Parameter Estimation for Rough Differential Equations

Anastasia Papavasiliou* and Christophe Ladroue†

May 10, 2010

Abstract

We construct the “expected signature matching” estimator for differential equations driven by rough paths and we prove its consistency and asymptotic normality. We use it to estimate parameters of a diffusion and a fractional diffusions, i.e. a differential equation driven by fractional Brownian motion.

Key words: Rough paths, Diffusions, Fractional Diffusions, Generalized Moment Matching, Parameter Estimation.

AMS subject classifications: 62F12, 62M99

1 Introduction

Statistical inference for stochastic processes is a huge field, both in terms of research output and importance. In particular, a lot of work has been done in the context of diffusions (see [20, 14, 11] for a general overview and [5] for some recent developments). Nevertheless, the problem of statistical inference for diffusions still poses many challenges, as for example constructing the Maximum Likelihood Estimator (MLE) for the general multi-dimensional diffusion. An alternative method in this case is that of the Generalized Moment Matching Estimator (GMME). While, in general, less efficient compared to the MLE, the GMME is usually easier to use, more flexible and has been successfully applied to general Markov processes (see [8]).

On the other hand, most methods of statistical inference in the context of non-Markovian continuous processes are restricted to models that depend linearly on the parameter. In the case of differential equations driven by fractional Brownian
motion, some recent results can be found in [9, 11, 22]. In [10], the author discusses the problem of parameter estimation for differential equations driven by Volterra type processes – which include fractional Brownian motion. In all these papers, the analysis is restricted to models that depend linearly on the parameter and for parameters appearing in the drift. Finally, for non-Markovian processes coming from stochastic delay equations, see [13, 21].

The theory of rough paths provides a general framework for making sense of differential equations driven by any type of noise modelled as a rough path – this includes diffusions, differential equations driven by fractional Brownian motion, delay equations and even delay equation driven by fractional Brownian motion (see [18]). The basic ideas have been developed in the '90s (see [17] and references within). However, the problem of statistical inference for differential equations driven by rough paths has not been addressed yet. This is exactly what we strive to do in this paper.

The exact setting of the statistical problem we consider is the following: we observe many independent copies of specific iterated integrals of the response \( \{Y_t, \ 0 < t < T\} \) of a differential equation

\[
  dY_t = f(Y_t; \theta) \cdot dX_t, \ Y_0 = y_0
\]

driven by the rough path \( X \). We will formally define what we mean by a rough path and a differential equation driven by it in section 2.1. Two examples of interest are \( X_t = (t, W_t) \) where \( W_t \) is Brownian motion and the differential equation is a Stratonovic stochastic differential equation and \( X_t = (t, B^H_t) \) where \( B^H_t \) is fractional Brownian motion. The iterated integrals are observed at a fixed time \( T \). However, if the response lives in more than one dimension, the iterated integrals will be functions of the whole path. For example, suppose that \( Y_t = (Y^{(1)}_t, Y^{(2)}_t) \) and we observe

\[
  \int \int_{0<u_1<u_2<T} dY^{(2)}_{u_1} dY^{(1)}_{u_2}
\]

for fixed time \( T \). We further assume that the vector field \( f(y; \theta) \) is polynomial in \( y \) and depends on the unknown parameter \( \theta \). Finally, we assume that we know the expected signature of the rough path \( X \) on the interval \([0, T]\). Again, the signature of a rough path will be formally defined later. For now, let’s just say that it is the set of all iterated integrals of \( X \) and its expectation fully describes the distribution of the rough path \( X \).

The first assumption is a bit unusual: it is much more common to assume that we observe one long path rather than many short ones. This setting is chosen for two reasons. The first is its simplicity: we develop here some basic tools for statistical inference of differential equation driven by rough paths. These can be generalized to other settings, such as observing one continuous path, provided that some ergodicity conditions are fulfilled. However, there are no general results on the ergodicity of differential equations driven by rough paths and ergodicity has to be checked for each case separately. For example, see [23] for some recent results on the ergodicity of differential equations driven by fractional Brownian motion.

The second reason was that such settings arise in the context of “equation-free” modelling of multiscale models (see [11]). Suppose that we have access to some
code that simulates the dynamics of a complex system, such as molecular dynamics. We treat the code as a “black box”. We are interested in the global behavior of a function of our system that “lives” in the slow scale, i.e. in some limit its dynamics follow a diffusion, which is, however, unknown. The basic idea of “equation-free” modelling is to run the code for a short time and use the output to locally estimate the parameters of the differential equation. This process is repeated several times with carefully chosen initial conditions, so as to get an estimate of the global dynamics. To summarize, in this problem:

(a) we observe many independent paths;

(b) time is short;

(c) we locally approximate the vector field by a polynomial.

Currently, the estimation is done using the MLE approach, pretending that the data comes from the diffusion rather than the multiscale model (see [2]). However, for short time $T$ we cannot expect the diffusion approximation to be a good one. We believe that in the scale of $T$, we can always approximate the dynamics by a differential equation driven by a rough path (see [19]).

The structure of the paper is the following: we start by reviewing some basic concepts and results from the theory of rough paths and we give a precise description of the problem we consider. In section 3, we describe the methodology. The idea is simple: we want to match the theoretical and the expected signatures of the response. However, in general we cannot expect to get an explicit formula for the theoretical expected signature, so we construct an approximation of it. We go on to give a precise definition of the “expected signature matching estimator” using this approximation and prove its consistency and asymptotic normality.

In section 4, we apply the method to two examples that represent the most common RDEs: diffusions and differential equations driven by fractional Brownian motion. We have written a package in Mathematica (available upon request) that can be used to recreate the examples we include in the paper or try out new ones.

2 Setting

2.1 Some basic results from the theory of Rough Paths

In this section, we review some of the basic results from the theory of rough paths. For more details, see [10] and references within. The goal of this theory is to give meaning to the differential equation

$$dY_t = f(Y_t) \cdot dX_t, \ Y_0 = y_0.$$ (1)

for very general continuous paths $X$. More specifically, we think of $X$ and $Y$ as paths on a Euclidean space: $X : I \to \mathbb{R}^n$ and $Y : I \to \mathbb{R}^m$ for $I := [0, T]$, so $X_t \in \mathbb{R}^n$ and $Y_t \in \mathbb{R}^m$ for each $t \in I$. Also, $f : \mathbb{R}^m \to L(\mathbb{R}^n, \mathbb{R}^m)$, where $L(\mathbb{R}^n, \mathbb{R}^m)$ is the space of
linear functions from \( \mathbb{R}^n \) to \( \mathbb{R}^m \) which is isomorphic to the space of \( m \times n \) matrices. For the sake of simplicity, we will assume that \( f(y) \) is a polynomial in \( y \) – however, the theory holds for more general \( f \). The path \( X \) is any path of finite \( p \)-variation, meaning that

\[
\sup_{D \subset [0,T]} \left( \sum_{\ell} \|X_{t_{\ell}} - X_{t_{\ell-1}}\|^p \right)^{1/p} < \infty,
\]

where \( D = \{t_{\ell}\}_\ell \) goes through all possible partitions of \([0,T]\) and \( \| \cdot \| \) is the Euclidean norm. Note that we will later define finite \( p \)-variation for multiplicative functionals, also to be defined later.

The fact the \( X \) is allowed to have any finite \( p \)-variation is exactly what makes this theory so general: Brownian motion is an example of a path that has finite \( p \)-variation for any \( p > 2 \) while fractional Brownian motion with Hurst index \( h \) has finite \( p \) variation for \( p > \frac{1}{h} \). We will define fractional Brownian motion formally in the corresponding example – for now, let us just say that it is Gaussian, self-similar but not Markovian except for \( h = 1/2 \) when it coincides with Brownian motion.

