

PATHWISE STABILITY OF LIKELIHOOD ESTIMATORS FOR DIFFUSIONS VIA ROUGH PATHS

BY JOSCHA DIEHL^{*,†,1}, PETER FRIZ^{*,‡,2} AND HILMAR MAI^{§,2}

TU Berlin^{*}, *University of California, San Diego*[†], *WIAS Berlin*[‡]
 and *ENSAE ParisTech*[§]

We consider the classical estimation problem of an unknown drift parameter within classes of nondegenerate diffusion processes. Using rough path theory (in the sense of T. Lyons), we analyze the Maximum Likelihood Estimator (MLE) with regard to its pathwise stability properties as well as robustness toward misspecification in volatility and even the very nature of the noise. Two numerical examples demonstrate the practical relevance of our results.

1. Introduction. Let W be d -dimensional Wiener process and $A \in \mathbb{V}$, some fixed finite-dimensional vector space. Consider sufficiently regular $h : \mathbb{R}^d \rightarrow L(\mathbb{V}, \mathbb{R}^d)$ and $\Sigma : \mathbb{R}^d \rightarrow L(\mathbb{R}^d, \mathbb{R}^d)$ so that³

$$(1.1) \quad dX_t = h(X_t)A dt + \Sigma(X_t) dW_t$$

has a unique solution, started from $X_0 = x_0$. Throughout we shall assume nondegeneracy of the matrix-valued diffusion coefficient Σ . The important example of multidimensional Ornstein–Uhlenbeck dynamics, for instance, falls in the class of diffusions considered here. We are interested in estimating the drift parameter A , given some observation sample path $\{X_t(\omega) = \omega_t : t \in [0, T]\}$. To this end, we consider the classical Maximum Likelihood Estimator (MLE), of the form

$$\hat{A}_T = \hat{A}_T(\omega) = \hat{A}_T(X(\omega)) \in \mathbb{V}$$

relative to the reference measure given by the law of the drift-free process. Note that \hat{A}_T is a functional on pathspace $C([0, T], \mathbb{R})$: for every (observation) sample path $X(\omega) = \omega$ one has a corresponding estimate $\hat{A}_T(X(\omega))$. Let us also recall that these MLEs are based on the Girsanov density of the pathspace measures, with- versus without-drift, respectively, see, for example, the standard text books of Kutoyants [12] or Liptser–Shiryaev [13]. It will be instructive to consider the simplest possible example with its fully explicit solution.

Received March 2015.

¹Supported by DFG, SPP 1324.

²Supported in part by European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement nr. 258237.

MSC2010 subject classifications. 62M05, 62F99; secondary 60H99.

Key words and phrases. Maximum likelihood estimation, robust estimation, rough paths analysis.

³ $L(V, W)$ denote the space of linear maps between two finite-dimensional vector spaces and may be identified with a space of matrices.

EXAMPLE 1 (Scalar Ornstein–Uhlenbeck process). In (1.1), take $d = 1$, $h(x) = x$, $\Sigma \equiv \sigma > 0$ and scalar drift parameter A . The MLE for this parameter is then given explicitly by

$$(1.2) \quad \hat{A}_T(X) = \frac{X_T^2 - x_0^2 - \sigma^2 T}{2 \int_0^T X_t^2 dt}.$$

Despite its simplicity, the above example exhibits a few interesting properties: First, it is not well defined for every possible path $X \in C([0, T], \mathbb{R})$, and indeed $X \equiv 0$ leaves us with an ill-defined division by zero. Second, provided we stay away from the zero-path, we have pathwise stability in the sense that two observations X and \tilde{X} which are uniformly close on $[0, T]$ give rise to close estimations $\hat{A}_T(X) \approx \hat{A}_T(\tilde{X})$. (In other words, the functional \hat{A}_T is continuous on $C([0, T], \mathbb{R}) - \{0\}$, with respect to the uniform topology.) At last, the estimator depends continuously on the parameter σ , despite the fact that pathspace measures associated to different values of σ are actually mutually singular.⁴

In order to understand such stability considerations in greater generality, we now review the MLE construction for a general diffusion as given in (1.1). To this end, recall that by Girsanov’s theorem, under the standing assumption that $C := \Sigma \Sigma^T$ is nondegenerate, the corresponding measures on pathspace, say $\mathbb{P}^{A, \Sigma}$ and $\mathbb{P}^0 := \mathbb{P}^{0, \Sigma}$, are absolutely continuous so that the MLE method is applicable. Standard computations, partially reviewed below, show that one has

$$I_T \hat{A}_T = S_T,$$

where

$$S_T = \int_0^T h(X_s)^T C^{-1}(X_s) dX_s \in \mathbb{V}^*$$

and

$$I_T = \int_0^T h(X_s)^T C^{-1}(X_s) h(X_s) ds \in L(\mathbb{V}, \mathbb{V}^*),$$

where the first integral above (against dX) is understood in Itô sense. Of course, degeneracy may be a problem, for instance, when $h \equiv 0$. On the other hand, for “reasonable” nondegenerate h [such as $h(x) = x$ in the Ornstein–Uhlenbeck model case] one can expect a.s. invertibility of I_T and thus an a.s. well-defined estimator

$$(1.3) \quad \hat{A}_T(\omega) = I_T^{-1} S_T \in \mathbb{V}.$$

Let us emphasize that S_T involves a stochastic (here: Itô) integral so that S_T is also only defined up to null-sets. At this stage, one has (at best) a measurable

⁴The laws of σW and $\tilde{\sigma} W$ are mutually singular when $\sigma \neq \tilde{\sigma}$; just compute the quadratic variation.

map $\hat{A}_T : C([0, T], \mathbb{R}^d) \rightarrow \mathbb{V}$ with the usual null-set ambiguity.⁵ The following questions then arise rather naturally:

(Q1) Under what conditions on h (and ω) is $I_T = I_T(X(\omega))$ actually invertible? [For \mathbb{P}^0 -a.e. $X(\omega) = \omega$, say, or provide a robust pathwise condition.]

(Q2) Assuming suitably invertibility of I_T , so that the estimator \hat{A}_T is well defined, do we have “robustness” of the estimate problem in the following sense: if $X \approx \tilde{X}$ (e.g., in the sense that $\sup_{t \in [0, T]} |X_t - \tilde{X}_t| \ll 1$ or perhaps a more complicated metric) is it true that

$$\hat{A}_T(X) \approx \hat{A}_T(\tilde{X})?$$

In other words, is the functional \hat{A}_T continuous in some topology?

(Q3) Write \hat{A}_T^σ to indicate the MLE under volatility specification $\Sigma = \sigma I$. Assume we are not entirely certain about the value of σ . Is it true—a rather sensible request from a user’s perspective—that

$$\sigma \approx \tilde{\sigma} \implies \hat{A}_T^\sigma \approx \hat{A}_T^{\tilde{\sigma}}?$$

We emphasize that (Q3) is a difficult question, last not least because the respective pathspace measures are singular whenever $\sigma \neq \tilde{\sigma}$. Hence, it is not even clear if one is allowed to speak “simultaneously” of \hat{A}_T^σ for all σ .⁶ The situation becomes even worse if one considers all possible volatility specifications in a class like

$$\{\Sigma : c^{-1}I \leq \Sigma \Sigma^T \leq cI\}.$$

Indeed, this space is infinite-dimensional, leaving no hope to “fix” things with Kolmogorov-type criteria. On the other hand, explicit computations (e.g., in the Ornstein–Uhlenbeck case, Example 1 and Section 7) show that \hat{A} is extremely well behaved in σ . Hence, we can certainly hope for some sort of robustness of the MLE with respect to the volatility specification.

The last question we would like to investigate is about misspecification of the noise W . The assumption of independent increments of W is a strong limitation in applications and a nontrivial dependence structure in time appears in many real data examples.

(Q4) Suppose that the model is misspecified in the sense that (1.1) is in fact driven by a fractional Brownian motion W^H with Hurst index H . Is the MLE \hat{A}_T robust in some sense (e.g., when $H \approx 1/2$) with respect to this change of the model?

⁵The situation is reminiscent of SDE theory: the Itô-map is also a measurable map on pathspace, in general only defined up to null-sets.

⁶The situation is reminiscent of stochastic flow theory: for each fixed starting point, SDE solution may be (well-) defined (up to null-sets), but it is far from clear—and not true in general in infinite dimension—that one can define solutions for all starting points on a common set of full measure. The financial theory of *uncertain volatility* (see [1] and [15]) also poses related problems.

