

MARKOVIAN APPROXIMATIONS OF STOCHASTIC VOLTERRA EQUATIONS WITH THE FRACTIONAL KERNEL

CHRISTIAN BAYER AND SIMON BRENEIS

ABSTRACT. We consider rough stochastic volatility models where the variance process satisfies a stochastic Volterra equation with the fractional kernel, as in the rough Bergomi and the rough Heston 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.

1. INTRODUCTION

Suppose we are given a stochastic Volterra equation of the form

$$(1.1) \quad X_t = x_0 + \int_0^t G(t-s)b(X_s)ds + \int_0^t G(t-s)\sigma(X_s)dW_s,$$

where $x_0 \in \mathbb{R}^d$, $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^{d \times d} \rightarrow \mathbb{R}^{d \times d}$ 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, $G(t) \simeq t^{H-1/2}$. In analogy to fractional Brownian motion (fBm), we call the parameter $0 < H < 1$ *Hurst parameter*. Indeed, the special case $\widetilde{W}_t := \sqrt{2H} \int_0^t (t-s)^{H-1/2} dW_s$ defines the so-called *Riemann–Liouville fBm*.

Models of this type are used in different applications – we refer to [Mar13] and [Che17] 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

$$dS_t = \sqrt{V_t}S_t dZ_t$$

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

Date: July 19, 2022.

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

Key words and phrases. Stochastic Volterra equation, Fractional kernel, Rough volatility model, Markovian approximation, Strong error.

C.B. and S.B. gratefully acknowledge the support of the DFG through the IRTG 2544. C.B. additionally acknowledges the support of the DFG through the FOR 2402. Both authors would also like to thank Ahmed Kebaier for the insightful discussions.

$r = 0$ for simplicity. As observed in [GJR18] (for time series of stock prices) and [BFG16] (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 [ER18], 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 [BFG16],

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

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 [CC98]. 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 μ 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

$$\widetilde{W}_t := \int_0^t G(t-s)dW_s = \int_0^\infty \int_0^t e^{-(t-s)x}dW_s\mu(dx) = \int_0^t Y_t^x\mu(dx), \quad Y_t^x := \int_0^t e^{-(t-s)x}dW_s,$$

where Y^x is an Ornstein–Uhlenbeck process, and, hence, a Markov process. Indeed, the infinite-dimensional process $\mathcal{Y}_t := (Y_t^x)_{x>0}$ is a Markov process with state space $L^1(\mathbb{R}_{>0}, \mu)$, see [CC98].

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

$$(1.2) \quad G(t) := \frac{t^{H-1/2}}{\Gamma(H+1/2)},$$

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

$$(1.3) \quad G(t) = c_H \int_0^\infty e^{-xt}x^{-H-1/2}dx, \quad c_H := \frac{1}{\Gamma(H+1/2)\Gamma(1/2-H)},$$

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., $G \approx \widehat{G}$ with

$$(1.4) \quad \widehat{G}(t) := \sum_{i=1}^N w_i e^{-x_i t},$$

we get

$$(1.5) \quad \widehat{X}_t = x_0 + \int_0^t \widehat{G}(t-s)b(\widehat{X}_s)ds + \int_0^t \widehat{G}(t-s)\sigma(\widehat{X}_s)dW_s.$$

Such approximations were already considered by, e.g., Abi Jaber and El Euch [AE19b], Alfonsi and Kebaier [AK21], and, for the particular case of fractional Brownian motion, by Harms [Har19]. Indeed, we have the following proposition, which shows that (1.5) reduces to an N -dimensional ordinary stochastic differential equation, instead of a stochastic Volterra equation.

Proposition 1.1. [AK21, Proposition 2.1]. *Let $x_0^1, \dots, x_0^N \in \mathbb{R}^d$ be such that $\sum_{i=1}^N w_i x_0^i = x_0$. Then the solution to (1.5) is given by $\sum_{i=1}^N w_i X_t^i$, where (X_t^1, \dots, X_t^N) is the solution to the $(N \times d)$ -dimensional SDE defined by*

$$(1.6) \quad X_t^i = x_0^i - \int_0^t x_i(X_s^i - x_0^i) ds + \int_0^t b \left(\sum_{j=1}^N w_j X_s^j \right) ds + \int_0^t \sigma \left(\sum_{j=1}^N w_j X_s^j \right) dW_s.$$

Perhaps unsurprisingly, the approximation error between X and the Markovian approximation \widehat{X} is controlled by the error between G and \widehat{G} , more precisely by the L^2 -error. The following bound is a slight modification of [AK21, Proposition 3.2].

Proposition 1.2. [AK21, Proposition 3.2]. *For every $T > 0$, there exists a constant C (depending on $T, |x_0|, b, \sigma$), such that*

$$\mathbb{E} |\widehat{X}_T - X_T|^2 \leq C \int_0^T |G(t) - \widehat{G}(t)|^2 dt.$$

Using this bound, Alfonsi and Kebaier in [AK21, Corollary 3.1] constructed a point set $(x_i)_{i=1, \dots, N}$ with weights $(w_i)_{i=1, \dots, N}$ by cutting off the integral in (1.3) 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 to almost 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 [BFG16]. 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 [Har19]. 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 (1.3). 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 [Har19, 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 [AK21] and [Har19]. 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 ∞ , 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 simplification.

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 A.

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 $A = A_H := \left(\frac{1}{H} + \frac{1}{3/2-H}\right)^{1/2}$, where $0 < H < 1/2$ denotes the Hurst index of the kernel G , we define

$$\begin{aligned} m &:= \frac{\beta}{A}\sqrt{N}, & n &:= \frac{A}{\beta}\sqrt{N} \ (\approx N/m), \\ \xi_0 &:= a \exp\left(-\frac{\alpha}{(3/2-H)A}\sqrt{N}\right), & \xi_n &:= b \exp\left(\frac{\alpha}{HA}\sqrt{N}\right), \\ \xi_i &:= \xi_0 \left(\frac{\xi_n}{\xi_0}\right)^{i/n}, & & \quad i = 0, \dots, n. \end{aligned}$$

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

$$(2.1) \quad w(x) := c_H x^{-H-1/2}, \quad c_H := \frac{1}{\Gamma(H+1/2)\Gamma(1/2-H)}.$$

In addition, we set $x_0 := 0$ and

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

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

$$(2.2) \quad \widehat{G}(t) := \sum_{i=0}^{nm} w_i e^{-x_i t}.$$

Theorem 2.1. *Let $x_0 \in \mathbb{R}^d$, and let $b : \mathbb{R}^d \rightarrow \mathbb{R}$ and $\sigma : \mathbb{R}^{d \times d} \rightarrow \mathbb{R}^{d \times d}$ be globally Lipschitz continuous functions. Let X be the solution of (1.1), and let \widehat{X} be the solution of (1.5), where we use the Gaussian rule of type $(H, N, \alpha, \beta, 1, 1)$ with $\alpha := 1.06418$ and $\beta := 0.4275$. Then,*

$$\mathbb{E}|X_T - \widehat{X}_T|^2 \leq C c_H^2 \left(\frac{T^3}{(3/2-H)^2} + \frac{3}{2H^2} + \frac{5\pi^3}{48} (e^{\alpha\beta} - 1)^2 \frac{A^{2-2H} T^{2H}}{\beta^{2-2H} H} N^{1-H} \right) \exp\left(-\frac{2\alpha}{A} \sqrt{N}\right),$$

where C is the constant from Proposition 1.2.

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

Remark 2.2. Theorem 2.1 is not only true for the specific values of α and β given above, but for an entire set $\Gamma \subset \mathbb{R}_+^2$ of pairs (α, β) . 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 (α, β) .

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 $x_0 = 0$ 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 $[0, \xi_0]$ 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 \widehat{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 [AE19b, Theorem 3.1].

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 $[\xi_0, \xi_n]$, 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 $[\xi_i, \xi_{i+1}]$ 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 A.

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

$$\left| c_H \int_a^b e^{-tx} x^{-H-1/2} dx - \sum_{i=1}^m \tilde{w}_i e^{-t\tilde{x}_i} \right| \leq \sqrt{\frac{5\pi^3}{18}} \frac{c_H}{2^{2m+1} m^H} t^{-1/2+H} \left(\frac{b}{a} - 1 \right)^{2m+1}.$$

Proof. By Lemma A.1 and Lemma A.2,

$$\begin{aligned} \left| c_H \int_a^b e^{-tx} x^{-H-1/2} dx - \sum_{i=1}^m \tilde{w}_i e^{-t\tilde{x}_i} \right| &= \left| \int_a^b \partial_x^{2m} e^{-tx} K_{2m}(x) dx \right| \\ &\leq t^{2m} \int_a^b e^{-tx} |K_{2m}(x)| dx \\ &\leq t^{2m} \int_a^b e^{-ta} \frac{(2\pi)^{2m}}{(2m)!} \left(\frac{b-a}{2} \right)^{2m} \sup_{y \in [-1, 1]} |B_{2m}(y)| \sup_{y \in [a, b]} |w(y)| dx. \end{aligned}$$

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

$$\begin{aligned} \left| c_H \int_a^b e^{-tx} x^{-H-1/2} dx - \sum_{i=1}^m \tilde{w}_i e^{-t\tilde{x}_i} \right| &\leq \\ &t^{2m} \int_a^b e^{-ta} \frac{(2\pi)^{2m}}{(2m)!} \left(\frac{b-a}{2} \right)^{2m} \frac{2\zeta(2m)}{(2\pi)^{2m}} c_H a^{-H-1/2} dx. \end{aligned}$$