When \( p \in [1,2) \), we say that \( Y \) is a solution of (1) if

\[
Y_t = Y_s + \int_s^t f(Y_u) \cdot dX_u, \quad \forall (s,t) \in \Delta_T,
\]

where \( \Delta_T := \{(s,t); 0 \leq s \leq t \leq T\} \). In this case, the integral is defined as the Young integral (see [17]). What does it mean for \( Y \) to be a solution of (1) when \( p \geq 2 \)? In order to answer this question, we first need to define the integral. To make this task possible, we re-write the integral so that the integrand is a function of the integrator: Set \( f_{y_0}(\cdot) := f(\cdot + y_0) \). Define \( h : \mathbb{R}^n \oplus \mathbb{R}^m \to \text{End}(\mathbb{R}^n \oplus \mathbb{R}^m) \) by

\[
h(x,y) := \begin{pmatrix} I_{n \times n} & 0_{n \times m} \\ f_{y_0}(y) & 0_{m \times m} \end{pmatrix}.
\]

Instead of defining \( \int_s^t f(Y_u) \cdot dX_u \), we will define the integral

\[
\int_s^t h(Z_u) \cdot dZ_u, \quad \forall (s,t) \in \Delta_T,
\]

where \( Z = (X,Y) \). Note that if \( f \) is a polynomial in \( y \), then \( h \) will also be a polynomial in \( z \). More generally, we will define this integral for any path \( Z \) in \( \mathbb{R}^{\ell_1} \) of finite \( p \)-variation and any polynomial \( h : \mathbb{R}^{\ell_1} \to L(\mathbb{R}^{\ell_1}, \mathbb{R}^{\ell_2}) \) of degree \( q \). Since \( h \) is a polynomial, its Taylor expansion will be a finite sum:

\[
h(z_2) = \sum_{k=0}^q h_k(z_1) \frac{(z_2 - z_1)^{\otimes k}}{k!}, \quad \forall z_1, z_2 \in \mathbb{R}^{\ell_1}
\]

where \( h_0 = h \) and \( h_k : \mathbb{R}^{\ell_1} \to L(\mathbb{R}^{\ell_1 \otimes k}, L(\mathbb{R}^{\ell_1}, \mathbb{R}^{\ell_2})) \) and for all \( z \in \mathbb{R}^{\ell_1} \), \( h_k(z) \) is a symmetric \( k \)-linear mapping from \( \mathbb{R}^{\ell_1} \) to \( L(\mathbb{R}^{\ell_1}, \mathbb{R}^{\ell_2}) \), for \( k \geq 1 \).
Suppose that $Z$ is a path of bounded variation (i.e. $p = 1$). Then, using the symmetry of $h_k(z)$ and the “shuffle product property”, we can write

$$h(Z_u) = \sum_{k=0}^q h_k(Z_s)Z_{s,u}^k, \quad \forall (s, u) \in \Delta_T$$

where for every $(s, t) \in \Delta_T$, $Z^0 \equiv 1 \in \mathbb{R}$ and $Z_{s,t}^k = \left\{ \int \ldots \int_{s < u_1 < \ldots < u_k < t} dZ_{u_1}^{(i_1)} \ldots dZ_{u_k}^{(i_k)} \right\}_{(i_1, \ldots, i_k) \in \{1, \ldots, n\}^k} \in \mathbb{R}^{\ell_1 \otimes k}$

More specifically, we use the notation

$$Z_{s,t}^{(i_1, \ldots, i_k)} := \int \ldots \int_{s < u_1 < \ldots < u_k < t} dZ_{u_1}^{(i_1)} \ldots dZ_{u_k}^{(i_k)}.$$

The “shuffle product property” says that for any $(s, u) \in \Delta_T$ and any “words” $\sigma_1, \sigma_2 \in \bigcup_{k \geq 0} \{1, \ldots, \ell_1\}^k$, we can write

$$Z_{s,u}^{\sigma_1}Z_{s,u}^{\sigma_2} = \sum_{\sigma \in \sigma_1 \sqcup \sigma_2} Z_{s,u}^\sigma,$$

(4)

where $\sigma_1 \sqcup \sigma_2$ is the shuffle product between the words $\sigma_1$ and $\sigma_2$, i.e. it is the set of all words (with repetition) that we can create by mixing up the letters of $\sigma_1$ and $\sigma_2$ without changing the order of letters within each word. For example, $(1, 2) \sqcup (2) = \{(1, 2, 2), (1, 2, 2), (2, 1, 2)\}$ (see [16]). This generalizes the “integration by parts” formula. Then, for all $(s, t) \in \Delta_T$,

$$\int_s^t h(Z_u)dZ_u = \sum_{k=0}^q h_k(Z_s)Z_{s,t}^{k+1}.$$

Example 1. Let us demonstrate what we have said so far with an example. Consider the ordinary differential equation

$$dY_t = Y_t dt + (Y_t^2 + 1)de^t, \quad Y_0 = 0.$$

Then, $X_t = (t, e^t)$ is a path in $\mathbb{R}^2$, $Y_t \in \mathbb{R}$ and $f(y) = (y, y^2 + 1) \in L(\mathbb{R}^2, \mathbb{R})$, which is polynomial of degree 2. In this case, $X$ is of bounded variation and $p = 1$. Following what we just mentioned, instead of defining the integral

$$\int_s^t f(Y_u)dX_u = \int_s^t (Y_u du + (Y_u^2 + 1)de^u)$$

directly, we set $Z_t = (X_t, Y_t)' = (t, e^t, Y_t)' \in \mathbb{R}^3$ and

$$h(Z_t) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Z_t^{(3)} & (Z_t^{(3)})^2 + 1 & 0 \end{pmatrix},$$
where \( Z^{(3)}_t = Y_t \) is the projection of \( Z_t \) to the third dimension. Then, the integral 
\[
\int_s^t h(Z_u)\,dZ_u = \left( 0, 0, \int_s^t f(Y_u)\,dX_u \right),
\]
so, defining \( \int_s^t h(Z_u)\,dZ_u \) is equivalent to defining \( \int_s^t f(Y_u)\,dX_u \). We now proceed to writing the integral as a linear combination of iterated integrals of \( Z \), using the fact that \( h \) is a quadratic polynomial. We define \( h_k \) as
\[
h_0(z) = h(z), \quad h_1(z) = \{\partial_i h(z)\}_{i=1}^3, \quad h_2(z) = \{\partial_{i_1,i_2} h(z)\}_{i_1,i_2=1}^3.
\]
Also, we note that
\[
((z_2 - z_1)^{\otimes 1})_i = z_2^{(i)} - z_1^{(i)} \quad \text{and} \quad ((z_2 - z_1)^{\otimes 2})_{i_1,i_2} = (z_2^{(i_1)} - z_1^{(i_1)})(z_2^{(i_2)} - z_1^{(i_2)})
\]
and thus the sum \( \sum_{k=0}^2 h_k(z_1)^{\otimes k}_{s,t} \) becomes
\[
\left( \begin{array}{c} 0 \\ 0 \\ z^{(3)}_1 + (z^{(3)}_1)^2 + 1 \end{array} \right) + \left( \begin{array}{c} 0 \\ 0 \\ (1 + 2z^{(3)}_1)(z^{(3)}_2 - z^{(3)}_1) \end{array} \right) + \left( \begin{array}{c} 0 \\ 0 \\ \frac{2(z^{(3)}_2 - z^{(3)}_1)^2}{2} \end{array} \right)
\]
which is equal to \( h(z_2) \). It is easy to see that for all \( 0 \leq s < t < T \),
\[
(z^{(3)}_t - z^{(3)}_s) = \int_s^t dz^{(3)}_u \quad \text{and} \quad \frac{(z^{(3)}_t - z^{(3)}_s)^2}{2} = \int_s^t \int_s^{u_1} dz^{(3)}_{u_1} dz^{(3)}_{u_2}.
\]
Thus, using the notation of the iterated integral, we write
\[
h(z_u) = h(z_s) + \partial_3 h(z_s) Z^{(3)}_{s,u} + \partial_3^2 h(z_s) Z^{(3,3)}_{s,u}
\]
and if we integrate once more we get
\[
\int_s^t h(z_u)\,du = h(z_s)Z^{(3)}_{s,t} + \partial_3 h(z_s)Z^{(3,3)}_{s,t} + \partial_3^2 h(z_s)Z^{(3,3,3)}_{s,t}.
\]
Note that in the above example, we did not use the shuffle product formula because \( m = 1 \) (\( Y_t \in \mathbb{R} \)). If the response \( Y \) lives in more than one dimensions, then the shuffle product formula is used, for example, to say that
\[
\frac{1}{2} (z_t - z_s)^{(i_2)} (z_t - z_s)^{(i_1)} = \frac{1}{2} Z^{(i_2)}_{s,t} Z^{(i_1)}_{s,t} = Z^{(i_1,i_2)}_{s,t} + Z^{(i_2,i_1)}_{s,t}.
\]
Below we give a concrete example to show how the shuffle product formula extends integration by parts.
Example 2. Let us give here an example of the shuffle product. Let \( z_t \) be a smooth path in \( \mathbb{R}^m \) for some \( m \geq 1 \). Then, for any pair \( i_1, i_2 \in \{1, \ldots, m\} \) using the integration by parts formula, we get
\[
Z^{(i_1, i_2)}_{s, t} = \int_s^t \int_s^u dz_u^{(i_1)} dz_u^{(i_2)} = \int_s^t (z_u^{(i_1)} - z_s^{(i_1)}) dz_u^{(i_2)} = \\
= \int_s^t z_u^{(i_1)} dz_u^{(i_2)} - z_s^{(i_1)} (z_t^{(i_2)} - z_s^{(i_2)}) = \\
= [z_u^{(i_1)} z_u^{(i_2)}]_s^t - \int_s^t z_u^{(i_2)} dz_u^{(i_1)} - z_s^{(i_1)} (z_t^{(i_2)} - z_s^{(i_2)}) = \\
= z_t^{(i_2)} (z_t^{(i_1)} - z_s^{(i_1)}) - \int_s^t z_u^{(i_2)} dz_u^{(i_1)} = \\
= (z_t^{(i_2)} - z_s^{(i_2)}) (z_t^{(i_1)} - z_s^{(i_1)}) - \int_s^t (z_u^{(i_2)} - z_s^{(i_2)}) dz_u^{(i_1)} = \\
= Z^{(i_1)}_{s, t} Z^{(i_2)}_{s, t} - Z^{(i_2, i_1)}_{s, t}
\]
which is in agreement with the shuffle product formula, since the shuffle product of two letters is \( (i_1) \cup (i_2) = \{(i_1, i_2), (i_2, i_1)\} \).

