Markovian approximations of stochastic Volterra equations with the fractional kernel

CHRISTIAN BAYER and SIMON BRENEIS

Weierstrass Institute, Mohrenstraße 39, Berlin 10117, Germany

(Received 19 October 2021; accepted 7 October 2022; published online 25 November 2022)

We consider rough stochastic volatility models where the variance process satisfies a stochastic Volterra equation with the fractional kernel, as in the rough Bergomi model. In particular, the variance process is therefore not a Markov process or semimartingale, and has quite low Hölder-regularity. In practice, simulating such rough processes thus often results in high computational cost. To remedy this, we study approximations of stochastic Volterra equations using an N-dimensional diffusion process defined as solution to a system of ordinary stochastic differential equation. If the coefficients of the stochastic Volterra equation are Lipschitz continuous, we show that these approximations converge strongly with superpolynomial rate in N. Finally, we apply this approximation to compute the implied volatility smile of a European call option under the rough Bergomi and the rough Heston model.

Keywords: Stochastic Volterra equation; Fractional kernel; Rough volatility model; Markovian approximation; Strong error

2020 Mathematics Subject Classification: 91G60, 60G22, 60H35

1. Introduction

Suppose we are given a stochastic Volterra equation of the form

\[ X_t = x_0 + \int_0^t G(t-s)b(X_s)\,ds + \int_0^t G(t-s)\sigma(X_s)\,dW_s, \]

where \( x_0 \in \mathbb{R}^d, b : \mathbb{R}^d \to \mathbb{R}^d \) and \( \sigma : \mathbb{R}^{d \times d} \to \mathbb{R}^{d \times d} \) are Lipschitz continuous, \( W \) is a \( d \)-dimensional Brownian motion. \( G \) denotes a—for simplicity—one-dimensional kernel assumed to be completely monotone. The kernel \( G \) is often a power law kernel, \( G(t) \sim t^{H-1/2} \). In analogy to fractional Brownian motion (fBM), we call the parameter \( 0 < H < 1 \) Hurst parameter. Indeed, the special case \( \tilde{W}_t := \int_0^t (t-s)^{H-1/2} \, dW_s \) defines the so-called Riemann–Liouville fBM.

Models of this type are used in different applications—we refer to Marie (2013) and Chevillard (2017) for applications in biology and turbulence, respectively. We are mostly motivated by the recent framework of rough stochastic volatility models in finance. In this context, we typically consider an asset price process

\[ dS_t = \sqrt{S_t} \, dZ_t \]

defined in terms a a Brownian motion (BM) \( Z \) and the stochastic variance process \( v \)—we consider dynamics under a risk neutral measure and assume interest rate \( r = 0 \) for simplicity. As observed in Gatheral et al. (2018) (for time series of stock prices) and Bayer et al. (2016) (for option prices), stochastic volatility models can provide excellent fits to market data when the stochastic variance follows a dynamic of stochastic Volterra type with fractional kernel an small \( H \approx 0 \)—suggesting the name ‘rough stochastic volatility’. A popular example of a rough stochastic volatility model is the rough Heston model, see El Euch and Rosenbaum (2018), given by the dynamics

\[ V_t = V_0 + \int_0^t G(t-s)(\theta - \lambda V_s)\,ds + \int_0^t G(t-s)\nu \sqrt{V_s} \, dB_s, \]

Another important example is the rough Bergomi model, see Bayer et al. (2016),

\[ V_t := \xi(t) \exp \left( \eta \sqrt{2H} \int_0^t (t-s)^{H-1/2} \, dB_s - \frac{\eta^2}{2} t^{2H} \right), \]

where \( \xi(t) \) is the forward variance curve. In both cases, the BM \( B \) driving the variance dynamics is correlated with the BM \( Z \) driving \( S \).
While very useful from a modeling point of view, the fact that rough stochastic volatility processes—or, more generally, stochastic Volterra processes—are no Markov processes leads to serious theoretical and numerical challenges. This motivates the use of Markovian approximations to the stochastic Volterra processes. A simple and flexible approach goes back to Carmona and Coutin (1998). It is well-known that any completely monotone kernel \( G \) allows for a representation in terms of a Laplace transform, i.e. there is a measure \( \mu \) on the positive half-line such that

\[
G(t) = \int_0^\infty e^{-tx} \mu(dx), \quad t > 0.
\]

Hence, the stochastic Fubini theorem implies that

\[
\tilde{W}_t := \int_0^t G(t-s) \, dW_s = \int_0^\infty \int_0^t e^{-\alpha(t-s)} \, dW_s \, \mu(dx)
= \int_0^t Y^\alpha_t \mu(dx), \quad Y^\alpha_t := \int_0^t e^{-\alpha(t-s)} \, dW_s,
\]

where \( Y^\alpha \) is an Ornstein–Uhlenbeck process, and, hence, a Markov process. Indeed, the infinite-dimensional process \( V \) is a Markov process, see Carmona and Coutin (1998).

In what follows, we will mostly restrict ourselves to the fractional kernel given by

\[
G(t) = \frac{t^{H-1/2}}{\Gamma(H+1/2)},
\]

where we will assume that \( H \in (0, 1/2) \). In this case, we have

\[
G(t) = c_H \int_0^\infty e^{-\alpha t} \alpha^{-H-1/2} \, d\alpha,
\]

\[
c_H := \frac{1}{\Gamma(H+1/2)\Gamma(1/2-H)},
\]

i.e. \( \mu(dx) = w(x) \, dx \) with \( w(x) := c_H x^{-H-1/2}, x > 0 \). Approximating this integral by a finite sum, i.e. \( G \approx \hat{G} \) with

\[
\hat{G}(t) := \sum_{i=1}^N w_i e^{-\alpha_i t},
\]

we get

\[
\hat{X}_t = x_0 + \int_0^t \hat{G}(t-s) b(\hat{X}_s) \, ds + \int_0^t \hat{G}(t-s) \sigma(\hat{X}_s) \, dW_s.
\]

Such approximations were already considered by, e.g. Abi Jaber and El Euch (2019b), Alfonsi and Kebaier (2021), and, for the particular case of fractional Brownian motion, by Harms (2019). Indeed, we have the following proposition, which shows that (5) reduces to an \( N \)-dimensional ordinary stochastic differential equation, instead of a stochastic Volterra equation.

**Proposition 1.1** (Alfonsi and Kebaier 2021, Proposition 2.1)

*Let \( x_0^1, \ldots, x_0^N \in \mathbb{R}^d \) be such that \( \sum_{i=1}^N w_i x_0^i = x_0 \). Then the solution to (5) is given by \( \sum_{i=1}^N w_i X^i_t \), where \( (X^1_t, \ldots, X^N_t) \) is the solution to the \((N \times d)\)-dimensional SDE defined by

\[
X^i_t = x^i_0 - \int_0^t x^i_s \, ds + \int_0^t \sigma(X^i_s) \, dW^i_s
+ \int_0^t b(X^i_s) \, ds,
\]

where \( i = 1, \ldots, N \).

Perhaps unsurprisingly, the approximation error between \( X \) and the Markovian approximation \( \hat{X} \) is controlled by the error between \( G \) and \( \hat{G} \), more precisely by the \( L^2 \)-error. The following bound is a slight modification of Alfonsi and Kebaier (2021, Proposition 3.2).

**Proposition 1.2** (Alfonsi and Kebaier 2021, Proposition 3.2)

*For every \( T > 0 \), there exists a constant \( C \) (depending on \( T, \|x_0\|, \|b\|, \|\sigma\|_2 \)) such that

\[
E|X_T - \hat{X}_T|^2 \leq C \int_0^T |G(t) - \hat{G}(t)|^2 \, dt.
\]

Using this bound, Alfonsi and Kebaier in Alfonsi and Kebaier (2021, Corollary 3.1) constructed a point set \( (x_j)_{j=1,\ldots,N} \) with weights \( (w_j)_{j=1,\ldots,N} \) by cutting off the integral in (3) at some large value \( K \), and then using the midpoint rule on linearly spaced intervals on \([0, K] \). They were able to prove a strong rate of convergence of \( N^{-2H/3} \). In the same paper, they were able to improve the rate of convergence almost to \( N^{-H} \) using some more sophisticated constructions. This is a very slow convergence rate, in particular as \( H \) is often found to be close to 0 in empirical studies, see Bayer et al. (2016). For instance, if \( H = 0.1 \), increasing the accuracy of the computation by one more significant digit would require increasing \( N \) by a factor \( 10^{10} \). The aim of this paper is to considerably improve the rate of convergence by using point sets adapted to the problem at hand.

In this we are motivated by Harms (2019). Rather than linearly spaced intervals, Harms used a geometric spacing. Additionally, he used Gaussian quadrature rules of arbitrary level \( m \) instead of the midpoint rule for the approximation of the integral in (3). Using this approach, Harms managed to prove a rate of convergence of almost \( n^{-2Hm/3} \), where \( n = N/m \) is the number of intervals used (see Harms 2019, Theorem 1). While this yields a convergence of arbitrary order, it is still apparent that \( m \) has to be chosen quite large to get a suitably large order if \( H \) is small. This can of course pose additional problems, as the constants in his bound likely grow in \( m \). It should also be noted that Harms did not give a suggestion on which \( m \) to use given a choice of \( N \), and also, that he chose a quite different approach and only proved this rate of convergence for the particular case of fractional Brownian motion. We also remark that, to the best of our knowledge, Harms was the first who considered a geometric spacing of the intervals for the approximation of the kernel.

We will combine the ideas of these papers by using the estimate of Proposition 1.2 by Alfonsi and Kebaier, and a modified version of the point set used by Harms. We obtain an error bound of the form \( \exp(-c\sqrt{N}) \). For more details on the rate of convergence together with an explicit choice of the points \( x_i \).
and weights \( w_i \) see Theorem 2.1. A further improvement (in terms of constants) is given in Theorem 2.15. These results are then applied to the rough Bergomi model in Section 3.2—in the context of simulation methods—and the rough Heston model in Section 3.3—regarding the characteristic function and Fourier pricing methods.

Numerical experiments using the various suggested point sets for the simulation of fractional Brownian motion are conducted in Section 4. Here, these approximations are also compared to the point sets suggested by Alfonsi and Kebaier (2021) and Harms (2019). Further improvements in the choices of points \( x_i \) and weights \( w_i \) are discussed in Section 4.2 based on empirical optimizations of constants. Finally, the point set suggested in Section 4.2 is applied for the approximation of the implied volatility smile under the rough Bergomi model in Section 4.3 and for the implied volatility smile under the rough Heston model in Section 4.4. For both models, we also compare our approximation to already existing ones.

It should be noted that similar results can be obtained for other completely monotone kernels \( G \) corresponding to other weight functions \( w \). Indeed, the rates of convergence observed in this paper are determined by the rate of explosion of \( w \) at 0 as well as the rate of decay at \( \infty \), as well as its regularity on the full domain. An abstract result of this form is provided in Theorem 2.11 below. However, for the delicate task of finding good point sets and weights, we prefer to work with the specific class of fractional kernel.

2. Main results

In this paper, we will write \( a \approx b \) for \( a \in \mathbb{N} \) and \( b \in \mathbb{R} \) if \( |a - b| < 1 \), i.e. if \( a \) can be obtained from \( b \) by rounding to one of the two nearest integers. This notation will be used in various error estimation or optimization procedures, when integer-valued variables will be relaxed to real ones for the full domain. An abstract result of this form is provided in Theorem 2.11 below. However, for the delicate task of finding good point sets and weights, we prefer to work with the specific class of fractional kernel.

