Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e. V.

Preprint ISSN 0946 – 8633

Pathwise stability of likelihood estimators for diffusions via rough paths

Joscha Diehl[1], Peter K. Friz [1,2], Hilmar Mai [2]

submitted: November 05, 2013

[1] Institute of Mathematics
TU Berlin
Strasse des 17. Juni 136
10623 Berlin
Germany
E-Mail: diehl@math.tu-berlin.de
friz@math.tu-berlin.de

[2] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: Peter.Friz@wias-berlin.de
Hilmar.Mai@wias-berlin.de

No. 1866
Berlin 2013

2010 *Mathematics Subject Classification.* 62F35, 62M99.

Key words and phrases. Rough paths, MLE for diffusions, pathwise stability, model misspecification.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

ABSTRACT. We consider the estimation problem of an unknown drift parameter within classes of non-degenerate diffusion processes. The Maximum Likelihood Estimator (MLE) is analyzed with regard to its pathwise stability properties and robustness towards misspecification in volatility and even the very nature of noise. We construct a version of the estimator based on rough integrals (in the sense of T. Lyons) and present strong evidence that this construction resolves a number of stability issues inherent to the standard MLEs.

1. INTRODUCTION

Let W be d-dimensional Wiener process and $A \in \mathbb{V} := L\left(\mathbb{R}^d, \mathbb{R}^d\right)$. Consider sufficiently regular $h : \mathbb{R}^d \to \mathbb{R}^d$ and $\Sigma : \mathbb{R}^d \to L\left(\mathbb{R}^d, \mathbb{R}^d\right)$ so that

$$dX_t = A\,h(X_t)\,dt + \Sigma(X_t)\,dW_t \tag{1.1}$$

has a unique solution, started from $X_0 = x_0$. The important example of multidimensional Ornstein-Uhlenbeck dynamics, for instance, falls in the class of diffusions considered here (take $h(x) = x, g = 0$ and constant, non-degenerate diffusion matrix Σ). We are interested in estimating the drift parameter A, given some observation sample path $\{X_t(\omega) = \omega_t : t \in [0,T]\}$. More precisely, we are looking for a Maximum Likelihood Estimator (MLE) of the form

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

relative to the reference measure given by the law of X, viewed as measure on pathspace, in the case $A \equiv 0$.

Example 1. *(Scalar Ornstein-Uhlenbeck process) Take $d = 1$, $h(x) = x$, $\Sigma \equiv \sigma > 0$ and $A = a \in \mathbb{R}$. Then it is well-known that*

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

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

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

$$I_T \hat{A}_T = S_T \in \mathbb{V}^*, \tag{1.3}$$

where $I_T \in L(\mathbb{V}, \mathbb{V}^*)$. (Of course, we may identify $\mathbb{R}^d$ and $\mathbb{V}$ with their duals; note also that $\mathbb{V}^*$ and $L(\mathbb{V}, \mathbb{V}^*)$ respectively can be thought of as $(1,1)$ resp. $(2,2)$-tensors.). In fact, in tensor notation (cf. Corollary 12) we find

$$\begin{aligned} \mathbb{V}^* \quad &\ni \quad S_T = \int_0^T h(X_s) \otimes C^{-1}(X_s)\,dX_s, \\ L(\mathbb{V}, \mathbb{V}^*) \quad &\ni \quad I_T = \int_0^T h(X_s) \otimes C^{-1}(X_s) \otimes h(X_s)\,ds, \end{aligned}$$

where the dX-integral is understood in Itô sense. Of course, degeneracy may be a problem, for instance when $h \equiv 0$. That being said, and although we are short of a reference, we believe it is folklore of the subject that, for reasonably non-degenerate h (such as $h(x) = x$ in the Ornstein-Uhlenbeck case) one has a.s. invertibility of I_T and thus an a.s. well-defined estimator

$$\hat{A}_T(\omega) = I_T^{-1} S_T. \tag{1.4}$$

May that be as it is, below we shall also give a simple sufficient condition on h under which this holds true. Let us also note 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 map $\hat{A}_T : C([0,T], \mathbb{R}^d) \to \mathbb{V}$ with the usual null-set ambiguity.[1] The following questions then arise rather naturally - and our attempt to answer them form the subject of this paper.

(Q1) Under what conditions on h (and ω) is $I_T = I_T(X(\omega))$ actually invertible? A minimum request would be that invertibility holds for $\mathbb{P}^0$-a.e. $X(\omega) = \omega$, but is there perhaps a pathwise condition?

