s,t)\in\Delta} \frac{d(\mathbf{x}_{s,t}, \mathbf{y}_{s,t})}{|t-s|^{\frac{1}{p}}}.$$

Note that $d_{\frac{1}{p}-\mathrm{Höl}}(\mathbf{x},0) = \|\mathbf{x}\|_{\alpha-\mathrm{Höl}}$ and $d_{p-\mathrm{var}}(\mathbf{x},0) = \|\mathbf{x}\|_{p-\mathrm{var}}$ where 0 denotes the constant path equal to the neutral element. These metrics are called *homogeneous* rough paths metrics. We define the following path spaces:

(i) $C_0^{p-\mathrm{var}}\left([0,T], G^N\left(\mathbb{R}^d\right)\right)$: the set of continuous functions $\mathbf{x}$ from $[0,T]$ into $G^N\left(\mathbb{R}^d\right)$ such that $\|\mathbf{x}\|_{p-\mathrm{var}} < \infty$ and $\mathbf{x}_0 = e$.

(ii) $C_0^{\alpha-\mathrm{Höl}}\left([0,T], G^N\left(\mathbb{R}^d\right)\right)$: the set of continuous functions $\mathbf{x}$ from $[0,T]$ into $G^N\left(\mathbb{R}^d\right)$ such that $\|\mathbf{x}\|_{\alpha-\mathrm{Höl}} < \infty$ and $\mathbf{x}_0 = e$.

(iii) $C_0^{0,p-\mathrm{var}}\left([0,T], G^N\left(\mathbb{R}^d\right)\right)$: the $d_{p-\mathrm{var}}$-closure of

$$\left\{S_N(x), x : [0,T] \to \mathbb{R}^d \text{ smooth}\right\}.$$

(iv) $C_0^{0,\alpha-\mathrm{Höl}}\left([0,T], G^N\left(\mathbb{R}^d\right)\right)$: the $d_{\alpha-\mathrm{Höl}}$-closure of

$$\left\{S_N(x), x : [0,T] \to \mathbb{R}^d \text{ smooth}\right\}.$$

If $N = \lfloor p \rfloor$, the elements of the spaces (i) and (ii) are called *weak geometric (Hölder) rough paths*, the elements of (iii) and (iv) are called *geometric (Hölder) rough paths*. Recall that if $V = (V_i)_{i=1,\dots,d}$ is a collection of $\mathrm{Lip}^\gamma(\mathbb{R}^e)$ vector fields (in the sense of Stein, cf. [FV10b]) for some $\gamma > p$ and $\mathbf{x}$ is a p-rough path, one can make sense of a unique solution $y\colon [0,T] \to \mathbb{R}^e$ of the equation

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

In this article, we will mainly be interested in *inhomogeneous* rough paths metrics which we aim to define now. First recall that a *control* is a function $\omega\colon \Delta \to \mathbb{R}_+$ which is continuous and super-additive in the sense that for all $s \le u \le t$ one has

$$\omega(s,u) + \omega(u,t) \le \omega(s,t).$$

If ω is a control, we define

$$\begin{aligned}
\|\mathbf{x}\|_{p-\omega;[s,t]} &:= \sup_{s\le u<v\le t} \frac{\|\mathbf{x}_{s,t}\|}{\omega(s,t)^{1/p}} \\
\rho^{(k)}_{p-\omega;[s,t]}(\mathbf{x},\mathbf{y}) &:= \sup_{s\le u<v\le t} \frac{|\pi_k(\mathbf{x}_{u,v} - \mathbf{y}_{u,v})|}{\omega(u,v)^{k/p}} \\
\rho_{p-\omega;[s,t]}(\mathbf{x},\mathbf{y}) &:= \max_{k=1,\dots,\lfloor p\rfloor} \rho^{(k)}_{p-\omega;[s,t]}(\mathbf{x},\mathbf{y}) \\
\rho^{(k)}_{p-\mathrm{var};[s,t]}(\mathbf{x},\mathbf{y}) &:= \sup_{(t_i)\subset[s,t]} \left(\sum_i \left|\pi_k\left(\mathbf{x}_{t_i,t_{i+1}} - \mathbf{y}_{t_i,t_{i+1}}\right)\right|^{p/k}\right)^{k/p} \\
\rho_{p-\mathrm{var};[s,t]}(\mathbf{x},\mathbf{y}) &:= \max_{k=1,\dots,\lfloor p\rfloor} \rho^{(k)}_{p-\mathrm{var};[s,t]}(\mathbf{x},\mathbf{y}).
\end{aligned}$$

Note that the metrics $d_{p-\mathrm{var}}$ and $\rho_{p-\mathrm{var}}$ both induce the same topology on the respective rough paths spaces, as do the metrics $d_{\frac{1}{p}-\mathrm{Höl}}$ and $\rho_{p-\omega;[s,t]}$ with the choice $\omega(s,t) = |t-s|$; cf. [FV10b] for more details.