It is now clear that in order to extend this construction to any path \( Z \) of finite \( p \)-variation, where \( p \geq 2 \), we will first need to define their iterated integrals \( Z_{s, t}^k \). These are not necessarily unique (for example, if \( Z \) is Brownian motion, then Itô and Stratonovici gave two different definitions for the integral). Then, we will need to find those integrals that respect the “shuffle product property”. Before going any further, we need to give some definitions:

**Definition 2.1.** Let \( \Delta_T := \{(s, t); 0 \leq s \leq t \leq T\} \). Let \( p \geq 1 \) be a real number. We denote by \( T^{(k)}(\mathbb{R}^{\ell_1}) \) the \( k^{\text{th}} \) truncated tensor algebra
\[
T^{(k)}(\mathbb{R}^{\ell_1}) := \mathbb{R} \oplus \mathbb{R}^{\ell_1} \oplus \mathbb{R}^{\ell_1 \otimes 2} \oplus \cdots \oplus \mathbb{R}^{\ell_1 \otimes k}.
\]
(1) Let \( Z : \Delta_T \to T^{(k)}(\mathbb{R}^{\ell_1}) \) be a continuous map. For each \( (s, t) \in \Delta_T \), denote by \( Z_{s, t} \) the image of \( (s, t) \) through \( Z \) and write
\[
Z_{s, t} = (Z^0_{s, t}, Z^1_{s, t}, \ldots, Z^k_{s, t}) \in T^{(k)}(\mathbb{R}^{\ell_1}), \text{ where } Z^j_{s, t} = \left\{ Z^{(i_1, \ldots, i_j)}_{s, t} \right\}_{i_1, \ldots, i_j = 1}^{\ell_1}.
\]
The function \( Z \) is called a multiplicative functional of degree \( k \) in \( \mathbb{R}^{\ell_1} \) if \( Z^0_{s, t} = 1 \) for all \( (s, t) \in \Delta_T \) and
\[
Z_{s, u} \otimes Z_{u, t} = Z_{s, t} \forall s, u, t \text{ satisfying } 0 \leq s \leq u \leq t \leq T,
\]
i.e. for every \( (i_1, \ldots, i_l) \in \{1, \ldots, \ell_1\}^l \) and \( l = 1, \ldots, k \):
\[
(Z_{s, u} \otimes Z_{u, t})^{(i_1, \ldots, i_l)} = \sum_{j=0}^l Z_{s, u}^{(i_1, \ldots, i_j)} Z_{u, t}^{(i_{j+1}, \ldots, i_l)}
\]
This is called Chen’s identity.
(2) A p-rough path $Z$ in $\mathbb{R}^{\ell_1}$ is a multiplicative functional of degree $[p]$ in $\mathbb{R}^{\ell_1}$ that has finite $p$-variation, i.e. $\forall i = 1, \ldots, [p]$ and $(s, t) \in \Delta_T$, it satisfies
\[
\|X_{s,t}^i\| \leq \frac{(M(t-s))^{i/\beta}}{\beta (\frac{i}{p})!},
\]
where $\| \cdot \|$ is the Euclidean norm in the appropriate dimension and $\beta$ a real number depending only on $p$ and $M$ is a fixed constant. The space of p-rough paths in $\mathbb{R}^{\ell_1}$ is denoted by $\Omega_p(\mathbb{R}^{\ell_1})$.

(3) A geometric p-rough path is a p-rough path that can be expressed as a limit of 1-rough paths in the $p$-variation distance $d_p$, defined as follows: for any $X, Y$ continuous functions from $\Delta_T$ to $T([p]) (\mathbb{R}^{\ell_1})$,
\[
d_p(X, Y) = \max_{1 \leq i \leq [p]} \sup_{D \subset [0,T]} \left( \sum_{t} \|X_{t_{t-1}}^{i} - Y_{t_{t-1}}^{i}\|^\frac{2}{p} \right),
\]
where $D = \{t_i\}_t$ goes through all possible partitions of $[0,T]$. The space of geometric p-rough paths in $\mathbb{R}^n$ is denoted by $\mathcal{G} \Omega_p(\mathbb{R}^{\ell_1})$.

One of the main results of the theory of rough paths is the following, called the "extension theorem":

**Theorem 2.2** (Theorem 3.7, [16]). Let $p \geq 1$ be a real number and $k \geq 1$ be an integer. Let $X : \Delta_T \to T^{(k)}(\mathbb{R}^n)$ be a multiplicative functional with finite $p$-variation. Assume that $k \geq [p]$. Then there exists a unique extension of $X$ to a multiplicative functional $X : \Delta_T \to T^{(k+1)}(\mathbb{R}^n)$.

Let $X : [0,T] \to \mathbb{R}^n$ be an $n$-dimensional path of finite $p$-variation for $n > 1$. One way of constructing a p-rough path is by considering the set of all iterated integrals of degree up to $[p]$. If $X_t = (X^{(1)}_t, \ldots, X^{(n)}_t)$, we define $X : \Delta_T \to T^{([p])}$ as follows:

\[
X^0 \equiv 1 \in \mathbb{R} \text{ and } X^k_{s,t} = \left\{ \int \ldots \int_{s < u_1 < \ldots < u_k < t} dX_{u_1}^{(i_1)} \ldots dX_{u_k}^{(i_k)} \right\}_{(i_1, \ldots, i_k) \in \{1, \ldots, n\}^k} \in \mathbb{R}^{n \otimes k}
\]
for $k = 1, \ldots, [p]$. Note that Chen's identity is an identity all iterated integrals satisfy. For example, for word $(i_1, i_2)$ Chen's identity says that
\[
(Z_{s,t})^{(i_1,i_2)} = (Z_{s,u})^{(i_1,i_2)} + (Z_{s,u})^{(i_1)} (Z_{u,t})^{(i_2)} + (Z_{u,t})^{(i_1,i_2)}.
\]

This follows by breaking the domain of integration $\{u_1, u_2 : s < u_1 < u_2 < t\}$ into three domains $\{u_1, u_2 : s < u_1 < u_2 < u\}, \{u_1, u_2 : u < u_1 < u_2 < t\}$ and $\{u_1, u_2 : s < u_1 < u \text{ and } u < u_2 < t\}$.