(Q2) Assuming suitably invertibility of I_T, is the estimation problem well-posed? In other words, if $X \approx \tilde{X}$ (e.g. in the sense that $\sup_{t\in[0,T]} |X_t - \tilde{X}_t| << 1$ or perhaps a more complicated metric) is it true that

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

(Q3) Write $\hat{A}_T^\sigma$ 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

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

From a stochastic analysis perspective, (Q3) is a difficult question also because the respective pathspace measures are singular whenever $\sigma \neq \tilde{\sigma}$. Hence, it is not even clear if one is allowed to speak ßimultaneouslyöf $\hat{A}_T^\sigma$ for all σ.[2] The situation becomes even worse if one considers all possible volatility specifications in a class like

$$\Xi := \left\{ \Sigma \in \mathrm{Lip}^2 : c^{-1} I \leq \Sigma\Sigma^T \leq cI \right\}.$$

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 6) 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. In applications the assumption of independent increments of W is a strong limitation and a non-trivial 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?

---

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

[2] The 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 [11]) also poses related problems.

Our main theorem provides reasonable answers to question (Q1) to (Q3) based on rough path theory [14, 15, 5], a short review of which will be given in section 2 below. Let us insist that one cannot obtain a similar result without rough path metrics and in Section 6.1 below we give an explicit counterexample. Question (Q4) will be addressed in Section 4.

Theorem 2. *(i) Define*

$$R_h := \left\{ X \in C\left([0,T], \mathbb{R}^d\right) : \operatorname{span}\left\{h(X_t) : t \in [0,T]\right\} = \mathbb{R}^d \right\}. \tag{1.5}$$

Assume that the set of critical points of h has no accumulation points (i.e. on every bounded set, there is only a finite set of points at which $\det Dh(x) = 0$). Then, for every fixed, non-degenerate volatility function Σ

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

As a consequence, $I_T = I_T(\omega)$ is $\mathbb{P}^{0,\Sigma}$-almost surely invertible so that $A_T = A_T(\omega) := I_T^{-1} S_T(\omega)$ is $\mathbb{P}^{0,\Sigma}$-almost surely well-defined.
(ii) 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, i.e. a random element in the rough path space $\mathscr{D}_g^\alpha\left([0,T], \mathbb{R}^d\right)$ (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.
(iii) Define $\mathbb{D} \subset \mathscr{D}_g^\alpha\left([0,T], \mathbb{R}^d\right)$ by

$$\mathbb{D} = \left\{ (X, \mathbb{X}) \in \mathscr{D}_g^\alpha : X \in R_h \right\}.$$

Then, under the assumption of (i), for every fixed, non-degenerate volatility function Σ,

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

(iv) There exists a deterministic, continuous [with respect to α-Hölder rough path metric] map

$$\hat{\mathbf{A}}_T : \begin{cases} \mathbb{D} & \to \mathbb{R}^{d \times d} \\ \mathbf{X} & \mapsto \hat{\mathbf{A}}_T(\mathbf{X}) \end{cases}$$

so that, for every fixed, non-degenerate volatility function Σ,

$$\mathbb{P}^{0,\Sigma}\left[\hat{\mathbf{A}}_T(\mathbf{X}(\omega)) = A_T(\omega)\right] = 1. \tag{1.6}$$

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

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

where

$$\mathbf{I}_T(X) := \int_0^T h(X_s) \otimes C^{-1}(X_s) \otimes h(X_s)\, ds,$$

$$\mathbf{S}_T(X, \mathbb{X})_{i,j} := \int_0^T h_i(X_s) \otimes C_{j\cdot}^{-1}(X_s) \circ dX_s - \frac{1}{2} \int_0^T \operatorname{Tr}[D(h_i C_{j\cdot}^{-1})(X_s) \Sigma(X_s) \Sigma(X_s)^T] ds$$

and the $\circ dX$ integral[3] is understood as a (deterministic) rough integration against $\mathbf{X} = (X, \mathbb{X})$.
(v) The map $\hat{\mathbf{A}}_T$ is also continuous with respect to the volatility specification. Indeed, fix $c > 0$ and set

$$\Xi := \left\{ \Sigma \in \mathrm{Lip}^2 : c^{-1} I \le \Sigma \Sigma^T \le c I \right\}.$$

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

[3] ... often written as d$\mathbf{X}$ integral in the literature on rough integration ...

Let us conclude this introduction with several remarks.