Our main theorem in Section 4 provides reasonable answers to question (Q1) to (Q3) based on T. Lyons' rough path theory [8, 9, 16, 18], a short review of which will be given in Section 3 below. It is worth emphasizing that the rough path ideas are pivotal to the pathwise robustness results obtained here: in Section 7, we give an explicit example illustrating the failure of robustness if one uses the usual uniform topology. Question (Q4) will be addressed in Section 6.

In Section 8, we present two numerical examples demonstrating the practical value of our theoretical results. The first one concerns an Ornstein–Uhlenbeck process driven by physical Brownian motion in a magnetic field. As was recently demonstrated in [7], physical Brownian motion in a magnetic field does *not* converge—in the zero mass limit—to standard Brownian motion on the level of rough paths; a correction term appears. Nonetheless, our main theorem tells us how to appropriately correct the estimator for the OU process driven by a standard Brownian motion in order to still get reasonable results. The second example concerns the Ornstein–Uhlenbeck process driven by fractional Brownian motion W^H with Hurst parameter H . For $H < 1/2$ naively applying the classical estimator is not well-posed, since the Itô integrals are not well defined. There exists, though, a canonical rough path lift for $H > 1/3$ and plugging this into the estimator of Theorem 8 leads to surprisingly good results even for $H \neq 1/2$. The theoretical background for this example is presented in Section 6; most importantly, the fact that the rough path lift is continuous in H .

The interplay of statistics and rough paths is very recent. The first and (to our knowledge) only paper is [20] where the authors consider general rough differential equations driven by random rough paths and propose parametric estimation of the coefficients based on Lyons' notion of expected signature. That said, the present paper constitutes the first use of rough path analysis toward robustness questions related to classical statistical estimation problems for diffusion processes.

2. A first step: Stratonovich estimators. Let us recall a few basic facts about convergence of discrete approximations of stochastic integrals. This is a central issue when applying the maximum likelihood estimators in the context of discrete observations and will be of importance for our numerical examples in Section 8.

Let X be a (possibly multi-dimensional) continuous semimartingale. Then for regular enough functions f :

- (i) the left-point Riemann sums converge to the Itô integral in probability

$$\sum_{[u,v] \in \mathcal{P}^n} f(X_u)[X_v - X_u] \xrightarrow{n \rightarrow \infty} \int f(X_r) dX_r,$$

for any sequence \mathcal{P}^n of partitions with mesh-size going to 0;

(ii) the trapezoidal Riemann sums converge to the Stratonovich integral in probability

$$\sum_{[u,v] \in \mathcal{P}^n} \frac{1}{2} [f(X_u) + f(X_v)] [X_v - X_u] \xrightarrow{n \rightarrow \infty} \int f(X_r) \circ dX_r,$$

for any sequence \mathcal{P}^n of partitions with mesh-size going to 0;

(iii) for any reasonable⁷ smooth approximations $X^n \xrightarrow{n \rightarrow \infty} X$ the corresponding classical Riemann–Stieltjes integrals converge to the Stratonovich integral in probability

$$\int f(X_s^n) dX_s^n = \int f(X_s^n) \dot{X}_s^n ds \xrightarrow{n \rightarrow \infty} \int f(X_r) \circ dX_r.$$

The first point illustrates how a MLE is usually used in practice, for discrete time observations: Since the process X is only known at a finite number of time points (discrete observations), the stochastic integrals are usually approximated by left-point Riemann sums. This is in fact a quite unstable procedure, as will be illustrated in Section 6.

On the other hand, looking at (iii), it is reasonable to expect, that any positive answer to (Q2) will start out with the Stratonovich formulation of the MLE:

$$(2.1) \quad I_T(X) \hat{A}_T(X) = S_T^{\text{Strat}}(X),$$

where

$$\begin{aligned} S_T^{\text{Strat}}(X)_{i,j} &= \int_0^T h_i(X_s)^T C_{j \cdot}^{-1}(X_s) \circ dX_s \\ &\quad - \frac{1}{2} \int_0^T \text{Tr}[D(h_i C_{j \cdot}^{-1})(X_s) \Sigma(X_s) \Sigma(X_s)^T] ds, \\ I_T(X) &= \int_0^T h(X_s)^T C^{-1}(X_s) h(X_s) ds \in L(\mathbb{V}, \mathbb{V}^*). \end{aligned}$$

There is, at first, only a notational difference between S^{Strat} and S , since we have just rewritten the Itô integral as a Stratonovich one. Taking a hint from point (iii) above though, we *define* from now on $S_T^{\text{Strat}}(X)$ for smooth paths X (a null-set under the diffusion measure) with the Stratonovich integrals replaced by Riemann–Stieltjes integrals. Before stating our first stability result which justifies this definition, we give the following well-posedness result on the estimator. We assume this result to be folklore in the statistical community, but were unable to find a relevant reference. The proof will be given in Section 5.

⁷For example, piecewise linear, mollified, etc.

PROPOSITION 2. *The MLE for A in equation (1.1) is characterized by (2.1). Moreover, if we define*

$$R_h := \{X \in C([0, T], \mathbb{R}^d) : \forall M \in \mathbb{V}, M \neq 0, \exists t \in [0, T] \text{ s.t. } h(X_t)M \neq 0\},$$

and assume that

$$\mathbb{P}^{0, \Sigma}(R_h) = 1,$$

then $I_T = I_T(X)$ is $\mathbb{P}^{0, \Sigma}$ -almost surely invertible so that $A_T = A_T(X) := I_T^{-1}S_T(X)$ is $\mathbb{P}^{0, \Sigma}$ -almost surely well defined.

We then have the following first stability result.

PROPOSITION 3. *Assume that $\mathbb{P}^{0, \Sigma}(R_h) = 1$, and let X^n be piecewise linear approximations to X such that R^h has full measure under the image measure of X^n for all n . Then in probability*

$$\hat{A}_T^{\text{Strat}}(X^n) := I_T^{-1}(X^n)S_T^{\text{Strat}}(X^n) \rightarrow_{n \rightarrow \infty} \hat{A}_T(X).$$

PROOF. The claimed stability (in probability) of Stratonovich integrals, which can be found, for example, in Section 6.6 in [10] (see [8], Section 9.2, for a modern proof), yields $S_T^{\text{Strat}}(X^n) \rightarrow S_T^{\text{Strat}}(X)$ as $n \rightarrow \infty$. Moreover I_T^{-1} is continuous in supremum norm in X on R_h , and hence the statement follows. \square

Note that the preceding result only concerns convergence in probability; it therefore does not provide a good answer to (Q2). To wit, for the Stratonovich estimator \hat{A}_T^{Strat} it is in general not true that paths that are uniformly close in supremum norm, that the resulting estimates will be close. We give an explicit (deterministic) counterexample in Section 7. A stochastic counterexample will be given in Section 8, in the setting of a physical Brownian motion in a magnetic field.

In order to fix this problem (and in order to answer the other questions), we will adopt a rough path perspective in the next section. To this end, we now give some recalls on rough path theory.

3. Brief review of rough paths. In this section, we introduce some basic notions from Lyons’ rough paths theory. Our notation here follows Friz–Hairer [8], which is also a source of much more on this material, together with the standard references [9, 16, 18].

We start by giving a definition of Hölder continuous rough paths that is suitable for our purpose. Let $X : [0, T] \rightarrow \mathbb{R}^d$ be a smooth path and define the second-order iterated integrals $\mathbb{X} : [0, T]^2 \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ of X by

$$\mathbb{X}_{s,t} := \int_s^t X_{s,r} \otimes dX_r,$$

where $X_{s,r} = X_r - X_s$ are the increments of X . Then the pair (X, \mathbb{X}) has the analytic property

$$(ANA)_\alpha : \begin{cases} |X_{s,t}| \lesssim |t - s|^\alpha, \\ |\mathbb{X}_{s,t}| \lesssim |t - s|^{2\alpha} \end{cases}$$

for any $\alpha \leq 1$ and satisfies the algebraic relation

$$(ALG) : \mathbb{X}_{s,t} + X_{s,t} \otimes X_{t,u} + \mathbb{X}_{t,u} = \mathbb{X}_{s,u},$$

$$(ALG') : 2 \text{Sym}(\mathbb{X}_{s,t}) = X_{s,t} \otimes X_{s,t},$$

for $s, t, u \in [0, T]$. More generally speaking, these two conditions are used to define a *rough path* in \mathbb{R}^d .

