Weak error rates for rough volatility numerics

Paul Gassiat

CEREMADE, Université Paris-Dauphine

Pau,
December 2022

Outline

Outline

1. Rough volatility
2. Weak error rates for left-point discretization
3. Weak error rates for the hybrid scheme

Rough volatility models

Stochastic volatility models, the (discounted) asset price has dynamics (under pricing measure)

$$dS_t = \sigma_t S_t dW_t,$$

and σ_t is a process with **rough** sample paths, typically modelled around a fractional Brownian motion W^H with $H \in (0, \frac{1}{2})$, e.g.

$$\sigma_t = f(t, W_t^H),$$

(where W^H and W are correlated).

Bayer-Friz-Gatheral '16 : rough Bergomi model, $f(t,x) = \zeta(t)\exp(\eta x)$.

fractional Brownian motion

W^H **fractional Brownian motion** (fBm) with **Hurst parameter** $H \in (0,1)$.

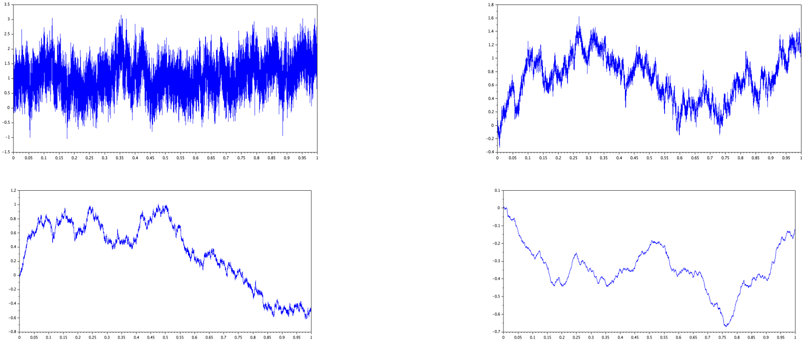
- Riemann-Liouville fBm :

$$W_t^H = C_H \int_0^t (t-s)^{H-\frac{1}{2}} dW_s,$$

  where W standard BM.
- Gaussian process, explicit covariance function.
- sample paths are $(H-\varepsilon)$-Hölder continuous ($H < \frac{1}{2}$: rough(er than standard BM) regime).
- NOT a semimartingale, NOT a Markov process (for $H \neq \frac{1}{2}$).

fractional Brownian motion : sample paths

W^H **fractional Brownian motion** (fBm) with **Hurst parameter** $H \in (0, 1)$.
Simulated sample paths with $H \in \{0.1, 0.35, 0.5, 0.75\}$:



Advantages of rough volatility models

- **historical time-series** of observed prices are consistent with a rough volatility process with $H \approx 0.1$ (Gatheral-Jaisson-Rosenbaum '16,...)
- can reproduce features of observed **volatility surfaces**, in particular explosion of ATM skew $\tau^{H-1/2}$, again $H \approx 0.1$ (Alos-Leon-Vives '07, Fukasawa '11)
- arises as scaling limit of **microstructure** models under natural assumptions (El Euch-Fukasawa-Rosenbaum '18, Jusselin-Rosenbaum '20).

A lot of research activity in the last 7 years
(cf. `http://sites.google.com/site/roughvol`)

However, lack of Markov (and semimartingale) property leads to complications, from both theoretical and practical perspectives.

Rough paths and rough volatility

• In general, does not fall in classical rough path setting. But can still have a rough path type approach

$$\int_0^T \sigma(\hat{W}_t) dW_t \approx \sum_i \sum_{0 \leqslant k \leqslant K} \frac{\sigma^{(k)}(\hat{W}_{t_i})}{k!} \left(\int_{t_i}^{t_{i+1}} \hat{W}^k_{t_i,s} dW_s \right)$$

cf. [Bayer, Friz, G., Martin, Stemper, *MF*, '20], see also Harang-Tindel-Wang, Bruned-Katsetsiadis, Fukasawa-Takano for related recent results.

• Useful to obtain **precise large deviation estimates**, cf. [Friz, G. , Pigato, *AAP*, '21], [Friz, G. , Pigato, *QF*, '22],

$$c(t, k_t) \sim_{t \to 0} \exp\left(- \frac{\Lambda(x)}{t^{2H}} \right) t^{1/2+2H} \frac{A(x)}{2\Lambda'(x)\sqrt{\Lambda(x)}\sqrt{\pi}}$$

where we combine Laplace method on Wiener space with rough path type methods (following Azencott, Ben Arous, Aida, Inahama,...).