Remark 3. *The continuity statement in (iv) and (v) also hold with respect to* p*-variation metric,* $p \in (2,3)$*. This and other rough path metrics are discussed in Section 2.*

Remark 4. *By* (1.6) *the well-known asymptotic properties of the maximum likelihood estimator like consistency and asymptotic normality (see for example* [9]*) directly apply to* $\hat{\mathbf{A}}_T$*.*

Remark 5. *We briefly discuss in what sense Theorem 1 provides answers to (Q1)-(Q3) above:*

- (Q1) *Theorem 1(i) gives a pathwise condition for existence of the MLE in terms of the drift coefficient* h*.*
- (Q2) *The discussion in Section 6.1 shows that the classical MLE violates the pathwise stability property that (Q2) asks for. In Theorem 1 (ii) to (iv) we show 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 1 (v) 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 6. *(Discrete and continuous observations as rough paths)*

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 6.1), it leaves the user with the question how exactly to understand discrete or continuous data as a rough path.

In essence, this amounts to measure the area associated to some (irregular) observation sample path. In this direction, one could imagine cases where the measurement of the area is feasible within the physical system under observation (see for instance [6]*, where the stochastic area to the trajectory of a Brownian particle with electric charge is linked to the presence of a magnetic field). That said, the understanding and classification of real world systems which allow measurements on the level of rough paths is a complex and difficult problem. Let us therefore adopt a more pragmatic point of view and even give up on the idea of continuous observation. Instead, we assume given discrete, but high-frequency, data, say* N *data points on some unit observation time horizon, say* $x = \{x_i : i = 0,...,N\}$*. There is a natural inclusion map* i *of such a data point into the space of Lipschitz continuous paths on* $[0,1]$*; simply by piecewise linear interpolation of the data observed at times* $0 = t_0 < t_1 .... < t_N = 1$*. This inclusion map is continuous and so is the resulting estimator*[4]

$$x \mapsto i(x) \mapsto \hat{A}(i(x)),$$

simply because all integrals appearing in the estimator depend now continuously (in the strong Lipschitz topology!) on $i(x)$*. More precisely, given* x *and* $\varepsilon > 0$ *there exists* δ *s.t.* $|x-y| < \delta$ *implies* $|\hat{A}(i(x)) - \hat{A}(i(y))| < \varepsilon$*. But in fact,* δ *will also depend on* $D = (t_i)$ *and in fact tend to zero as* $mesh(D) \to 0$*. In other words, the continuity properties are getting worse and worse* [5] *as* $N \to \infty$ *(or* $mesh(D) \to 0$*), which is of course consistent with the lack of (pathwise) continuity in the continuous time setting. The point is that, with discrete (high-frequency) data, one has continuous estimators in principle, but with potentially terrible modulus of continuity in practice. One may then be better off to construct* $(i(X), \int i(X) \otimes di(X))$ *as rough path, whose* α*-Hölder regularity, with* $\alpha < 1/2$*, is uniformly bounded, as* $mesh(D) \to 0$*.*

[4]... as given in (iv) of Theorem 2, but with $\circ dX = di(x)$ understood as Riemann-Stieltjes differential, well-defined since $i(x)$ has bounded variation.

[5]An explicit computation in this spirit is found in [17].

Remark 7. *The interplay of statistics and rough paths is very recent. The first and (to our knowledge) only paper is* [16] *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. It would then appear that the present paper constitutes the first attempt to use rough path analysis towards robustness questions related to statistical estimation of classical diffusion processes.*

Acknowledgement 1. *J. Diehl is supported by SPP 1324. P.K. Friz has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement nr. 258237.*

2. Brief review of rough paths

In this section we introduce some basic notions from rough paths theory. For a detailed presentation in a much more general setting we refer to [14, 15, 5]. We start by giving a definition of Hölder continuous rough paths that is suitable for our purpose. Let $X : [0,T] \to \mathbb{R}^d$ be a smooth path and define the second order iterated integrals $\mathbb{X} : [0,T]^2 \to \mathbb{R}^d \otimes \mathbb{R}^d$ 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. Smoothness of X is understood such that the integral in the definition of $\mathbb{X}$ is well-defined. 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

$$\begin{aligned} (ALG) \quad &: \quad \mathbb{X}_{s,t} + X_{s,t} \otimes X_{t,u} + \mathbb{X}_{t,u} = \mathbb{X}_{s,u}, \\ (ALG') \quad &: \quad 2\,\mathrm{Sym}\,(\mathbb{X}_{s,t}) = X_{s,t} \otimes X_{s,t}, \end{aligned}$$

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 8. *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 $\mathscr{D}^\alpha\left([0,T], \mathbb{R}^d\right)$ and $\mathscr{D}_g^\alpha\left([0,T], \mathbb{R}^d\right)$ 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 $\mathbf{B} = (B, \mathbb{B}) \in \mathscr{D}^\alpha\left([0,T], \mathbb{R}^d\right)$ 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 $\mathscr{D}^\alpha\left([0,T], \mathbb{R}^d\right)$. It turns out that an adequate metric on $\mathscr{D}^\alpha\left([0,T], \mathbb{R}^d\right)$ can be defined from $(ANA)_\alpha$ as follows.