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

$$\begin{aligned} \left| c_H \int_a^b e^{-tx} x^{-H-1/2} dx - \sum_{i=1}^m \tilde{w}_i e^{-t\tilde{x}_i} \right| &\leq t^{2m} e^{-ta} \frac{1}{(2m)!} \frac{(b-a)^{2m+1}}{2^{2m}} \frac{\pi^2}{3} c_H a^{-H-1/2} \\ &= \frac{\pi^2 c_H}{3 \cdot 2^{2m} (2m)!} t^{2m} e^{-ta} a^{2m+1/2-H} \left(\frac{b}{a} - 1 \right)^{2m+1}. \end{aligned}$$

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

$$\begin{aligned}
& \left| c_H \int_a^b e^{-tx} x^{-H-1/2} dx - \sum_{i=1}^m \tilde{w}_i e^{-t\tilde{x}_i} \right| \\
& \leq \frac{\pi^2 c_H}{3 \cdot 2^{2m} 2\sqrt{\pi m}} \left(\frac{e}{2m}\right)^{2m} t^{2m} \left(\frac{2m+1/2-H}{e}\right)^{2m+1/2-H} \times \\
& \quad \times (ta)^{-2m-1/2+H} a^{2m+1/2-H} \left(\frac{b}{a} - 1\right)^{2m+1} \\
& = \frac{\pi^2 c_H}{3 \cdot 2^{2m+1} \sqrt{\pi m}} \left(\frac{2m+1/2-H}{2m}\right)^{2m} \left(\frac{2m+1/2-H}{e}\right)^{1/2-H} t^{-1/2+H} \left(\frac{b}{a} - 1\right)^{2m+1} \\
& \leq \frac{\pi^2 c_H}{3 \cdot 2^{2m+1} \sqrt{\pi m}} e^{1/2-H} \left(\frac{2m+1/2-H}{e}\right)^{1/2-H} t^{-1/2+H} \left(\frac{b}{a} - 1\right)^{2m+1} \\
& \leq \frac{\pi^2 c_H}{3 \cdot 2^{2m+1} m^H} \left(\frac{5/2}{\pi}\right)^{1/2} t^{-1/2+H} \left(\frac{b}{a} - 1\right)^{2m+1} \\
& \leq \sqrt{\frac{5\pi^3}{18}} \frac{c_H}{2^{2m+1} m^H} t^{-1/2+H} \left(\frac{b}{a} - 1\right)^{2m+1}. \quad \square
\end{aligned}$$

The previous lemma is an error estimate of our integration error on an interval $[\xi_i, \xi_{i+1}]$. We will now combine the errors over all such intervals.

Lemma 2.9. *In the setting of Theorem 2.1, we have*

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

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

$$\begin{aligned}
& \int_0^T \left| c_H \int_{\xi_0}^{\xi_n} e^{-tx} x^{-H-1/2} dx - \sum_{i=1}^N w_i e^{-tx_i} \right|^2 dt \\
& \leq \int_0^T \left(\sum_{i=0}^{n-1} \sqrt{\frac{5\pi^3}{18}} \frac{c_H}{2^{2m+1} m^H} t^{-1/2+H} \left(\frac{\xi_{i+1}}{\xi_i} - 1\right)^{2m+1} \right)^2 dt.
\end{aligned}$$

Recall that

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

Thus,

$$\begin{aligned}
& \int_0^T \left| c_H \int_{\xi_0}^{\xi_n} e^{-tx} x^{-H-1/2} dx - \sum_{i=1}^N w_i e^{-tx_i} \right|^2 dt \\
& \leq \int_0^T \left(\sum_{i=0}^{n-1} \sqrt{\frac{5\pi^3}{18}} \frac{c_H}{2^{2m+1} m^H} t^{-1/2+H} (e^{\alpha\beta} - 1)^{2m+1} \right)^2 dt \\
& = \int_0^T \left(n \sqrt{\frac{5\pi^3}{18}} \frac{c_H}{2^{2m+1} m^H} t^{-1/2+H} (e^{\alpha\beta} - 1)^{2m+1} \right)^2 dt \\
& = \frac{5\pi^3}{36} \frac{c_H^2 T^{2H}}{H} \frac{n^2}{m^{2H}} \left(\frac{1}{2} (e^{\alpha\beta} - 1) \right)^{4m+2}. \quad \square
\end{aligned}$$

We now move to the proof of Theorem 2.1.

Proof of Theorem 2.1. By Proposition 1.2, we have

$$\mathbb{E}|X_T - \widehat{X}_T|^2 \leq C \int_0^T |G(t) - \widehat{G}(t)|^2 dt.$$

Hence,

$$\begin{aligned}
\mathbb{E}|X_T - \widehat{X}_T|^2 & \leq C \int_0^T \left| c_H \int_0^\infty e^{-tx} x^{-H-1/2} dx - \sum_{i=0}^N w_i e^{-tx_i} \right|^2 dt \\
& \leq 3C \int_0^T \left(\left| c_H \int_0^{\xi_0} e^{-tx} x^{-H-1/2} dx - w_0 \right|^2 \right. \\
& \quad \left. + \left| c_H \int_{\xi_0}^{\xi_n} e^{-tx} x^{-H-1/2} dx - \sum_{i=1}^N w_i e^{-tx_i} \right|^2 \right. \\
& \quad \left. + \left| c_H \int_{\xi_n}^\infty e^{-tx} x^{-H-1/2} dx \right|^2 \right) dt.
\end{aligned}$$

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

$$\left| c_H \int_0^{\xi_0} e^{-tx} x^{-H-1/2} dx - w_0 \right| = c_H \int_0^{\xi_0} (1 - e^{-tx}) x^{-H-1/2} dx.$$

Since $e^y \geq 1 + y$, we have

$$\begin{aligned}
c_H \int_0^{\xi_0} (1 - e^{-tx}) x^{-H-1/2} dx & \leq c_H \int_0^{\xi_0} tx x^{-H-1/2} dx \\
& = \frac{c_H}{3/2 - H} t \xi_0^{3/2-H} \\
& = \frac{c_H}{3/2 - H} t \exp\left(-\frac{\alpha}{A} \sqrt{N}\right).
\end{aligned}$$

Squaring and integrating gives

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

Now, consider the last summand. We have

$$\begin{aligned}
\int_0^T \left(c_H \int_{\xi_n}^{\infty} e^{-tx} x^{-H-1/2} dx \right)^2 dt &= c_H^2 \int_0^T \int_{\xi_n}^{\infty} \int_{\xi_n}^{\infty} e^{-t(x+y)} x^{-H-1/2} y^{-H-1/2} dy dx dt \\
&\leq c_H^2 \int_{\xi_n}^{\infty} \int_{\xi_n}^{\infty} \int_0^{\infty} e^{-t(x+y)} dt x^{-H-1/2} y^{-H-1/2} dy dx \\
&= c_H^2 \int_{\xi_n}^{\infty} \int_{\xi_n}^{\infty} \frac{x^{-H-1/2} y^{-H-1/2}}{x+y} dy dx \\
&\leq \frac{c_H^2}{2} \int_{\xi_n}^{\infty} \int_{\xi_n}^{\infty} \frac{x^{-H-1/2} y^{-H-1/2}}{\sqrt{xy}} dy dx \\
&= \frac{c_H^2}{2H^2} \xi_n^{-2H} \\
&= \frac{c_H^2}{2H^2} \exp\left(-\frac{2\alpha}{A} \sqrt{N}\right).
\end{aligned}$$

Putting these estimates and Lemma 2.9 together, we get

$$\begin{aligned}
\mathbb{E}|X_T - \widehat{X}_T|^2 &\leq 3C \left(\frac{c_H^2 T^3}{3(3/2-H)^2} \exp\left(-\frac{2\alpha}{A} \sqrt{N}\right) \right. \\
&\quad \left. + \frac{5\pi^3}{36} \frac{c_H^2 T^{2H}}{H} \frac{n^2}{m^{2H}} \left(\frac{1}{2}(e^{\alpha\beta} - 1)\right)^{4m+2} \right. \\
&\quad \left. + \frac{c_H^2}{2H^2} \exp\left(-\frac{2\alpha}{A} \sqrt{N}\right) \right) \\
(2.3) \quad &= C c_H^2 \left(\left(\frac{T^3}{(3/2-H)^2} + \frac{3}{2H^2} \right) \exp\left(-\frac{2\alpha}{A} \sqrt{N}\right) \right. \\
&\quad \left. + \frac{5\pi^3}{12} \frac{A^{2-2H} T^{2H}}{\beta^{2-2H} H} N^{1-H} \left(\frac{1}{2}(e^{\alpha\beta} - 1)\right)^{4m+2} \right).
\end{aligned}$$

Notice that

$$(2.4) \quad \left(\frac{1}{2}(e^{\alpha\beta} - 1)\right)^{4m} = \exp\left(\log\left(\frac{1}{2}(e^{\alpha\beta} - 1)\right) \frac{4\beta}{A} \sqrt{N}\right).$$

Our next goal is to choose α and β in such a way as to maximize the rate of convergence in N . The rate of convergence in the first term of (2.3) is of course larger the larger α is, indicating that we would like to choose α as large as possible. However, larger α at the same time leads to slower rate of convergence in the second term of (2.3), which is equivalent to the right-hand side of (2.4). Hence, to maximize the overall rate, we first set the rates of the first term in (2.3) equal to the rate of the second term in (2.3), and afterwards, we maximize over α . In all that, we ignore the factor N^{1-H} in the second term of (2.3), as it is comparatively negligible. We therefore have to solve the optimization problem

$$(2.5) \quad \alpha \rightarrow \max!, \text{ subject to } -\frac{2\alpha}{A} \sqrt{N} = \log\left(\frac{1}{2}(e^{\alpha\beta} - 1)\right) \frac{4\beta}{A} \sqrt{N},$$

where we maximize over $\alpha, \beta > 0$.