Outline

1. Rough volatility
2. Weak error rates for left-point discretization
3. Weak error rates for the hybrid scheme

Motivation : rough volatility pricing

Rough Bergomi (type) model :

$$dS_t/S_t = \sigma_t(\rho dW_t + \bar{\rho} d\bar{W}_t), \quad \sigma_t = f(t, \hat{W}_t),$$

$$\hat{W}_t = \int_0^t K(t,s) dW_s, \quad K(t,s) = (t-s)_+^{H-1/2}.$$

Call prices in this model given by (Romano-Touzi formula)

$$\mathbb{E}\left[C_{BS}\left(S_0 \exp\left(\rho \int_0^T f(t,\hat{W}_t) dW_t - \frac{\rho^2}{2}\int_0^T f(t,\hat{W}_t)^2 dt \right), K, \frac{\bar{\rho}^2}{2}\int_0^T f(t,\hat{W}_t)^2 dt \right) \right]$$

$$=: \mathbb{E}\left[\Phi\left(\mathcal{I}, \mathcal{J}\right)\right]$$

where

$$\mathcal{I} = \int_0^T f(t,\hat{W}_t) dW_t, \quad \mathcal{J} = \int_0^T f(t,\hat{W}_t)^2 dt$$

Numerical simulation of $\mathcal{I}$

$$\mathcal{I} = \int_0^T f(\hat{W}_t)dW_t, \quad \hat{W}_t = \int_0^t (t-s)^{H-1/2} dW_s.$$

Discretization : N time steps, $h = \frac{T}{N}$, $t_k = kh$, $k = 0, \ldots, N$.

$$\mathcal{I} \;\approx\; \hat{\mathcal{I}}^N := \sum_{k=0}^{N-1} f(\hat{W}_{t_k}) \left(W_{t_{k+1}} - W_{t_k} \right).$$

The Gaussian vector

$$\left(\hat{W}_{t_k}, W_{t_k} \right)_{k=1,\ldots,N}$$

has explicit covariance and can be simulated exactly by the Cholesky method (cost : $O(N^2)$ + preliminary $O(N^3)$).

Strong error vs weak error

$$\mathcal{I} = \int_0^T f(\hat{W}_t) dW_t, \quad \hat{\mathcal{I}} = \sum_k f(\hat{W}_{t_k}) \left(W_{t_{k+1}} - W_{t_k}\right) = \int_0^T f(\hat{W}_{\eta(t)}) dW_t$$

with $\eta(t) = h\lfloor t/h \rfloor$.
Strong error

$$\mathbb{E}\left[\left(\mathcal{I} - \hat{\mathcal{I}}\right)^2\right] = \int_0^T \mathbb{E}\left[\left(f(\hat{W}_t) - f(\hat{W}_{\eta(t)})\right)^2\right] dt \sim N^{-2H}$$

$\rightarrow$ strong order H, very slow convergence for H close to 0 !

But what about **weak** error ? i.e. given test function Φ,

$$\mathcal{E}_\Phi := \mathbb{E}\left[\Phi(\mathcal{I})\right] - \mathbb{E}\left[\Phi(\hat{\mathcal{I}})\right] \lesssim N^{-??}$$

Recall classical case $H = 1/2$, strong rate $= 1/2$, weak rate $= 1$ (classical works of Talay and co-authors).
For general H ? Guesses : $2H$, $H + 1/2$, 1, ... ?

A heuristic computation suggesting rate $2H$

Asymptotic distribution of the strong error :

$$\Delta\mathcal{I} = \mathcal{I} - \hat{\mathcal{I}} = \int_0^T \left(f(\hat{W}_t) - f(\hat{W}_{\eta(t)}) \right) dW_t \approx V_{f,H}(\omega) Z,$$

where $V_{f,H}(\omega) = C_H (\int_0^T f'(\hat{W}_t)^2 dt)^{1/2} N^{-H}$ and $Z \sim \mathcal{N}(0,1)$, $Z \perp\!\!\!\perp W$.
(stable convergence, cf. Rootzen '80,...).
This leads to

$$\begin{aligned}\mathbb{E}\left[\Phi(\mathcal{I})\right] - \mathbb{E}\left[\Phi(\hat{\mathcal{I}})\right] &\approx \mathbb{E}\left[\Phi'(\mathcal{I})\Delta\mathcal{I}\right] + \frac{1}{2}\mathbb{E}\left[\Phi''(\hat{\mathcal{I}})(\Delta\mathcal{I})^2\right] \\ &\approx \mathbb{E}\left[\Phi'(\mathcal{I})V_{f,H}\right]\mathbb{E}\left[Z\right] + \frac{1}{2}\mathbb{E}\left[\Phi''(\mathcal{I})V_{f,H}^2\right]\mathbb{E}\left[Z^2\right] \\ &\sim C'_{f,H,\Phi} N^{-2H}\end{aligned}$$

This suggests rate $2H$...which turns out to be wrong !