Definition 9. *For* $\mathbf{X},\mathbf{Y} \in \mathscr{D}^\alpha\left([0,T],\mathbb{R}^d\right)$ *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 10. *In the original formulation of rough paths theory in* [12] *paths were measured in* p*-variation instead of the* α*-Hölder distance that we use here. Note that the results in this work can equivalently be formulated in the* p*-variation setting. This holds true in particular for the continuity of the map* $\hat{\mathbf{A}}_T$ *in Theorem 2(iv) and (v). We have chosen here the* α*-Hölder formulation, since most readers will already be familiar with classical Hölder spaces.*

We conclude this section with rough integrals and its relation to stochastic integration. Let $\mathscr{P}$ be a partition of $[0,T]$ and denote by $|\mathscr{P}|$ the length of its largest element. For $\mathbf{X} = (X,\mathbb{X}) \in \mathscr{D}^\alpha\left([0,T],\mathbb{R}^d\right)$ and $\alpha > 1/3$ we aim at integrating $F(X)$ for $F \in \mathscr{C}_b^2(\mathbb{R}^d,\mathscr{L}(\mathbb{R}^d,\mathbb{R}^m))$ 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 $X \in \mathscr{C}^\alpha$ 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

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

exists and is called a rough integral (cf. [12, 5]). By taking $\mathbf{X} = \mathbf{B}$ to be Itô enhanced Brownian motion we recover with probability one the Itô integral in a path-wise sense. The rough integral will be crucial for us to define a robust version of the MLE in Section 5.

3. MLE FOR DIFFUSION DRIFT PARAMETERS

3.1. **Basics.** In this section we prove part (i) of the main theorem. In fact, we find it notationally convenient to consider a slightly more general setup. Namely, let W be d-dimensional Wiener process on $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geq 0},\mathbb{P})$, $A \in \mathbb{V}$ (some fixed finite dimensional vector space),

$$f : \mathbb{R}^d \to L\left(\mathbb{V},\mathbb{R}^d\right),\ \Sigma : \mathbb{R}^d \to L\left(\mathbb{R}^d,\mathbb{R}^d\right)$$

say Lipschitz continuous, so that the stochastic differential equation

$$\begin{aligned} dX_t &= f(X_t)Adt + \Sigma(X_t)\, dW_t, \quad t \in \mathbb{R}_+, \\ X_0 &= x_0. \end{aligned} \tag{3.1}$$

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

Theorem 11. *Write* $\mathbb{P} = \mathbb{P}^A$ *for the path-space measure induced by the solution* X *to* (3.1)*. Assume* $C = \Sigma\Sigma^T$ *is (everywhere in space) non-degenerate (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*

$$I_T\hat{A}_T = S_T \tag{3.2}$$

where

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

and

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

Proof. The statement follows from standard theory of likelihood inference for diffusion processes (see e.g [9] and [10]). ☐

This immediately leads to the identification of the MLE in the setting of our main theorem.

Corollary 12. *Consider $h : \mathbb{R}^d \to \mathbb{R}^d$ Lipschitz, $A \in \mathbb{V} := L\left(\mathbb{R}^d, \mathbb{R}^d\right)$ and Σ as before so that* (1.1) *has a unique solution X started from $X_0 = x_0$. Then the MLE is characterized by*

$$I_T \hat{A}_T = S_T, \tag{3.3}$$

where

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

and

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

Proof. Consider $A = (A^i_j) \in L\left(\mathbb{R}^d, \mathbb{R}^d\right) = \mathbb{V}$. Then it suffices to apply the previous result with $f = I \otimes h$, in coordinates

$$(f_i^{k,j}) = \left(h^j \delta_i^k\right),$$

so that (with summation over up-down indices)

$$f_i^{k,j}(x) A^i_j = A^k_j h^j(x)\,.$$

We think of I_T as quadratic form (say Q) on $\mathbb{V}$, in coordinates

$$\langle A, I_T A\rangle = \sum_{\substack{i,k\\ j,l}} A^i_k \underbrace{\int_0^T h^k C^{-1}_{i,j} h^l ds}_{=:Q^{k,l}_{i,j}} A^j_l =: \left\langle A, \int_0^T \left(h \otimes C^{-1} \otimes h\right) ds A \right\rangle.$$