When $p \in [1,2)$, the iterated integrals are uniquely defined as Young integrals. However, as we already mentioned, when $p \geq 2$ there will be more than one way of defining them. What the extension theorem says is that if the path has finite
p-variation and we define the first \([p]\) iterated integrals, the rest will be uniquely defined. So, if the path is of bounded variation \((p = 1)\) we only need to know its increments, while for an \(n\)-dimensional Brownian path, we need to define the second iterated integrals by specifying the rules on how to construct them. In general, we can think of a \(p\)-rough path as a path \(X : [0, T] \to \mathbb{R}^n\) of finite \(p\)-variation, together with a set of rules on how to define the first \([p]\) iterated integrals. Once we know how to construct the first \([p]\), we know how to construct all of them.

**Definition 2.3.** Let \(X : [0, T] \to \mathbb{R}^n\) be a path. The set of all iterated integrals is called the signature of the path and is denoted by \(S(X)\).

We can now proceed to define the integral \((3)\) when \(Z\) is a path of finite \(p\)-variation with \(p \geq 2\). First, it is clear that in order for the integral to be uniquely defined, we should define the first \([p]\) iterated integrals, so we define the integral not with respect to \(Z\) but a corresponding \(p\)-rough path \(Z\). To extend the previous construction, we also need that \(Z\) satisfies the “shuffle product property”. It is not hard to see that geometric \(p\)-rough paths do satisfy this property since they are limits of paths of bounded variation and for paths of bounded variation the property follows from the usual integration by parts formula (see also \([17]\)). So, we will define \(\int h(Z)\,dZ\), where \(Z\) is a geometric \(p\)-rough path in \(\mathbb{R}^{\ell_1}\), i.e. \(Z \in G\Omega_p(\mathbb{R}^{\ell_1})\).

By definition, there exists a sequence \(Z(r) \in \Omega_1(\mathbb{R}^{\ell_1})\) such that \(d_p(Z(r), Z) \to 0\) as \(r \to \infty\). Then, for each \(r > 0\), we define \(\tilde{Z}(r) := \int h(Z(r))\,dZ(r)\). These are also a \(1\)-rough paths in \(\mathbb{R}^{\ell_2}\) and thus, their higher iterated integrals are uniquely defined. In addition, it is possible to show that the map \(\int h : \Omega_1(\mathbb{R}^{\ell_1}) \to \Omega_1(\mathbb{R}^{\ell_2})\) sending \(Z(r)\) to \(\tilde{Z}(r)\) is continuous in the \(p\)-variation topology.

We define \(\tilde{Z} := \int h(Z)\,dZ\) as the limit of the \(\tilde{Z}(r)\) with respect to \(d_p\) – this will also be a geometric \(p\)-rough path. In other words, the continuous map \(\int h\) can be extended to a continuous map from \(G\Omega_p(\mathbb{R}^{\ell_1})\) to \(G\Omega_p(\mathbb{R}^{\ell_2})\), which are the closures of \(\Omega_1(\mathbb{R}^{\ell_1})\) and \(\Omega_1(\mathbb{R}^{\ell_2})\) respectively (see Theorem 4.12, \([16]\)).

Note that this construction of the integral can be extended for any \(h \in \text{Lip}(\gamma - 1)\) for \(\gamma > p\) (see \([16]\)).

**Remark 2.4.** We say that a sequence \(Z(r)\) of \(p\)-rough paths converges to a \(p\)-rough path \(Z\) in \(p\)-variation topology if there exists an \(M \in \mathbb{R}\) and a sequence \(a(r)\) converging to zero when \(r \to \infty\), such that

\[
\|Z(r)_{s,t}\|, \|Z'_{s,t}\| \leq (M(t - s))^{\frac{1}{p}}, \text{ and } \|Z(r)_{s,t} - Z'_{s,t}\| \leq a(r)(M(t - s))^{\frac{1}{p}}
\]

for \(i = 1, \ldots, [p]\) and \((s, t) \in \Delta_T\). Note that this is not exactly equivalent to convergence in \(d_p\): while convergence in \(d_p\) implies convergence in the \(p\)-variation topology, the opposite is not true. Convergence in the \(p\)-variation topology implies that there is a subsequence that converges in \(d_p\).

We can now give the precise meaning of the solution of \((1)\), when driven not by a path \(X\) but a geometric \(p\)-rough path \(\tilde{X}\):
Definition 2.5. Consider $X \in G\Omega_p(\mathbb{R}^n)$ and $y_0 \in \mathbb{R}^m$. Set $f_{y_0}(\cdot) := f(\cdot + y_0)$ and define $h : \mathbb{R}^n \oplus \mathbb{R}^m \to \text{End}(\mathbb{R}^n \oplus \mathbb{R}^m)$ as in (2). We call $Z \in G\Omega_p(\mathbb{R}^n \oplus \mathbb{R}^m)$ a solution of (1) if the following two conditions hold:

(i) $Z = \int h(Z) dZ$.

(ii) $\pi_{\mathbb{R}^n}(Z) = X$, where by $\pi_{\mathbb{R}^n}$ we denote the projection of $Z$ to $\mathbb{R}^n$.

As in the case of ordinary differential equations ($p = 1$), it is possible to construct the solution using Picard iterations: we define $Z(0) := (X, e)$, where by $e$ we denote the trivial rough path $e = (1, 0_{\mathbb{R}^n}, 0_{\mathbb{R}^n \otimes 2}, \ldots)$. Then, for every $r \geq 1$, we define $Z(r) = \int h(Z(r - 1)) dZ(r - 1)$. The following theorem, called the “Universal Limit Theorem”, gives the conditions for the existence and uniqueness of the solution to (1). The theorem holds for any $f \in \text{Lip}(\gamma)$ for $\gamma > p$ but we will assume that $f$ is a polynomial. The proof is based on the convergence of the Picard iterations.

Theorem 2.6 (Theorem 5.3, [16]). Let $p \geq 1$. For all $X \in G\Omega_p(\mathbb{R}^n)$ and all $y_0 \in \mathbb{R}^m$, equation (1) admits a unique solution $Z = (X, Y) \in G\Omega_p(\mathbb{R}^n \oplus \mathbb{R}^m)$, in the sense of definition 2.5. This solution depends continuously on $X$ and $y_0$ and the mapping $I_f : G\Omega_p(\mathbb{R}^n) \to G\Omega_p(\mathbb{R}^m)$ which sends $(X, y_0)$ to $Y$ is continuous in the $p$-variation topology.

The rough path $Y$ is the limit of the sequence $Y(r)$, where $Y(r)$ is the projection of the $r$th Picard iteration $Z(r)$ to $\mathbb{R}^m$. For all $\rho > 1$, there exists $T_\rho \in (0, T]$ such that

$$\|Y(r)_{s,t}^i - Y(r + 1)_{s,t}^i\| \leq 2^i \rho^{-r} (M(t - s))^{\frac{i}{p}}, \forall (s, t) \in \Delta T_\rho, \forall i = 0, \ldots, [p].$$

The constant $T_\rho$ depends only on $f$ and $p$.