Weak error rate for quadratics ($\Phi(x) = x^2$)

(Computation taken from Bayer-Hall-Tempone '20, attributed to Neuenkirch)

$$\mathbb{E}[\mathcal{I}^2] = \int_0^T \mathbb{E}[f(\hat{W}_t)^2]dt = \int_0^T \phi_f(t)dt$$

where $\phi_f(t) = \mathbb{E}[f^2(c_H t^H Z)]$, $Z \sim \mathcal{N}(0,1)$.
Similarly,

$$\mathbb{E}[\hat{\mathcal{I}}^2] = \int_0^T \phi_f(\eta(t))dt.$$

But ϕ is locally Lipschitz, more precisely $\partial_t \phi_f(t) \lesssim t^{H-1}$ and

$$\mathbb{E}[\mathcal{I}^2] - \mathbb{E}[\hat{\mathcal{I}}^2] \leqslant \frac{T}{N} \int_0^T |\partial_t \phi_f(t)| dt \leqslant \frac{C}{N}$$

For quadratics, weak rate is $= 1$, much better than $2H$!

Main result : weak rate is $(3H + 1/2) \wedge 1$ (in some cases)

Theorem (G. '22)

Assume that either :

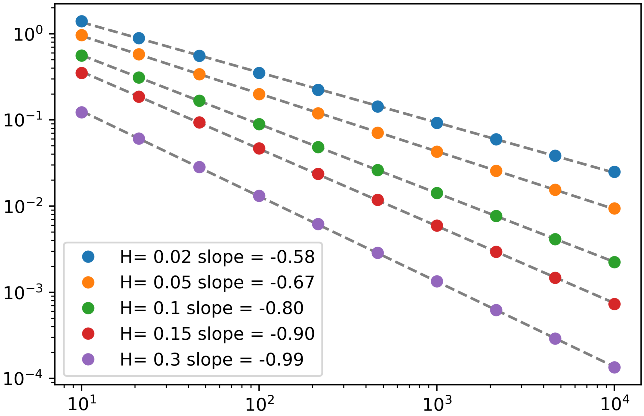
1. $f(x) = x$, *and* $\Phi \in C_b^{\lceil \frac{1}{2H} \rceil + 4}$,
2. $f \in C_b^2$, $\Phi(x) = x^3$.

Then it holds that

$$\mathcal{E}_\Phi \leqslant C N^{-(3H+1/2)\wedge 1}$$

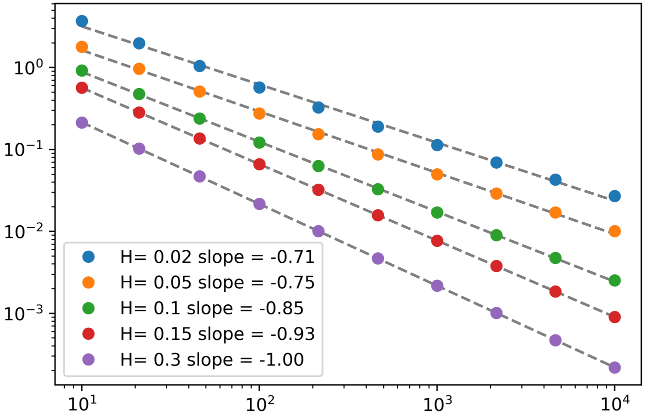
- Bayer-Hall-Tempone '20 had already proved that (in case (1)), weak rate was at least $H + 1/2$, using PDE methods.
- The proof uses a direct method based on Malliavin integration by parts, following Clément-Kohatsu-Higa-Lamberton '06.
- Recently, case (2) has been generalized to Φ arbitrary polynomial by Friz-Salkeld-Wagenhofer. They also prove optimality of the rate.

Numerical illustration 1



log-log plot of $\mathcal{E}_\Phi$ as function of N for $\Phi(x) = x^3$, $f(x) = x$.
(Note : $\mathbb{E}[\mathcal{I}^3]$, $\mathbb{E}[\hat{\mathcal{I}}^3]$ can be exactly computed in this case)

Numerical illustration 2



log-log plot of $\mathcal{E}_\Phi$ as function of N for $\Phi(x) = (x+2)^3$, $f(x) = x$.