☐

4. 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 non-trivial dependence structure in time. In fact, for the sake of argument, we shall consider (1.1) with fractional Brownian noise. For further simplicity assume $\Sigma \equiv I$ so that the dynamics are

$$dX^H_t = A\, h\left(X^H_t\right) dt + dW^H_t, \tag{4.1}$$

started from a fixed starting point x_0, with W^H a multi-dimensional Volterrra fractional Brownian motion with Hurst index $H \in (0,1)$, i.e.

$$W^H_t = \int_0^t K^H(t,s) dW_s$$

where W is a standard Brownian motion, $K^H(t,s) = (t-s)^{H-1/2}$ is the Volterra kernel.[6] Note that $W^H|_{H=1/2} = W$ is a standard Brownian motion and that $X^H \to X$, e.g. in probability uniformly on $[0,T]$ as $H \to 1/2$, where

$$dX_t = A\,h(X_t)\,dt + dW_t. \tag{4.2}$$

(Thanks to addivity of the noise in (4.1) this is a truly elementary statement.) Suppose now that the *true* dynamics correspond to (4.1) with $H = 1/2 - \varepsilon$. Clearly, for $\varepsilon << 1$, the model (4.2), mathematically much easier, is still a good description of the true dynamics. In particular, we can perform classical MLE estimation on (4.2) and write down the estimator $\hat{A}_T = I_T^{-1} S_T$ as was done in (1.4). Recall that this estimator will involve, in general and through S_T, Itô integrals, defined as limit of left-point Riemann-Stieltjes approximations. But unfortunately, such Riemann-Stieltjes approximations may blow up when applied to fractional Brownian sample paths "rougher"than Brownian motion.[7]. Our proposed solution here is to use the rough path estimator $\hat{\mathbf{A}}_T$. Not only does it remain well-defined when $H = 1/2 - \varepsilon$, but also behaves continuously in H. This is spelled out fully in the following theorem.

Theorem 13. *Suppose that $H \in (1/3,1)$. Then, for every $\alpha \in (1/3,H)$, there exists a geometric α-Hölder rough path lift $\mathbf{X^H} = (X^H, \mathbb{X}^{\mathbb{H}})$ 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, $\hat{\mathbf{A}}_T(\mathbf{X^H})$ is well-defined and robust with respect to the Hurst parameter,*

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

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

Proof. Without loss of generality $T = 1$. It is a well-known fact (Section 15 in [5]) that for fixed $H \in (1/3,1]$, X^H can be lifted to an α-Hölder rough path $\mathbf{X^H} = (X^H, \mathbb{X}^{\mathbb{H}})$

We will apply Kolmogorov's continuity theorem to construct $\mathbf{W}^H$ that is almost surely continuous in H. First

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

We can now apply Remark 15.38 in [5] to get

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

[6] The results of this section also hold true for classical fractional Brownian motion, using the kernel given in [3]. The only difference is that the estimates in the proof of Theorem 13 become more technical.

[7] This is well known and in fact easy to see: just consider the left-point Riemann Stieltjes approximations to the Ito-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.

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 an rough differential equation driven by $\mathbf{W}^H$, i.e. 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 $\hat{\mathbf{A}}_T(X^H, \mathbb{X}^{\mathbb{H}})$ follows now from Theorem 2(iv). □

5. Proof of main result (Theorem 1)

Proof of (i). We need to understand when I_T is non-degenerate. To this end, pick any non-zero

$M = \left(M^i_k\right) \in \mathbb{V}$. Then, with $g = Mh$ (in coordinates, $g^i = \sum_k M^i_k h^k$, also $\left\langle g, C^{-1} g\right\rangle = \sum_{i,j} g^i C^{-1}_{i,j} g^j$) we have

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

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

$$\left\langle g, C^{-1} g\right\rangle = \left\langle Mh(X_\cdot), C^{-1}(X_\cdot) Mh(X_\cdot)\right\rangle \equiv 0$$

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

$$Mh(X_\cdot) \equiv 0$$

on $[0,T]$ which is equivalent to (M was non-zero, hence $\ker M \subsetneq \mathbb{R}^d$)

$$\{h(X_t) : t \in [0,T]\} \subset \ker M \subsetneq \mathbb{R}^d.$$

This leads us to the following (pathwise) condition.