2.1. Superpolynomial rate of convergence

We now describe the nodes and weights used in our approximation of the kernel \( G \). They essentially follow a Gaussian quadrature rule. A reminder of Gaussian quadrature is given in Appendix 1.

Let \( N \in \mathbb{N} \) the total number of nodes and \( \alpha, \beta, a, b \in (0, \infty) \) be parameters of the scheme. Together with \( A = A_H := \left( \frac{1}{2} + \frac{1}{\sqrt{2H}} \right)^{1/2} \), where \( 0 < H < 1/2 \) denotes the Hurst index of the kernel \( G \), we define

\[
m := \frac{\beta}{A} \sqrt{N}, \quad n := \frac{A}{\beta} \sqrt{N} \quad (\approx N/m),
\]

\[
\xi_0 := a \exp \left( \frac{-\alpha}{(3/2 - H)A} \sqrt{N} \right), \quad \xi_\ell := b \exp \left( \frac{\alpha}{HA} \sqrt{N} \right),
\]

\[
\xi_i := \xi_0 \left( \frac{\xi_n}{\xi_0} \right)^{i/n}, \quad i = 0, \ldots, n.
\]

Then, we define the Gaussian rule of type \((H, N, \alpha, \beta, a, b)\) to be the set of nodes \((x_i)_{i=1}^n\) with weights \((w_i)_{i=1}^n\) of the Gaussian quadrature rule of level \( m \) applied to the intervals \([\xi_i, \xi_{i+1}]\) for \( i = 0, \ldots, n - 1 \) with the weight function

\[
w(x) := c_H x^{-H/2}, \quad c_H := \frac{1}{\Gamma(H + 1/2)\Gamma(1/2 - H)}. \tag{7}
\]

In addition, we set \( x_0 := 0 \) and

\[
w_0 := c_H \int_0^{\xi_0} x^{-H/2} \, dx = \frac{c_H}{1/2 - H} \xi_0^{1/2 - H}.
\]

Whether in practice we round up or down in the definition of \( m \) depends on the context and will always be stated explicitly. The value of \( n \) is then chosen such that \( nm \) is as close as possible to \( N \). The approximation of \( G \) is then given by

\[
\hat{G}(t) := \sum_{i=0}^{nm} w_i e^{-x_i t}.
\]

Theorem 2.1 Let \( x_0 \in \mathbb{R}^d \), and let \( b : \mathbb{R}^d \to \mathbb{R} \) and \( \sigma : \mathbb{R}^d \to \mathbb{R}^d \) be globally Lipschitz continuous functions. Let \( X \) be the solution of (1), and let \( \hat{X} \) be the solution of (5), where we use the Gaussian rule of type \((H, N, \alpha, \beta, 1, 1)\) with \( \alpha := 1.06418 \) and \( \beta := 0.4275 \). Then,

\[
\mathbb{E}[X_T - \hat{X}_T]^2 \leq C_H \left( \left( \frac{T^3}{(3/2 - H)^2} + \frac{3}{2H^2} \right) + \frac{5\pi^3}{48} \left( \frac{\alpha^2}{\beta^2} - 1 \right) \frac{A^2 - 2H + 2 - 2H}{\beta^2 - 2H} - H \right)^{1/2} \times \exp \left( \frac{2\alpha}{A} \sqrt{N} \right),
\]

where \( C \) is the constant from Proposition 1.2.

Note that we have \( \alpha/A \sim \alpha \sqrt{H} \) as \( H \to 0 \). For \( H = 0.1 \), \( \alpha/A \approx 0.3251 \).

Remark 2.2 Theorem 2.1 is not only true for the specific values of \( \alpha \) and \( \beta \) given above, but for an entire set \( \Gamma \subset \mathbb{R}_+^2 \) of pairs \((\alpha, \beta)\). The theorem thus tells us that in particular, \((1.06418, 0.4275) \in \Gamma \). While this parameter choice is the best for which we were able to show the theorem, we will observe in Section 4.2 that in practice we may achieve even better results with a different choice of \((\alpha, \beta)\).

Remark 2.3 While we apply Gaussian quadrature rules of level \( m \) on the intervals \([\xi_i, \xi_{i+1}]\), we only use a Riemann-type approximation on \([0, \xi_0]\). This has a couple of reasons. First, it simplifies the proof. Second, we are interested in rough volatility models, where one usually needs to jointly simulate the Volterra process and the underlying Brownian motion (for the stock price). As taking \( m = 0 \) corresponds to no mean reversion, adding this node often does not increase the computational cost at all. Third, while using a Gaussian quadrature rule on \([0, \xi_0]\) could potentially improve the error up to

\[
\exp \left( -\alpha \sqrt{H} \sqrt{N} \right),
\]

this is really only a minor asymptotic speedup. Indeed, for \( H = 0.1 \), we have \( \alpha \sqrt{H} \approx 0.3365 \), which is only insignificantly larger than \( \alpha/A \approx 0.3251 \). At the same time, this would
add \( m \) additional nodes that are (compared to the node \( x_0 = 0 \)) non-trivial, increasing the computational cost especially for small \( N \). Numerical experiments indicate that using a Gaussian quadrature rule on \([0, \xi]\) only becomes worthwhile for sufficiently large \( N \), say \( N > 100 \).

Remark 2.4 It is well-known that the weights obtained by Gaussian quadrature are non-negative. This guarantees that the approximated kernel \( \tilde{G} \) is completely monotone, just like the original kernel \( G \). In particular, this ensures the non-negativity of the approximating volatility process for the rough Heston model, see Abi Jaber and El Euch (2019b).

Remark 2.5 The bound in Theorem 2.1 is completely non-asymptotic. For simplicity, we assume that \( m \) and \( n \) are real-valued, rather than integer-valued. One can still take the suggested point set of the theorem by simply rounding \( m \) and \( n \). As the theorem seems to overestimate the quadrature error (see Section 4.1), resulting in artificially small intervals \([\xi, \xi]\), we recommend always rounding \( m \) up to the next integer. (Note that we do not make the same recommendation for the point set suggested in Section 4.2.)

Remark 2.6 We are not sure whether the geometric choice of the quadrature intervals \([\xi, \xi]\) is optimal. While with this choice of intervals some expressions simplify nicely in the proof of the upcoming Lemma 2.9, it might be possible to achieve a better rate of convergence with a smarter choice of intervals.

Remark 2.7 In practice, we actually propose using a weight \( w_0 \) different from the one stated in Theorem 2.1, see Remark 2.14. The weight \( w_0 \) suggested above is only used to simplify the proof of the theorem.

Before proceeding to the proof of Theorem 2.1, we prove some auxiliary lemmas on the errors of Gaussian quadrature. In these proofs, we will make use of results in Appendix 1.

Lemma 2.8 Let \( \tilde{w}_i \) be the weights and \( \tilde{x}_i \) be the nodes of the Gaussian quadrature rule for \( i = 1, \ldots, m \) on the interval \([a, b]\) with respect to the weight function \( w(x) = c_H x^{-H-1/2} \). Then,

\[
\begin{align*}
\left| c_H \int_a^b e^{-tx} x^{-H-1/2} dx - \sum_{i=1}^m \tilde{w}_i e^{-\tilde{x}_i} \right| &\leq \sqrt{\frac{5\pi}{18}} c_H \frac{mH}{2^{m+1}} t^{-1+2H} \left( \frac{b}{a} - 1 \right)^{2m+1}.
\end{align*}
\]

Proof By Lemmas A.1 and A.2,

\[
\begin{align*}
\left| c_H \int_a^b e^{-tx} x^{-H-1/2} dx - \sum_{i=1}^m \tilde{w}_i e^{-\tilde{x}_i} \right| &= \left| \int_a^b \tilde{w}_i e^{-\tilde{x}_i} K_{2m}(x) dx \right| \\
&\leq \tilde{c}_m \int_a^b e^{-\tilde{x}_i} K_{2m}(x) dx \\
&\leq \tilde{c}_m \int_a^b e^{-tx} K_{2m}(x) dx \\
&\leq \tilde{c}_m \int_a^b e^{-tx} \frac{(2\pi)^m}{(2m)!} \left( \frac{b-a}{2} \right)^{2m} \sup_{y \in [a,b]} |B_{2m}(y)| \\
&\leq \frac{5\pi}{18} c_H \frac{mH}{2^{m+1}} t^{-1+2H} \left( \frac{b}{a} - 1 \right)^{2m+1}.
\end{align*}
\]

Note that for even \( s, B_s \) is an even function, implying that \( \sup_{y \in [-1,1]} |B_s(y)| = \sup_{y \in [0,1]} |B_s(y)| \). Hence, we can apply Lemma A.3 to get

\[
\begin{align*}
\left| c_H \int_a^b e^{-tx} x^{-H-1/2} dx - \sum_{i=1}^m \tilde{w}_i e^{-\tilde{x}_i} \right| &\leq \tilde{c}_m \int_a^b e^{-ta} (2\pi)^m \left( \frac{b-a}{2} \right)^{2m} \frac{2\xi(2m)}{(2\pi)^{2m}} c_H a^{-H-1/2} dx.
\end{align*}
\]

Using that \( \xi(2m) \leq \xi(2) = \pi^2/6 \) yields

\[
\begin{align*}
\left| c_H \int_a^b e^{-tx} x^{-H-1/2} dx - \sum_{i=1}^m \tilde{w}_i e^{-\tilde{x}_i} \right| &\leq \frac{\pi^2 c_H}{3} \frac{mH}{2^{m+1}} \frac{2m+1}{2m} \left( \frac{b-a}{2} \right)^{2m+1}.
\end{align*}
\]

Stirling’s formula \( k! \geq \sqrt{2\pi k} (\frac{k}{e})^k \) and the estimate \( e^{-x} \leq (\frac{x}{e})^\eta \), \( x, \eta \geq 0 \) for \( \eta = 2m+1/2-H \) yield

\[
\begin{align*}
\left| c_H \int_a^b e^{-tx} x^{-H-1/2} dx - \sum_{i=1}^m \tilde{w}_i e^{-\tilde{x}_i} \right| &\leq \frac{\pi^2 c_H}{3} \frac{mH}{2^{m+1}} \frac{2m+1}{2m} \left( \frac{b-a}{2} \right)^{2m+1} \\
&\times \left( \frac{2m+1/2-H}{e} \right)^{1/2-H} \left( \frac{b}{a} - 1 \right)^{2m+1}.
\end{align*}
\]

The previous lemma is an error estimate of our integration error on an interval \([\xi, \xi]\). We will now combine the errors over all such intervals.
**Lemma 2.9** In the setting of Theorem 2.1, we have

\[
\left| \int_0^T \left( c_H \int_{\xi_0}^{\xi_n} e^{-x t} x^{-H-1/2} \, dx - \sum_{i=1}^N w_i e^{-\xi_i t} \right)^2 \, dt \right| \leq \frac{5\pi^3}{36} \frac{c_H^2 T^{2H}}{H} \frac{n^2}{m^2H} \left( \frac{1}{2} (\alpha^2 - 1) \right)^{4m+2}.
\]

**Proof** Since we split up the interval \([\xi_0, \xi_n]\) into \(n\) subintervals and apply a Gaussian quadrature rule on each of them, we have with the triangle inequality and Lemma 2.8

\[
\int_0^T \left| c_H \int_{\xi_0}^{\xi_n} e^{-x t} x^{-H-1/2} \, dx - \sum_{i=1}^N w_i e^{-\xi_i t} \right|^2 \, dt \\
\leq \int_0^T \left( \sum_{i=1}^{n-1} \left[ \frac{5\pi^3}{18} \frac{c_H^2 T^{2H}}{2^{2m+1}m^3H} r^{1/2+H} \times \left( \frac{\xi_{i+1}}{\xi_i} - 1 \right)^{2m+1} \right]^2 \right. \, dt.
\]