Elements of proof : Case (1)

Some notations : $\eta(t) = \lfloor nt \rfloor / n$,

$$K(t,s) = (t-s)_+^{H-1/2}, \quad K'(t,s) = K(\eta(t), s), \;\; \Delta K = K' - K$$

$$\mathcal{I} = \int_0^1 \hat{W}_t dW_t, \quad \hat{\mathcal{I}} = \int_0^1 \hat{W}_{\eta(t)} dW_t,$$

$$\Delta\mathcal{I} = \hat{\mathcal{I}} - \mathcal{I} = \int_0^1 \Delta\hat{W}_t dW_t,$$

$$\text{and for } \theta \in [0,1], \;\; \mathcal{I}^\theta = (1-\theta)\hat{\mathcal{I}} - \theta\mathcal{I} = \int_0^1 \hat{W}_t^\theta dW_t.$$

We have

$$\mathcal{E}_\Phi = \int_0^1 d\theta \, \mathbb{E}\left[\Phi'(\mathcal{I}^\theta)\Delta\mathcal{I}\right].$$

Applying the Malliavin integration by parts twice (note $\Delta\mathcal{I}$ is a double Wiener integral) :

$$\begin{aligned}\mathbb{E}\left[\Phi'(\mathcal{I}^\theta)\Delta\mathcal{I}\right] &= \int_0^T dt \int_0^t ds\ \mathbb{E}\left[D_s D_t \Phi'(\mathcal{I}^\theta)\right] \Delta K(t,s) \\ &= \int_0^T dt \int_0^t ds\ \mathbb{E}\left[\Phi^{(3)}(\mathcal{I}^\theta)(D_s\mathcal{I}^\theta)(D_t\mathcal{I}^\theta)\right] \Delta K(t,s) \\ &\quad + \int_0^T dt \int_0^t ds\ \mathbb{E}\left[\Phi''(\mathcal{I}^\theta)\right] K^\theta(t,s)\Delta K(t,s)\end{aligned}$$

Using the continuity (on average) properties of

$$(s,t) \mapsto \mathbb{E}\left[\Phi^{(3)}(\mathcal{I}^\theta)(D_s\mathcal{I}^\theta)(D_t\mathcal{I}^\theta)\right],$$

the first term is of order $N^{-3H-1/2}$.

The second term (integrated in θ) is treated by an induction procedure , using

$$\int\int (\Delta K(t,s))^2 ds dt \lesssim N^{-2H}, \quad \int\int \Delta(K^2)(t,s) ds dt \lesssim N^{-1}.$$

Elements of proof : Case (2)

Relies on the identity

$$\mathbb{E}\left(\int_0^T f(\hat{W}_t)dW_t\right)^3 = \int_0^T dt \int_0^t ds\, \mathbb{E}\left[f(\hat{W}_s)(ff')(\hat{W}_t)\right] K(t,s),$$

and again, the continuity properties in (s,t) of the expectation in the integral.

Outline

1. Rough volatility
2. Weak error rates for left-point discretization
3. Weak error rates for the hybrid scheme

Hybrid scheme

Cholesky method is slow. ($O(N^2)$)
Bennedsen-Lunde-Pakkanen '16 proposed a faster **hybrid scheme**.
The idea is to approximate for a grid-point $t = t_k$,

$$\hat{W}_t = \int_0^t (t-s)^{H-1/2} dW_s$$

by

$$\check{W}_t = \int_{t-\kappa h}^t (t-s)^{H-1/2} dW_s + \sum_{j=0}^{k-\kappa-1} \check{K}_{k-j} \left(\int_{t_j}^{t_{j+1}} dW_s \right)$$

where κ is a fixed (small) parameter, and $\check{K}_{k-j} \approx (t_k - s)^{H-1/2}$ for $s \in [t_j, t_{j+1}]$
(This requires to simulate N independent copies of a $\kappa + 1$-dimensional Gaussian, and perform a convolution $\rightarrow O(N \log N)$).
We then consider

$$\check{\mathcal{I}} = \int_0^T f(\check{W}_{\eta(t)}) dW_t.$$

Need to choose $\check{K}_\ell$ to approximate $r^{H-1/2}$, $r \in [(\ell-1)h, \ell h]$.
Possible choice of weights :