DEFINITION 4. Fix $\alpha \in (1/3, 1/2]$. Any $\mathbf{X} = (X, \mathbb{X})$ for which $(ANA)_\alpha + (ALG)$ holds is called (weak α -Hölder) rough path. If also (ALG') is satisfied call it geometric. The space of α -Hölder rough paths and its subset of geometric rough paths are denoted by $\mathcal{C}^\alpha([0, T], \mathbb{R}^d)$ and $\mathcal{C}_g^\alpha([0, T], \mathbb{R}^d)$, respectively.

Rough paths arise naturally as sample paths of stochastic processes. The basic example is a d -dimensional Brownian motion B enhanced with its iterated integrals

$$\mathbb{B}_{s,t} := \int_s^t B_{s,r} \otimes dB_r \in \mathbb{R}^{d \times d},$$

where the integral on the right-hand side can be understood in Itô or Stratonovich sense leading to Itô or Stratonovich enhanced Brownian motion, respectively. Then with probability one $\mathbf{B} = (B, \mathbb{B}) \in \mathcal{C}^\alpha([0, T], \mathbb{R}^d)$ for any $\alpha \in (1/3, 1/2)$ and $T > 0$. We also say that we can lift B to a rough path \mathbf{B} by adding the second-order terms \mathbb{B} . A similar rough paths lift is given in our main result for the solution of (1.1).

To investigate stability questions for the parameter estimation problem in a pathwise sense, we need suitable metric on $\mathcal{C}^\alpha([0, T], \mathbb{R}^d)$. It turns out that an adequate metric on $\mathcal{C}^\alpha([0, T], \mathbb{R}^d)$ can be defined from $(ANA)_\alpha$ as follows.

DEFINITION 5. For $\mathbf{X}, \mathbf{Y} \in \mathcal{C}^\alpha([0, T], \mathbb{R}^d)$ the α -Hölder rough path metric is given by

$$\rho_\alpha(\mathbf{X}, \mathbf{Y}) := \sup_{s \neq t \in [0, T]} \frac{|X_{s,t} - Y_{s,t}|}{|t - s|^\alpha} + \sup_{s \neq t \in [0, T]} \frac{|\mathbb{X}_{s,t} - \mathbb{Y}_{s,t}|}{|t - s|^{2\alpha}}.$$

REMARK 6. In the original formulation of rough paths theory in [17], distance was measured in p -variation norm instead of the α -Hölder norm used here. The results in this work can be rephrased without difficulty in a p -variation setting. This applies in particular to the continuity of the map \hat{A}_T in Theorem 8(ii) and (iii) below.

We conclude this section with rough integrals and its relation to stochastic integration. Let \mathcal{P} be a partition of $[0, T]$ and denote by $|\mathcal{P}|$ the length of its largest element. For $\mathbf{X} = (X, \mathbb{X}) \in \mathcal{C}^\alpha([0, T], \mathbb{R}^d)$ and $\alpha > 1/3$ we aim at integrating $F(X)$ for $F \in \mathcal{C}_b^2(\mathbb{R}^d, \mathcal{L}(\mathbb{R}^d, \mathbb{R}^m))$ against \mathbf{X} . It is well known that classical Young integration is possible for expressions of the form

$$\int_0^T F(X_t) dX_t$$

only if $X \in C^\alpha$ for $\alpha > 1/2$. This excludes, for example, paths of Brownian motion which are of order $\alpha < 1/2$. This barrier was overcome by rough paths theory by taking into account “second order” terms. Indeed, one can show that the limit in

$$\int_0^T F(X_s) d\mathbf{X}_s := \lim_{|\mathcal{P}| \rightarrow 0} \sum_{(s,t) \in \mathcal{P}} F(X_s) X_{s,t} + DF(X_s) \mathbb{X}_{s,t}$$

exists and is called a (Lyons’) rough integral [17]. Most importantly for us, rough integrals depend continuously in rough path metric on \mathbf{X} and by taking $\mathbf{X} = \mathbf{B}$ to be (Stratonovich) enhanced Brownian motion one recovers with probability one the Stratonovich integral.

We shall need the following standard result; see, for example, Friz–Victoir [9], Section 13, or Friz–Hairer [8], Section 10.

PROPOSITION 7. *Fix $\alpha \in (1/3, 1/2)$. Then, $\mathbb{P}^{0,\Sigma}$ -almost surely, $X(\omega)$ lifts to a (random) geometric α -Hölder rough path, that is, a random element in the rough path space $\mathcal{C}_g^\alpha([0, T], \mathbb{R}^d)$ (as reviewed in the next section), via the (existing) limit in probability*

$$\mathbf{X}(\omega) := (X(\omega), \mathbb{X}(\omega)) := \lim_n \left(X^n, \int X^n \otimes dX^n \right),$$

where X^n denotes dyadic piecewise linear approximations to X .

4. Main result. We are now ready to formulate our main result. By constructing an estimator on rough path space, we resolve the pathwise stability problem that is inherent to the Stratonovich estimator (compare Proposition 3 and Section 7).

THEOREM 8. *Assume that $\mathbb{P}^{0,\Sigma}(R_h) = 1$, so that the MLE \hat{A} is well defined by Proposition 2.*

(i) Define $\mathbb{D} \subset \mathcal{C}_g^\alpha([0, T], \mathbb{R}^d)$ by

$$\mathbb{D} = \{(X, \mathbb{X}) \in \mathcal{C}_g^\alpha : X \in R_h\}.$$

Then for every fixed, nondegenerate volatility function Σ ,

$$\mathbb{P}^{0,\Sigma}[\mathbf{X}(\omega) \in \mathbb{D}] = 1.$$

(ii) *There exists a deterministic, continuous (with respect to α -Hölder rough path metric; see Definition 5) map*

$$\hat{\mathbf{A}}_T : \begin{cases} \mathbb{D} \rightarrow \mathbb{V}, \\ \mathbf{X} \mapsto \hat{\mathbf{A}}_T(\mathbf{X}) \end{cases}$$

so that, for every fixed, nondegenerate volatility function Σ ,

$$(4.1) \quad \mathbb{P}^{0, \Sigma}[\hat{\mathbf{A}}_T(\mathbf{X}(\omega)) = \hat{A}_T(\omega)] = 1.$$

In fact, $\hat{\mathbf{A}}_T$ is explicitly given, for $(X, \mathbb{X}) \in \mathbb{D} \subset \mathcal{C}_g^\alpha$, by

$$\hat{\mathbf{A}}(X, \mathbb{X}) := \mathbf{I}_T^{-1}(X) \mathbf{S}_T(X, \mathbb{X}),$$

where

$$\begin{aligned} \mathbf{I}_T(X) &:= \int_0^T h(X_s)^T C^{-1}(X_s) h(X_s) ds, \\ \mathbf{S}_T(X, \mathbb{X})_{i,j} &:= \int_0^T h_i(X_s)^T C_j^{-1}(X_s) d\mathbf{X}_s \\ &\quad - \frac{1}{2} \int_0^T \text{Tr}[D(h_i C_j^{-1})(X_s) \Sigma(X_s) \Sigma(X_s)^T] ds \end{aligned}$$

and the $d\mathbf{X}$ integral is understood as a (deterministic) rough integration against $\mathbf{X} = (X, \mathbb{X})$.

(iii) *The map $\hat{\mathbf{A}}_T$ is also continuous with respect to the volatility specification. Indeed, fix $c > 0$ and set*

$$\mathbb{E} := \{ \Sigma \in C_b^2 : c^{-1} I \leq \Sigma \Sigma^T \leq c I \}.$$

Then $\hat{\mathbf{A}}_T$ viewed as map from $\mathbb{D} \times \mathbb{E} \rightarrow \mathbb{R}^d$ is continuous.

EXAMPLE 9. The case of the d -dimensional Ornstein–Uhlenbeck process

$$dX_t = Af(X_t) dt + \Sigma(X_t) dW_t,$$

with $A \in L(\mathbb{R}^d, \mathbb{R}^d)$, $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is covered by our setting by taking $\mathbb{V} = L(\mathbb{R}^d, \mathbb{R}^d)$ and $h = I \otimes f$, in coordinates

$$(h_i^{k,j}) = (f^j \delta_i^k),$$

so that (with summation over up-down indices)

$$h_i^{k,j}(x) A_j^i = A^k f^j(x).$$

In this case, the nondegeneracy condition in point (i) is, for example, satisfied if the set of critical points of f has no accumulation points [i.e., on every bounded set, there is only a finite set of points at which $\det Df(x) = 0$], which can be seen by an application of the (functional) law of the iterated logarithm for diffusions (Strassen’s law), for example, Proposition 4.1 in [3].