Recall that

\[
\frac{\xi_{i+1}}{\xi_i} = \left( \frac{\xi_0}{\xi_0} \right)^{1/n} = \exp \left( \frac{\alpha \sqrt{N}}{A n} \left( \frac{1}{H} + \frac{1}{3/2-H} \right) \right) = \exp \left( \frac{\alpha A \sqrt{N}}{n} \right) = \exp \left( \frac{\alpha B A \sqrt{N}}{A \sqrt{N}} \right) = e^{\alpha B}.
\]

Thus,

\[
\int_0^T \left| c_H \int_{\xi_0}^{\xi_n} e^{-x t} x^{-H-1/2} \, dx - \sum_{i=1}^N w_i e^{-\xi_i t} \right|^2 \, dt \\
\leq \int_0^T \left( \sum_{i=1}^{n-1} \left[ \frac{5\pi^3}{18} \frac{c_H^2 T^{2H}}{2^{2m+1}m^3H} r^{1/2+H} (e^{\alpha B} - 1)^{2m+1} \right]^2 \right. \, dt \\
= \int_0^T \left( \left[ \frac{5\pi^3}{18} \frac{c_H^2 T^{2H}}{2^{2m+1}m^3H} r^{1/2+H} (e^{\alpha B} - 1)^{2m+1} \right]^2 \right. \, dt \\
= \frac{5\pi^3}{36} \frac{c_H^2 T^{2H}}{H} \frac{n^2}{m^2H} \left( \frac{1}{2} (e^{\alpha^2} - 1) \right)^{4m+2}.
\]

We now move to the proof of Theorem 2.1.

**Proof of Theorem 2.1.** By Proposition 1.2, we have

\[
E|X_T - \tilde{X}_T|^2 \leq C \int_0^T \left| G(t) - \tilde{G}(t) \right|^2 \, dt.
\]

Hence,

\[
E|X_T - \tilde{X}_T|^2 \\
\leq C \int_0^T \left| c_H \int_{\xi_0}^{\xi_n} e^{-x t} x^{-H-1/2} \, dx - \sum_{i=1}^N w_i e^{-\xi_i t} \right|^2 \, dt \\
\leq 3C \int_0^T \left( \left| c_H \int_{\xi_0}^{\xi_n} e^{-x t} x^{-H-1/2} \, dx - w_0 \right|^2 \\
+ \left| c_H \int_{\xi_0}^{\xi_n} e^{-x t} x^{-H-1/2} \, dx - \sum_{i=1}^N w_i e^{-\xi_i t} \right|^2 \\
+ \left| c_H \int_{\xi_0}^{\xi_n} e^{-x t} x^{-H-1/2} \, dx \right|^2 \right) \, dt.
\]

Let us consider the first summand first. By the choice of \(w_0\), we have

\[
\left| c_H \int_{\xi_0}^{\xi_n} e^{-x t} x^{-H-1/2} \, dx - w_0 \right|^2 = c_H \left| \int_{\xi_0}^{\xi_n} (1 - e^{-\xi t}) x^{-H-1/2} \, dx \right|^2.
\]

Since \(e^r \geq 1 + r\), we have

\[
c_H \int_{\xi_0}^{\xi_n} (1 - e^{-\xi t}) x^{-H-1/2} \, dx \leq c_H \int_{\xi_0}^{\xi_n} t x x^{-H-1/2} \, dx \\
= \frac{c_H}{3/2-H} \int_{\xi_0}^{\xi_n} e^{3/2-H} \, dx \\
= \frac{c_H}{3/2-H} \exp \left( \frac{\alpha}{A} \sqrt{N} \right).
\]

Squaring and integrating gives

\[
\int_0^T \left| c_H \int_{\xi_0}^{\xi_n} e^{-x t} x^{-H-1/2} \, dx - w_0 \right|^2 \, dt \\
\leq \frac{c_H^2 T^3}{3(3/2-H)^2} \exp \left( \frac{2\alpha}{A} \sqrt{N} \right).
\]

Now, consider the last summand. We have

\[
\int_0^T \left( c_H \int_{\xi_0}^{\xi_n} e^{-x t} x^{-H-1/2} \, dx \right)^2 \, dt \\
= c_H^2 \int_0^T \int_{\xi_0}^{\xi_n} \int_{\xi_0}^{\xi_n} e^{-\xi x+y} x^{-H-1/2} y^{-H-1/2} \, dx \, dy \, dt \\
\leq c_H^2 \int_0^T \int_{\xi_0}^{\xi_n} \int_{\xi_0}^{\xi_n} e^{-\xi x+y} x^{-H-1/2} y^{-H-1/2} \, dx \, dy \\
= \frac{c_H^2}{2} \int_0^T \int_{\xi_0}^{\xi_n} \frac{x^{-H-1/2} y^{-H-1/2}}{\sqrt{xy}} \, dx \, dy \\
= \frac{c_H^2}{2} \int_0^T \int_{\xi_0}^{\xi_n} \frac{x^{-H-1/2} y^{-H-1/2}}{\sqrt{xy}} \, dx \, dy \\
= \frac{c_H^2}{2} \exp \left( \frac{2\alpha}{A} \sqrt{N} \right).
\]

Putting these estimates and Lemma 2.9 together, we get

\[
E|X_T - \tilde{X}_T|^2 \leq 3C \left( \frac{c_H^2 T^3}{3(3/2-H)^2} \exp \left( \frac{2\alpha}{A} \sqrt{N} \right) \right).
\]
\[ \alpha \text{ and term in (9), and afterwards, we maximize over the rates of the first term in (9) equal to the rate of the second term in (10). Hence, to maximize the overall rate, we first set } \]
\[ \alpha \text{ in the second term of (9), which is equivalent to the right-hand side of (10). Notice that } \]
\[ \text{we ignore the factor } \]
\[ \text{of the theorem. For these values, we have } \]
\[ \text{optimization problem } \]
\[ \alpha \to \text{max!, subject to } -2 \alpha A \sqrt{N} \]
\[ = \log \left( \frac{1}{2} (e^{\alpha \beta} - 1) \right) \frac{4 \beta A}{\sqrt{N}}, \]  
\[ \text{(11)} \]
\[ \text{where we maximize over } \alpha, \beta > 0. \]
\[ \text{As one can see, this optimization problem is completely 'dimensionless', so the solutions } \alpha, \beta \text{ are actually numbers independent of any parameters, and are given as in the statement of the theorem. For these values, we have } \]
\[ \mathbb{E} |X_T - \bar{X}_T|^2 \leq C \alpha^2 \left( \frac{T^3}{(3/2 - H)^2} + \frac{3}{2H^2} + \frac{5\pi^3}{48} (e^{\alpha \beta} - 1)^2 \right) \frac{A^{2-2H}T^{2H}}{\beta^{2-2H}H} N^{1-H} \exp \left( -\frac{2\alpha}{A} \sqrt{N} \right). \]
\[ \text{This proves the theorem.} \]

2.2. Markovian approximation for general completely monotone kernels

We remark that a proof similar to the one of Theorem 2.1 can be done for more general completely monotone kernels. In Theorem 2.11 we give one such generalization. A rough sketch of the proof and the precise choice of nodes is given in Appendix 2. We consider the following class of kernels.

**Definition 2.10** A completely monotone kernel \( G \) is said to be of type \((\gamma, \delta)\) if
\[ G(t) = \int_0^\infty e^{-\gamma x} w(x) \, dx \]
for some non-negative weight function \( w \) that is continuous on \((0, \infty)\) and satisfies \( w(x) = O(x^{-\gamma}) \), \( x \to 0 \), and \( w(x) = O(x^{-\delta}) \), \( x \to \infty \). Furthermore, we define \( A_{\gamma, \delta} := \left( \frac{1}{1 - \gamma^2} + \frac{1}{\delta^2} \right)^{1/2} \).

**Theorem 2.11** Let \( x_0 \in \mathbb{R}^d \), and let \( b : \mathbb{R}^d \to \mathbb{R} \) and \( \sigma : \mathbb{R}^{d \times d} \to \mathbb{R}^{d \times d} \) be globally Lipschitz continuous functions. Let \( G \) be of type \((\gamma, \delta)\) with \( 1/2 < \delta < 3/2 \). Let \( X \) be the solution of (1), and let \( \tilde{X} \) be the solution of (5) for the choice of nodes and weights in Appendix 2. Then,
\[ \mathbb{E} |X_T - \tilde{X}_T|^2 \leq CN^{1-(3/2-\gamma\delta)} \exp \left( -\frac{2\alpha}{A_{\gamma, \delta}} \sqrt{N} \right), \]
where \( C \) does not depend on \( N \), and \( \alpha \) is chosen as in Theorem 2.1.

2.3. Optimizing over constants

As mentioned in Remark 2.7, we recommend choosing a weight \( w_0 \) different from the one stated in Theorem 2.1. We will now analyze what the optimal choice of \( w_0 \) is. For this, we need the following exact error representation, which follows immediately by expanding the square.

**Proposition 2.12** Consider \( \tilde{G}(t) = \sum_{i=0}^{N} w_i e^{-\alpha_i T} \), where \( x_0 = 0 \) and \( x_i > 0 \) for \( i = 1, \ldots, N \). Then,
\[ \int_0^T (G(t) - \tilde{G}(t))^2 \, dt = \frac{T^{2H}}{2H(H + 1/2)^2} + w_0^2 T^2 + 2w_0 \sum_{i=1}^{N} w_i x_i (1 - e^{-\alpha_i T}) \]
\[ + \sum_{i,j=1}^{N} \frac{w_i w_j}{x_i + x_j} (1 - e^{-2(\alpha_i + \alpha_j) T}) - 2w_0 T^{H+1/2} \]
\[ - \frac{2}{\Gamma(H + 3/2)} \frac{2}{\Gamma(H + 1/2)} \]
\[ \times \sum_{i=1}^{N} \frac{w_i}{x_i} T^{H+1/2} \int_0^T t^{H-1/2} e^{-t} \, dt. \]  
\[ \text{(12)} \]

**Remark 2.13** The expression in (12) is almost explicit, except for the integral on the right hand side. This is however an incomplete gamma function, which can be computed efficiently.

**Remark 2.14** Note that the right hand side of (12) is a quadratic polynomial in \( w_0 \). This polynomial can easily be minimized in closed form. Indeed, we propose using the \( w_0 \) minimizing equation (12) instead of the \( w_0 \) chosen in Theorem 2.1 or Theorem 2.15. This results in a slight improvement in the error.
In Theorem 2.1 we have chosen Gaussian rules of type $(H, N, \alpha, \beta, a, b)$, where we have set $a = b = 1$. Our goal is now to choose different $a$ and $b$ that improve the constants and the polynomial rate in Theorem 2.1. We get Theorem 2.15 below. The values of $a$ and $b$ in this theorem are chosen as the solution to some optimization problem. Since the proof of this theorem is somewhat technical and not very enlightening, we defer it to Appendix 3.