- left-point : $K_\ell = (\ell h)^{H-1/2}$
- mid-point : $K_\ell = ((\ell - 1/2)h)^{H-1/2}$
- minimizing MSE : $K_\ell = h^{-1} \int_{(\ell-1)h}^{\ell h} r^{H-1/2} dr$.
- matching the second moment : $K_\ell = \left(h^{-1} \int_{(\ell-1)h}^{\ell h} (r^{H-1/2})^2 dr \right)^{1/2}$

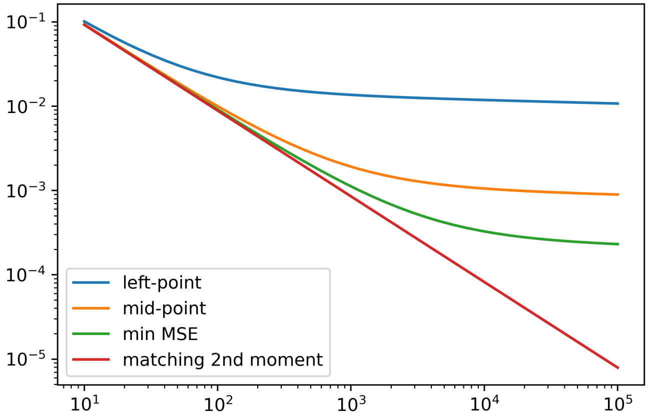
For the first three choices, $\mathbb{E}[\check{W}_t^2] - \mathbb{E}[\hat{W}_t^2] \sim N^{-2H}$, which leads to weak error rate of order $2H$ for quadratics $\to$ not good !

For the last choice however, $\mathbb{E}[\check{W}_t^2] = \mathbb{E}[\hat{W}_t^2]$ for grid-points t, and, like in the Cholesky case, this gives

$$\mathbb{E}[\mathcal{I}^2] - \mathbb{E}[\check{\mathcal{I}}^2] \lesssim N^{-1}.$$

(This choice of weights was first proposed in Horvath-Jacquier-Muguruza '17)

Choice of weights : Numerical illustration



log-log plot of $\mathcal{E}_{x^2}$ as a function of N for $H = 0.02$, $\kappa = 1$.

Weak rate for hybrid scheme is $H + 1/2$

Theorem (G. '22)

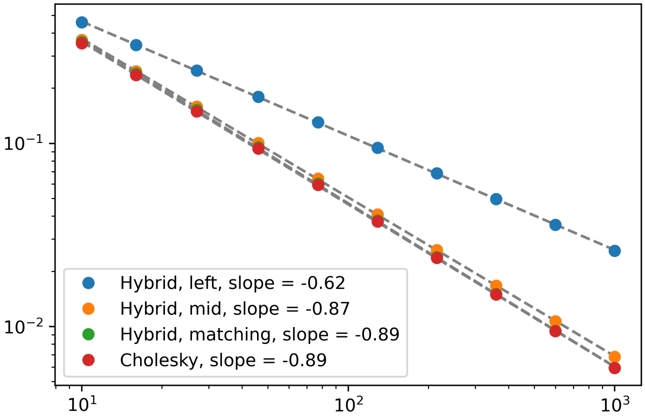
Assume that either :

1. $f(x) = x$, *and* $\Phi \in C_b^{\lceil \frac{1}{2H} \rceil + 4}$,
2. $f \in C_b^3$, $\Phi(x) = x^3$,

Then, with second moment matching weights, it holds that

$$\mathbb{E}[\Phi(\mathcal{I})] - \mathbb{E}[\Phi(\check{\mathcal{I}})] \leqslant CN^{-(H+1/2)}$$

- The proof is similar as before
- The weaker rate comes from the fact that it is not possible to match both second and first moments of the kernel.
- Practical remark : for cubic test functions, the constant in front of $N^{-(H+1/2)}$ can be computed and is very small. In that case, for practical values of N the weak error of the hybrid scheme seems roughly the same as that of the Cholesky method.



log-log plot of $\mathcal{E}_{x^3}$ as a function of N for $H = 0.15$, $\kappa = 1$.

Conclusion

Discretization error of rough volatility models :

- Strong rate : always H
- Weak rate : $(3H + 1/2) \wedge 1$ for exact left-point discretization, $H + 1/2$ for the hybrid scheme with well-chosen weights, $2H$ for the hybrid scheme with "bad" choice of weights.

Future work and open questions :

- Weak error for less smooth test functions ?
- Practical implications ?
- Can we obtain (weak rate) $>>$ $2\times$(strong rate) in other contexts ?