REMARK 10. Note that Proposition 3 can be regarded as a corollary of Theorem 8 and Proposition 7.

REMARK 11. The continuity statements in (ii) and (iii) also hold with respect to p -variation metric, $p \in (2, 3)$. This and other rough path metrics are discussed in Section 3.

REMARK 12. By (4.1) the well-known asymptotic properties of the maximum likelihood estimator \hat{A}_T like consistency and asymptotic normality (see, e.g., [12]) also hold for $\hat{\mathbf{A}}_T$.

REMARK 13. We briefly discuss in what sense Theorem 8 provides answers to (Q1)–(Q3) above:

(Q1) Proposition 2 gives a pathwise condition for existence of the MLE in terms of the drift coefficient h .

(Q2) The discussion in Section 7 shows that the classical MLE violates the pathwise stability property that (Q2) asks for. Theorem 8 shows that by considering the signal X as a rough path we can construct a continuous estimator $\hat{\mathbf{A}}_T$ that overcomes this difficulty.

(Q3) The question of stability in the volatility coefficient σ can also be solved by moving to a rough path space. Indeed, Theorem 8(iii) shows that $\hat{\mathbf{A}}_T^\sigma$ is continuous with respect to the observations and the volatility coefficient. Here, the pathwise approach is crucial, since in the classical setting it is not even clear how to define the estimator as a mapping in both variables whereas in the rough paths approach this is an obvious consequence.

REMARK 14. While our answer to (Q2) above is best possible, in the sense that one cannot hope for pathwise stability without introducing rough paths (see the explicit counterexample in Section 7), it leaves the user with the question of how to exactly understand discrete or continuous data as a rough path.

In essence, this amounts to measuring the Lévy area associated to an observed path. In this direction, there are in fact cases where the measurement of the area is feasible within the physical system under observation; see [2].

5. Proof of the main result. To recall, let W be d -dimensional Wiener process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, $A \in \mathbb{V}$ (some fixed finite-dimensional vector space) and

$$h : \mathbb{R}^d \rightarrow L(\mathbb{V}, \mathbb{R}^d), \quad \Sigma : \mathbb{R}^d \rightarrow L(\mathbb{R}^d, \mathbb{R}^d)$$

are Lipschitz continuous coefficients, so that the stochastic differential equation

$$(5.1) \quad \begin{aligned} dX_t &= h(X_t)A dt + \Sigma(X_t) dW_t, & t \in \mathbb{R}_+, \\ X_0 &= x_0, \end{aligned}$$

has a unique solution. We are interested in estimation of A , as function of some observed sample path $X = X(\omega) : [0, T] \rightarrow \mathbb{R}^d$ when the coefficients f and Σ are known.

LEMMA 15. Write $\mathbb{P} = \mathbb{P}^{A, \Sigma}$ for the path-space measure induced by the solution X to (5.1). Assume $C = \Sigma \Sigma^T$ is nondegenerate (say $c^{-1}I \leq C^{-1} \leq cI$ for some $c > 0$). Then the \mathbb{V} -valued MLE (relative to \mathbb{P}^0), $A = \hat{A}_T$, is characterized by

$$(5.2) \quad I_T \hat{A}_T = S_T,$$

where

$$S_T = \int_0^T h(X_s)^T C^{-1}(X_s) dX_s \in \mathbb{V}^*$$

and

$$I_T = \int_0^T h(X_s)^T C^{-1}(X_s) h(X_s) ds \in L(\mathbb{V}, \mathbb{V}^*).$$

PROOF. The statement follows from Girsanov’s theorem; see, for example, [11], Theorem III.5.34. \square

PROOF OF PROPOSITION 2. The first statement follows from Lemma 15. Now we need to understand when I_T is nondegenerate. To this end, pick any nonzero $M \in \mathbb{V}$. Then, with $g = hM$ we have

$$\langle M, I_T M \rangle = \int_0^T \langle g, C^{-1}g \rangle ds \geq 0$$

and since $\langle g, C^{-1}g \rangle \geq 0$ we see that $\langle M, I_T M \rangle$ vanishes iff

$$\langle g, C^{-1}g \rangle = \langle h(X_\cdot)M, C^{-1}(X_\cdot)h(X_\cdot)M \rangle \equiv 0$$

on $[0, T]$. Thanks to (assumed) nondegeneracy of C this happens iff

$$h(X_\cdot)M \equiv 0$$

on $[0, T]$. Hence, for every path in R_h , I_T is invertible. \square

PROOF OF THEOREM 8. (i) Follows as combination of Proposition 2 and Proposition 7.

(ii) Recall that for $(X, \mathbb{X}) \in \mathcal{C}_g^\alpha$ we have

$$\hat{A}(X, \mathbb{X}) := \mathbf{I}_T^{-1}(X) \mathbf{S}_T(X, \mathbb{X}),$$

where

$$\begin{aligned} \mathbf{I}_T(X) &:= \int_0^T h(X_s)^T C^{-1}(X_s) h(X_s) ds, \\ \mathbf{S}_T(X, \mathbb{X})_{i,j} &:= \sum_k \int_0^T h_i(X_s)^T C_{jk}^{-1}(X_s) d\mathbf{X}_s^k, \\ &- \sum_k \frac{1}{2} \int_0^T \sum_{n,m} [h_i(X_s) \partial_{x_n} C_{jk}^{-1}(X_s) + \partial_{x_n} h_i(X_s) C_{jk}^{-1}(X_s)] \Sigma_{n,m}(X_s) \Sigma_{k,m}(X_s) ds \\ &= \sum_k \int_0^T h_i(X_s)^T C_{j\cdot}^{-1}(X_s) d\mathbf{X}_s \\ &- \frac{1}{2} \int_0^T \text{Tr}[D(h_i C_{j\cdot}^{-1})(X_s) \Sigma(X_s) \Sigma(X_s)^T] ds, \end{aligned}$$

where the dX integral is understood as a rough path integral (Section 3). Note that in the definition of \mathbf{S}_T we have formally used the Stratonovich form S_T^{Strat} , which is sensible since rough path lift given in Proposition 7 is the Stratonovich lift of X .

Now $\mathbf{S}_T(X, \mathbb{X})$ is continuous in rough path metric by the just mentioned references. Moreover, $\mathbf{I}_T(X)$ is obviously continuous in supremum metric, and hence is its inverse [everywhere defined on \mathbb{D} by (i)].

Finally, by Proposition 17.1 in [9], $\mathbf{S}_T(X, \mathbb{X})|_{\mathbf{X}=\mathbf{X}(\omega)}$ coincides with $S_T(\omega)$. $\mathbf{I}_T(X)|_{\mathbf{X}=\mathbf{X}(\omega)}$ trivially coincides with $I_T(\omega)$, since it only depends on the path (the first level of the rough path). Hence, $\hat{\mathbf{A}}_T(\mathbf{X}(\omega)) = A_T(\omega)$ a.s. under $\mathbb{P}^{0,\Sigma}$.

(iii) This boils down to continuity of the rough integrals as functions of integrand 1-form; see, for example, Theorem 10.47 in [9]. \square

6. Misspecification of the noise. In this section, we investigate the behavior of the MLE under misspecification of the noise W in the sense that we suppose that the true model has a driving process with nontrivial dependence structure in time. For the sake of argument, we shall consider (1.1) with fractional Brownian noise. Fractional noise was first used in stochastic modeling by Mandelbrot and van Nees in their seminal paper [19] and is now heavily used in such diverse fields as the study of turbulence or mathematical finance, see, for example, [4, 23].