2.2 The problem

We now describe the problem that we are going to study in the rest of the paper. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to G\Omega_p(\mathbb{R}^n)$ a random variable, taking values in the space of geometric $p$-rough paths endowed with the $p$-variation topology. For each $\omega \in \Omega$, the rough path $X(\omega)$ drives the following differential equation

$$dY_t(\omega) = f(Y_t(\omega); \theta) \cdot dX_t(\omega), \quad Y_0 = y_0$$

where $\theta \in \Theta \subseteq \mathbb{R}^d$, $\Theta$ being the parameter space and for each $\theta \in \Theta$. As before, $f : \mathbb{R}^m \times \Theta \to L(\mathbb{R}^n, \mathbb{R}^m)$ and $f_\theta(y) := f(y; \theta)$ is a polynomial in $y$ for each $\theta \in \Theta$. According to theorem 2.6, we can think of equation (5) as a map

$$I_{f_\theta,y_0} : G\Omega_p(\mathbb{R}^n) \to G\Omega_p(\mathbb{R}^m),$$

sending a geometric $p$-rough path $X$ to a geometric $p$-rough path $Y$ and is continuous with respect to the $p$-variation topology. Consequently,

$$Y := I_{f_\theta,y_0} \circ X : \Omega \to G\Omega_p(\mathbb{R}^m)$$
is also a random variable, taking values in $G_{\Omega_p}(\mathbb{R}^m)$ and if $\mathbb{P}^T$ is the distribution of $X_{0,T}$, the distribution of $Y_{0,T}$ will be

$$Q^T_\theta = \mathbb{P}^T \circ I_{f_{\theta,y_0}}^{-1},$$

(7)

Suppose that we know the expected signature of $X$ at $[0,T]$, i.e. we know

$$\mathbb{E} \left( X_{0,T}^{(i_1,\ldots,i_k)} \right) := \mathbb{E} \left( \int \cdots \int_{0<u_1<\ldots<u_k<T} dX_{u_1}^{(i_1)} \cdots dX_{u_k}^{(i_k)} \right),$$

for all $i_j \in \{1,\ldots,n\}$ where $j = 1,\ldots,k$ and $k \geq 1$. Our goal will be to estimate $\theta$, given several realizations of $Y_{0,T}$, i.e. $\{Y_{0,T}(\omega_i)\}_{i=1}^N$.

3 Method

In order to estimate $\theta$, we are going to use a method that is similar to the “Method of Moments”. The idea is simple: we will try to (partially) match the empirical expected signature of the observed $p$-rough path with the theoretical one, which is a function of the unknown parameters. Remember that the data we have available is several realizations of the $p$-rough path $Y_{0,T}$ described in section 2.2. To make this more precise, let us introduce some notation: let

$$E^\tau(\theta) := \mathbb{E}_\theta(Y_{0,T}^\tau)$$

(8)

be the theoretical expected signature corresponding to parameter value $\theta$ and word $\tau$ and

$$M_N^\tau := \frac{1}{N} \sum_{i=1}^N Y_{0,T}^\tau(\omega_i)$$

(9)

be the empirical expected signature, which is a Monte Carlo approximation of the actual one. The word $\tau$ is constructed from the alphabet $\{1,\ldots,m\}$, i.e. $\tau \in W_m$ where $W_m := \bigcup_{k \geq 0} \{1,\ldots,m\}^k$. The idea is to find $\hat{\theta}$ such that

$$E^\tau(\hat{\theta}) = M_N^\tau, \forall \tau \in V \subset W_m$$

for some choice of a set of words $V$. Then, $\hat{\theta}$ will be our estimate.

**Remark 3.1.** When $m = 1$, the expected signature of $Y$ is equivalent to its moments, since

$$Y_{0,T}^{(1,\ldots,1)}(\omega_i) = \frac{1}{m!}(Y_T - Y_0)^m.$$  

When $m = 2$, one example is to consider the word $\tau = (1,2)$. Then, one needs to compute the iterated integral (or an approximation of, if the path is discretely observed)

$$Y_{0,T}^{(1,2)}(\omega_i) = \int_0^T \int_0^s dY_{u}^{(1)}(\omega_i)dY_{s}^{(2)}(\omega_i)$$
for each path $Y_t(\omega_i) = \left( Y_t^{(1)}(\omega_i), Y_t^{(2)}(\omega_i) \right)$, for $i = 1, \ldots, N$. Then

$$M^{(1,2)}_N := \frac{1}{N} \sum_{i=1}^{N} Y^{(1,2)}_{0,T}(\omega_i).$$

Note that this is closely related to the correlation of the two one-dimensional paths $\{Y_t^{(1)}\}_{t \in [0,T]}$ and $\{Y_t^{(2)}\}_{t \in [0,T]}$ since, by the shuffle product,

$$Y^{(1,2)}_{0,T}(\omega_i) + Y^{(2,1)}_{0,T}(\omega_i) = Y^{(1)}_{0,T}(\omega_i) Y^{(2)}_{0,T}(\omega_i)$$

and by the law of large numbers,

$$\lim_{N \to \infty} \left( M^{(1,2)}_N + M^{(2,1)}_N \right) = \mathbb{E} \left( (Y^{(1)}_{T} - Y^{(1)}_{0}) (Y^{(2)}_{T} - Y^{(2)}_{0}) \right).$$

Several questions arise:

(i) How can we get an analytic expression for $E^\tau(\theta)$ as a function of $\theta$?

(ii) What is a good choice for $V$ or, for $m = 1$, how do we choose which moments to match?

(iii) How good is $\hat{\theta}$ as an estimate?

We will try to answer these questions below.

### 3.1 Computing the Theoretical Expected Signature

We want to get an analytic expression for the expected signature of the $p$-rough path $Y$ at $(0,T)$, where $Y$ is the solution of (5) in the sense described above. In other words, we want to compute (8). We are given the expected signature of the $p$-rough path $X$ which is driving the equation, again at $(0,T)$, i.e. we are given

$$\mathbb{E} \left( X^{(k)}_{0,T} \right), \quad \forall \sigma \in \{1, \ldots, n\}^k, k \in \mathbb{N}.$$

In addition, we know the vector field $f_\theta(y) = f(y; \theta)$ in (5), up to parameter $\theta$ and we know that it is polynomial.

It turns out that we cannot compute (8), in general. We need to make one more approximation since the solution $Y$ will not usually be available: we will approximate the solution by the $r$th Picard iteration $Y(r)$, described in the Universal Limit Theorem (Theorem 2.6). Finally, we will approximate the expected signature of the solution corresponding to a word $\tau$, $E^\tau(\theta)$, by the expected signature of the $r$th Picard iteration at $\tau$, which we will denote by $E^\tau_r(\theta)$:

$$E^\tau_r(\theta) := \mathbb{E}_\theta(Y(r)_{0,T}^\tau).$$

The good news is that when $f_\theta$ is a polynomial of degree $q$ on $y$, for any $q \in \mathbb{N}$, the $r$th Picard iteration of the solution is a linear combination of iterated integrals
of the driving force $X$. More specifically, for any realization $\omega$ and any time interval $(s, t) \in \Delta_T$, we can write:

$$Y(r)_{s,t} = \sum_{|\sigma| \leq |r| \frac{q-1}{q}} \alpha_{r,\sigma}(y_0, s; \theta) X_{s,t}^\sigma,$$

(11)

where $\alpha_{r,\sigma}(y; \theta)$ is a polynomial in $y$ of degree $q^r$ and $|\cdot|$ gives the length of a word. Thus,

$$E_r^\tau(\theta) = \sum_{|\sigma| \leq |r| \frac{q-1}{q}} \alpha_{r,\sigma}(y_0, s; \theta) E(X_{s,t}^\sigma),$$

(12)

We will prove (11), first for $p = 1$ and then for any $p \geq 1$ by taking limits with respect to $d_p$. We will need the following lemma.

**Lemma 3.2.** Suppose that $X \in G\Omega_1(\mathbb{R}^n)$, $Y \in G\Omega_1(\mathbb{R}^m)$ and it is possible to write

$$Y_{s,t}^{(j)} = \sum_{\sigma \in W_n, \quad q_1 \leq |\sigma| \leq q_2} \alpha_\sigma^{(j)}(y_s) X_{s,t}^\sigma, \quad \forall (s, t) \in \Delta_T \quad \text{and} \quad \forall j = 1, \ldots, m$$

(13)

where $\alpha_\sigma^{(j)} : \mathbb{R}^m \to L(\mathbb{R}, \mathbb{R})$ is a polynomial of degree $q$ with $q, q_1, q_2 \in \mathbb{N}$ and $q_1 \geq 1$. Then,

$$Y_{s,t}^\tau = \sum_{\sigma \in W_n, \quad |\tau| q_1 \leq |\sigma| \leq |\tau| q_2} \alpha_\sigma^\tau(y_s) X_{s,t}^\sigma,$$

(14)

for all $(s, t) \in \Delta_T$ and $\tau \in W_m$. $\alpha_\sigma^\tau : \mathbb{R}^m \to L(\mathbb{R}, \mathbb{R})$ are polynomials of degree $\leq q|\tau|$.