**Theorem 2.15** Let $x_0 \in \mathbb{R}^d$, and let $b : \mathbb{R}^d \to \mathbb{R}$ and $\sigma : \mathbb{R}^{d \times d} \to \mathbb{R}^{d \times d}$ be globally Lipschitz continuous functions. Let $X$ be the solution of (1), and let $\tilde{X}$ be the solution of (5), where we use the Gaussian rule of type $(H, N, \alpha, \beta, a, b)$ with $\alpha := 1.06418$ and $\beta := 0.4275$,

\[
a = T^{-1} \left( \frac{9 - 6H}{2H} \right)^{\frac{\epsilon_0}{\epsilon_0(2H-1)}} \times \left( \frac{5\pi^3}{768} \left( e^{\alpha g} - 1 \right) \left( \frac{A^2 - 2H}{\beta^2 - 2H} \right)^{N^{1-H}} \right)^2,
\]

\[
b = T^{-1} \left( \frac{9 - 6H}{2H} \right)^{\frac{\epsilon_0}{\epsilon_0(2H-1)}} \times \left( \frac{5\pi^3}{1152} \left( e^{\alpha g} - 1 \right) \left( \frac{A^2 - 2H}{\beta^2 - 2H} \right)^{N^{1-H}} \right)^2 \gamma,
\]

where $\gamma := \left( \frac{1}{\epsilon_0(2H-1)} + 6H - 4H^2 \right)^{-1}$ and we assume that $N \geq 2$. Then,

\[
\mathbb{E} \left| X_T - \tilde{X}_T \right|^2 \leq C(N) \exp \left( -\frac{2\alpha}{A} \sqrt{N} \right),
\]

and $C(N)$ can be given explicitly as

\[
C(N) = C_N^2 T^{2H} \left( \frac{1}{2H} + 8 \frac{e^{\alpha g} - 1}{e^{\alpha g} + 1} + \frac{1}{3 - 2H} \right) \times \left( \frac{3}{H} \right)^{\frac{\epsilon_0(2H-1)}{\epsilon_0(2H-1)}} \times \left( \frac{5\pi^3}{384} \right)^{\frac{\epsilon_0}{\epsilon_0(2H-1)}} \left( e^{\alpha g} - 1 \right) \times \left( \frac{A^2 - 2H}{\beta^2 - 2H} \right)^{N^{1-H}} \times \left( \frac{1}{3/2 - H} \right)^{\gamma},
\]

where $C$ is the constant from Proposition 1.2.

**Remark 2.16** The polynomial order of $C(N)$ in $N$ is rather intransparent. Taking the maximum over $H \in (0, 1/2)$, we get that for any $H$ in this interval,

\[
C(N) = O(N^{0.42}).
\]

3. Examples

We apply the kernel approximation to three different use cases: fractional Brownian motion of Riemann–Liouville type, the rough Bergomi model, and finally, the rough Heston model.

### 3.1. The case of fractional Brownian motion

Suppose we want to simulate a modified Riemann–Liouville fBm, i.e., $X_t = \frac{1}{m H} \int_0^t (t - s)^{H-1/2} \, dW_s$. It is not hard to see that the constant $C$ appearing in Proposition 1.2 can be chosen as $C = 1$. Hence, Theorem 2.15 holds with $C = 1$. To get a better feeling for the resulting constants and rates, let us specialize to $H = 0.1$ and $T = 1$. Then we get the following result.

**Corollary 3.1** For $H = 0.1$, consider the approximation with $N + 1$ nodes, one of them being at $x_0 = 0$, suggested above. Write $N = nm$, where $m$ is the level of the Gaussian quadrature rule and $n$ is the number of intervals. Choose

\[
m \approx 0.1306 \sqrt{N}, \quad n = N/m \approx 7.6568 \sqrt{N},
\]

and

\[
\xi_0 = 4.3679 N^{0.1135} e^{-0.2322 \sqrt{N}},
\]

\[
\xi_n = 0.1421 N^{-1.5889} e^{0.2511 \sqrt{N}}.
\]

Then, we have

\[
\mathbb{E} \left| X_T - \tilde{X}_T \right|^2 \leq 33.6483 N^{0.3178} e^{-0.6502 N}.
\]

### 3.2. The case of the rough Bergomi model

The following presentation of the rough Bergomi model is largely inspired by Bayer et al. (2016). The rough Bergomi model is given by the system

\[
S_t = S_0 \exp \left( \int_0^t \sqrt{\rho} \, dW_s \right) + \sqrt{1 - \rho^2} \, dB_s - \frac{1}{2} \int_0^t V_s \, ds,
\]

\[
V_t = V_0 \exp \left( \eta \sqrt{2H} \int_0^t (t - s)^{-H-1/2} \, dW_s - \frac{\eta^2}{2} \int_0^t \right).
\]

The rough Bergomi model was first studied by Bayer, Friz and Gatheral in Bayer et al. (2016), and can be seen as a rough, non-Markovian generalization of the Bergomi model by Bergomi (2005). Here, the Hurst parameter $H$ controls the decay of the term structure of volatility skew for very short expirations and is usually around 0.1, $\rho$ is the correlation between the Brownian motion driving the volatility process $V$ and the stock price $S$, and is usually negative, and the product $\eta \eta \eta$ sets the level of the ATM skew for longer expirations. More information on how to choose these parameters can be found in Bayer et al. (2016).

We can apply the above Markovian approximation of stochastic Volterra processes to the process $V$, or, more precisely, $\log V$. What we really care about, however, is the simulation of the stock price $S$. Assuming w.l.o.g. that $S_0 = 1$, we can rewrite (13) as follows.

\[
S_t = 1 + \int_0^t S_0 \exp(X_s) \, dW_s,
\]
\[ \text{d}X_t = d\left( \frac{1}{2} \log(V_t) \right) \right) \frac{\eta \sqrt{H}}{\sqrt{2}} \int_0^t (t-s)^{H-1/2} \text{d}\tilde{W}_s. \quad (14) \]

Now it is apparent that \( X \) is the solution of a stochastic Volterra process with \( b \equiv 0 \) and \( \sigma \equiv \eta \sqrt{H} / \sqrt{2} \).

We recall the following Lemma by Harms (2019).

**Lemma 3.2 Harms 2019, Lemma 3** Let \( X, \tilde{X}, \tilde{S} : [0, T] \times \Omega \to \mathbb{R} \) be continuous stochastic processes with \( X_0 = \tilde{X}_0 = 0 \) and

\[ S_t = 1 + \int_0^t S_r \exp(X_r) \text{d}W_r, \quad \tilde{S}_t = 1 + \int_0^t \tilde{S}_r \exp(\tilde{X}_r) \text{d}W_r, \]

and let \( f : (0, \infty) \to \mathbb{R} \) be a function such that \( f \circ \exp \) is Lipschitz continuous with Lipschitz constant \( L \). Then,

\[ |E[f(S_T)] - E[f(\tilde{S}_T)]| \leq L(\sqrt{T} + 6) \times \left( \sup_{t \in [0,T]} |\exp(2|X_t|) + \exp(2|\tilde{X}_t|)|^2 \right)^{1/2} \times \left( \int_0^T \mathbb{E} \left| X_r - \tilde{X}_r \right|^2 \, \text{d}r \right)^{1/2}. \]

The following corollary is immediate.

**Corollary 3.3** Let \( f : (0, \infty) \to \mathbb{R} \) be a function such that \( f \circ \exp \) is Lipschitz continuous. Let \( (S, X) \) be the solution of the stochastic Volterra equation (14), and let \( (\tilde{S}, \tilde{X}) \) be the solution to the Markovian approximation using the point sets proposed in Theorem 2.15 with \( N + 1 \) points. Then, for some \( C \) that can be chosen independent of \( N \), we have

\[ |E[f(S_T)] - E[f(\tilde{S}_T)]| \leq C N^{0.21} \exp \left( -\frac{\alpha}{A} \sqrt{N} \right), \]

with \( \alpha \) and \( A \) as in Theorem 2.15.

**Proof** Note that \( X \) is, essentially, fractional Brownian motion, and that \( \tilde{X} \) is an Ornstein-Uhlenbeck process. Hence, exponential moments of \( \sup_t |X_t| \) and \( \sup_t |\tilde{X}_t| \) exist, and we have by Lemma 3.2 that

\[ |E[f(S_T)] - E[f(\tilde{S}_T)]| \leq C_1 \left\| X - \tilde{X} \right\|_{L^2([0,T] \times \Omega)}. \]

Now, by Theorem 2.15, (where we note that the constant \( C(N) \) in the bound of the theorem can be chosen independent of \( t \)),

\[ \left\| X - \tilde{X} \right\|_{L^2([0,T] \times \Omega)} = \int_0^T \mathbb{E} |X_r - \tilde{X}_r|^2 \, \text{d}r \leq \int_0^T C(N) \exp \left( -\frac{2\alpha}{A} \sqrt{N} \right) \, \text{d}r \leq C_2 N^{0.42} \exp \left( -\frac{2\alpha}{A} \sqrt{N} \right). \]

In the last step, we used Remark 2.16. Combining these estimates and taking the square root gives the result. \( \blacksquare \)

**Remark 3.4** It should be noted that neither \( S \) nor \( \tilde{S} \) can be simulated exactly, both require time discretization of the SDE (14) defining \( S \) or \( \tilde{S} \), respectively. So far, a proper weak error analysis of the Euler scheme for \( S \) is missing—but see Bayer et al. (2020) for partial results, indicating weak rate 1/2 + \( H \). On the other hand, \( (\tilde{S}, \tilde{X}) \) is a standard diffusion process. Hence, an Euler discretization of \( \tilde{S} \) —obviously, \( \tilde{X} \) does not need to be discretized further—converges with weak rate 1, for any fixed \( N \). This does not contradict the weak rate 1/2 + \( H \) reported in Bayer et al. (2020), as the constant for the weak error will depend on \( N \) as well as \( H \), and might explode as \( N \to \infty \).

### 3.3. The case of the rough Heston model

The following presentation of the rough Heston model is largely inspired by Abi Jaber and El Euch (2019b). The rough Heston model is given by

\[ \text{d}S_t = S_t \sqrt{V_t} \text{d}W_t, \quad V_t = V_0 + \int_0^t G(t-s) (\theta(s) - \lambda V_s) \, \text{d}s + \int_0^t G(t-s) \nu \sqrt{V_s} \, \text{d}B_s, \quad (15) \]

where \((W, B)\) is a two-dimensional correlated Brownian motion with correlation \( \rho \in [-1, 1] \), and \( G \) is the usual rough kernel for some \( H \in (0, 1/2) \). Moreover, \( \theta(s) \) is a deterministic, continuous function satisfying some conditions, as explained in El Euch and Rosenbaum (2018, Definition 2.1). This time-dependent mean-reversion level is used to fit the forward variance curve \((\mathbb{E}V_t)_{t \in T} \), as explained in Abi Jaber and El Euch (2019b, Section 2). Also in Abi Jaber and El Euch (2019b, Section 2), the existence of a weak non-negative solution to a generalization of (16) with Hölder-regularity \( H - s \) for all \( s > 0 \) is shown.

Note that due to the square root in (16), the rough Heston model does not satisfy the Lipschitz assumptions of our framework. However, \((S, V)\) turns out to be an affine process. More precisely, as was shown in Abi Jaber and El Euch (2019a), El Euch et al. (2019), El Euch and Rosenbaum (2018) and El Euch and Rosenbaum (2019), the characteristic function of \( \log(S_t/S_0) \) can be given in terms of the solution of a fractional Riccati equation, i.e.

\[ \mathbb{E} \exp \left( z \log(S_t/S_0) \right) = \exp \left( \int_0^t F(z, \psi(t-s, z)) g(s) \, \text{d}s \right). \]

Here,

\[ g(t) = V_0 + \int_0^t G(t-s) \theta(s) \, \text{d}s, \]

and \( \psi(., z) \) is the unique continuous solution to the fractional Riccati equation

\[ \psi(t, z) = \int_0^t G(t-s) F(z, \psi(s, z)) \, \text{d}s, \quad t \in [0, T], \quad (17) \]

where

\[ F(z, x) := \frac{1}{2} \left( z^2 - z \right) + (\rho v z - \lambda) x + \frac{\nu^2}{2} x. \]
While equation (17) cannot be solved explicitly, it can be solved numerically using the Adams scheme developed in Diethelm et al. (2002), Diethelm et al. (2004), Diethelm and Freed (1998) and El Euch and Rosenbaum (2019).