For further simplicity, assume $\Sigma \equiv I$ so that the dynamics are

$$(6.1) \quad dX_t^H = h(X_t^H) A dt + dW_t^H,$$

started from a fixed starting point x_0 , with W^H a multi-dimensional Volterra fractional Brownian motion with Hurst index $H \in (0, 1)$, that is,

$$(6.2) \quad W_t^H = \int_0^t K^H(t, s) dW_s,$$

where W is a standard Brownian motion, $K^H(t, s) = (t - s)^{H-1/2}$ is the Volterra kernel.⁸ Note that $W^H|_{H=1/2} = W$ is a standard Brownian motion and that $X^H \rightarrow X$, for example, in probability uniformly on $[0, T]$ as $H \rightarrow 1/2$, where

$$(6.3) \quad dX_t = h(X_t)A dt + dW_t.$$

[Thanks to additivity of the noise in (6.1) this is a truly elementary statement, namely a consequence of the continuity of the Itô-map as detailed below.] Suppose now that the *true* dynamics correspond to (6.1) with $H = 1/2 - \varepsilon$. Clearly, for very small $\varepsilon > 0$, the model (6.3), mathematically much easier, is still an excellent description of the true dynamics. Indeed, it is well known that in the additive noise case (6.1) or (6.3) the solution map

$$(W_t : t \in [0, T]) \mapsto (X_t^H : t \in [0, T])$$

is locally Lipschitz continuous with respect to sup-norm (see, e.g., [6], page 188). We can then try to perform classical MLE estimation using the wrong model (6.3) and write down the estimator $\hat{A}_T = I_T^{-1}S_T$ as was done in (1.3).

If we use the Itô form of the estimator, the Itô integrals appearing blow up when applied to fractional Brownian sample paths “rougher” than Brownian motion.⁹ As pointed out in Section 2, the Stratonovich version of the estimator is much more stable. Using rough path theory, and in particular our rough path estimator \hat{A}_T , we can show not only that the estimator remains well defined when $H = 1/2 - \varepsilon$, but also that it behaves continuously in H . This is spelled out fully in the following theorem.

THEOREM 16. *Suppose that $H \in (1/3, 1)$. Then, for every $\alpha \in (1/3, H)$, there exists a geometric α -Hölder rough path lift $\mathbf{X}^H = (X^H, \mathbb{X}^H)$ of X^H (natural in the sense that \mathbf{X}^H is the common rough path limit, in probability, of piecewise linear, mollifier or Karhunen–Loeve approximations to X^H). Moreover, there is a continuous modification of $\mathbf{X}^H : H \in (1/3, 1)$. As a consequence, $\hat{A}_T(\mathbf{X}^H)$ is well defined and robust with respect to the Hurst parameter,*

$$\hat{A}_T(X^H, \mathbb{X}^H) \rightarrow \hat{A}_T(X, \mathbb{X})$$

almost surely, as $H \rightarrow 1/2$, where (X, \mathbb{X}) is the lift \mathbf{X} of X from Theorem 8.

⁸The results of this section also hold true for classical fractional Brownian motion, using the kernel given in [5]. The only difference is that the estimates in the proof of Theorem 16 become more technical.

⁹This is well known and in fact easy to see: just consider the left-point Riemann–Stieltjes approximations to the Itô-integral $\int_0^1 W^H dW^H$ where W^H is a scalar fractional Brownian motion. When $H > 1/2$ one has convergence to the Young integral [actually equal to $(1/2)(W_1^H)^2$]. When $H = 1/2$ one has convergence to the Itô integral. When $H < 1/2$ the approximations diverge, as may be seen by computing their (exploding) variance.

PROOF. Without loss of generality $T = 1$. It is a well-known fact (Section 15 in [9]) that for fixed $H \in (1/3, 1]$, X^H can be lifted to an α -Hölder rough path $\mathbf{X}^H = (X^H, \mathbb{X}^H)$.

We will apply Kolmogorov’s continuity theorem to construct \mathbf{W}^H that is almost surely continuous in H . First, using (6.2)

$$\begin{aligned} R_{W^H - W^{H'}}(s, t) &= \mathbb{E}[(W_s^H - W_s^{H'})(W_t^H - W_t^{H'})] \\ &\leq \sup_{t \in [0, 1]} \mathbb{E}[(W_t^H - W_t^{H'})^2] \\ &= \sup_{t \in [0, 1]} \int_0^t (|t - r|^{H-1/2} - |t - r|^{H'-1/2})^2 dr \\ &= \int_0^1 (r^{H-1/2} - r^{H'-1/2})^2 dr \\ &= O(|H - H'|^2). \end{aligned}$$

We can now apply Remark 15.38 in [9] to get

$$\mathbb{E}[\rho_\alpha(\mathbf{W}^H, \mathbf{W}^{H'})^q] \leq C|H - H'|^\theta,$$

for some q, C large enough and $\theta > 0$ small enough. Applying Kolmogorov’s continuity criterion, we get a version of \mathbf{W}^H that is continuous in H . Since \mathbf{X}^H is the solution to a rough differential equation driven by \mathbf{W}^H , that is, the continuous image of \mathbf{W}^H , it is clear that \mathbf{X}^H is also continuous in H (with respect to α -Hölder rough path topology). The convergence of $\hat{\mathbf{A}}_T(X^H, \mathbb{X}^H)$ follows now from Theorem 8(ii). \square

7. Failure of continuity for the classical MLE. We consider the two-dimensional Ornstein–Uhlenbeck process. This class of processes was first used by Ornstein and Uhlenbeck to describe the movement of a particle due to random impulses known as physical Brownian motion (see [7] for a detailed analysis in a rough path context). Later these dynamics were applied extensively in finance, for example, to model commodity prices [22] or interest rates, where it is called the Vasicek model [24].

More precisely, let $A \in \mathbb{V} := L(\mathbb{R}^2, \mathbb{R}^2)$, $h(x) = x$ for all x , $g \equiv 0$ and $\Sigma = I$ and consider the model¹⁰

$$\begin{aligned} dX_t &= AX_t dt + dW_t, \\ X_0 &= x_0 \in \mathbb{R}^2. \end{aligned}$$

¹⁰...which of course fits in the framework of this paper, as pointed out in Example 9.

By Lemma 15, the (classical) likelihood estimator $\hat{A}_T \in \mathbb{R}^{2 \times 2}$ is obtained from the relation

$$(7.1) \quad I_T \hat{A}_T = S_T.$$

A straightforward calculation gives

$$(7.2) \quad \hat{A}_T^{i,j}(\omega) = \frac{1}{U(X)} \left(\int_0^T (X_s^{\bar{j}})^2 ds \int_0^T X_s^j dX_s^i - \int_0^T X_s^i X_s^{\bar{j}} ds \int_0^T X_s^{\bar{j}} dX_s^i \right),$$

where $\bar{i} := 3 - i, \bar{j} := 3 - j, U(X) = \int_0^T (X_r^1)^2 dr \int_0^T (X_r^2)^2 dr - (\int_0^T X_r^1 X_r^2 dr)^2$ and all stochastic integrals are understood in Itô sense. Note that this allows us to see the precise dependence of the MLE on the iterated integrals of the observation. The Stratonovich version reads as

$$\begin{aligned} \hat{A}_T^{\text{Strat},i,j}(\omega) &= \frac{1}{U(X)} \left(\int_0^T (X_s^{\bar{j}})^2 ds \left(\int_0^T X_s^j \circ dX_s^i - \delta_{i,j} \frac{T}{2} \right) \right. \\ &\quad \left. - \int_0^T X_s^i X_s^{\bar{j}} ds \left(\int_0^T X_s^{\bar{j}} \circ dX_s^i - \delta_{j,i} \frac{T}{2} \right) \right). \end{aligned}$$

As shown in Section 2 this estimator, *defined* on smooth path by replacing Stratonovich with Riemann–Stieltjes integrals, possesses a certain continuity in probability.

We now show that *pathwise* stability fails for this MLE.¹¹ To this end, it suffices to consider the case $i = j = 1$ and we construct a sequence of observations paths X_n that converges uniformly to some limit X , but

$$|\hat{A}_T^{\text{Strat},1,1}(X_n) - \hat{A}_T^{\text{Strat},1,1}(X)| \rightarrow \infty,$$

as $n \rightarrow \infty$. This means that observations can be arbitrarily close in uniform norm, but the corresponding estimates for A diverge. At the core of this robustness problem lies, as we will see below, the fact that multi-dimensional iterated integrals (as the ones appearing in \hat{A}_T) are discontinuous in sup-norm.

We modify the usually given example of “spinning fast enough around the origin” (see, e.g., Section 1.5.2 in [18]), since we want the limiting path to yield an $I_T(X)$ that is invertible.

We start with the path $X : [0, 1] \rightarrow \mathbb{R}^2$ that goes, at constant speed, clockwise, through the square with corners $(0, 0)$ and $(1, 1)$. This path lies in the set R_h of Theorem 8 (see Remark 9 for the definition of h).