As one can see, this optimization problem is completely “dimensionless”, so the solutions α, β are actually numbers independent of any parameters, and are given as in the statement of the theorem. For these values, we have

$$\mathbb{E}|X_T - \widehat{X}_T|^2 \leq C c_H^2 \left(\frac{T^3}{(3/2 - H)^2} + \frac{3}{2H^2} + \frac{5\pi^3}{48} (e^{\alpha\beta} - 1)^2 \frac{A^{2-2H} T^{2H}}{\beta^{2-2H} H} N^{1-H} \right) \exp\left(-\frac{2\alpha}{A} \sqrt{N}\right).$$

This proves the theorem. \square

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 B. We consider the following class of kernels.

Definition 2.10. A completely monotone kernel G is said to be of type (γ, δ) if

$$G(t) = \int_0^\infty e^{-xt} w(x) dx$$

for some non-negative weight function w that is continuous on $(0, \infty)$ and satisfies $w(x) = \mathcal{O}(x^{-\gamma})$, $x \rightarrow 0$, and $w(x) = \mathcal{O}(x^{-\delta})$, $x \rightarrow \infty$. Furthermore, we define

$$A_{\gamma, \delta} := \left(\frac{1}{\delta - 1/2} + \frac{1}{2 - \gamma} \right)^{1/2}.$$

Theorem 2.11. Let $x_0 \in \mathbb{R}^d$, and let $b : \mathbb{R}^d \rightarrow \mathbb{R}$ and $\sigma : \mathbb{R}^{d \times d} \rightarrow \mathbb{R}^{d \times d}$ be globally Lipschitz continuous functions. Let G be of type (γ, δ) with $1/2 < \delta < 3/2$. Let X be the solution of (1.1), and let \widehat{X} be the solution of (1.5) for the choice of nodes and weights in Appendix B. Then,

$$\mathbb{E}|X_T - \widehat{X}_T|^2 \leq C N^{3/2 - (\gamma \wedge \delta)} \exp\left(-\frac{2\alpha}{A_{\gamma, \delta}} \sqrt{N}\right),$$

where C does not depend on N , and α 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 analyse 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 $\widehat{G}(t) = \sum_{i=0}^N w_i e^{-x_i t}$, where $x_0 = 0$ and $x_i > 0$ for $i = 1, \dots, N$. Then,

$$\begin{aligned} \int_0^T |G(t) - \widehat{G}(t)|^2 dt &= \frac{T^{2H}}{2H\Gamma(H+1/2)^2} + w_0^2 T + 2w_0 \sum_{i=1}^N \frac{w_i}{x_i} (1 - e^{-x_i T}) \\ (2.6) \quad &+ \sum_{i,j=1}^N \frac{w_i w_j}{x_i + x_j} (1 - e^{-(x_i + x_j)T}) \\ &- \frac{2w_0 T^{H+1/2}}{\Gamma(H+3/2)} - \frac{2}{\Gamma(H+1/2)} \sum_{i=1}^N \frac{w_i}{x_i^{H+1/2}} \int_0^{x_i T} t^{H-1/2} e^{-t} dt. \end{aligned}$$

Remark 2.13. The expression in (2.6) 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 (2.6) 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 (2.6) 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 C.

Theorem 2.15. *Let $x_0 \in \mathbb{R}^d$, and let $b : \mathbb{R}^d \rightarrow \mathbb{R}$ and $\sigma : \mathbb{R}^{d \times d} \rightarrow \mathbb{R}^{d \times d}$ be globally Lipschitz continuous functions. Let X be the solution of (1.1), and let \widehat{X} be the solution of (1.5), where we use the Gaussian rule of type $(H, N, \alpha, \beta, a, b)$ with $\alpha := 1.06418$ and $\beta := 0.4275$,*

$$a = T^{-1} \left(\left(\frac{9 - 6H}{2H} \right)^{\frac{e^{\alpha\beta}}{8(e^{\alpha\beta} - 1)}} \left(\frac{5\pi^3}{768} e^{\alpha\beta} (e^{\alpha\beta} - 1) \frac{A^{2-2H} (3 - 2H)}{\beta^{2-2H} H} N^{1-H} \right)^{2H} \right)^\gamma,$$

$$b = T^{-1} \left(\left(\frac{9 - 6H}{2H} \right)^{\frac{e^{\alpha\beta}}{8(e^{\alpha\beta} - 1)}} \left(\frac{5\pi^3}{1152} e^{\alpha\beta} (e^{\alpha\beta} - 1) \frac{A^{2-2H}}{\beta^{2-2H}} N^{1-H} \right)^{2H-3} \right)^\gamma,$$

where $\gamma := \left(\frac{3e^{\alpha\beta}}{8(e^{\alpha\beta} - 1)} + 6H - 4H^2 \right)^{-1}$ and we assume that $N \geq 2$. Then,

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

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

$$C(N) = C c_H^2 T^{2H} \left(\frac{1}{2H} + 8 \frac{e^{\alpha\beta} - 1}{e^{\alpha\beta}} + \frac{1}{3 - 2H} \right) \times$$

$$\times \left(\left(\frac{3}{H} \right)^{\frac{e^{\alpha\beta} (3-2H)}{8(e^{\alpha\beta} - 1)}} \left(\frac{5\pi^3}{384} e^{\alpha\beta} (e^{\alpha\beta} - 1) \frac{A^{2-2H}}{\beta^{2-2H} H} N^{1-H} \right)^{6H-4H^2} \left(\frac{1}{3/2 - H} \right)^{\frac{e^{\alpha\beta} H}{4(e^{\alpha\beta} - 1)}} \right)^\gamma,$$

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., $X_t = \frac{1}{\Gamma(H+1/2)} \int_0^t (t-s)^{H-1/2} dW_s$. 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

$$\xi_0 = 4.3679N^{0.1135}e^{-0.2322\sqrt{N}}, \quad \xi_n = 0.1421N^{-1.5889}e^{3.2511\sqrt{N}}.$$

Then, we have

$$\mathbb{E}|X_T - \widehat{X}_T|^2 \leq 33.6483N^{0.3178}e^{-0.6502\sqrt{N}}.$$

3.2. The case of the rough Bergomi model. The following presentation of the rough Bergomi model is largely inspired by [BFG16]. The rough Bergomi model is given by the system

$$(3.1) \quad \begin{aligned} S_t &= S_0 \exp\left(\int_0^t \sqrt{V_s}(\rho dW_s + \sqrt{1-\rho^2}dB_s) - \frac{1}{2}\int_0^t V_s ds\right), \\ V_t &= V_0 \exp\left(\eta\sqrt{2H}\int_0^t (t-s)^{H-1/2}dW_s - \frac{\eta^2}{2}t^{2H}\right). \end{aligned}$$

The rough Bergomi model was first studied by Bayer, Friz and Gatheral in [BFG16], and can be seen as a rough, non-Markovian generalization of the Bergomi model by Bergomi [Ber05]. 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, ρ is the correlation between the Brownian motion driving the volatility process V and the stock price S , and is usually negative, and the product $\rho\eta$ sets the level of the ATM skew for longer expirations. More information on how to choose these parameters can be found in [BFG16].

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 (3.1) as follows.

$$(3.2) \quad \begin{aligned} S_t &= 1 + \int_0^t S_s \exp(X_s) dW_s, \\ dX_t &= d\left(\frac{1}{2}\log(V_t)\right) = \frac{\eta\sqrt{H}}{\sqrt{2}} \int_0^t (t-s)^{H-1/2} d\widehat{W}_s. \end{aligned}$$

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

We recall the following Lemma by Harms [Har19].

Lemma 3.2. *[Har19, Lemma 3]. Let $X, \widehat{X}, S, \widehat{S} : [0, T] \times \Omega \rightarrow \mathbb{R}$ be continuous stochastic processes with $X_0 = \widehat{X}_0 = 0$ and*

$$S_t = 1 + \int_0^t S_s \exp(X_s) dW_s, \quad \widehat{S}_t = 1 + \int_0^t \widehat{S}_s \exp(\widehat{X}_s) dW_s,$$

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

$$\begin{aligned} |\mathbb{E}f(S_T) - \mathbb{E}f(\widehat{S}_T)| &\leq L(\sqrt{T} + 6) \left(\mathbb{E} \sup_{t \in [0, T]} \left| \exp(2|X_t|) + \exp(2|\widehat{X}_t|) \right|^2 \right)^{1/2} \\ &\quad \times \left(\int_0^T \mathbb{E} |X_t - \widehat{X}_t|^2 dt \right)^{1/2}. \end{aligned}$$

The following corollary is immediate.