Now, we again replace the kernel $G$ by an approximation $\hat{G}$ of the type (4). We denote the solution of the corresponding ordinary stochastic differential equations similar to equations (15) and (16) by $(\hat{S}, \hat{V})$. As was noted in Abi Jaber and El Euch (2019b, Section 4.1), we again have

$$E\exp(z\log(\hat{S}/S_0)) = \exp\left(\int_0^T F(z, \hat{\psi}(t-s, z))\hat{G}(s) \, ds\right).$$

Here,

$$\hat{G}(t) = V_0 + \int_0^t \hat{G}(t-s)\theta(s) \, ds,$$

and $\hat{\psi}(\cdot, z)$ is the unique continuous solution to the Riccati equation

$$\hat{\psi}(t, z) = \int_t^\infty \hat{G}(t-s)F(z, \hat{\psi}(s, z)) \, ds, \quad t \in [0, T].$$

Also in Abi Jaber and El Euch (2019b, Section 4.1), it is illustrated that $\hat{\psi}$ is given as the solution of an $N$-dimensional system of ordinary Riccati equations, which can be solved numerically by usual numerical integrators for ODEs. Abi Jaber and El Euch were able to show the following two results in Abi Jaber and El Euch (2019b).

**Proposition 3.5** Abi Jaber and El Euch (2019b, Theorem 4.1) There exists a constant $C > 0$ such that for all $a \in [0, 1]$, $b \in \mathbb{R}$ and $N \geq 1$, we have

$$\sup_{t \in [0, T]} |\hat{\psi}(t, a + ib) - \psi(t, a + ib)|$$

$$\leq C(1 + b^4) \int_0^T |\hat{G}(s) - G(s)| \, ds.$$

**Proposition 3.6** (Abi Jaber and El Euch 2019b, Proposition 4.3) Let $C(k, T)$ denote the price of the call option in the rough Heston model with maturity $T > 0$ and log-moneyness $k \in \mathbb{R}$. We denote by $\tilde{C}(k, T)$ the price in the Markovian approximation of the rough Heston model. If $|\rho| < 1$, then there exists a constant $c > 0$ such that

$$|C(k, T) - \tilde{C}(k, T)| \leq c N^{0.21} \exp\left(-\frac{\alpha}{A}\sqrt{N}\right).$$

The following two corollaries are immediate. We only prove the first one, as the proof for the second corollary is completely analogous.

**Corollary 3.7** Assume that we take the point set proposed in Theorem 2.15 with $N + 1$ points. Then, there exists a constant $C > 0$ such that for all $a \in [0, 1]$, $b \in \mathbb{R}$ and $N \geq 0$, we have

$$\sup_{t \in [0, T]} |\hat{\psi}(t, a + ib) - \psi(t, a + ib)|$$

$$\leq C(1 + b^4)N^{0.21} \exp\left(-\frac{\alpha}{A}\sqrt{N}\right).$$

**Proof** Using Proposition 3.5, Jensen’s inequality, Theorem 2.15 and Remark 2.16, we get

$$\sup_{t \in [0, T]} |\hat{\psi}(t, a + ib) - \psi(t, a + ib)|$$

$$\leq C_1(1 + b^4) \int_0^T |\hat{G}(s) - G(s)| \, ds$$

$$\leq C_1 \sqrt{T}(1 + b^4) \left(\int_0^T |\hat{G}(s) - G(s)|^2 \, ds\right)^{1/2}$$

$$\leq C_1 \sqrt{T}(1 + b^4) \left(C(N) \exp\left(-\frac{2\alpha}{A}\sqrt{N}\right)\right)^{1/2}$$

$$\leq C(1 + b^4)N^{0.21} \exp\left(-\frac{\alpha}{A}\sqrt{N}\right).$$

**Corollary 3.8** Let $C(k, T)$ denote the price of the call option in the rough Heston model with maturity $T > 0$ and log-moneyness $k \in \mathbb{R}$. We denote by $\tilde{C}(k, T)$ the price in the Markovian approximation of the rough Heston model, where we used $N + 1$ points as suggested by Theorem 2.15. If $|\rho| < 1$, then there exists a constant $c > 0$ that can be chosen independent of $N$, such that

$$|C(k, T) - \tilde{C}(k, T)| \leq c N^{0.21} \exp\left(-\frac{\alpha}{A}\sqrt{N}\right).$$

The main advantage of using the Markovian approximations above instead of the actual rough Heston model is of course that we have to solve an ordinary (multidimensional) Riccati equation instead of a fractional Riccati equation, which can usually be done faster. But using an ordinary Riccati equation instead of a fractional one has the additional advantage that the rate of convergence of the discretization of the Riccati equation is higher. Indeed, it was proved in Li and Tao (2009) that the rate of convergence of the Adams scheme that is usually used for solving the fractional Riccati equation is $O(\Delta t)$, where $\Delta t$ is the step size. At the same time, we can solve the ordinary Riccati equation with a simple predictor corrector method, for example, which has rate of convergence $O(\Delta t^2)$.

4. Numerics

4.1. Optimal choices of $m$, $\xi_0$ and $\xi_n$ for $H = 0.1$

Suppose we take the approach as before to choose $N + 1$ points, one of them at $x_0 = 0$ with the optimal weight given by Remark 2.14, and the other $N$ points are chosen by applying Gaussian quadrature of level $m$ on $n$ intervals that are geometrically spaced on $[\xi_0, \xi_n]$, and such that $mn \approx N$. Theorems 2.1 and 2.15 give us possible choices of $m$, $\xi_0$ and $\xi_n$, but what are the optimal choices? Using the exact error representation given by Proposition 2.12, we can minimize this error over $\xi_0$ and $\xi_n$ for each $m$. More precisely, for each $m = 1, \ldots, 10$, we try to minimize the right-hand side of Proposition 2.12 as a function of $\xi_0$ and $\xi_n$, where for each step in the optimization procedure, we compute the nodes $(x_i)$ and the weights $(w_i)$ corresponding to the current choice of $\xi_0$ and $\xi_n$ according to our definition of Gaussian rules in the beginning of
Section 2.1. One may of course ask why we do not directly optimize over the \((x_i)\) and \((w_i)\). The reason is that the resulting optimization problem would be very high-dimensional, and non-convex. Conversely, using Gaussian rules, we only have to optimize over 2 (or 3 if we include \(m\)) parameters. Although we are only optimizing over two parameters, the problem is still quite delicate, and we may not have gotten the exact global optima. Nonetheless, we expect our results to be very close to the optima. Figure 1 illustrates the optimized \(L^2\)-errors (i.e. where we have taken the square root of the expression in Proposition 2.12) for the choice \(H = 0.1\) and \(T = 1\). We can see how, the larger \(N\) gets, we should also choose larger \(m\) to get the best results.

In table 1 we compare the point sets for various different parameter choices.

Note that the optimal choices of \(- \log \xi_0\) and \(\log \xi_n\) are larger than the choices made in the theorem. This indicates that the theorem is overestimating the quadrature errors, which leads to artificially small intervals \([\xi_i, \xi_{i+1}]\). Hence, as indicated in Remark 2.5, we recommend always rounding \(m\) up when applying Theorem 2.1 or Theorem 2.15, to slightly improve the results. Also, a comparison of the columns ‘Optimal \(\xi_0\) and \(\xi_n\)’ and ‘Optimal \(m\), \(\xi_0\), \(\xi_n\)’ shows that the correct choice of \(m\) is not extremely crucial, as long as it is of the right order of magnitude.

Figure 2 illustrates the errors reported in table 1.

4.2. Learning a better rate of convergence

While Theorem 2.15 produced decent results, figure 2 illustrates further potential improvements. Using optimization to get the optimal values of \(m\), \(\xi_0\) and \(\xi_n\) for any \(N\) and \(H\), and for \(\xi_0\) and \(\xi_n\) as specified in Corollary 3.1, we find that the actual error for these choices, computed using Proposition 2.12.

| \(N\) | \(m\) | \(n\) | \(- \log \xi_0\) | \(\log \xi_n\) | Error | \(\log \xi_0\) | \(\log \xi_n\) | Error |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.62 | 1 | 1.268 | 0.8891 | 0.0393 | 21.394 | 0.098625 | 0.3569 |
| 2 | 1 | 2 | 1.242 | 0.99649 | 0.9690 | 4.19075 | 0.683678 | 0.683678 |
| 3 | 1 | 3 | 1.219 | 0.975001 | 0.948925 | 1.4621 | 9.1800 | 0.52823 | 0.52823 |
| 4 | 1 | 4 | 1.176 | 0.9897 | 0.937328 | 0.7776 | 14.455 | 0.346109 | 0.346109 |
| 5 | 1 | 5 | 1.135 | 0.9503 | 0.8739 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 6 | 1 | 6 | 1.085 | 0.9034 | 0.72598 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 7 | 1 | 7 | 1.037 | 0.8524 | 0.57103 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 8 | 1 | 8 | 0.989 | 0.8024 | 0.46846 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 9 | 1 | 9 | 0.942 | 0.75286 | 0.32859 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 10 | 1 | 10 | 0.896 | 0.7043 | 0.19724 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 11 | 1 | 11 | 0.842 | 0.6567 | 0.06648 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 12 | 1 | 12 | 0.789 | 0.6098 | 0.03671 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 13 | 1 | 13 | 0.738 | 0.5641 | 0.00777 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 14 | 1 | 14 | 0.688 | 0.5193 | 0.0000 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 15 | 1 | 15 | 0.639 | 0.4761 | 0.0000 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 16 | 1 | 16 | 0.591 | 0.4343 | 0.0000 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 17 | 1 | 17 | 0.543 | 0.3943 | 0.0000 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 18 | 1 | 18 | 0.496 | 0.3558 | 0.0000 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 19 | 1 | 19 | 0.449 | 0.3188 | 0.0000 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
| 20 | 1 | 20 | 0.403 | 0.2833 | 0.0000 | 0.5037 | 21.394 | 0.098625 | 0.098625 |
then applying regression, we get the following approximations of the optimal parameter values as a function of $N$.

$$
\xi_0 = 0.65 T^{-1} e^{3H} \exp \left( -\frac{1.8}{(3/2 - HA) N} \sqrt{N} \right),
$$

$$
\xi_n = T^{-1} e^{3H - 0.4} \exp \left( \frac{1.8}{HA} \sqrt{N} \right),
$$

$$
m = \frac{0.9}{A} \sqrt{N}.
$$

We even get a corresponding error estimate, given by

$$
\left( \int_0^T |G(t) - \hat{G}(t)|^2 \right)^{1/2} \approx T^{H} e^{0.065H - 1.1} \exp \left( -\frac{1.8}{A} \sqrt{N} \right)
$$

where ‘$\approx$’ has the usual meaning of ‘is approximately equal to’. Some further explanation on how we achieved these results can be found in Appendix 4.

Note that when applying the values of $\xi_0$, $\xi_n$ and $m$ suggested above, we do not recommend always rounding $m$ up, in contrast to Remark 2.5. Instead, we recommend rounding $m$ to the closest positive integer.