Now we attach a fast spinning loop at the end as follows:

$$\begin{aligned} X_n(t) &:= X\left(\frac{n}{n-1}t\right), & t \in [0, (n-1)/n], \\ X_n(t) &:= \frac{1}{n} \left(e^{i2\pi n^3(t-(n-1)/n)} - 1 \right), & t \in [(n-1)/n, 1]. \end{aligned}$$

¹¹In fact, more is true: by the result in [14], there exists no continuous functional F on pathspace, such that $\hat{A}_T^{\text{Strat},i,j}(\omega) = F(X)$.

Evaluating the upper left component of the likelihood estimator \hat{A}_T from (7.2) for $X = (X^{(1)}, X^{(2)})$ and $T = 1$ yields

$$\begin{aligned} \hat{A}_T^{\text{Strat},1,1}(X) &= \frac{1}{U(X)} \left(\int_0^1 X_r^{(2)} X_r^{(2)} dr \left(\int_0^1 X_r^{(1)} dX_r^{(1)} - \frac{T}{2} \right) \right. \\ &\quad \left. - \int_0^1 X_r^{(1)} X_r^{(2)} dr \left(\int_0^1 X_r^{(2)} dX_r^{(1)} - \frac{T}{2} \right) \right). \end{aligned}$$

The prefactor $U(X)$, consisting only of Riemann integrals, is continuous in supremum and so $U(X^{(n)})$ converges to a finite limit as $n \rightarrow \infty$. The same holds true for the first factor in the large bracket (the stochastic integral is seen to be continuous by an application of Itô’s formula) and the factor $\int_0^1 X_r^{(1)} X_r^{(2)} dr$ in the last term. Now for $\int_0^1 X_r^{(2)} dX_r^{(1)}$, note first that

$$\begin{aligned} \int_0^1 X^{(2)}(r) dX^{(1)}(r) &= \int_0^{(n-1)/n} X^{(2)}(r) dX^{(1)}(r) \\ &\quad + \int_{(n-1)/n}^1 X^{(2)}(r) - X^{(2)}((n-1)/n) dX^{(1)}(r). \end{aligned}$$

Moreover, since the $X^{(n)}$ have the same value at $t = 0, (n-1)/n, 1$ it is easy to see that

$$\begin{aligned} \int_0^{(n-1)/n} X^{(2)}(r) dX^{(1)}(r) &= \mathcal{A}_{0,(n-1)/n}(X^{(n)}), \\ \int_{(n-1)/n}^1 X^{(2)}(r) dX^{(1)}(r) &= \mathcal{A}_{(n-1)/n,1}(X^{(n)}), \end{aligned}$$

where $\mathcal{A}_{s,t}(X)$ is (two times) the area between that curve $\{X(r) \equiv (X_1(r), X_2(r)) : s \leq r \leq t\}$ and the chord from $X(t)$ to $X(s)$.¹² Hence, $\mathcal{A}_{0,(n-1)/n}(X^{(n)}) \equiv -2$ and $\mathcal{A}_{(n-1)/n,1}(X^{(n)}) = -\pi n$ and, therefore, as desired,

$$|\hat{A}_T^{\text{Strat},1,1}(X_n)| \rightarrow \infty.$$

In conclusion, the estimation problem is *not well-posed* if one measures distance of paths in supremum norm. Let us note that by working in stronger topologies on pathspace, say α -Hölder with $\alpha < 1/2$ (so that they support Wiener measure), the situation does not change; see, for example, [14].

On the other hand, as was seen in Section 4, the “rough” estimator \hat{A}_T built in Theorem 8 restores continuity. In the present example, this reduces to the fact that the iterated stochastic integrals which appear in $\hat{A}_T(\omega)$ may be rewritten in terms of rough integrals against the rough path lift of X .

¹²It is given as

$$\mathcal{A}_{s,t}(X) = \int_s^t (X_1(r) - X_1(s)) dx_2(r) - \int_s^t (X_2(r) - X_2(s)) dx_1(r).$$

8. Numerical examples. We illustrate our theoretical results in two numerical examples. The first example uses a fractional Brownian motion with Hurst parameter H as driving noise. The Itô integral is not even well defined for Hurst parameter $H \neq \frac{1}{2}$, but we show that the estimator using the rough path lift of the fractional Brownian motion performs well in this setting. We use Stratonovich-type Riemann sums to approximate the rough path lift (and hence are strictly speaking only performing the robustification laid out in Section 2), and hence this example can foremost be seen as a strong encouragement to use them over Itô-type approximations.

In the second example, the driving noise is replaced by a physical Brownian motion in a magnetic field. On the level of the path this is known to converge to Brownian motion, but its lifted rough path does not (cf. [7]). We demonstrate that the classical MLE breaks down in this setting and how a deterministic correction on the second level of the rough path leads nonetheless to good estimation results.

8.1. *Fractional Ornstein–Uhlenbeck process.*

8.1.1. *One-dimensional.* In this section, we demonstrate in the setting of Example 1 the instability of the MLE due to the Itô integrals, if one does not get rid of the stochastic integral via integration by parts. Furthermore, we show that by using Stratonovich-type approximations as suggested in Section 2 we obtain a stable estimator.

We simulate samples from a one-dimensional fractional Ornstein–Uhlenbeck process defined by

$$(8.1) \quad dX_t^H = AX_t^H dt + dW_t^H, \quad t \in [0, T].$$

We use an exact simulation scheme to draw equidistant samples

$$X_\Delta^H, X_{2\Delta}^H, \dots, X_{n\Delta}^H \quad \text{for } \Delta > 0$$

from X^H such that $T = n\Delta$. The discretized maximum likelihood estimator \hat{A}_T for A in this model is given by

$$\hat{A}_T^n(X^H) = \frac{\sum_{i=1}^{n-1} X_{i\Delta}^H (X_{(i+1)\Delta}^H - X_{i\Delta}^H)}{\sum_{i=1}^{n-1} (X_{i\Delta}^H)^2 \Delta}.$$

From Theorem 8, we obtain the discretized rough MLE

$$\hat{\mathbf{A}}_T^n(\mathbf{X}^H) = \frac{\sum_{i=1}^{n-1} X_{i\Delta}^H (X_{(i+1)\Delta}^H - X_{i\Delta}^H) + \mathbb{X}_{i\Delta} - \mathbb{X}_{(i+1)\Delta}}{\sum_{i=1}^{n-1} (X_{i\Delta}^H)^2 \Delta}.$$

In Figure 1, Monte Carlo estimates of variances of \hat{A}_T^n and the rough MLE $\hat{\mathbf{A}}_T^n$ are depicted for varying Hurst index from each 500 Monte Carlo iterations. The sample size is $n = 100$ (i.e., the time mesh size of observation is $1/n$) and the time horizon $T = 1$. We clearly see that the variance increases when H moves away

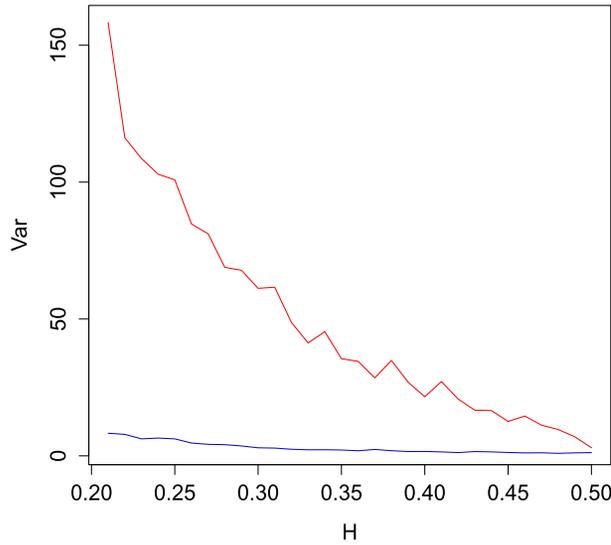


FIG. 1. Monte Carlo estimate of the variance of the classical MLE (red) and rough MLE (blue) for different Hurst indices H .

from $1/2$ and explodes for H going to 0. On the contrary, the rough MLE remains stable on the whole range of H values with an almost constant variance.

Note that the rate of convergence for the variance of \hat{A}_T^n is proportional to $n(1/2 - H)$ so that the effect that can be seen in Figure 1 becomes more severe with growing sample size. This connection is depicted in Figure 2 where we see Monte Carlo estimates of the variance of \hat{A}_T^n for increasing sample size. The number of Monte Carlo iterations for each n is $N = 100$, the time horizon $T = 1$ and the Hurst index $H = 0.35$.