We now recall the basic facts about Gaussian rough paths. If $I = [a,b]$ is an interval, a *dissection of* I is a finite subset of points of the form $\{a = t_0 < \ldots < t_m = b\}$. The family of all dissections of I is denoted by $\mathcal{D}(I)$. Let $I \subset \mathbb{R}$ be an interval and $A = [a,b]\times[c,d] \subset I \times I$ be a rectangle. If $f\colon I \times I \to V$ is a function, mapping into a normed vector space V, we define the *rectangular increment* $f(A)$ by setting

$$f(A) := f\begin{pmatrix} a,b \\ c,d \end{pmatrix} := f\begin{pmatrix} b \\ d \end{pmatrix} - f\begin{pmatrix} a \\ d \end{pmatrix} - f\begin{pmatrix} b \\ c \end{pmatrix} + f\begin{pmatrix} a \\ c \end{pmatrix}.$$

Definition 24. *Let $p \geq 1$ and $f\colon I \times I \to V$. For $[s,t] \times [u,v] \subset I \times I$, set*

$$V_p(f;[s,t]\times[u,v]) := \left(\sup_{\substack{(t_i)\in\mathcal{D}([s,t]) \\ (t'_j)\in\mathcal{D}([u,v])}} \sum_{t_i,t'_j} \left| f\begin{pmatrix} t_i, t_{i+1} \\ t'_j, t'_{j+1} \end{pmatrix} \right|^p \right)^{\frac{1}{p}}.$$

If $V_p(f, I\times I) < \infty$, we say that f has finite (2D) p-variation. Similarly, we set

$$V_{1,p}(f;[s,t]\times[u,v]) := \left(\sup_{\substack{(t_i)\in\mathcal{D}([s,t]) \\ (t'_j)\in\mathcal{D}([u,v])}} \sum_{t'_j} \left(\sum_{t_i} \left| f\begin{pmatrix} t_i, t_{i+1} \\ t'_j, t'_{j+1} \end{pmatrix} \right| \right)^p \right)^{\frac{1}{p}}.$$

and call this the (mixed, right) $(1,p)$-variation of f.

Let $X = (X^1,\dots,X^d)\colon [0,T] \to \mathbb{R}^d$ be a centered Gaussian process. Then the covariance function $R_X(s,t) := \mathrm{Cov}_X(s,t) = E(X_s \otimes X_t)$ is a map $R_X\colon I\times I \to \mathbb{R}^{d\times d}$. Next, we cite the fundamental existence result about Gaussian rough paths. For a proof, cf. [FV10a] or [FV10b, Chapter 15].

Theorem 25. *Let $X\colon [0,T] \to \mathbb{R}^d$ be a centered Gaussian process with continuous sample paths and independent components. Assume that there is a $\rho \in [1,2)$ such that $V_\rho(R_X;[0,T]^2) < \infty$. Then X admits a lift $\mathbf{X}$ to a process whose sample paths are geometric p-rough paths for any $p > 2\rho$, i.e. with sample paths in $C_0^{0,p-var}([0,T], G^{\lfloor p\rfloor}(\mathbb{R}^d))$ and $\pi_1(\mathbf{X}_{s,t}) = X_t - X_s$ for any $s<t$. $\mathbf{X}$ is a natural lift of X in the sense that if X^h is a suitable approximation of X (cf.* [FV10b, chapter 15] *or section 2.2 for the exact definition), then*

$$\left| d_{p-var}(S_{\lfloor p\rfloor}(X^h), \mathbf{X}) \right|_{L^r} \to 0$$

for $h\to 0$ and all $r \geq 1$. Moreover, if $V_\rho(R_X;[s,t]^2) \lesssim |t-s|^{\frac{1}{\rho}}$ holds for all $s<t$, then the sample paths of X can be lifted to $\frac{1}{p}$-Hölder rough paths and the L^r-convergence holds for the $d_{\frac{1}{p}-Höl}$ metric.

Remark 26. *The condition $V_\rho(R_X;[s,t]^2) \lesssim |t-s|^{\frac{1}{\rho}}$ can be checked for many Gaussian processes. It holds, for instance, for Brownian motion, the Ornstein-Uhlenbeck process and the Brownian bridge with the choice $\rho = 1$. Moreover, it holds for the fractional Brownian motion B^H with $\frac{1}{2\rho} = H$, where H denotes the Hurst parameter, even in the stronger form of mixed $(1,\rho)$-variation, cf.* [FGR13, Theorem 6]. *Theorem 25 implies that B^H has a lift in the sense of Theorem 25 as long as $H > \frac{1}{4}$.*

References

[BSD13] Denis Belomestny, John Schoenmakers, and Fabian Dickmann, *Multilevel dual approach for pricing american style derivatives*, Finance and Stochastics (2013), to appear.

[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.

[CLL] Thomas Cass, Christian Litterer, and Terry J. Lyons, *Integrability estimates for gaussian rough differential equations*, to appear in Ann. of Prob.

[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.

[FGR13] Peter K. Friz, Benjamin Gess, and Sebastian Riedel, *On the Variational Regularity of Cameron–Martin paths*, arXiv:1305.2943 (2013).

[FR] Peter K. Friz and Sebastian Riedel, *Convergence rates for the full Gaussian rough paths*, To appear in Ann. Inst. Henri Poincaré Probab. Stat.

[FR13] ———, *Integrability of (non-)linear rough differential equations and integrals*, Stoch. Anal. Appl. **31** (2013), no. 2, 336–358.

[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.

[GS12] Michael B. Giles and Lukasz Szpruch, *Antithetic multilevel Monte Carlo estimation for multi-dimensional sdes without Lévy area simulation*, arXiv:1202.6283 (2012).

[HP] Martin Hairer and Natesh S. Pillai, *Regularity of laws and ergodicity of hypoelliptic SDEs driven by rough paths*, to appear in Ann. of Prob.

[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.

[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.

[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