$$\operatorname{span}\{h(X_t) : t \in [0,T]\} = \mathbb{R}^d.$$

It remains to see that this happens with $\mathbb{P}^{0,\Sigma}$-probability one, given the stated non-degeneracy condition on h. This is certainly true when h is the identity map on $\mathbb{R}^d$, for the non-degeneracy of C will guarantee that with probability one the process explores every neighborhood of every point of his trajectory. This follows for example from the (functional) law of the iterated logarithm for diffusions (Strassen's law), e.g. Proposition 4.1 in [2]. By assumption, critical points of h do not accumulate so a.s. there exists times t^* at which $\det Dh(X_{t^*}) \neq 0$. But h is a diffeomorphism in a neighborhood of X_{t^*}, so that $\{h(X_t) : t \in [t^*, t^* + \varepsilon]\}$ also explores its neighborhood a.s. and hence cannot be confined in a (linear) subspace. And it follows that, $\forall T > 0$,

$$\mathbb{P}^{0,\Sigma}\left(\operatorname{span}\{h(X_t) : t \in [0,T]\} = \mathbb{R}^d\right) = 1.$$

(A determinist understanding of what it means to explore every neighborhood can be given in terms of "true roughness"[8, 4].)

Proof of (ii, iii). (ii) The construction of a Stratonovich lift associated to the diffusion processes under consideration is standard in rough path theory, see for example Section 14 in [5].

(iii) Then follows as combination of points (i) and (ii).

Proof of (iv). Recall that for $(X,\mathbb{X}) \in \mathscr{D}_g^\alpha$ we have

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

where

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

where the dX integral is understood as a rough path integral ([12], Section 10.6 in [5] or [7]). Note that in the definition of $\mathbf{S}_T$ we have formally rewritten the Ito integrals in S_T in terms of Stratonovich integrals.

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 [5], $\mathbf{S}_T(X,\mathbb{X})|_{\mathbf{X}=\mathbf{X}(\omega)}$ 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 $\mathbb{P}^{0,\Sigma}$.

Proof of (v). This boils down to continuity of the rough integrals as functions of integrand 1-form, see for example Theorem 10.47 in [5].

6. Explicit computations for Ornstein-Uhlenbeck dynamics

As our main example 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 [6] for a detailed analysis in a rough path context). Later these dynamics were applied in finance in several different contexts to model interest rates (Vasicek model), currency exchange rates and commodity prices.

Our goal in this section is twofold. First of all we calculate $\hat{A}_T$ in explicit form in order to see its dependence on iterated integrals of the observed path. Then we give a counterexample that demonstrates the stability problems that the classical MLE $\hat{A}_T$ exhibits.

Let $A \in L(\mathbb{R}^2,\mathbb{R}^2)$, $h(x) = x$ for all x, $g \equiv 0$ and $\Sigma = I$ such that we obtain the following model

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

According to Corollary 12 the likelihood estimator $\hat{A}_T \in \mathbb{R}^{2\times 2}$ is characterized by

$$I_T \hat{A}_T = S_T, \tag{6.1}$$

with

$$I_T = \int_0^T X_s \otimes I \otimes X_s ds$$
$$S_T = \int_0^T X_s \otimes I dX_s.$$

For explicit computations we consider $\hat{A}_T = (\hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{a}_4)^\top$ as element in $\mathbb{R}^4$. Then (6.1) translates to

$$M\hat{A}_T = b,$$

where

$$M := \begin{pmatrix} \int_0^T X_r^{(1)} X_r^{(1)} dr & \int_0^T X_r^{(1)} X_r^{(2)} dr & 0 & 0 \\ \int_0^T X_r^{(1)} X_r^{(2)} dr & \int_0^T X_r^{(2)} X_r^{(2)} dr & 0 & 0 \\ 0 & 0 & \int_0^T X_r^{(1)} X_r^{(1)} dr & \int_0^T X_r^{(1)} X_r^{(2)} dr \\ 0 & 0 & \int_0^T X_r^{(1)} X_r^{(2)} dr & \int_0^T X_r^{(2)} X_r^{(2)} dr \end{pmatrix}$$

and

$$b := \begin{pmatrix} \int_0^T X_r^{(1)} dX_r^{(1)} \\ \int_0^T X_r^{(2)} dX_r^{(1)} \\ \int_0^T X_r^{(1)} dX_r^{(2)} \\ \int_0^T X_r^{(2)} dX_r^{(2)} \end{pmatrix}.$$