In Table 1, the mean and standard deviation of the RMLE \hat{A}_T^n are estimated for the fractional Ornstein–Uhlenbeck model and for various Hurst indices. Each estimate consists of 1000 Monte Carlo iterations and the true parameter was $A = 2$. We find that already for quite moderate sample size of $n = 100$ the estimator performs very well. We also observe that when T grows a slight discretization bias appears that is typically of the order Δ . Surprisingly, the RMLE gives accurate results even when the Hurst parameter is far away from the classical case at $H = 1/2$.

8.1.2. *Two-dimensional.* Here, we give numerical examples for the two-dimensional Ornstein–Uhlenbeck dynamics. We apply a Euler scheme to draw an equidistant sample $X_\Delta^H, X_{2\Delta}^H, \dots, X_{n\Delta}^H$ for $\Delta > 0$ from the process X solving

$$dX_t = AX_t dt + dW_t^H,$$

$$X_0 = x_0 \in \mathbb{R}^2,$$

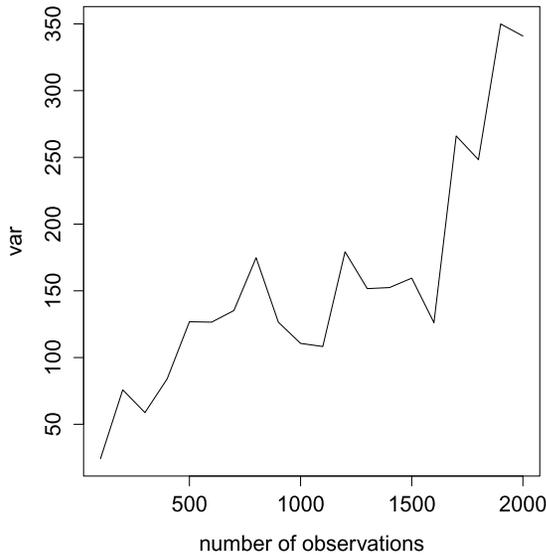


FIG. 2. Estimated variance of the classical MLE for varying sample size.

where W^H is a two-dimensional fractional Brownian motion with Hurst parameter $H \in (0, 1)$ and A is given by

$$A = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}.$$

The expression for the classical maximum likelihood estimator [see (7.2)] is of course only valid for $H = 1/2$. Moreover, for $H < 1/2$ the Itô integrals appearing in that estimator are in general not even well defined. Nonetheless, since we simu-

TABLE 1
Mean and standard deviation of the RMLE \hat{A}_T^n for varying Hurst indices in the misspecified model

a	T	n	$H = 0.5$		$H = 0.4$		$H = 0.3$	
			Mean	Std dev	Mean	Std dev	Mean	Std dev
2	1	100	2.0	0.20	2.0	0.18	2.0	0.33
		200	2.0	0.17	2.0	0.18	2.0	0.20
		500	2.0	0.16	2.0	0.19	2.0	0.19
2	2	100	2.0	0.023	2.0	0.023	2.0	0.026
		200	2.0	0.025	2.0	0.026	2.0	0.024
		500	2.0	0.022	2.0	0.022	2.0	0.051
5	5	100	2.1	8.1e-05	2.1	8.3e-05	2.1	9.2e-05
		200	2.1	6.7e-05	2.1	6.6e-05	2.1	7.0e-05
		500	2.0	5.6e-05	2.0	5.6e-05	2.0	6.4e-05

late on a discrete time grid, we can calculate its discretized version, replacing the stochastic integrals by Itô-type Riemann sums.

On the other hand, for every $H > 1/3$ fractional Brownian motion possesses a natural rough path lift (see, e.g., [9], Chapter 15), so the expression for the “rough” MLE (8.3) is at least well defined, also for $H \neq 1/2$. Since we deal with a simulation at discrete timepoints, we have to approximate this rough path lift. We shall use Stratonovich-type Riemann sums, which are well known to converge (see, e.g., [8], Chapter 10, and the references therein). We then plug the result into the rough path estimator (8.3).

We give the estimation results for the upper right coordinate of A with true value equal to 2, on a discrete grid for varying number of observations n and observation length T .

In Table 2, the estimated mean and standard deviation for the discretized classical MLE (top) and the discretized “rough” MLE (bottom) are given. Each value is based on 100 Monte Carlo runs of the estimator.

TABLE 2
We consider the “true” parameter value $A^{1,2} = 2$ and give estimates of the mean and standard deviation of the classical MLE $\hat{A}_{T,n}^{1,2}$ (top) and the “rough” MLE $\hat{A}_{T,n}^{1,2}$ (bottom) based on 100 Monte Carlo iterations for varying Hurst indices for the 2-dim. OU process. Here, n is the number of time-steps in the Euler approximation

T	n	$H = 0.5$		$H = 0.4$		$H = 0.35$	
		Mean	Std dev	Mean	Std dev	Mean	Std dev
1	100	2.7	1.6	3.4	2.0	5.5	2.9
	200	2.7	1.8	3.7	2.6	6.9	3.2
	500	2.7	1.8	4.8	2.7	9.7	4.6
5	100	2.1	0.90	2.7	1.13	3.9	1.1
	200	2.2	0.99	3.1	1.1	4.9	1.4
	500	2.2	1.06	3.9	1.3	6.5	1.9
10	100	2.0	0.61	2.5	0.70	3.0	0.43
	200	2.0	0.73	2.7	0.75	3.9	0.54
	500	2.1	0.74	3.3	0.90	5.1	0.90
1	100	2.7	1.6	2.3	1.6	2.1	2.2
	200	2.8	1.6	2.4	2.0	2.2	1.9
	500	2.7	1.6	2.3	1.9	2.2	2.2
5	100	2.2	1.0	2.1	1.2	1.8	1.1
	200	2.3	1.0	1.9	1.1	1.8	1.2
	500	2.2	1.0	1.9	1.1	1.9	1.3
10	100	2.0	0.64	1.8	0.71	1.8	0.83
	200	2.1	0.75	1.9	0.86	1.8	0.87
	500	2.0	0.75	2.0	0.91	1.9	0.89

We find that for $H = 1/2$ the classical MLE performs well if the observation length T is large enough. When H moves away from $1/2$ the instability of the estimator becomes apparent. The standard deviation increases significantly and the estimator is strongly biased.

In contrast to that the “rough” MLE $\hat{A}_{T,n}^{1,2}$ performs equally well over the whole range of Hurst indices. For $H = 1/2$, both estimators give similar results as expected from our results in Theorem 8 whereas in the dependent regime $\hat{A}_{T,n}^{1,2}$ clearly outperforms the classical MLE.

8.2. *Physical Brownian motion in a magnetic field.* The dynamics of a two-dimensional physical Brownian motion $\bar{W}^{\alpha,m}$ in a magnetic field are given by (see [7])

$$d\bar{W}_t^{\alpha,m} = \frac{1}{m} P_t^{\alpha,m} dt,$$

$$dP_t^{\alpha,m} = -\frac{1}{m} M P_t^{\alpha,m} dt + dW_t.$$

Here,

$$M = M_\alpha = \begin{pmatrix} 1 & \alpha \\ -\alpha & 1 \end{pmatrix},$$

with strength of the magnetic field given by the scalar parameter α and mass $m > 0$ of the particle, assumed to carry unit charge.

As in the preceding section, we observe the realization of an Ornstein–Uhlenbeck process, but now driven by the physical Brownian motion and with covariance matrix M :

(8.2)
$$dX_t^{\alpha,m} = AX_t^{\alpha,m} dt + M d\bar{W}_t^{\alpha,m}.$$

Now, it is quite easy to show (see, e.g., [21], Section 11.7.7) that $M\bar{W}^{\alpha,m} \rightarrow W$ in supremum norm, as $m \rightarrow 0$. In the one-dimensional case (where the MLE is continuous in supremum norm, as we saw in Example 1), it automatically follows that

$$\hat{A}_T(X^{\alpha,m}) \xrightarrow{m \rightarrow 0} \hat{A}_T(X),$$

where X solves the classical OU equation $dX = AX dt + dW$ and \hat{A}_T is the MLE for the latter. On the other hand, in dimension $d \geq 2$, and in presence of a magnetic field $\alpha \neq 0$ we still have $M\bar{W}^{\alpha,m} \rightarrow W$ but the desired convergence *fails*, that is,

$$\hat{A}_T(X^{\alpha,m}) \not\xrightarrow{m \rightarrow 0} \hat{A}_T(X).$$

The underlying reason is failure of convergence at the level of iterated integrals to the Stratonovich iterated integrals of W . Instead, as was shown in detail in [7], one has

$$\left(M\bar{W}_s^{\alpha,m}, \int_s^t M\bar{W}_{s,r}^{\alpha,m} \otimes dM\bar{W}_r \right) \rightarrow \left(W_s, \int_s^t W_{s,r} \otimes \circ W_r + (t - s)D \right),$$

with correction term

$$D := \frac{1}{2} \text{Anti}[M] \text{Sym}[M]^{-1} = \frac{1}{2} \begin{pmatrix} 0 & \alpha \\ -\alpha & 0 \end{pmatrix}.$$