**Proof.** We will prove (14) by induction on $|\tau|$, i.e. the length of the word. By hypothesis, it is true when $|\tau| = 1$. Suppose that it is true for any $\tau \in W_m$ such that $|\tau| = k \geq 1$. First, note that from (13), we get that

$$dY_u^{(j)} = \sum_{\sigma \in W_n, \quad q_1 \leq |\sigma| \leq q_2} \alpha_\sigma^{(j)}(y_s) X_{s,u}^{\sigma-} dX_u^{\sigma_\ell}, \quad \forall u \in [s, t]$$

where $\sigma-$ is the word $\sigma$ without the last letter and $\sigma_\ell$ is the last letter. For example, if $\sigma = (i_1, \ldots, i_{b-1}, i_b)$, then $\sigma- = (i_1, \ldots, i_{b-1})$ and $\sigma_\ell = i_b$. Note that this cannot be defined when $\sigma$ is the empty word $\emptyset$ ($b = 0$). Now suppose that $|\tau| = k + 1$, so $\tau = (j_1, \ldots, j_k, j_{k+1})$ for some $j_1, \ldots, j_{k+1} \in \{1, \ldots, m\}$. Then

$$Y_{s,t}^\tau = \int_s^t Y_{s,u}^{\sigma-} dY_u^{(j_{k+1})} =$$

$$= \int_s^t \left( \sum_{k_1 \leq |\sigma_1| \leq k_2} \alpha_{\sigma_1}^{\tau-}(y_s) X_{s,u}^{\sigma_1} \right) \sum_{q_1 \leq |\sigma_2| \leq q_2} \alpha_{\sigma_2}^{(j_{k+1})}(y_s) X_{s,u}^{\sigma_2-} dX_u^{\sigma_2_\ell} =$$

$$= \sum_{k_1 \leq |\sigma_1| \leq k_2, q_1 \leq |\sigma_2| \leq q_2} \left( \alpha_{\sigma_1}^{\tau-}(y_s) \alpha_{\sigma_2}^{(j_{k+1})}(y_s) \right) \int_s^t X_{s,u}^{\sigma_1} X_{s,u}^{\sigma_2-} dX_u^{\sigma_2_\ell}. $$
Now we use the fact that for any geometric rough path $X$ and any $(s, u) \in \Delta_T$, we can write
\[
X_{s,u}^{\sigma_1} X_{s,u}^{\sigma_2} = \sum_{\sigma \in \sigma_1 \shuffle (\sigma_2^-)} X_{s,u}^\sigma
\]
(15)
where $\sigma_1 \shuffle (\sigma_2^-)$ is the shuffle product between the words $\sigma_1$ and $\sigma_2^-$. Applying (15) above, we get
\[
Y_{s,t}^{\tau} = \sum_{\sigma \in W_n, \ (k+1)q_1 \leq |\sigma| \leq (k+1)q_2} \alpha_\sigma^\tau(y_s) X_{s,t}^\sigma,
\]
where
\[
\alpha_\sigma^\tau(y_s) = \sum_{(\sigma_1 \shuffle (\sigma_2^-)) \ni \sigma^-} \alpha_{\sigma_1}^\tau(y_s) \alpha_{\sigma_2}^{\tau_\ell}(y_s)
\]
is a polynomial of degree $\leq kq + q = (k + 1)q$. Note that the above sum is over all $\sigma_1, \sigma_2 \in W_n$ such that $kq_1 \leq |\sigma_1| \leq kq_2$ and $q_1 \leq |\sigma_1| \leq q_2$.

We now prove (11) for $p = 1$.

**Lemma 3.3.** Suppose that $X \in \Omega_1(\mathbb{R}^n)$ is driving system (11), where $f : \mathbb{R}^m \to L(\mathbb{R}^n, \mathbb{R}^m)$ is a polynomial of degree $q$. Let $Y(r)$ be the projection of the $r$-th Picard iteration $Z(r)$ to $\mathbb{R}^m$, as described above. Then, $Y(r) \in \Omega_1(\mathbb{R}^m)$ and it satisfies
\[
Y_{s,t}^{\tau} = \sum_{|\sigma| \leq r} \alpha_{r,\sigma}^{\tau}(y_0, s) X_{s,t}^\sigma,
\]
(16)
for all $(s, t) \in \Delta_T$ and $\tau \in W_m$. $\alpha_{r,\sigma}^{\tau}(y_0, s)$ is a polynomial of degree $\leq |\tau|q^r$ in $y$.

**Proof.** For every $r \geq 0$, $Z(r) \in \Omega_1(\mathbb{R}^{n+m})$ since $Z(0) := (X, e)$, $X \in \Omega_1(\mathbb{R}^n)$ and integrals preserve the roughness of the integrator. So, $Y(r) \in \Omega_1(\mathbb{R}^m)$. We will prove the claim by induction on $r$.

For $r = 0$, $Y(0) = e$ and thus (16) becomes
\[
Y_{s,t}^{0} = \alpha_{0,0}^{0}(y_0, s)
\]
and it is true for $\alpha_{0,0}^{0} \equiv 1$ and $\alpha_{0,0}^{0} \equiv 0$ for every $\tau \in W_m$ such that $|\tau| > 0$.

Now suppose it is true for some $r \geq 0$. Remember that $Z(r) = (X, Y(r))$ and that $Z(r + 1)$ is defined by
\[
Z(r + 1) = \int h(Z(r)) dZ(r)
\]
where $h$ is defined in (2) and $f_{y_0}(y) = f(y_0 + y)$. Since $f$ is a polynomial of degree $q$, $h$ is also a polynomial of degree $q$ and, thus, it is possible to write
\[
h(z_2) = \sum_{k=0}^{q} h_k(z_1) \frac{(z_2 - z_1)^{\otimes k}}{k!}, \quad \forall z_1, z_2 \in \mathbb{R}^\ell,
\]
(17)
where $\ell = n + m$. Then, the integral is defined to be
\[
Z_{s,t}^{r+1} = \int_{s}^{t} h(Z(r)) dZ(r) = \sum_{k=0}^{q} h_k(Z(r)) Z_{s,t}^{k+1} \forall (s, t) \in \Delta_T.
\]
By definition, \( h_j \in \{ \text{contains any letters in} \quad \text{and by re-organizing the sums, we get} \quad \text{i} \quad \text{and thus, for every} \quad W \quad \text{where} \quad \alpha \quad \text{If} \quad q \quad \text{of all words of length between} \quad m \quad \text{for every} \quad \tau \quad \text{The result follow by applying lemma 3.2. Notice that (in the notation of lemma 3.2)} \quad R \quad \text{is the projection of} \quad Z(r + 1) \quad \text{onto} \quad \mathbb{R}^m. \quad \text{So, for each} \quad j \in \{1, \ldots, m\}, \quad Y(r + 1)_{s,t}^{(j)} = Z(r + 1)_{s,t}^{(n+j)} = \sum_{k=0}^{q} \left( h_k \left(Z(r)_{s}\right) Z(r)_{s,t}^{k+1} \right)^{(n+j)} \\
= \sum_{i=1}^{\ell} \sum_{\tau \in W_m(0,q)} \partial_{\tau+n} h_{n+i,j} \left(Z(r)_{s}\right) Z(r)_{s,t}^{(\tau+n,i)} \\
= \sum_{i=1}^{\ell} \sum_{\tau \in W_m(0,q)} \partial_{\tau} f_{j,i} \left(y_0 + Y(r)_{s}\right) Y(r)_{s,t}^{(\tau,i)}, \quad (18) \]
where \( W_m(k_1,k_2) = \{ \tau \in W_m \mid k_1 \leq |\tau| \leq k_2 \} \) for any \( k_1, k_2 \in \mathbb{N} \), i.e. it is the set of all words of length between \( k_1 \) and \( k_2 \). By the induction hypothesis, we know that for every \( \tau \in W_m \),
\[
Z(r)_{s,t}^{\tau+n} = Y(r)_{s,t}^{\tau} = \sum_{|\sigma| \leq |\tau| \frac{q^{n-1}}{q^r}} \alpha_{r,\sigma}^{\tau} (y_0, s) X_{s,t}^\sigma
\]
and thus, for every \( i = 1, \ldots, n \),
\[
Z(r)_{s,t}^{(\tau+n,i)} = \sum_{|\sigma| \leq |\tau| \frac{q^{n-1}}{q^r}} \alpha_{r,\sigma}^{\tau} (y_0, s) X_{s,t}^{(\sigma,i)}. \quad (19) \]
Putting this back to the equation above, we get
\[
Y(r + 1)_{s,t}^{(j)} = \sum_{i=1}^{n} \sum_{|\tau| \leq q} \partial_{\tau} f_{j,i} \left(y_0 + Y(r)_{s}\right) \sum_{|\sigma| \leq |\tau| \frac{q^{n-1}}{q^r}} \alpha_{r,\sigma}^{\tau} (y_0, s) X_{s,t}^{(\sigma,i)}
\]
and by re-organizing the sums, we get
\[
Y(r + 1)_{s,t}^{(j)} = \sum_{|\sigma| \leq q^{n+1} \frac{q^{n-1}}{q^r} + 1 = n^r + 1 \frac{q^{n-1}}{q^r}} \alpha_{r+1,\sigma}^{(j)} (y_0, s) X_{s,t}^{\sigma}, \quad (20) \]
where \( \alpha_{r+1,\emptyset}^{(j)} \equiv 0 \) and for every \( \sigma \in W_n - \emptyset \),
\[
\alpha_{r+1,\sigma}^{(j)} (y_0, s) = \sum_{\sum_{q^r | (\sigma-1) \leq |\tau| \leq q}} \partial_{\tau} f_{j,\sigma} \left(y_0 + Y(r)_{s}\right) \alpha_{r,\sigma-}^{\tau} (y_0, s).
\]
If \( \alpha_{r,\sigma}^{\tau} \) are polynomials of degree \( \leq |\tau| q^r \), then \( \alpha_{r,\sigma}^{(j)} \) are polynomials of degree \( \leq q^r \). The result follow by applying lemma 3.2. Notice that (in the notation of lemma 3.2) \( q_1 \geq 1 \) since \( \alpha_{r,\sigma}^{(j)} \equiv 0. \)
We will now prove (11) for any \( p \geq 1 \).