For the case $H = 0.1$ and $T = 1$, we compare the different approximations and estimates in table 2. The original method of Alfonsi and Kebaier and the one taken from Alfonsi and Kebaier (2021, Corollary 3.1), while the improved one is taken from Alfonsi and Kebaier (2021, Table 6, left column), where we remark that both versions exhibit a polynomial convergence rate of $0.8H$. Harms is the method in Harms (2019) with Gaussian quadrature level $m = 1$ and $m = 10$, respectively. Theorem refers to the point set given in Corollary 3.1, Numerical estimates refers to the numerically inferred point set above, and Optimum is the optimized error we achieved before.

Figure 3 plots the errors reported in table 2. Notice how close the errors of using the numerical estimates of the optimal choices of $m$, $\xi_0$ and $\xi_n$ that we got in this section are to the optimal errors.

Finally, in table 3 we give the number of points required to achieve a certain error in the $L^2$-approximation of fractional Brownian motion for different Hurst parameters and time $T = 1$.
Recall the definition of the rough Bergomi model in (13). Here, we replace the integral
\[
\int_0^t (t - s)^{H - 1/2} \, dW_s
\]
in the exponent of the volatility by our Markovian approximation. Also, note that the \( r^H \) in the exponent of the volatility corresponds to the square of the \( L^2 \)-norm of the aforementioned stochastic integral. Hence, we also replace \( r^H \) by the square of the \( L^2 \)-norm of the Markovian approximation. We obtain the system
\[
S_t = S_0 \exp \left( \int_0^t \sqrt{V_s} \left( \rho \, dW_s + \sqrt{1 - \rho^2} \, dB_s \right) \right)
- \frac{1}{2} \int_0^t V_s \, ds ,
\]
\[
V_t = V_0 \exp \left( \eta \sqrt{2H} \Gamma (H + 1/2) \sum_{i=0}^N w_i \int_0^t e^{-\kappa(t-s)} \, dW_s \right)
- 2\eta^2 H \Gamma (H + 1/2) \sum_{i=0}^N \left( \sum_{j=0}^N w_i e^{-\kappa(t-s)} \right) ^2 \, ds .
\]
Now we discretize this system in time. Let \( 0 = t_0 < \cdots < t_m = T \) be a uniform partition with step size \( \Delta t = T/m \). Initialize the \( N + 3 \) components \( \mathbb{S}, \nabla, \) and \( (\nabla^i)_{i=0}^N \) by
\[
\mathbb{S}_0 = S_0, \quad \nabla_0 = V_0, \quad \nabla_0^i = 0.
\]
Here, \( V_j^i \) corresponds to the integral
\[
\int_0^t e^{-\kappa(t-s)} \, dW_s .
\]
If we have already simulated the system for time \( t_j \), we then simulate the Gaussian vector
\[
\left( \int_{t_j}^{t_{j+1}} e^{-\kappa(t,s)} \, dW_s, \ldots, \right.
\]
\[
\int_{t_j}^{t_{j+1}} e^{-\kappa(t,s)} \, dW_s, W_{t_{j+1}} - W_{t_j}, \right)
\]
If the grid was chosen uniformly in time, the distribution of this vector is of course independent of \( j \), and we only have to compute a Cholesky decomposition once. Then, we set
\[
\mathbb{S}_{t_{j+1}} = \mathbb{S}_j \exp \left( \sqrt{\nabla_j} \left( \rho (W_{t_{j+1}} - W_j) + \sqrt{1 - \rho^2} (B_{t_{j+1}} - B_j) \right) - \frac{\Delta t}{2} \nabla_j \right),
\]
\[
\nabla^i_{t_{j+1}} = e^{-\kappa \Delta t} \nabla_j^i + \int_{t_j}^{t_{j+1}} e^{-\kappa(t,s)} \, dW_s,
\]
\[
W_{t_{j+1}} = V_0 \exp \left( \frac{\eta \sqrt{2H} \Gamma (H + 1/2)}{\nabla_j} \sum_{i=0}^N w_i \nabla_{t_{j+1}}^i \right).
\]

| N   | 1   | 2   | 4   | 8   | 16  | 32  | ∞    |
|-----|-----|-----|-----|-----|-----|-----|------|
| Time| 78  | 123 | 230 | 490 | 1318| 5563| 2579 |

**Notes:** Here, \( N = \infty \) refers to the method in Bayer et al. (2016). Here, the final deterministic integral can easily be computed in closed form.

Using the parameters \( T = 0.9, H = 0.07, \eta = 1.9, \rho = -0.9, S_0 = 1, V_0 = 0.235^2 \), and using 2000 time steps and \( 10^6 \) samples for the Monte Carlo estimates, we get the implied volatility smiles shown in figure 4(a). The number of time steps and samples were chosen to ensure that the time discretization and Monte Carlo errors are sufficiently small. In the figure, the dashed black lines are the 95% confidence interval of the volatility smile where the differential equations were only discretized in time, as suggested by Bayer, Friz and Gatheral in Bayer et al. (2016). The confidence intervals of the approximations have similar sizes and shapes, but were not drawn for better interpretability. Note how the smile generated for \( N = 16 \) is already almost indistinguishable to the human eye from the true smile. We recall that by Lemma 3.2, the errors in the smile are directly related to the \( L^2 \)-errors of the approximation of \( K \). Finally, we remark that the parameter choice we used was the one suggested in Section 4.2.

The time in seconds it took to generate these smiles is given in table 4. Here, \( N = \infty \) is the non-Markovian method described in Bayer et al. (2016), which can also be formally attained by taking \( N \to \infty \).

We can see in figure 4(a) that the choice \( N = 16 \) already produced good results. We can now compare our approximation with 16 nodes to the approximations given in Alfonsi and Kebaier (2021) and Harms (2019) with an equal number of nodes. Since the approximations are all of the same nature, and only the choice of nodes and weights differs, keeping the number of nodes constant means keeping the computational cost constant. Hence, in figure 4(b) we see again the volatility smile generated using the non-Markovian approximation from Bayer et al. (2016) including the 95% confidence intervals, as well as the volatility smiles generated using our point set, the point set in Alfonsi and Kebaier (2021, Corollary 3.1), and the point set in Harms (2019) with Gaussian quadrature levels 1 and 8.

### 4.4. Implied volatility smile of the rough Heston model

Finally, we apply our approximation to the rough Heston model, see Section 3.3. We choose the same parameters as in Abi Jaber and El Euch (2019b, Section 4.2), i.e.

\[ \lambda = 0.3, \quad \rho = -0.7, \quad \nu = 0.3, \quad H = 0.1, \]
\[ V_0 = 0.02, \quad \theta = 0.02, \quad T = 1, \quad S_0 = 1. \]
on the interval \([2019b, Section 4.1]\), we have to solve the system of fractional Riccati equations. As explained in Abi Jaber and El Euch (2019), Section 4.1), essentially, this scheme is a predictor-corrector method.

We now explain how we solve the ordinary multidimensional Riccati equations. As explained in Abi Jaber and El Euch (2019b, Section 4.1), we have to solve the system

\[
\tilde{\psi}(t, z) = \sum_{i=1}^{N} w_i \psi_i(t, z),
\]

\[
\partial_t \psi(t, z) = -x_i \psi(t, z) + F(z, \tilde{\psi}(t, z)), \quad \psi(0, z) = 0,
\]

where

\[
F(z, x) = \frac{1}{2} (-z^2 - x) + \lambda (\nu z - 1) x + \frac{\lambda y^2}{2} x^2.
\]

For comparability, we would also like to solve this system using a predictor-corrector method. However, the usual Euler method with the trapezoidal rule will not work here. The reason is that the mean reversion \(\lambda\) is large, making any Euler approximation completely intractable. Hence, we only discretize the second summand in the right-hand side of (18) and keep the first summand as it is.

More precisely, choose a step size \(\Delta\) and set \(t_k = k\Delta\). Given an approximation \(\tilde{\psi}(t_k, z)\) of \(\tilde{\psi}(t, z)\), approximate

\[
\partial_t \psi(t, z) = -x \psi(t, z) + F(z, \tilde{\psi}(t, z))
\]

on the interval \([t_k, t_{k+1}]\) by

\[
\partial_t \psi(t, z) = x \psi(t, z) + F(z, \tilde{\psi}(t_k, z)).
\]

This is now an ordinary, one-dimensional differential equation of the form

\[
\frac{d}{dt} y(t) = ay(t) + b.
\]

The solution to this ODE is

\[
y(t_{k+1}) = \frac{b}{a} \left( e^{a\Delta} - 1 \right) + y(t_k) e^{a\Delta}.
\]

Using this, we apply the following predictor-corrector method. Given \(\tilde{\psi}(t_k, z)\) and \(\hat{\psi}(t_k, z)\), we set

\[
\tilde{\psi}^{t_{k+1}} = F(z, \tilde{\psi}(t_k, z)) x (1 - e^{-a\Delta}) + \tilde{\psi}(t_k, z) e^{-a\Delta},
\]

\[
\hat{\psi}^{t_{k+1}} = \frac{1}{2} \left( \tilde{\psi}(t_k, z) + \sum_{i=0}^{N} w_i \tilde{\psi}^{t_i} \right),
\]

\[
\tilde{\psi}(t_{k+1}, z) = \sum_{i=0}^{N} w_i \hat{\psi}^{t_i}.
\]

Having described our schemes for the solutions of the Riccati equations, we now describe our approximations of the Fourier inversion. To that end, recall the Fourier inversion formula for suitable functions \(f\) and random variables \(X\).

\[
\mathbb{E}[X] = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(u + iR) \tilde{f}(u + iR) du
\]

for some integrability conditions on \(f\) and \(\tilde{f}\). We use this inversion formula with

\[
X = \log(S_T) / \log(S_0), \quad f(x) = (e^x - K)^+. \quad \text{Moreover, we choose the parameter } R = 2. \text{ We then need to approximate the integral in (19). To this end, we somewhat arbitrarily truncate the integral and restrict ourselves to } [-50, 50]. \text{ On this interval, we then use the trapezoidal rule on an equidistant grid with 10000 points. Finally, for}

\[
\text{Figure 4. Implied volatility smiles on the rough Bergomi model for the approximation given in Bayer et al. (2016) and the Markovian approximations. (a) Approximations based on Theorem 2.15 for different } N \text{ and (b) Comparison with other methods from the literature, } N = 16.
\]
both the fractional and the ordinary multidimensional Riccati equations we use 3000 time steps.

Figure 5(a) shows the implied volatility smile of the rough Heston model under the above mentioned parameters. The black line is the non-Markovian scheme as described in El Euch and Rosenbaum (2019), while the other lines use our Markovian approximation for different numbers $N$ of nodes with the parameters given Section 4.2. Note how, even for $N = 6$, our approximation is barely distinguishable from the exact model as approximated by the Adam scheme. We further compare our approximation with $N = 16$ to the approximation given in Abi Jaber and El Euch (2019b) with a varying number of nodes, see figure 5(b).

Since the approximations are all already quite accurate, the lines are barely distinguishable. However, it turns out that our approximation with $N = 6$ (that is, using 7 nodes if one counts the free node at 0) is better than the approximation in Abi Jaber and El Euch (2019b) with 1024 points. Since we were able to prove a superpolynomial rate of convergence in Corollary 3.8, and the point set in Abi Jaber and El Euch (2019b) has only a convergence rate of $N^{-4H/5}$, we expect our point set to do comparatively even better for larger $N$.