Which gives

$$\hat{A}_T = \begin{pmatrix} \hat{a}_T^1 \\ \hat{a}_T^2 \\ \hat{a}_T^3 \\ \hat{a}_T^4 \end{pmatrix} = \frac{1}{\int_0^T X_r^{(1)} X_r^{(1)} dr \int_0^T X_r^{(2)} X_r^{(2)} dr - \int_0^T X_r^{(1)} X_r^{(2)} dr \int_0^T X_r^{(1)} X_r^{(2)} dr} \tag{6.2}$$
$$\times \begin{pmatrix} \int_0^T X_r^{(2)} dX_r^{(1)} \int_0^T X_r^{(1)} X_r^{(2)} dr - \int_0^T X_r^{(2)} dX_r^{(1)} \int_0^T X_r^{(1)} X_r^{(2)} dr \\ -\int_0^T X_r^{(2)} dX_r^{(1)} \int_0^T X_r^{(1)} X_r^{(1)} dr + \int_0^T X_r^{(1)} dX_r^{(1)} \int_0^T X_r^{(1)} X_r^{(2)} dr \\ \int_0^T X_r^{(2)} dX_r^{(2)} \int_0^T X_r^{(1)} X_r^{(2)} dr - \int_0^T X_r^{(1)} dX_r^{(2)} \int_0^T X_r^{(2)} X_r^{(2)} dr \\ -\int_0^T X_r^{(2)} dX_r^{(2)} \int_0^T X_r^{(1)} X_r^{(1)} dr + \int_0^T X_r^{(1)} dX_r^{(2)} \int_0^T X_r^{(1)} X_r^{(2)} dr \end{pmatrix}.$$

6.1. **Failure of continuity for the MLE.** For $d \geq 2$ pathwise stability (i.e. in supremum norm) fails in general for the MLE $\hat{A}_T$. We demonstrate this in the setting of the 2-dimensional OU process of the previous section. We construct a path x and a sequence of paths $(x^{(n)})$ such that $x^{(n)} \to x$ uniformly, but

$$|\hat{A}_T(x^{(n)}) - \hat{A}_T(x)| \to \infty$$

as $n \to \infty$. 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 start with an example of $x \in C([0,1], \mathbb{R}^2)$, piecewise smooth, with $\operatorname{span}\{x_t : 0 \leq t \leq 1\} = \mathbb{R}^2$, together with uniform approximation $x^{(n)}$ for which

$$\int x^{(n)} \otimes dx^{(n)} \nrightarrow \int x \otimes dx.$$

Consider $x(t) = (t,0)$ on $[0,1/2]$, and then $x(t) = (1/2, t-1/2)$ on $(1/2,1]$. (The "full span"condition (1.5) is then satisfied.) Write $\mathbf{x} = (x, \int x \otimes dx)$. Consider then dyadic piecewise linear approximations to x, e.g. on a dyadic level $n = 2$ we have $2^n = 4$ intervals and, trivially, $x^n(t) = (t,0)$ on $I_1 = [0,1/4]$ and $x^n(t) = (t,0)$ on $I_2 = [1/4,1/2]$ and so on. Trivially ($x^n \equiv x\ \forall n \geq 1$), $\mathbf{x}^n = (x^n, \int x^n \otimes dx^n) \to \mathbf{x}$, even in rough path metrics. Attach now a loop at the end of each dyadic interval. To this end run at double speed, leaving time for the loop. That is, still with $n = 2$,

$$\begin{aligned} x^{(n)}(t) &= (2t,0) \text{ on } [0,1/8] \\ x^{(n)}(t) &= (1/4,0) + r_n\left(e^{2\pi it} - 1\right) \text{ on } [1/8,1/4]. \end{aligned}$$

This way, at time $1/4$, the end point of I_1, we are at the same point as before, $x^n(t)|_{t=1/4} = x^{(n)}(t)|_{t=1/4} = x(t)|_{t=1/4}$. And so on.

More generally, for arbitrary $n \geq 1$, $x^{(n)}(t) = x(t)$ for all dyadic times $t = 1/2^n$. And, as long as $r_n \to 0$, $x^{(n)} \to x$ uniformly on $[0,1]$. A necessary condition for (rough path) convergence of $\mathbf{x}^{(n)} \to \mathbf{x}$ is the (pointwise) convergence

$$\mathscr{A}_{0,1/2}(x^{(n)}) \to \mathscr{A}_{0,1/2}(x) = 0.$$

where the ärea"$\mathscr{A}_{s,t}(x)$ of $x = (x_1,x_2)$ on (s,t) is given as

$$\mathscr{A}_{s,t}(x) = \int_s^t (x_1(r) - x_1(s))\,dx_2(r) - \int_s^t (x_2(r) - x_2(s))\,dx_1(r).$$