Corollary 3.3. *Let $f : (0, \infty) \rightarrow \mathbb{R}$ be a function such that $f \circ \exp$ is Lipschitz continuous. Let (S, X) be the solution of the stochastic Volterra equation (3.2), and let $(\widehat{S}, \widehat{X})$ 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*

$$|\mathbb{E}f(S_T) - \mathbb{E}f(\widehat{S}_T)| \leq CN^{0.21} \exp\left(-\frac{\alpha}{A}\sqrt{N}\right),$$

with α and A as in Theorem 2.15.

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

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

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),

$$\begin{aligned} \|X - \widehat{X}\|_{L^2([0, T] \times \Omega)}^2 &= \int_0^T \mathbb{E} |X_t - \widehat{X}_t|^2 dt \\ &\leq \int_0^T C(N) \exp\left(-\frac{2\alpha}{A}\sqrt{N}\right) dt \\ &\leq C_2 N^{0.42} \exp\left(-\frac{2\alpha}{A}\sqrt{N}\right). \end{aligned}$$

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

Remark 3.4. It should be noted that neither S nor \widehat{S} can be simulated exactly, both require time discretization of the SDE (3.2) defining S or \widehat{S} , respectively. So far, a proper weak error analysis of the Euler scheme for S is missing – but see [BHT20] for partial results, indicating weak rate $1/2 + H$. On the other hand, $(\widehat{S}, \widehat{X})$ is a standard diffusion process. Hence, an Euler discretization of \widehat{S} – obviously, \widehat{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 [BHT20], as the constant for the weak error will depend on N as well as H , and might explode as $N \rightarrow \infty$.

3.3. The case of the rough Heston model. The following presentation of the rough Heston model is largely inspired by [AE19b]. The rough Heston model is given by

$$(3.3) \quad dS_t = S_t \sqrt{V_t} dW_t,$$

$$(3.4) \quad V_t = V_0 + \int_0^t G(t-s)(\theta(s) - \lambda V_s) ds + \int_0^t G(t-s) \nu \sqrt{V_s} dB_s,$$

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 [ER18, Definition 2.1]. This time-dependent mean-reversion level is used to fit the forward variance curve $(\mathbb{E}V_t)_{t \leq T}$, as explained in [AE19b, Section 2]. Also in [AE19b, Section 2], the existence of a weak non-negative solution to a generalization of (3.4) with Hölder-regularity $H - \varepsilon$ for all $\varepsilon > 0$ is shown.

Note that due to the square root in (3.4), 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 [AE19a; EGR19; ER18; ER19], the characteristic function of $\log(S_t/S_0)$ can be given in terms of the solution of a fractional Riccati equation, i.e.

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

Here,

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

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

$$(3.5) \quad \psi(t, z) = \int_0^t G(t-s) F(z, \psi(s, z)) ds, \quad t \in [0, T],$$

where

$$F(z, x) := \frac{1}{2}(z^2 - z) + (\rho \nu z - \lambda)x + \frac{\nu^2}{2}x.$$

While equation (3.5) cannot be solved explicitly, it can be solved numerically using the Adams scheme developed in [DFF02; DFF04; DF98; ER19].

Now, we again replace the kernel G by an approximation \widehat{G} of the type (1.4). We denote the solution of the corresponding ordinary stochastic differential equations similar to equations (3.3) and (3.4) by $(\widehat{S}, \widehat{V})$. As was noted in [AE19b, Section 4.1], we again have

$$\mathbb{E} \exp(z \log(\widehat{S}_t/S_0)) = \exp\left(\int_0^t F(z, \widehat{\psi}(t-s, z)) \widehat{g}(s) ds\right).$$

Here,

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

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

$$\widehat{\psi}(t, z) = \int_0^t \widehat{G}(t-s) F(z, \widehat{\psi}(s, z)) ds, \quad t \in [0, T].$$

Also in [AE19b, Section 4.1], it is illustrated that $\widehat{\psi}$ 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 [AE19b].

Proposition 3.5. [AE19b, 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*

$$\sup_{t \in [0, T]} |\widehat{\psi}(t, a + ib) - \psi(t, a + ib)| \leq C(1 + b^4) \int_0^T |\widehat{G}(s) - G(s)| ds.$$

Proposition 3.6. [AE19b, 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 $\widehat{C}(k, T)$ the price in the Markovian approximation of the rough Heston model. If $|\rho| < 1$, then there exists a constant $c > 0$ such that*

$$|C(k, T) - \widehat{C}(k, T)| \leq c \int_0^T |G(t) - \widehat{G}(t)| dt.$$

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*

$$\sup_{t \in [0, T]} |\widehat{\psi}(t, a + ib) - \psi(t, a + ib)| \leq C(1 + b^4) N^{0.21} \exp\left(-\frac{\alpha}{A} \sqrt{N}\right).$$

Proof. Using Proposition 3.5, Jensen's inequality, Theorem 2.15 and Remark 2.16, we get

$$\begin{aligned} \sup_{t \in [0, T]} |\widehat{\psi}(t, a + ib) - \psi(t, a + ib)| &\leq C_1(1 + b^4) \int_0^T |\widehat{G}(s) - G(s)| ds \\ &\leq C_1 \sqrt{T} (1 + b^4) \left(\int_0^T |\widehat{G}(s) - G(s)|^2 ds \right)^{1/2} \\ &\leq C_1 \sqrt{T} (1 + b^4) \left(C(N) \exp\left(-\frac{2\alpha}{A} \sqrt{N}\right) \right)^{1/2} \\ &\leq C(1 + b^4) N^{0.21} \exp\left(-\frac{\alpha}{A} \sqrt{N}\right). \quad \square \end{aligned}$$

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 $\widehat{C}(k, T)$ 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*

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

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 [LT09] that the rate of convergence of the Adams scheme that is usually used for solving the fractional Riccati equation is $O(\Delta t)$, where Δt 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 , ξ_0 and ξ_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 , ξ_0 and ξ_n , but what are the optimal choices? Using the exact error representation given by Proposition 2.12, we can minimize this error over ξ_0 and ξ_n for each m . More precisely, for each $m = 1, \dots, 10$, we try to minimize the right-hand side of Proposition 2.12 as a function of ξ_0 and ξ_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 ξ_0 and ξ_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 ξ_0, ξ_n ” and “Optimal m, ξ_0, ξ_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 , ξ_0 and ξ_n for any N and H , and then applying regression, we get the following approximations of the optimal parameter values as a function of N .

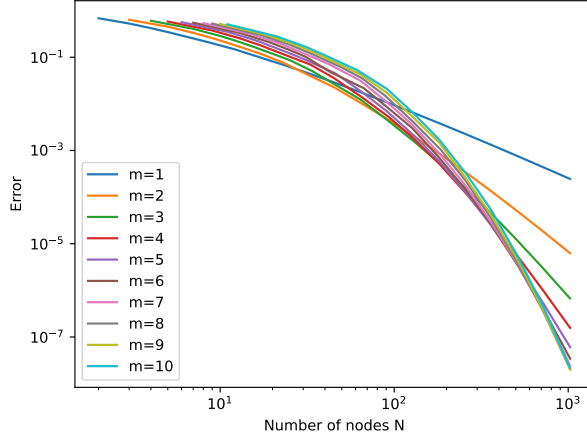


FIGURE 1. L^2 -error for the approximation of fractional Brownian motion with $H = 0.1$ and $T = 1$ for different quadrature levels m as a function of the number of nodes N , using the optimal values for ξ_0 and ξ_n

N	m	n	Theorem				Optimal ξ_0, ξ_n			Optimal m, ξ_0, ξ_n				
			$-\log \xi_0$	$\log \xi_n$	Error	Bound	$-\log \xi_0$	$\log \xi_n$	Error	m	n	$-\log \xi_0$	$\log \xi_n$	Error
1	1	1	-1.2421	1.2999	0.996490	4.190757	83.372	4.8778	0.683687	1	1	83.372	4.8778	0.683687
2	1	2	-1.2246	1.5452	0.975001	4.089250	1.4621	9.1800	0.528237	1	2	1.4621	9.1800	0.528237
4	1	4	-1.1672	2.3483	0.899764	3.773728	0.7776	14.455	0.346109	1	4	0.7776	14.455	0.346109
8	1	8	-1.0535	3.9403	0.757286	3.218395	0.5037	21.394	0.199291	1	8	0.5037	21.394	0.199291
16	1	16	-0.8602	6.6478	0.571030	2.455046	1.6463	28.971	0.098625	1	16	1.6463	28.971	0.098625
32	1	32	-0.5541	10.933	0.372303	1.599424	2.3096	37.865	0.043699	2	16	1.8629	36.893	0.039571
64	2	32	-0.0887	17.450	0.195570	0.833625	2.7007	51.739	0.010167	2	32	2.7007	51.739	0.010167
128	2	64	0.6021	27.121	0.075222	0.316916	4.4629	68.195	0.002037	3	43	3.9939	70.067	0.001559
256	3	85	1.6115	41.256	0.018481	0.077112	6.5970	93.266	0.000158	4	64	6.2656	95.048	0.000123
512	3	171	3.0718	61.701	0.002405	0.009982	8.2385	120.50	1.14e-05	6	85	9.4066	130.22	3.50e-06
1024	5	205	5.1694	91.071	0.000128	0.000529	12.162	172.37	6.03e-08	9	114	13.029	180.30	1.98e-08

TABLE 1. L^2 -approximation errors of fractional Brownian motion with $H = 0.1$ and $T = 1$ for various choices of N , m , n , ξ_0 and ξ_n . The column labeled “Theorem” denotes the values of m , n , $-\log \xi_0$ and $\log \xi_n$ suggested by Corollary 3.1. The error column denotes the actual error for these choices, computed using Proposition 2.12. The bound column gives the error bound in Corollary 3.1. The column labelled “Optimal ξ_0, ξ_n ” uses the same values for m and n as suggested by Corollary 3.1, but takes the corresponding optimal values for ξ_0 and ξ_n that can be found using an optimization algorithm. The error column again shows the errors for these choices. The column labelled “Optimal m, ξ_0, ξ_n ” takes the optimal value for m and the corresponding value for n given N , and then also the optimal ξ_0 and ξ_n as calculated by optimization. Finally, again the error is given in the error column.