Since the smiles are barely distinguishable, we also plot the errors of our method and the one used in Abi Jaber and El Euch (2019b). Here, we simply use the error functional

$$\text{err} = \left( \int_{-0.5}^{0.3} |\sigma(k, T) - \hat{\sigma}(k, T)|^2 \, dk \right)^{1/2},$$

where $\sigma$ and $\hat{\sigma}$ are the implied volatility for the European option with the above mentioned parameters under the rough Heston model and its approximation, respectively. The parameter $k$ refers to the log-moneyness. The errors are shown in figure 6.

The time in seconds it took to generate these smiles for different values of $N$ is given in table 5. Here $N = \infty$ refers to the non-Markovian model. This notation is reasonable, as the Markovian approximations converge to the non-Markovian model as $N \to \infty$. The times for our approximation and the approximation by Abi Jaber and El Euch were essentially indistinguishable.

**Acknowledgments**

Both authors would also like to thank Ahmed Kebaier for the insightful discussions. Both authors are affiliated with the WIAS Berlin.
Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

C.B. and S.B. gratefully acknowledge the support of the DFG (Deutsche Forschungsgemeinschaft) through [grant number IRTG 2544], C.B. additionally acknowledges the support of the DFG (Deutsche Forschungsgemeinschaft) through [grant number FOR 2402].

ORCID

Christian Bayer https://orcid.org/0000-0002-9116-0039
Simon Brenes https://orcid.org/0000-0002-7944-7686

References

Abi Jaber, E. and El Euch, O., Markovian structure of the Volterra Heston model. Stat. Probab. Lett., 2019a, 149, 63–72.
Abi Jaber, E. and El Euch, O., Multifactor approximation of rough volatility models. SIAM J. Financ. Math., 2019b, 10, 309–349.
Alfonsi, A. and Kebaier, A., Approximation of stochastic Volterra equations with kernels of completely monotone type, 2021.
Bayer, C., Friz, P. and Gatheral, J., Pricing under rough volatility. Quant. Finance, 2016, 16, 887–904.
Bayer, C., Hall, E.J. and Tempone, R., Weak error rates for option pricing under the rough Bergomi model. arXiv preprint arXiv:2009.01219, 2020.
Bergomi, L., Smile dynamics II. Risk October, pp. 67–73, 2005.
Brass, H. and Hämmerlin, G., Numerical Integration IV (Oberwolfach, 1992), 1993 (Birkhäuser Verlag: Basel, Boston, Berlin).
Brass, H. and Petras, K., Quadrature Theory: The Theory of Numerical Integration on a Compact Interval, 178, 2011 (American Mathematical Soc: Providence, RI).
Carmona, P. and Coutin, L., Fractional Brownian motion and the Markov property. Electron. Commun. Probabil., 1998, 3, 95–107.
Chevillard, L., Regularized fractional Ornstein–Uhlenbeck processes and their relevance to the modeling of fluid turbulence. Phys. Rev. E., 2017, 96, 033111.
Diethelm, K. and Freed, A.D., The FracPECE subroutine for the numerical solution of differential equations of fractional order. Forschung Und Wissenschaftliches Rechnen, 1998, 1999, 57–71.
Diethelm, K., Ford, N.J. and Freed, A.D., A predictor-corrector approach for the numerical solution of fractional differential equations. Nonlinear Dyn., 2002, 29, 3–22.
Diethelm, K., Ford, N.J. and Freed, A.D., Detailed error analysis for a fractional Adams method. Numer. Algorithms, 2004, 36, 31–52.
El Euch, O. and Rosenbaum, M., Perfect hedging in rough Heston models. Ann. Appl. Probab., 2018, 28, 3813–3856.
El Euch, O. and Rosenbaum, M., The characteristic function of rough Heston models. Math. Finance, 2019, 29, 3–38.
El Euch, O., Gatheral, J. and Rosenbaum, M., Roughening Heston. Risk, 2019, 29, 84–89.
Gatheral, J., Jaisson, T. and Rosenbaum, M., Volatility is rough. Quant. Finance, 2018, 18, 933–949.
Harms, P., Strong convergence rates for Markovian representations of fractional Brownian motion. arXiv preprint arXiv:1902.01471, 2019.
Lehmer, D.H., On the maxima and minima of Bernoulli polynomials. Ann. Math. Mon., 1940, 47, 533–538.
Li, C. and Tao, C., On the fractional Adams method. Comput. Math. Appl., 2009, 58, 1573–1588.
Marie, N., A pathwise fractional one compartment intra-venous bolus model. arXiv preprint arXiv:1311.1647, 2013.

Appendices

Appendix 1. Gaussian quadrature

In this section, we recall some basic facts about Gaussian quadrature. A more extensive treatment can be found in many books on numerical integration, e.g. in Brass and Petras (2011).

Let $[a, b]$ be a finite, non-degenerate interval, let $w : [a, b] \to \mathbb{R}$ be a continuous positive weight function, and let $f : [a, b] \to \mathbb{R}$ be a function that we wish to integrate. Then, we consider the approximation

$$
\sum_{i=1}^{m} w_i f(x_i) \approx \int_{a}^{b} f(x)w(x)\,dx,
$$

(A1)

using a quadrature rule with nodes $(x_i)_{i=1}^{m}$ that lie in $[a, b]$, and weights $(w_i)_{i=1}^{m}$, where ‘$\approx$’ has the usual meaning of ‘is approximately equal to’. Note that we have $2m$ degrees of freedom in the choice of our nodes and weights. The Gaussian quadrature rule of level $m$ is the unique choice of nodes and weights that integrates all polynomials of degree at most $2m − 1$ exactly. This means that for all polynomials of degree at most $2m − 1$, we have equality in (A1). The Gaussian quadrature rule exists for all continuous weight functions $w$. We now give a rough sketch how the nodes and weights can be computed. First, one needs to find orthogonal polynomials $p_n$ of degree $n$. Orthogonality is meant with respect to the scalar product

$$
\langle \phi, \psi \rangle = \int_{a}^{b} \phi(x)\psi(x)w(x)\,dx.
$$

Such orthogonal polynomials can be found using a recurrence relation, e.g.

$$
q_{r+1}(x) = x - \frac{(xp_r, p_r)}{\langle p_r, p_r \rangle} p_r(x) - \sum_{j=0}^{r-1} \frac{(xp_r, p_j)}{\langle p_r, p_j \rangle} p_j(x),
$$

where $p_0(x) = 1$. The nodes for the level $m$ quadrature rule are then the $m$ roots of $p_m$. It can be shown that these roots are all real and lie in $[a, b]$. The weight $w_i$ corresponding to the node $x_i$ can then be computed as

$$
w_i = \frac{(p_m, p_{m-1})}{p_m'(x_i)p_{m-1}(x_i)}.
$$

An efficient algorithm for the computation of these nodes and weights is, for example, the Golub-Welsch algorithm with a complexity of $O(m^2)$.

Of course, good error bounds for numerical integration in general, and Gaussian quadrature in particular, are of central importance. Below we give such an error bound. The idea is essentially, that if the function $f$ we wish to integrate is $(2m)$-times continuously differentiable, then we can approximate $f$ by its Taylor expansion of order $2m$, and the corresponding Taylor polynomial of order $2m − 1$ will be integrated exactly by the definition of Gaussian quadrature. In the results below, this is illustrated by the use of the Peano kernel, which, loosely speaking, is the error of the Gaussian quadrature rule in approximating the integral of $x \mapsto x^{2m}$.

LEMMA A.1 (Brass and Petras 2011, Theorem 4.2.3) Let $f : [a, b] \to \mathbb{R}$ be a $(2m)$-times continuously differentiable function, and let $\tilde{w}_i$ be

\begin{align*}
\sum_{i=1}^{m} \tilde{w}_i f(x_i) \approx \int_{a}^{b} f(x)w(x)\,dx,
\end{align*}

(A2)

where $(\tilde{w}_i)_{i=1}^{m}$ are weights. Then, the error of the Gaussian quadrature rule (A1) satisfies

\begin{align*}
\left| \int_{a}^{b} f(x)w(x)\,dx - \sum_{i=1}^{m} \tilde{w}_i f(x_i) \right| \leq C_{2m} h^2 m^{-2m} \| f^{2m} \|_{\infty}.
\end{align*}

A1
the weights and $\tilde{\xi}_i$ be the nodes of the Gaussian quadrature rule for $i = 1, \ldots, m$. Then,
\[
c_H \int_a^b f(x)x^{-H-1/2} \, dx - \sum_{i=1}^m \tilde{w}_i f(\tilde{\xi}_i) = \int_a^b f^{(2m)}(x)K_{2m}(x) \, dx,
\]
where $K_{2m}$ is the Peano kernel (corresponding to the weight function $w$).

**Lemma A.2 (Brass and Hämmerlin 1993, Theorem 2)** The above Peano kernel $K_{2m} : [a,b] \to \mathbb{R}$ satisfies
\[
\sup_{x \in [a,b]} |K_{2m}(x)| \leq \left( \frac{(2\pi)^m}{(2m)!} \right)^{1/2},
\]
where $B_{2m}$ is the Bernoulli function given by
\[
B_{2m}(x) = -2 \sum_{k=1}^{\infty} \cos \left( \frac{2\pi kx - \pi}{2} \right),
\]
where $\xi$ is the Riemann zeta function.

**Appendix 2. Proof of Theorem 2.11**

First, we give the choice of the point set in Theorem 2.11. It is of the same type as the point set in Theorem 2.1, except that we now choose
\[
A := A_{H, \delta} := \frac{1}{\delta - 1/2} \left( 1 - \frac{1}{2 - \gamma} \right) \sqrt{N},
\]
\[
\xi_0 := \exp \left( -\frac{\alpha}{(2 - \gamma)A} \sqrt{N} \right),
\]
\[
\xi_n := b \exp \left( \frac{\alpha}{(b - 1/2)A} \sqrt{N} \right).
\]
Furthermore, $\alpha$ and $\beta$ are as in Theorem 2.1. Note that this agrees with the choice of Theorem 2.1 in the case $\gamma = \delta = H + 1/2$.

The proof of Theorem 2.11 is now essentially analogous to the proof of Theorem 2.1, except that we slightly modify Lemma 2.8 and Lemma 2.9. First, Lemma 2.8 is replaced by the following lemma.

**Lemma A.4** Let $\tilde{w}_i$ be the weights and $\tilde{\xi}_i$ be the nodes of the Gaussian quadrature rule for $i = 1, \ldots, m$ on the interval $[a, b]$ with respect to the weight function $w$ of type $(\gamma, \delta)$. Then we have the following error bounds.

1. If $b \leq e^{\alpha \beta}$, then
\[
\left| \int_a^b e^{-\xi w(x)} \, dx - \sum_{i=1}^m \tilde{w}_i e^{-\tilde{\xi}_i} \right| \leq C_1 \sqrt{\pi} \left( \frac{b - a}{a} \right)^{1/2}.
\]
2. If $a \geq e^{-\alpha \beta}$, then
\[
\left| \int_a^b e^{-\xi w(x)} \, dx - \sum_{i=1}^m \tilde{w}_i e^{-\tilde{\xi}_i} \right| \leq C_2 \sqrt{\pi \beta} \left( \frac{b - a}{a} \right)^{1/2}.
\]

**Appendix 3. Proof of Theorem 2.15**

This proof is split in three steps. First, we determine the dependence of $a$ and $b$ on $T$. In the second step, we optimize $a$ and $b$ over $H$ and $N$. In this step, we make some additional assumptions, which we verify in the last step.