It is easy to see that $\mathscr{A}_{0,1/2}(x^{(x)})$ is the sum of the areas of all the (2^{n-1}) loops (over $[0,1/2]$) upon which $x^{(n)}$ was constructed from x^n. That is,

$$\mathscr{A}_{0,1/2}(x^{(x)}) = 2^{n-1} r_n^2 \pi.$$

Evaluating the first component $\hat{a}^1$ of the likelihood estimator $\hat{A}_T$ from (6.2) for $x = (x_1,x_2)$ and $T = 1$ yields

$$\begin{aligned} \hat{a}_T^1(x) &= \frac{\int_0^T x_1(r)x_2(r)dr}{\int_0^T x_1(r)x_1(r)dr \int_0^T x_2(r)x_2(r)dr - \int_0^T x_1(r)x_2(r)dr \int_0^T x_1(r)x_2(r)dr} \mathscr{A}_{0,T}(x) \\ &=: U(x)\mathscr{A}_{0,T}(x). \end{aligned}$$

The prefactor $U(x)$, consisting only of Riemann integrals, is continuous in supremum and also $U(x^{(n)})$ converges to a finite limit as $n \to \infty$. Taking now $r_n = 2^{-n/4}$ we obtain for the distance of the corresponding estimates

$$\left| \hat{a}_T^1(x^{(n)}) - \hat{a}_T^1(x) \right| = \left| U(x^{(n)})\mathscr{A}_{0,T}(x^{(n)}) - U(x) \right| \sim 2^{n/2-1} \to \infty$$

as $n \to \infty$. The estimation problem is hence not well-posed if one measures distance of paths in supremum norm. We emphasize that stronger pathspace norms such as α-Hölder with $\alpha < 1/2$ will not help; see [13]. For the desired stability, it is crucial to use rough path spaces.

REFERENCES

[1] Avellaneda, M.; Levy, A.; Parás, A. "Pricing and hedging derivative securities in markets with uncertain volatilities.Äpplied Mathematical Finance 2.2 (1995): 73-88.

[2] Caramellino, Lucia. SStrassenŠs law of the iterated logarithm for diffusion processes for small time.SStochastic processes and their applications 74.1 (1998): 1-19.

[3] Decreusefond, L; Üstünel A. SStochastic analysis of the fractional Brownian motion."Potential analysis 10.2 (1999): 177-214.

[4] Friz, Peter, and Atul Shekhar. "Doob–Meyer for rough paths.ärXiv preprint arXiv:1205.2505 (2012).

[5] Friz, P.; Victoir, N.: Multidimensional Stochastic Processes as Rough Paths, Cambridge Studies in Advanced Mathematics Vol 120, Cambridge University Press, 2010

[6] Friz, P.; Gassiat, P.; Lyons, T.: Physcial Brownian motion in magnetic field. arXiv-preprint and accepted at Trans. AMS. (2013)

[7] Gubinelli, Massimiliano. "Controlling rough paths."Journal of Functional Analysis 216.1 (2004): 86-140.

[8] Hairer, Martin, and Natesh S. Pillai. "Regularity of laws and ergodicity of hypoelliptic SDEs driven by rough paths.ärXiv preprint arXiv:1104.5218 (2011).

[9] Kutoyants, Yury A.: Statistical inference for ergodic diffusion processes. Springer Series in Statistics. London: Springer 2004.

[10] Liptser, Robert S.; Shiryaev, Albert N.: Statistics of Random Processes, Vol I+II, Springer 2001

[11] Lyons, T.: Uncertain volatility and the risk-free synthesis of derivatives, Applied Mathematical Finance 2, no. 2, 117-133, 1995

[12] Lyons, T.: Differential equations driven by rough signals, Rev. Mat. Iberoamericana 14, no. 2, 215–310, 1998

[13] Lyons, Terry. Ön the non-existence of path integrals."Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences 432.1885 (1991): 281-290.

[14] Lyons, T.; Qian, Z.: System Control and Rough Paths, Oxford University Press, 2002

[15] Lyons, T.; Caruana, M.; Lévy, T.: Lectures on Probability Theory and Statistics (Saint-Flour, 1996), Vol. 1908 of Lecture Notes in Mathematics, Springer, 2007

[16] Papavasiliou, Anastasia; Ladroue, Christophe: Parameter estimation for rough differential equations. Ann. Statist. 39 (2011), no. 4, 2047Ð2073.

[17] L. C. Young. An inequality of Hňolder type connected with Stieltjes integration. Acta Math., (67):251Ð282, 1936.