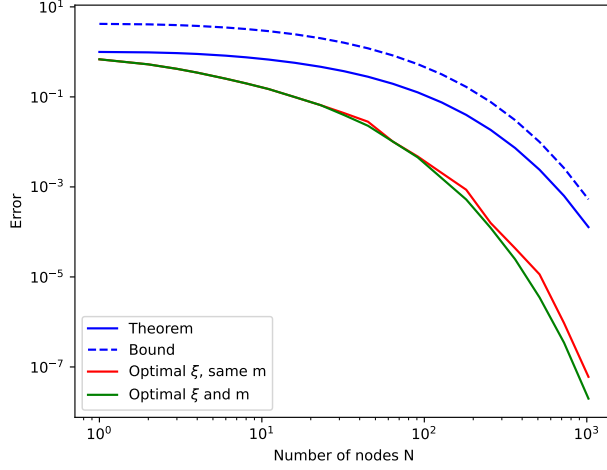


FIGURE 2. L^2 -approximation errors for fractional Brownian with $H = 0.1$ and $T = 1$. “Theorem” is the true error using the values of m , n , ξ_0 and ξ_n specified in Corollary 3.1, while “Bound” is the corresponding error bound of Corollary 3.1. “Optimal ξ , same m ” uses the same m as the one suggested in Corollary 3.1, but the corresponding optimized values for ξ_0 and ξ_n . “Optimal ξ and m ” uses both the optimal values for m and for ξ_0 and ξ_n .

$$\begin{aligned}\xi_0 &= 0.65T^{-1}e^{3.1H} \exp\left(-\frac{1.8}{(3/2-H)A}\sqrt{N}\right), \\ \xi_n &= T^{-1}e^{3H-0.4} \exp\left(\frac{1.8}{HA}\sqrt{N}\right), \\ m &= \frac{0.9}{A}\sqrt{N}.\end{aligned}$$

We even get a corresponding error estimate, given by

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

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 D.

Note that when applying the values of ξ_0 , ξ_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 is the one taken from [AK21, Corollary 3.1], while the improved one is taken from [AK21, Table 6, left column], where we remark that both versions exhibit a polynomial convergence rate of $0.8H$. Harms is the method in [Har19] 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.

N	Alfonsi, Kebaier		Harms		Theorem		Numerical estimates		Optimum
	Original	Improved	$m = 1$	$m = 10$	Error	Bound	Error	Bound	Error
1	1.093956				0.996490	4.190757	0.917761	1.307860	0.683687
2	1.038641	0.656022	1.408506		0.975001	4.089250	0.697745	1.041449	0.528237
4	0.987654	0.471510	1.318878		0.899764	3.773728	0.389907	0.754638	0.346109
8	0.940801	0.290318	1.235692		0.757286	3.218395	0.211681	0.478511	0.199291
16	0.897258	0.151103	1.160442	0.836713	0.571030	2.455046	0.098789	0.251243	0.098625
32	0.856277	0.068764	1.092909	0.632356	0.372303	1.599424	0.041534	0.101018	0.039571
64	0.817406	0.028360	1.032075	0.398268	0.195570	0.833625	0.010345	0.027849	0.010167
128	0.780402	0.010653	0.976870	0.238136	0.075222	0.316916	0.001611	0.004502	0.001559
256	0.745117	0.003793	0.926378	0.150077	0.018481	0.077112	0.000124	0.000342	0.000123
512	0.711447	0.001506	0.879856	0.095788	0.002405	0.009982	3.72e-06	8.94e-06	3.50e-06
1024	0.679308	0.000933	0.836708	0.060342	0.000128	0.000529	2.24e-08	5.16e-08	1.98e-08

TABLE 2. L^2 errors for the approximation of fractional Brownian motion with $H = 0.1$ and $T = 1$ using different point sets. Since the point sets of Harms are only well-defined if $n \geq 2$, there are some gaps in the table. Also, the number N of nodes is only approximate. For example, Harms' point set with $n = 2$ and $m = 10$ has 20 points, but is in the row $N = 16$, as this the value closest to 20.

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 , ξ_0 and ξ_n that we got in this section are to the optimal errors.

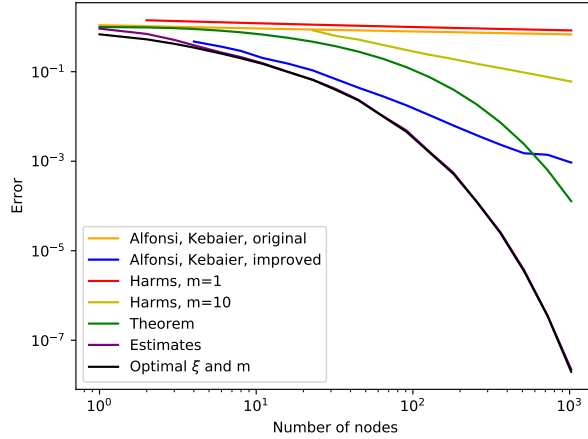


FIGURE 3. L^2 errors for the approximation of fractional Brownian motion with $H = 0.1$ and $T = 1$ using different point sets. “Estimates” refers to the point sets obtained in this section.

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 $T = 1$.

H	Error tolerance	Theorem bound	Theorem	Numerical estimates
0.05	0.1	524	242	50
0.05	0.01	1093	661	161
0.1	0.1	230	108	16
0.1	0.01	512	325	66
0.2	0.1	90	36	4
0.2	0.01	228	139	26

TABLE 3. Number of nodes needed to achieve a certain error for a given Hurst parameter. “Theorem bound” refers to the bound in Theorem 2.15, “Theorem” uses the actual errors of the point sets of Theorem 2.15, and “Numerical estimates” refers uses the point sets with the learned values for m , ξ_0 and ξ_n .

4.3. Implied volatility smile of the rough Bergomi model. We now apply our approximation to the rough Bergomi model, see Section 3.2. We compute the implied volatility using Monte Carlo simulation. We give the algorithm for simulating the Markovian approximation below.

Recall the definition of the rough Bergomi model in (3.1). 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 t^{2H} in the exponent of the volatility corresponds to the square of the L^2 -norm of the aforementioned stochastic integral. Hence, we also replace t^{2H} by the square of the L^2 -norm of the Markovian approximation. We obtain the system

$$\begin{aligned} S_t &= S_0 \exp \left(\int_0^t \sqrt{V_s} \left(\rho dW_s + \sqrt{1-\rho^2} dB_s \right) - \frac{1}{2} \int_0^t V_s ds \right), \\ V_t &= V_0 \exp \left(\eta \sqrt{2H} \Gamma(H+1/2) \sum_{i=0}^N w_i \int_0^t e^{-x_i(t-s)} dW_s \right. \\ &\quad \left. - 2\eta^2 H \Gamma(H+1/2)^2 \int_0^t \left(\sum_{i=0}^N w_i e^{-x_i(t-s)} \right)^2 ds \right). \end{aligned}$$

Now we discretize this system in time. Let $0 = t_0 < \dots < t_m = T$ be a uniform partition with step size $\Delta t = T/m$. Initialize the $N+3$ components \bar{S} , \bar{V} , and $(\bar{V}^i)_{i=0}^N$ by

$$\bar{S}_0 = S_0, \quad \bar{V}_0 = V_0, \quad \bar{V}_0^i = 0.$$

Here, V_t^i corresponds to the integral

$$\int_0^t e^{-x_i(t-s)} dW_s.$$

If we have already simulated the system for time t_j , we then simulate the Gaussian vector

$$\left(\int_{t_j}^{t_{j+1}} e^{-x_0(t_{j+1}-s)} dW_s, \dots, \int_{t_j}^{t_{j+1}} e^{-x_N(t_{j+1}-s)} dW_s, W_{t_{j+1}} - W_{t_j} \right).$$

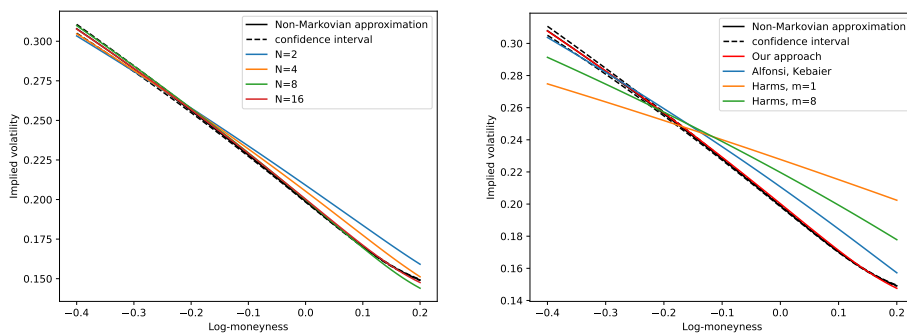
If the grid was chosen uniformly in time, the distribution of this vector is of course independent of j , and we only have to compute a Cholesky decomposition once.