**Theorem 3.4.** The result of lemma 3.3 still holds when \( X \in G\Omega_p(\mathbb{R}^n) \), for any \( p \geq 1 \).

**Proof.** Since \( X \in G\Omega_p(\mathbb{R}^n) \), there exists a sequence \( \{X(k)\}_{k \geq 0} \) in \( G\Omega_1(\mathbb{R}^n) \), such that \( X(k) \xrightarrow{k \to \infty} X \) in the \( p \)-variation topology. We denote by \( Z(k, r) \) and \( Z(r) \) the \( r \)-th Picard iteration corresponding to equation (1) driven by \( X(k) \) and \( X \) respectively.

First, we show that \( Z(k, r) \xrightarrow{k \to \infty} Z(r) \) and consequently \( Y(k, r) \xrightarrow{k \to \infty} Y(r) \) in the \( p \)-variation topology, for every \( r \geq 0 \). It is clearly true for \( r = 0 \). Now suppose that it is true for some \( r \geq 0 \). By definition, \( Z(r + 1) = \int h(Z(r))dZ(r) \). Remember that the integral is defined as the limit in the \( p \)-variation topology of the integrals corresponding to a sequence of 1-rough paths that converge to \( Z(r) \) in the \( p \)-variation topology. By the induction hypothesis, this sequence can be \( Z(k, r) \). It follows that \( Z(k, r + 1) = \int h(Z(k, r))dZ(k, r) \) converges to \( Z(r + 1) \), which proves the claim. Convergence of the rough paths in \( p \)-variation topology implies convergence of each of the iterated integrals, i.e.

\[
Y(k, r)_{s,t}^\tau \xrightarrow{k \to \infty} Y(r)_{s,t}^\tau
\]

for all \( r \geq 0 \), \((s, t) \in \Delta_T \) and \( \tau \in W_m \).

By lemma 3.3 since \( X(k) \in G\Omega_1(\mathbb{R}^n) \) for every \( k \geq 1 \), we can write

\[
Y(k, r)_{s,t}^\tau = \sum_{|\sigma| \leq |\tau|} \alpha_{r, \sigma}^{\tau}(y_0, s)X(k)^\sigma_{s,t},
\]

for every \( \tau \in W_m \), \((s, t) \in \Delta_T \) and \( k \geq 1 \). Since \( X(k) \xrightarrow{k \to \infty} X \) in the \( p \)-variation topology and the sum is finite, it follows that

\[
Y(k, r)_{s,t}^\tau \xrightarrow{k \to \infty} \sum_{|\sigma| \leq |\tau|} \alpha_{r, \sigma}^{\tau}(y_0, s)X^\sigma_{s,t}.
\]

The statement of the theorem follows. \( \square \)

### 3.2 The Expected Signature Matching Estimator

We can now give a precise definition of the estimator, which we will formally call the **Expected Signature Matching Estimator** (ESME): suppose that we are in the setting of the problem described in section 2.2 and \( M_N^r \) and \( E_r^\theta \) are defined as in (9) and (10) respectively, for every \( \tau \in W_m \). Let \( V \subset W_m \) be a set of \( d \) words constructed from the alphabet \( \{1, \ldots, m\} \). For each such \( V \), we define the ESME \( \hat{\theta}_r^V \) as the solution to

\[
E_r^\theta = M_N^r, \forall \tau \in V.
\]

This definition requires that (21) has a unique solution. This will not be true in general. Let \( \mathcal{V}_r \) be the set of all \( V \) containing \( d \) words, such that \( E_r^\theta(\theta) = M, \forall \tau \in V \) has a unique solution for all \( M \in S_\tau \subseteq \mathbb{R} \) where \( S_\tau \) is the set of all possible values of \( M_N^r \), for any \( N \geq 1 \). We will assume the following:
Assumption 1 (Observability). The set $V_r$ is non-empty and known (at least up to a non-empty subset).

Then, $\hat{\theta}_{r,N}^{V}$ can be defined for every $V \in V_r$.

Remark 3.5. In order to achieve uniqueness of the estimator, we might need some extra information that we could get by looking at time correlations. We can fit this into our framework by considering scaled versions of (5) together with the original one: for example consider the equation

$$
\frac{dY_t(\omega)}{dY(c)_t(\omega)} = f(Y_t(\omega); \theta) \cdot dX_t(\omega), Y_0 = y_0
$$

for some appropriate constant $c$. Then, $Y(c)_t = Y_{ct}$ and the expected signature at $[0, T]$ will also contain information about $E\left(Y_{T}^{(j_1)}Y_{cT}^{(j_2)}\right)$ for any $j_1, j_2 = 1, \ldots, m$.

It is very difficult to say anything about the solutions of system (21), as it is very general. However, if we assume that $f$ is also a polynomial in $\theta$, then (21) becomes a system of polynomial equations.

Note that one can also create a Generalized Expected Signature Matching Estimator as the solution of

$$
P_{\alpha}(E_{\tau}^\tau(\theta)) = P_{\alpha}(M_{N}^\tau), \text{ for } \alpha \in A
$$

where $P_{\alpha}$ are polynomials of (empirical or theoretical) expected values of iterated integrals corresponding to words $\tau$ and $A$ an appropriate index set.

Remark 3.6. In the case where $y_t$ is a Markov process, the Generalized Moment Matching Estimator can be seen as a special case of the Generalized Expected Signature Matching Estimator. In that case, the question of identifiability has been studied in detail (see [3]), but without considering the extra approximation of the theoretical moments by Picard iteration.