As an easy consequence,

$$\left(X_s^{\alpha,m}, \int_s^t X_{s,r}^{\alpha,m} \otimes dX_r^{\alpha,m} \right) \rightarrow \left(X_s, \int_s^t X_{s,r} \otimes \circ X_r + (t-s)D \right),$$

and so, by Theorem 8, we have a “modified” MLE stability of the form

$$\hat{A}_T \left(\left(X_s^{\alpha,m}, \int_s^t X_{s,r}^{\alpha,m} \otimes dX_r^{\alpha,m} - (t-s)D \right) \right) \rightarrow \hat{A}_T \left(\left(X_s, \int_s^t X_{s,r} \otimes \circ X_r \right) \right),$$

where, with $\bar{i} := 3 - i, \bar{j} := 3 - j$ and $U(X) = \int_0^T (X_r^1)^2 dr \int_0^t (X_r^2)^2 dr - (\int_0^T X_r^1 X_r^2 dr)^2$,

$$(8.3) \quad \hat{A}_T^{i,j}(X, \mathbb{X}) = \frac{1}{U(X)} \left(\int_0^T (X_s^{\bar{j}})^2 ds \left\{ \mathbb{X}_{0,T}^{i,j} - \delta_{i,j} \frac{T}{2} \right\} - \int_0^T X_s^i X_s^{\bar{j}} ds \left\{ \mathbb{X}_{0,T}^{j,i} - \delta_{j,i} \frac{T}{2} \right\} \right).$$

In summary, it is *perfectly justified*, in the small mass regime $m \ll 1$, to consider the effective dynamics $dX_t = AX_t dt + dW_t$ as approximation for (8.2). However, it would be *wrong* to use the resulting MLE estimator on the realizations of realization of $X^{\alpha,m}$, even in the limit $m \rightarrow 0$. Instead, the estimation procedure based on $X^{\alpha,m}$ *must* take account of the correction term D we exhibited above. At last, we support our findings with concrete numerical results, taking

$$A = \begin{pmatrix} -3 & 2 \\ 0 & -4 \end{pmatrix},$$

and 100 Monte Carlo simulations. The force of the magnetic field is chosen as $\alpha = 1.0$, the mass of the particle as $m = 0.01$ and discretization is done on a time grid of 10^5 equidistant points.

The results are shown in Table 3. As is clearly visible, the corrected estimator yields good results, with decreasing standard variation for increasing time hori-

TABLE 3
Mean and standard deviation of $\hat{A}_T^{1,2}$ over 100 Monte Carlo runs for the physical Brownian motion model. The correct value is 2.0

	T = 1.0		T = 3.0		T = 10.0		T = 30.0	
	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
w/correction	2.0	0.84	2.0	0.72	2.1	0.59	2.2	0.45
w/o correction	1.7	0.79	1.1	0.64	0.3	0.49	-0.4	0.35

zons. The uncorrected estimator on the other hand yields useless results for times larger than 3.0.

REFERENCES

- [1] AVELLANEDA, M., LEVY, A. and PARÁS, A. (1995). Pricing and hedging derivative securities in markets with uncertain volatilities. *Appl. Math. Finance* **2** 73–88.
- [2] BAILLEUL, I. and DIEHL, J. (2014). Recovering a signal. Preprint. Available at [arXiv:1407.2768](https://arxiv.org/abs/1407.2768).
- [3] CARAMELLINO, L. (1998). Strassen’s law of the iterated logarithm for diffusion processes for small time. *Stochastic Process. Appl.* **74** 1–19. [MR1624072](#)
- [4] COMTE, F., COUTIN, L. and RENAULT, E. (2012). Affine fractional stochastic volatility models. *Ann. Finance* **8** 337–378. [MR2922801](#)
- [5] DECREUSEFOND, L. and ÜSTÜNEL, A. S. (1999). Stochastic analysis of the fractional Brownian motion. *Potential Anal.* **10** 177–214. [MR1677455](#)
- [6] DEMBO, A. and ZEITOUNI, O. (1993). *Large Deviations Techniques and Applications*. Jones and Bartlett Publishers, Boston, MA. [MR1202429](#)
- [7] FRIZ, P., GASSIAT, P. and LYONS, T. (2015). Physical Brownian motion in a magnetic field as a rough path. *Trans. Amer. Math. Soc.* **367** 7939–7955. [MR3391905](#)
- [8] FRIZ, P. K. and HAIRER, M. (2014). *A Course on Rough Paths: With an Introduction to Regularity Structures*. Universitext. Springer, Cham. [MR3289027](#)
- [9] FRIZ, P. K. and VICTOIR, N. B. (2010). *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge Studies in Advanced Mathematics **120**. Cambridge Univ. Press, Cambridge. [MR2604669](#)
- [10] IKEDA, N. and WATANABE, S. (1981). *Stochastic Differential Equations and Diffusion Processes*. Elsevier, Amsterdam. [MR0637061](#)
- [11] JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] **288**. Springer, Berlin. [MR0959133](#)
- [12] KUTOYANTS, Y. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer Series in Statistics. Springer, London. [MR2144185](#)
- [13] LIPTSER, R. S. and SHIRYAEV, A. N. (2001). *Statistics of Random Processes, Vol. I, II*. Springer, Berlin.
- [14] LYONS, T. (1991). On the nonexistence of path integrals. *Proc. Roy. Soc. London Ser. A* **432** 281–290. [MR1116958](#)
- [15] LYONS, T. (1995). Uncertain volatility and the risk-free synthesis of derivatives. *Appl. Math. Finance* **2** 117–133.
- [16] LYONS, T. and QIAN, Z. (2002). *System Control and Rough Paths*. Oxford Mathematical Monographs. Oxford Univ. Press, Oxford. [MR2036784](#)
- [17] LYONS, T. J. (1998). Differential equations driven by rough signals. *Rev. Mat. Iberoam.* **14** 215–310. [MR1654527](#)
- [18] LYONS, T. J., CARUANA, M. and LÉVY, T. (2007). *Differential Equations Driven by Rough Paths*. Lecture Notes in Math. **1908**. Springer, Berlin. [MR2314753](#)
- [19] MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437. [MR0242239](#)
- [20] PAPAVALIOU, A. and LADROUE, C. (2011). Parameter estimation for rough differential equations. *Ann. Statist.* **39** 2047–2073. [MR2893861](#)
- [21] PAVLIOTIS, G. A. and STUART, A. M. (2008). *Multiscale Methods: Averaging and Homogenization*. Texts in Applied Mathematics **53**. Springer, New York. [MR2382139](#)

- [22] SCHWARTZ, E. S. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. *J. Finance* **52** 923–973.
- [23] SHAO, Y. (1995). The fractional Ornstein–Uhlenbeck process as a representation of homogeneous Eulerian velocity turbulence. *Phys. D* **83** 461–477. [MR1334532](#)
- [24] VASICEK, O. (1977). An equilibrium characterisation of the term structure. *Journal of Financial Economics* **5** 177–188.

J. DIEHL
TU BERLIN
STRASSE DES 17. JUNI 136
10623 BERLIN
GERMANY
AND
DEPARTMENT OF MATHEMATICS—MC 0112
UNIVERSITY OF CALIFORNIA, SAN DIEGO
9500 GILMAN DRIVE
LA JOLLA, CALIFORNIA 92093
USA
E-MAIL: joscha.diehl@gmail.com

P. FRIZ
TU BERLIN
STRASSE DES 17. JUNI 136
10623 BERLIN
GERMANY
AND
WEIERSTRASS INSTITUTE FOR
APPLIED ANALYSIS AND STOCHASTICS
MOHRENSTRASSE 39
10117 BERLIN
GERMANY
E-MAIL: friz@math.tu-berlin.de
friz@wias-berlin.de

H. MAI
CREST ENSAE PARISTECH
3, AV. PIERRE LAROUSSE
92240 MALAKOFF
FRANCE
E-MAIL: hilmar.mai@ensae.fr