Then, we set

$$\begin{aligned} \bar{S}_{t_{j+1}} &= \bar{S}_{t_j} \exp \left(\sqrt{\bar{V}_{t_j}} \left(\rho(W_{t_{j+1}} - W_{t_j}) + \sqrt{1 - \rho^2}(B_{t_{j+1}} - B_{t_j}) \right) - \frac{\Delta t}{2} \bar{V}_{t_j} \right), \\ \bar{V}_{t_{j+1}}^i &= e^{-x_i \Delta t} \bar{V}_{t_j}^i + \int_{t_j}^{t_{j+1}} e^{-x_i(t_{j+1}-s)} dW_s, \\ \bar{V}_{t_{j+1}} &= V_0 \exp \left(\eta \sqrt{2H} \Gamma(H + 1/2) \sum_{i=0}^N w_i \bar{V}_{t_{j+1}}^i \right. \\ &\quad \left. - 2\eta^2 H \Gamma(H + 1/2)^2 \int_0^t \left(\sum_{i=0}^N w_i e^{-x_i(t-s)} \right)^2 ds \right). \end{aligned}$$

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 4a. 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 [BFG16]. 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.



(A) Approximations based on Theorem 2.15 for different N . (B) Comparison with other methods from the literature, $N = 16$.

FIGURE 4. Implied volatility smiles on the rough Bergomi model for the approximation given in [BFG16] and the Markovian approximations.

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 [BFG16], which can also be formally attained by taking $N \rightarrow \infty$.

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

TABLE 4. Time in seconds to generate the implied volatility smiles for different values of N . Here, $N = \infty$ refers to the method in [BFG16].

We can see in Figure 4a that the choice $N = 16$ already produced good results. We can now compare our approximation with 16 nodes to the approximations given in [AK21] and [Har19] 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 4b we see again the volatility smile generated using the non-Markovian approximation from [BFG16] including the 95% confidence intervals, as well as the volatility smiles generated using our point set, the point set in [AK21, Corollary 3.1], and the point set in [Har19] 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 [AE19b, Section 4.2], i.e.

$$\lambda = 0.3, \quad \rho = -0.7, \quad \nu = 0.3, \quad H = 0.1, \quad V_0 = 0.02, \quad \theta = 0.02, \quad T = 1, \quad S_0 = 1.$$

The implied volatility smiles are computed using Fourier inversion, in the spirit of Section 3.3. As described above, we need to solve (fractional or ordinary multidimensional) Riccati equations. For the fractional Riccati equation we use the Adams scheme, which is well explained in [ER19, Section 5.1]. Essentially, this scheme is a predictor-corrector method.

We now explain how we solve the ordinary multidimensional Riccati equations. As explained in [AE19b, Section 4.1], we have to solve the system

$$\begin{aligned} \widehat{\psi}(t, z) &= \sum_{i=1}^n w_i \psi^{x_i}(t, z), \\ (4.1) \quad \partial_t \psi^{x_i}(t, z) &= -x_i \psi^{x_i}(t, z) + F(z, \widehat{\psi}(t, z)), \quad \psi^{x_i}(0, z) = 0, \end{aligned}$$

where

$$F(z, x) = \frac{1}{2}(-z^2 - iz) + \lambda(\rho\nu iz - 1)x + \frac{(\lambda\nu)^2}{2}x^2.$$

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 x_i can be absurdly large, making any Euler approximation completely intractable. Hence, we only discretize the second summand in the right-hand side of (4.1) and keep the first summand as it is.

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

$$\partial_t \psi^x(t, z) = -x \psi^x(t, z) + F(z, \widehat{\psi}(t, z))$$

on the interval $[t_k, t_{k+1}]$ by

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

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 $\tilde{\psi}^x(t_k, z)$, we set

$$\begin{aligned} \tilde{\psi}^{x,P} &= \frac{F(z, \tilde{\psi}(t_k, z))}{x} \left(1 - e^{-x\Delta} \right) + \tilde{\psi}^x(t_k, z) e^{-x\Delta}, \\ \tilde{\psi}^P &= \frac{1}{2} \left(\tilde{\psi}(t_k, z) + \sum_{i=0}^N w_i \tilde{\psi}^{x_i,P} \right), \\ \tilde{\psi}^x(t_{k+1}, z) &= \frac{F(z, \tilde{\psi}^P)}{x} \left(1 - e^{-x\Delta} \right) + \tilde{\psi}^x(t_k, z) e^{-x\Delta}, \\ \tilde{\psi}(t_{k+1}, z) &= \sum_{i=0}^N w_i \tilde{\psi}^{x_i}. \end{aligned}$$

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

$$(4.2) \quad \mathbb{E}f(X) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(u + iR) \hat{f}(u + iR) du$$

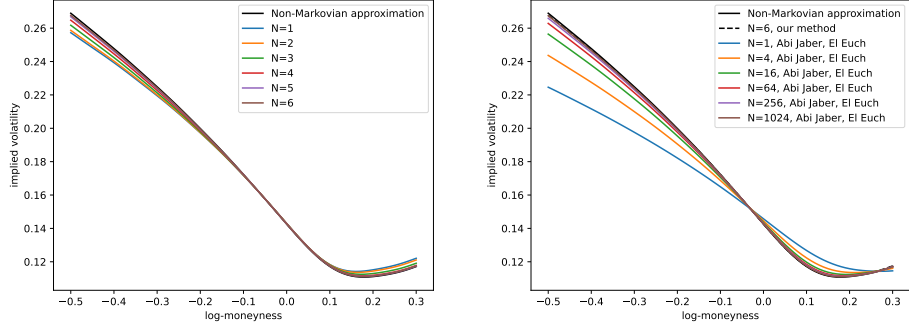
for suitable functions f and random variables X . Here, $\varphi_X(u) = \mathbb{E}e^{iuX}$ is the characteristic function of X , and \hat{f} is the Fourier transform of f . Furthermore, $R \in \mathbb{R}$ is a parameter that has to be chosen suitably to ensure existence of $\varphi_X(u + iR)$ and some integrability conditions on f and \hat{f} . We use this inversion formula with

$$X = \log(S_T)/\log(S_0), \quad \text{and} \quad f(x) = (e^x - K)^+.$$

Moreover, we choose the parameter $R = 2$. We then need to approximate the integral in (4.2). To this end, we somewhat arbitrarily truncate the integral and restrict ourselves to $[-50, 50]$. On this interval, we then use the trapezoidal rule on an equidistant grid with 10000 points. Finally, for both the fractional and the ordinary multidimensional Riccati equations we use 3000 time steps.

Figure 5a 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 [ER19], 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 [AE19b] with a varying numbers of nodes, see Figure 5b.

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,



(A) Approximations based on Theorem 2.15 for different N . (B) Comparison with other methods from the literature.

FIGURE 5. Implied volatility smiles for the rough Heston model using different approximations.

N	1	2	4	8	16	32	64	128	256	512	1024	∞
Time	3105	3119	3102	3135	3241	3393	3592	3866	4387	5458	7384	18567

TABLE 5. Time in seconds to generate the implied volatility smiles for different values of N . Here, $N = \infty$ refers to the method in [ER19].

using 7 nodes if one counts the free node at 0) is better than the approximation in [AE19b] with 1024 points. Since we were able to prove a superpolynomial rate of convergence in Corollary 3.8, and the point set in [AE19b] 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 [AE19b]. 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 σ 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 $N \rightarrow \infty$. The times for our approximation and the approximation by Abi Jaber and El Euch were essentially indistinguishable.

APPENDIX A. 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 [BP11].

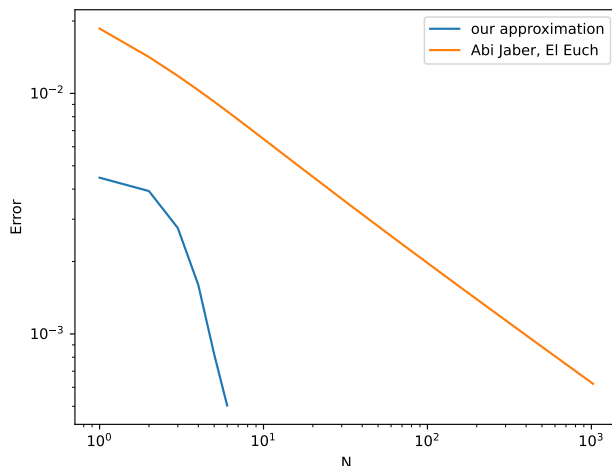


FIGURE 6. Comparison of the approximations errors of the implied volatility smile using our point set and the point set in [AE19b].

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

$$(A.1) \quad \sum_{i=1}^m w_i f(x_i) \approx \int_a^b f(x) w(x) dx,$$

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 (A.1).

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

$$\langle f, g \rangle = \int_a^b f(x) g(x) w(x) dx.$$

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

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

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

$$w_i = \frac{\langle p_{m-1}, p_{m-1} \rangle}{p'_m(x_i) p_{m-1}(x_i)}.$$

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

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. *[BP11, Theorem 4.2.3]. Let $f : [a, b] \rightarrow \mathbb{R}$ be a $(2m)$ -times continuously differentiable function, and let \tilde{w}_i be the weights and \tilde{x}_i be the nodes of the Gaussian quadrature rule for $i = 1, \dots, m$. Then,*

$$c_H \int_a^b f(x) x^{-H-1/2} dx - \sum_{i=1}^m \tilde{w}_i f(\tilde{x}_i) = \int_a^b f^{(2m)}(x) K_{2m}(x) dx,$$

where K_{2m} is the Peano kernel (corresponding to the weight function w).