**Step 1:** Let us first consider the dependence of $a$ and $b$ on $T$. Considering the proof of Theorem 2.1, it is no so difficult to see that we have
\[
\mathbb{E}|X_T - \tilde{X}_T|^2 \leq C_{H}^2 \left( \frac{T^3}{(3/2 - H)^2} a^{-2H} + \frac{3}{2H^2} b^{-2H} \right) \times \exp \left( -\frac{2a \sqrt{N}}{A} \right) + \frac{5a^3}{12} \beta^{-3/2} N^{1/2} H^{-3/2} \times \left( \frac{1}{2} \left( \frac{b}{\beta} \right)^{1/2} \right)^{4m+2}.
\]

This bound can be made homogeneous in $T$ by choosing $a = a_1 T^{-1}$ and $b = b_1 T^{-1}$. With some abuse of notation, we again call $a_1$ and $b_1$ $a$ and $b$, respectively.

**Step 2:** Now, we want to determine the dependence of $a$ and $b$ on other variables. Denote $c := \left( \frac{\xi}{2} \right)^{1/2}$. Then,
\[
\left( \frac{1}{2} \left( ce^{\alpha \beta} - 1 \right) \right)^{4m+2} = \left( \frac{1}{2} \left( ce^{\alpha \beta} - e^{\beta} - 1 \right) \right)^{4m+2} = \left( \frac{1}{2} \left( e^{\alpha \beta} - 1 \right) \right)^{4m+2}.
\]

Therefore, by our choice of $a$ and $\beta$,
\[
\mathbb{E}|X_T - \tilde{X}_T|^2 \leq C_{H}^2 \left( \frac{1}{(3/2 - H)^2} a^{-2H} + \frac{3}{2H^2} b^{-2H} \right) \times \left( \frac{5a^3}{48} \beta^{-3/2} N^{1/2} H^{-3/2} \times \left( \frac{1}{2} \left( \frac{b}{\beta} \right)^{1/2} \right)^{4m+2} \right) \exp \left( -\frac{2a \sqrt{N}}{A} \right).
\]
Note that

\[
\left( \frac{e^{a+b} - 1}{e^{a+b} - 1} \right)^{4m+2} = \left( 1 + (c-1) \frac{e^{a+b} - 1}{e^{a+b} - 1} \right)^{4m+2}.
\]

A simple argument shows that it is advisable to choose \( c < 1 \), i.e. \( b < a \). Then, as \( c < 1 \), we have \( \log c < 0 \), and, hence,

\[
\epsilon = e^{\log c} \leq 1 + \log c + \frac{(\log c)^2}{2}.
\]

Then,

\[
\left( 1 + (c-1) \frac{e^{a+b} - 1}{e^{a+b} - 1} \right)^{4m+2} \leq \left( 1 + \frac{e^{a+b} - 1}{e^{a+b} - 1} \left( \log c + \frac{(\log c)^2}{2} \right) \right)^{4m+2} = \left( 1 + \frac{e^{a+b} - 1}{e^{a+b} - 1} \left( \frac{b}{a} \log b - \frac{a}{2A} \log \left( \frac{b}{a} \right)^2 \right) \right)^{4m+2}.
\]

Assume that \( N \) is so large that

\[
-\frac{\beta}{A\sqrt{N}} \log \frac{b}{a} \geq \frac{\beta^2}{2A^2N} \left( \log \frac{b}{a} \right)^2,
\]

i.e.

\[
N \geq \left( \frac{\beta}{A} \log \frac{b}{a} \right) \frac{b}{a}.
\]

Then,

\[
\left( 1 + \frac{e^{a+b} - 1}{e^{a+b} - 1} \left( \frac{b}{a} \log b - \frac{a}{2A} \log \left( \frac{b}{a} \right)^2 \right) \right)^{4m+2} \leq \left( 1 + \frac{e^{a+b} - 1}{8(e^{a+b} - 1) A\sqrt{N}} \log \frac{b}{a} \right)^{4m+2} \leq \exp \left( \frac{e^{a+b} - 1}{8(e^{a+b} - 1) A\sqrt{N}} \log \frac{b}{a} \right)^{4m+2} \leq \left( \frac{b}{a} \right)^{4m+2}.
\]

Thus,

\[
\mathbb{E} | \tilde{X}_T - \tilde{X}_T | \leq C_{q_1} T^{2H} \left( \frac{1}{(3/2 - H)A} a^{-3/2H} + \frac{3}{2H} b^{-2H} \right.
\]

\[
+ \left. \frac{5\pi^3}{48} \left( e^{a+b} - 1 \right)^{2} \frac{A^{2-2H}}{\beta^{2-2H} H} \right) N^{-1/2} \left( \frac{b}{a} \right)^{4m+2} \exp \left( \frac{2\alpha}{A} \log \frac{b}{a} \right).
\]

We now want to solve the optimization problem

\[
\min_{a,b \geq 0} A_1 b^{-q_1} + A_2 \left( \frac{b}{a} \right)^{q_2} + A_3 a^{q_3}.
\]

This is minimized in

\[
a = \left( (q_1 A_1)^{2/3} (q_2 A_2)^{2/3} (q_3 A_3)^{2/3} \right)^{\frac{1}{2/3 + 1/3}} = \left( \frac{q_1 A_1}{q_2 A_2} \right)^{2/3} \frac{1}{(2/3 + 1/3) q_3 A_3},
\]

\[
b = \left( (q_1 A_1)^{2/3} (q_2 A_2)^{2/3} (q_3 A_3)^{2/3} \right)^{\frac{1}{2/3 + 1/3}} = \left( \frac{q_1 A_1}{q_2 A_2} \right)^{2/3} \frac{1}{(2/3 + 1/3) q_3 A_3},
\]

where,

\[
q_1 = 2H, \quad q_2 = \frac{e^{a+b}}{8(e^{a+b} - 1)}, \quad q_3 = 3 - 2H,
\]

\[
A_1 = \frac{3}{2H^2}, \quad A_2 = \frac{5\pi^3}{48} \left( e^{a+b} - 1 \right)^{2} \frac{A^{2-2H}}{\beta^{2-2H} H} A^{1-H}, \quad A_3 = \frac{1}{(3/2 - H)^2}.
\]

Plugging in these values, we get precisely the statement of the theorem.

**Step 3:** We now need to verify that \( b < a \), and that (A2) holds. We start with \( b < a \). Indeed,

\[
b = \left( \frac{5\pi^3}{1152} \left( e^{a+b} - 1 \right)^{2} \frac{A^{2-2H}}{\beta^{2-2H} H} A^{1-H} \right)^{\frac{1}{3}} \frac{H}{(3/2 - H)} \right)^{2H} \leq 1
\]

if and only if

\[
N \geq \left( \frac{5\pi^3}{1152} \left( e^{a+b} - 1 \right)^{2} \frac{A^{2-2H}}{\beta^{2-2H} H} A^{1-H} \right)^{\frac{1}{3}} \frac{H}{(3/2 - H)} \right)^{\frac{2H}{3}}.
\]

The right hand side can be uniformly (over \( H \)) bounded by 2. Hence, it suffices to choose \( N \geq 2 \).

Furthermore, we have made the assumption

\[
N \geq \left( \frac{\beta}{A} \log \frac{b}{a} \right)^2
\]

in (A2), which we also need to verify. We have

\[
\left| \frac{\beta}{A} \log \frac{b}{a} \right| = \left| \frac{\beta}{A} \frac{1}{\exp(e^{a+b} - 1) + 6H - 4H^2} \right|
\]

\[
\times \log \left( \left( \frac{5\pi^3}{1152} \left( e^{a+b} - 1 \right)^{2} \frac{A^{2-2H}}{\beta^{2-2H} H} A^{1-H} \right)^{\frac{1}{3}} \frac{H}{(3/2 - H)} \right)^{2H} \right)
\]

\[
\leq \frac{3}{A} \frac{\log \left( \frac{5\pi^3}{1152} \left( e^{a+b} - 1 \right)^{2} \frac{A^{2-2H}}{\beta^{2-2H} H} A^{1-H} \right)^{\frac{1}{3}} \frac{H}{(3/2 - H)} \right)^{2H} \right)
\]

\[
\times \log \left( \frac{5\pi^3}{1152} \left( e^{a+b} - 1 \right)^{2} \frac{A^{2-2H}}{\beta^{2-2H} H} A^{1-H} \right)^{\frac{1}{3}} \frac{H}{(3/2 - H)} \right)^{2H} \right)
\]

\[
\leq 8 \frac{\beta(e^{a+b} - 1)}{A(e^{a+b} - 1)} \log \left( \frac{5\pi^3}{1152} \left( e^{a+b} - 1 \right)^{2} \frac{A^{2-2H}}{\beta^{2-2H} H} A^{1-H} \right)^{\frac{1}{3}} \frac{H}{(3/2 - H)} \right)^{2H} \right)
\]

\[
+ \frac{8 \beta(e^{a+b} - 1)}{A(e^{a+b} - 1)} \log N.
\]

Plugging in our values of \( \alpha \) and \( \beta \), we get

\[
\left( \frac{\beta}{A} \log \frac{b}{a} \right)^2 \leq \left( \frac{1.26}{A} \log (0.67A^2) + \frac{1.26}{A} \log N \right)^2
\]

\[
\leq (0.8 + 0.73 \log N)^2.
\]

Now,

\[
N \geq (0.8 + 0.73 \log N)^2
\]

holds for all \( N \geq 1 \). This shows the theorem.
Appendix 4. Learning a better rate of convergence

We describe here how we got the results in Section 4.2. We make the ansatz

\[ \xi_0 = C_1 T^{-1} \exp \left( -\frac{\alpha}{(3/2 - H)A} \sqrt{N} \right), \]  
\[ (A3) \]

\[ \xi_n = C_2 T^{-1} \exp \left( -\frac{\alpha}{H A} \sqrt{N} \right), \]  
\[ (A4) \]

\[ m = \frac{\beta}{A} \sqrt{N}, \]  
\[ (A5) \]

\[ \int_0^T |G(t) - \tilde{G}(t)|^2 \, dt = C_3 T^{2H} \exp \left( -\frac{2\alpha}{A} \sqrt{N} \right), \]  
\[ (A6) \]

and our goal is to learn \( \alpha \) and \( \beta \) in particular, and get some approximation for \( C_1, C_2 \) and \( C_3 \), which may depend on \( H \) and \( N \). Here, (A6) philosophically has a different meaning, as the left-hand side, i.e. the true error, is determined by our choice for \( \xi_0, \xi_n \) and \( m \), and need not be of the form of the right-hand side of (A6). Nonetheless, we still choose to use the errors as well to estimate \( \alpha \) with equation (A6).

To do so, we proceed as follows. We take the error representation in Proposition 2.12 for \( T = 1 \). Then, we pick

\[ H \in [0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45] \]

and \( N \) in a range from 1 to 1024. For each of these choices of \( H \) and \( N \), we compute the optimal \( m, \xi_0 \) and \( \xi_n \), as well as the corresponding error, using an optimization algorithm. Then we use this data to estimate \( \alpha \) and \( \beta \).

First, using the optimal values of \( m \), we can determine the corresponding values of \( \beta \) by reformulating equation (A5). Empirically, we observed that paths of the estimated values of \( \beta \) for different choices of \( H \) and \( N \) indeed converge to the same number, independent of \( H \), and we estimate \( \beta = 0.9 \).

For \( \alpha \), we can use \( \xi_0, \xi_n \) and the error as estimators. In other words, we set \( C_i = 1 \), and solve equations (A3), (A4), and (A6) for \( \alpha \). Again we empirically observe that all the paths seem to converge to some uniform number, which we estimate as \( \alpha = 1.8 \).

We can now plug these values for \( \alpha \) and \( \beta \) into equations (A3)–(A6) and apply regression to get estimates for the constants \( C_1, C_2 \) and \( C_3 \). We get the results of Section 4.2.