Lemma A.2. *[BH93, Theorem 2]. The above Peano kernel $K_{2m} : [a, b] \rightarrow \mathbb{R}$ satisfies*

$$\sup_{x \in [a, b]} |K_{2m}(x)| \leq \frac{(2\pi)^{2m}}{(2m)!} \left(\frac{b-a}{2} \right)^{2m} \sup_{x \in [-1, 1]} |B_{2m}(x)| \sup_{x \in [a, b]} |w(x)|,$$

where B_{2m} is the Bernoulli function given by

$$B_s(x) = -2 \sum_{k=1}^{\infty} \frac{\cos(2\pi kx - \frac{\pi s}{2})}{(2\pi k)^s}.$$

Using the fact that the Bernoulli functions agree with the Bernoulli polynomials on $[0, 1]$, we have the following lemma.

Lemma A.3. *[Leh40, Theorem 1]. For even s , we have*

$$\sup_{x \in [0, 1]} |B_s(x)| = \frac{2\zeta(s)}{(2\pi)^s},$$

where ζ is the Riemann zeta function.

APPENDIX B. 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

$$A := A_{\gamma, \delta} := \left(\frac{1}{\delta - 1/2} + \frac{1}{2 - \gamma} \right)^{1/2}, \quad \xi_0 := \exp \left(- \frac{\alpha}{(2 - \gamma)A} \sqrt{N} \right), \quad \xi_n := b \exp \left(\frac{\alpha}{(\delta - 1/2)A} \sqrt{N} \right).$$

Furthermore, α and β are as in Theorem 2.1. Note that this agrees with the choice in 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 B.1. *Let \tilde{w}_i be the weights and \tilde{x}_i be the nodes of the Gaussian quadrature rule for $i = 1, \dots, m$ on the interval $[a, b]$ with respect to the weight function w of type (γ, δ) . Then we have the following error bounds.*

(1) *If $b \leq e^{\alpha\beta}$, then*

$$\left| \int_a^b e^{-tx} w(x) dx - \sum_{i=1}^m \tilde{w}_i e^{-t\tilde{x}_i} \right| \leq Ct^{\gamma-1} \frac{m^{1/2-\gamma}}{2^{2m}} \left(\frac{b}{a} - 1 \right)^{2m+1}.$$

(2) *If $a \geq e^{-\alpha\beta}$, then*

$$\left| \int_a^b e^{-tx} w(x) dx - \sum_{i=1}^m \tilde{w}_i e^{-t\tilde{x}_i} \right| \leq Ct^{\delta-1} \frac{m^{1/2-\delta}}{2^{2m}} \left(\frac{b}{a} - 1 \right)^{2m+1}.$$

Here, C is a constant depending only on w .

The proof of Lemma B.1 is analogous to the proof of Lemma 2.8. Additionally, Lemma 2.9 is replaced by the following lemma.

Lemma B.2. *In the setting of Theorem 2.11, we have*

$$\int_0^T \left| \int_{\xi_0}^{\xi_n} e^{-tx} w(x) dx - \sum_{i=1}^N w_i e^{-t\tilde{x}_i} \right|^2 dt \leq C \frac{n^2 m^{1-2(\gamma \wedge \delta)}}{2^{4m}} \left(e^{\alpha\beta} - 1 \right)^{4m+2}.$$

The proof of Lemma B.2 is again almost the same as the proof of Lemma 2.9. The only difference is that one first needs to note that we always have $\xi_{i+1}/\xi_i = e^{\alpha\beta}$, and that thus, for all intervals $[\xi_i, \xi_{i+1}]$, one of the two cases of Lemma B.1 applies. Afterwards, we proceed as in the proof of Lemma 2.9.

Finally, the proof of Theorem 2.11 follows the same lines as the proof of Theorem 2.1. We also remark that an optimization procedure similar to the one in Theorem 2.15 can be performed to improve the polynomial rate in N in Theorem 2.11.

APPENDIX C. PROOF OF THEOREM 2.15

This proof is split in three steps. First, we determine the dependence of a , 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 not so difficult to see that we have

$$\begin{aligned} \mathbb{E}|X_T - \hat{X}_T|^2 &\leq Cc_H^2 \left(\left(\frac{T^3}{(3/2 - H)^2} a^{3-2H} + \frac{3}{2H^2} b^{-2H} \right) \exp \left(-\frac{2\alpha}{A} \sqrt{N} \right) \right. \\ &\quad \left. + \frac{5\pi^3}{12} \frac{A^{2-2H} T^{2H}}{\beta^{2-2H} H} N^{1-H} \left(\frac{1}{2} \left(\left(\frac{b}{a} \right)^{1/n} e^{\alpha\beta} - 1 \right) \right)^{4m+2} \right). \end{aligned}$$

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 $c := \left(\frac{b}{a}\right)^{1/n}$. Then,

$$\begin{aligned} \left(\frac{1}{2}\left(ce^{\alpha\beta} - 1\right)\right)^{4m+2} &= \left(\frac{1}{2}\left(\frac{ce^{\alpha\beta} - 1}{e^{\alpha\beta} - 1}\left(e^{\alpha\beta} - 1\right)\right)\right)^{4m+2} \\ &= \left(\frac{1}{2}\left(e^{\alpha\beta} - 1\right)\right)^{4m+2} \left(\frac{ce^{\alpha\beta} - 1}{e^{\alpha\beta} - 1}\right)^{4m+2}. \end{aligned}$$

Therefore, by our choice of α and β ,

$$\begin{aligned} \mathbb{E}|X_T - \widehat{X}_T|^2 &\leq Cc_H^2 T^{2H} \left(\frac{1}{(3/2 - H)^2} a^{3-2H} + \frac{3}{2H^2} b^{-2H} \right. \\ &\quad \left. + \frac{5\pi^3}{48} \left(e^{\alpha\beta} - 1\right)^2 \frac{A^{2-2H}}{\beta^{2-2H}H} N^{1-H} \left(\frac{ce^{\alpha\beta} - 1}{e^{\alpha\beta} - 1}\right)^{4m+2} \right) \exp\left(-\frac{2\alpha}{A}\sqrt{N}\right). \end{aligned}$$

Note that

$$\left(\frac{ce^{\alpha\beta} - 1}{e^{\alpha\beta} - 1}\right)^{4m+2} = \left(1 + (c-1)\frac{e^{\alpha\beta}}{e^{\alpha\beta} - 1}\right)^{4m+2}.$$

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,

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

Then,

$$\begin{aligned} \left(1 + (c-1)\frac{e^{\alpha\beta}}{e^{\alpha\beta} - 1}\right)^{4m+2} &\leq \left(1 + \frac{e^{\alpha\beta}}{e^{\alpha\beta} - 1} \left(\log c + \frac{(\log c)^2}{2}\right)\right)^{4m+2} \\ &= \left(1 + \frac{e^{\alpha\beta}}{e^{\alpha\beta} - 1} \left(\frac{\beta}{A\sqrt{N}} \log \frac{b}{a} + \frac{\beta^2}{2A^2N} \left(\log \frac{b}{a}\right)^2\right)\right)^{4m+2}. \end{aligned}$$

Assume that N is so large that

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

i.e.

$$(C.1) \quad N \geq \left(\frac{\beta}{A} \log \frac{b}{a}\right)^2.$$

Then,

$$\begin{aligned} &\left(1 + \frac{e^{\alpha\beta}}{e^{\alpha\beta} - 1} \left(\frac{\beta}{A\sqrt{N}} \log \frac{b}{a} + \frac{\beta^2}{2A^2N} \left(\log \frac{b}{a}\right)^2\right)\right)^{4m+2} \\ &\leq \left(1 + \frac{e^{\alpha\beta}}{8(e^{\alpha\beta} - 1)} \frac{4\beta}{A\sqrt{N}} \log \frac{b}{a}\right)^{\frac{4\beta}{A}\sqrt{N}+2} \\ &\leq \exp\left(\frac{e^{\alpha\beta}}{8(e^{\alpha\beta} - 1)} \log \frac{b}{a}\right) \left(1 + \frac{e^{\alpha\beta}}{8(e^{\alpha\beta} - 1)} \frac{4\beta}{A\sqrt{N}} \log \frac{b}{a}\right)^2 \\ &\leq \left(\frac{b}{a}\right)^{\frac{e^{\alpha\beta}}{8(e^{\alpha\beta} - 1)}}. \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{E}|X_T - \widehat{X}_T|^2 &\leq C c_H^2 T^{2H} \left(\frac{1}{(3/2 - H)^2} a^{3-2H} + \frac{3}{2H^2} b^{-2H} \right. \\ &\quad \left. + \frac{5\pi^3}{48} (e^{\alpha\beta} - 1)^2 \frac{A^{2-2H}}{\beta^{2-2H} H} N^{1-H} \left(\frac{b}{a} \right)^{\frac{e^{\alpha\beta}}{8(e^{\alpha\beta}-1)}} \right) \exp\left(-\frac{2\alpha}{A} \sqrt{N}\right). \end{aligned}$$

We now want to solve the optimization problem

$$\min_{0 < b \leq a} A_1 b^{-q_1} + A_2 \left(\frac{b}{a} \right)^{q_2} + A_3 a^{q_3}.$$

This is minimized in

$$\begin{aligned} a &= \left((q_1 A_1)^{q_2} (q_2 A_2)^{q_1} (q_3 A_3)^{-(q_1+q_2)} \right)^{\frac{1}{q_1 q_2 + q_1 q_3 + q_2 q_3}}, \\ b &= \left((q_1 A_1)^{q_2+q_3} (q_2 A_2)^{-q_3} (q_3 A_3)^{-q_2} \right)^{\frac{1}{q_1 q_2 + q_1 q_3 + q_2 q_3}}, \end{aligned}$$

where,

$$\begin{aligned} q_1 &= 2H, & q_2 &= \frac{e^{\alpha\beta}}{8(e^{\alpha\beta} - 1)}, & q_3 &= 3 - 2H, \\ A_1 &= \frac{3}{2H^2}, & A_2 &= \frac{5\pi^3}{48} (e^{\alpha\beta} - 1)^2 \frac{A^{2-2H}}{\beta^{2-2H} H} N^{1-H}, & A_3 &= \frac{1}{(3/2 - H)^2}. \end{aligned}$$

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

Step 3: We now need to verify that $b < a$, and that (C.1) holds. We start with $b < a$. Indeed,

$$\frac{b}{a} = \left(\left(\frac{5\pi^3}{1152} e^{\alpha\beta} (e^{\alpha\beta} - 1) \frac{A^{2-2H}}{\beta^{2-2H}} N^{1-H} \right)^{-3} \left(\frac{H}{3(3/2 - H)} \right)^{2H} \right)^{\frac{1}{\frac{3e^{\alpha\beta}}{8(e^{\alpha\beta}-1)} + 6H - 4H^2}} \leq 1$$

if and only if

$$N \geq \left(\frac{5\pi^3}{1152} e^{\alpha\beta} (e^{\alpha\beta} - 1) \frac{A^{2-2H}}{\beta^{2-2H}} \right)^{-3} \left(\frac{H}{3(3/2 - H)} \right)^{\frac{2H}{3-3H}}.$$

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 (C.1), which we also need to verify. We have

$$\begin{aligned} \left| \frac{\beta}{A} \log \frac{b}{a} \right| &= \left| \frac{\beta}{A} \frac{1}{\frac{3e^{\alpha\beta}}{8(e^{\alpha\beta}-1)} + 6H - 4H^2} \log \left(\left(\frac{5\pi^3}{1152} e^{\alpha\beta} (e^{\alpha\beta} - 1) \frac{A^{2-2H}}{\beta^{2-2H}} N^{1-H} \right)^{-3} \left(\frac{H}{3(3/2 - H)} \right)^{2H} \right) \right| \\ &\leq \frac{\beta}{A} \frac{3}{\frac{3e^{\alpha\beta}}{8(e^{\alpha\beta}-1)} + 6H - 4H^2} \log \left(\frac{5\pi^3}{1152} e^{\alpha\beta} (e^{\alpha\beta} - 1) \frac{A^{2-2H}}{\beta^{2-2H}} N \right) \\ &\leq \frac{8\beta(e^{\alpha\beta} - 1)}{Ae^{\alpha\beta}} \log \left(\frac{5\pi^3}{1152} e^{\alpha\beta} (e^{\alpha\beta} - 1) \frac{A^2}{\beta^2} \right) + \frac{8\beta(e^{\alpha\beta} - 1)}{Ae^{\alpha\beta}} \log N. \end{aligned}$$

Plugging in our values of α and β , 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 D. LEARNING A BETTER RATE OF CONVERGENCE

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

$$(D.1) \quad \xi_0 = C_1 T^{-1} \exp\left(-\frac{\alpha}{(3/2 - H)A} \sqrt{N}\right),$$

$$(D.2) \quad \xi_n = C_2 T^{-1} \exp\left(-\frac{\alpha}{HA} \sqrt{N}\right),$$

$$(D.3) \quad m = \frac{\beta}{A} \sqrt{N},$$

$$(D.4) \quad \int_0^T |G(t) - \widehat{G}(t)|^2 dt = C_3 T^{2H} \exp\left(-\frac{2\alpha}{A} \sqrt{N}\right),$$

and our goal is to learn α and β in particular, and get some approximation for C_1 , C_2 and C_3 , which may depend on H and N . Here, (D.4) philosophically has a different meaning, as the left-hand side, i.e. the true error, is determined by our choice for ξ_0 , ξ_n and m , and need not be of the form of the right-hand side of (D.4). Nonetheless, we still choose to use the errors as well to estimate α with equation (D.4).

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 , ξ_0 and ξ_n , as well as the corresponding error, using an optimization algorithm. Then we use this data to estimate α and β .

First, using the optimal values of m , we can determine the corresponding values of β by reformulating equation (D.3). Empirically, we observed that paths of the estimated values of β for different choices of H and N indeed converge to the same number, independent of H , and we estimate $\beta = 0.9$.

For α , we can use ξ_0 , ξ_n and the error as estimators. In other words, we set $C_i = 1$, and solve equations (D.1), (D.2), and (D.4) for α . 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 α and β into equations (D.1)-(D.4) and apply regression to get estimates for the constants C_1 , C_2 and C_3 . We get the results of Section 4.2.

REFERENCES

- [AE19a] Eduardo Abi Jaber and Omar El Euch. ‘‘Markovian structure of the Volterra Heston model’’. In: *Statistics & Probability Letters* 149 (2019), pp. 63–72.

- [AE19b] Eduardo Abi Jaber and Omar El Euch. “Multifactor approximation of rough volatility models”. In: *SIAM Journal on Financial Mathematics* 10.2 (2019), pp. 309–349.
- [AK21] Aurélien Alfonsi and Ahmed Kebaier. “Approximation of Stochastic Volterra Equations with kernels of completely monotone type”. In: (2021). arXiv: 2102.13505 [math.PR].
- [Ber05] Lorenzo Bergomi. “Smile dynamics II”. In: *Risk October* (2005), pp. 67–73.
- [BFG16] Christian Bayer, Peter Friz, and Jim Gatheral. “Pricing under rough volatility”. In: *Quantitative Finance* 16.6 (2016), pp. 887–904.
- [BH93] H Brass and G Hämmerlin. *Numerical integration IV (Oberwolfach, 1992)*. Birkhäuser Verlag, 1993.
- [BHT20] Christian Bayer, Eric Joseph Hall, and Raúl Tempone. “Weak error rates for option pricing under the rough Bergomi model”. In: *arXiv preprint arXiv:2009.01219* (2020).
- [BP11] Helmut Brass and Knut Petras. *Quadrature theory: the theory of numerical integration on a compact interval*. 178. American Mathematical Soc., 2011.
- [CC98] Philippe Carmona and Laure Coutin. “Fractional Brownian motion and the Markov property”. In: *Electronic Communications in Probability* 3 (1998), pp. 95–107.
- [Che17] Laurent Chevillard. “Regularized fractional Ornstein-Uhlenbeck processes and their relevance to the modeling of fluid turbulence”. In: *Physical review E* 96.3 (2017), p. 033111.
- [DF98] Kai Diethelm and Alan D Freed. “The FracPECE subroutine for the numerical solution of differential equations of fractional order”. In: *Forschung und wissenschaftliches Rechnen 1999* (1998), pp. 57–71.
- [DF02] Kai Diethelm, Neville J Ford, and Alan D Freed. “A predictor-corrector approach for the numerical solution of fractional differential equations”. In: *Nonlinear Dynamics* 29.1 (2002), pp. 3–22.
- [DF04] Kai Diethelm, Neville J Ford, and Alan D Freed. “Detailed error analysis for a fractional Adams method”. In: *Numerical algorithms* 36.1 (2004), pp. 31–52.
- [EGR19] Omar El Euch, Jim Gatheral, and Mathieu Rosenbaum. “Roughening heston”. In: *Risk* (2019), pp. 84–89.
- [ER18] Omar El Euch and Mathieu Rosenbaum. “Perfect hedging in rough Heston models”. In: *The Annals of Applied Probability* 28.6 (2018), pp. 3813–3856.
- [ER19] Omar El Euch and Mathieu Rosenbaum. “The characteristic function of rough Heston models”. In: *Mathematical Finance* 29.1 (2019), pp. 3–38.
- [GJR18] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. “Volatility is rough”. In: *Quantitative finance* 18.6 (2018), pp. 933–949.
- [Har19] Philipp Harms. “Strong convergence rates for Markovian representations of fractional Brownian motion”. In: *arXiv preprint arXiv:1902.01471* (2019).
- [Leh40] Derrick H Lehmer. “On the maxima and minima of Bernoulli polynomials”. In: *The American Mathematical Monthly* 47.8 (1940), pp. 533–538.

- [LT09] Changpin Li and Chunxing Tao. “On the fractional Adams method”. In: *Computers & Mathematics with Applications* 58.8 (2009), pp. 1573–1588.
- [Mar13] Nicolas Marie. “A pathwise fractional one compartment intra-veinous bolus model”. In: *arXiv preprint arXiv:1311.1647* (2013).

WEIERSTRASS INSTITUTE, MOHRENSTRASSE 39, 10117 BERLIN, GERMANY
Email address: christian.bayer@wias-berlin.de

WEIERSTRASS INSTITUTE, MOHRENSTRASSE 39, 10117 BERLIN, GERMANY
Email address: simon.breineis@wias-berlin.de