3.3 Properties of the ESME

It is possible to show that the ESME defined as the solution of (21) will converge to the true value of the parameter and will be asymptotically normal. More precisely, the following holds:

Theorem 3.7. Let $\hat{\theta}_{r,N}^{V}$ be the Expected Signature Matching Estimator for the system described in section 2.2 and $V \in V_r$. Assume that the expected signature of $Y_{0,T}$ is finite and that $f(y; \theta)$ is a polynomial of degree $q$ with respect to $y$ and twice differentiable with respect to $\theta$. Let $\theta_0$ be the ‘true’ parameter value, meaning that the distribution of the observed signature $Y_{0,T}$ is $Q_{\theta_0}$, defined in (7). Set

$$
D^V_{r}(\theta)_{i,\tau} = \frac{\partial}{\partial \theta_i} E^\tau_r(\theta) \quad \text{and} \quad \Sigma_V(\theta_0)_{\tau,\tau'} = \text{cov}\left(Y^\tau_{0,T}, Y^\tau'_{0,T}\right)
$$

(22)
and assume that 
\[ \inf_{r>0, \theta \in \Theta} \| D^V_r(\theta) \| > 0, \] i.e. \( D^V_r(\theta) \) is uniformly non-degenerate with respect to \( r \) and \( \theta \). Then, for \( r \propto \log N \) and \( \mathcal{T} \) are sufficiently small,
\[
\hat{\theta}^V_{r,N} \to \theta_0, \quad \text{with probability 1}, \tag{23}
\]
and
\[
\sqrt{N} \Phi_V(\theta_0)^{-1} \left( \hat{\theta}^V_{r,N} - \theta_0 \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I), \tag{24}
\]
as \( N \to \infty \), where
\[
\Phi_V(\theta_0) = D^V_r(\theta_0)^{-1} \Sigma_V(\theta_0)^{1/2}. \tag{25}
\]

**Proof.** By theorem 3.4 and the definition of \( E^\tau_r(\theta) \),
\[
E^\tau_r(\theta) = \sum_{|\sigma| \leq |\tau|} \alpha^\tau_{r,\sigma}(y_0; \theta) \mathbb{E}(X^\sigma_{0,T})
\]
where functions \( \alpha^\tau_{r,\sigma}(y_0; \theta) \) are constructed recursively, as in lemmas 3.2 and 3.3. Since \( f \) is twice differentiable with respect to \( \theta \), functions \( \alpha \) and consequently \( E^\tau_r \) will also be twice differentiable with respect to \( \theta \). Thus, we can write
\[
E^\tau_r(\theta_0) - E^\tau_r(\theta_0) = D^V_r(\hat{\theta}^V_{r,N})^{-1} \left( E^V_r(\hat{\theta}^V_{r,N}) - E^V_r(\theta_0) \right)
\]
where \( E^V_r(\theta) = \{ E^\tau_r(\theta) \} \tau \in \mathcal{V} \). By definition
\[
E^V_r(\hat{\theta}^V_{r,N}) = \{ M^\tau_N \} \tau \in \mathcal{V} = \left\{ \frac{1}{N} \sum_{i=1}^{N} Y^\tau_{0,T}(\omega_i) \right\} \tau \in \mathcal{V}
\]
where \( Y^\tau_{0,T}(\omega_i) \) are independent realizations of the random variable \( Y_{0,T} \). Suppose that \( T \) is small enough, so that the above Monte-Carlo approximation satisfies both the Law of Large Numbers and the Central Limit Theorem, i.e. the covariance matrix satisfies \( 0 < \| \Sigma_V(\theta_0) \| < \infty \). Then, for \( N \to \infty \)
\[
| E^\tau_r(\hat{\theta}^V_{r,N}) - E^\tau(\theta_0) | = | E^\tau_r(\hat{\theta}^V_{r,N}) - \mathbb{E}(Y^\tau_{0,T}) | \to 0, \quad \forall \tau \in \mathcal{V}
\]
with probability 1. Note that the convergence does not depend on \( r \). Also, for \( r \to \infty \)
\[
E^\tau_r(\theta_0) \to E^\tau(\theta_0)
\]
as a result of theorem 2.6. Thus, for \( r \propto \log N \)
\[
| E^\tau_r(\hat{\theta}^V_{r,N}) - E^\tau_r(\theta_0) | \to 0, \quad \text{with probability 1, } \forall \tau \in \mathcal{V}.
\]
Combining this with (26) and the uniform non-degeneracy of $D^V$, we get (23). From (23) and the continuity and uniform non-degeneracy of $D^V$, we conclude that

$$D^V_r(\theta_0)D^V_r(\hat{\theta}^V_{r,N})^{-1} \to I,$$

with probability 1 provided that $T$ is small enough, so that $E^V_r(\theta_0) < \infty$. Now, since

$$\Phi^V_r(\theta_0)^{-1} (\hat{\theta}^V_{r,N} - \theta_0) = \Sigma^V_r(\theta_0)^{-1/2} \left( D^V_r(\theta_0)D^V_r(\hat{\theta}^V_{r,N})^{-1} \right) \left( E^V_r(\hat{\theta}^V_{r,N}) - E^V_r(\theta_0) \right)$$

to prove (24) it is sufficient to prove that

$$\sqrt{N} \Sigma^V_r(\theta_0)^{-1/2} \left( E^V_r(\hat{\theta}^V_{r,N}) - E^V_r(\theta_0) \right) \overset{L}{\to} \mathcal{N}(0, I)$$

It follows directly from (27) that

$$\sqrt{N} \Sigma^V_r(\theta_0)^{-1/2} \left( E^V_r(\hat{\theta}^V_{r,N}) - E^V_r(\theta_0) \right) \overset{L}{\to} \mathcal{N}(0, I).$$

It remains to show that

$$\sqrt{N} \Sigma^V_r(\theta_0)^{-1/2} \left( E^V_r(\theta_0) - E^V_r(\theta_0) \right) \to 0.$$ 

It follows from theorem 2.6 that

$$\|E^V_r(\theta_0) - E^V_r(\theta_0)\| \leq C \rho^{-r}$$

for any $\rho > 1$ and sufficiently small $T$. The constant $C$ depends on $V, p$ and $T$. Suppose that $r = a \log N$ for some $a > 0$ and choose $\rho > \exp(\frac{1}{2e})$. Then

$$\sqrt{N} \|E^V_r(\theta_0) - E^V_r(\theta_0)\| \leq C N^{(\frac{1}{2} - c \log \rho)}$$

which proves the claim.

**Remark 3.8.** We have now completed the discussion of the questions set in section 3: we provided a way for getting an analytic expression for an approximation of $E^r(\theta)$. Also, the asymptotic variance of the estimator can be used to compare different choices of $V$ and to assess the quality of the estimator.

In the case of diffusions and the GMM estimator, a discussion on how to optimally choose which moments to work on can be found in [4].

### 4 Examples

In this section, we use the ESME in a specific example of a diffusion and a fractional diffusion.
4.1 Diffusions

First, we apply the ESME to estimate the parameters of the following Stratonovich SDE:

\[ dY_t = a(1 - Y_t)dX^{(1)}_t + bY_t^2 dX^{(2)}_t, \quad Y^{(1)}_0 = 0, \]

where \(X^{(1)}_t = t\) and \(X^{(2)}_t = W_t\). We chose an SDE because the expected signature of \((t, W_t)\) can easily be computed explicitly.

After three Picard iterations and replacing the expected signature of \((t, W_t)\) by its value (see [12]), we get

\[ \mathbb{E}(Y^{(3)}_{0,t}) = at - \frac{a^2}{2}t^2 + \frac{a^3}{6}t^3 - \frac{a^3b^2}{4}t^4 - \frac{a^4b^2}{10}t^5 \]

\[ \mathbb{E}(2Y^{(3)}_{0,t}) = a^2t^2 - a^3t^3 + \frac{7a^4}{12}t^4 - \left(\frac{a^5}{6} - \frac{7a^4b^2}{10}\right)t^5 + \left(\frac{a^6}{36} - \frac{17a^5b^2}{20}\right)t^6 + \frac{191a^6b^2}{420}t^7 \]

\[ + \left(-\frac{11a^7b^2}{105} + \frac{21a^6b^4}{80}\right)t^8 + \left(\frac{a^8b^2}{144} - \frac{43a^7b^4}{180}\right)t^9 + \frac{33a^8b^4}{700}t^{10} + \frac{a^8b^6}{50}t^{11}. \]

This gives us an approximation of the moments of the solution as polynomials of the parameters.

The empirical moments are computed from the data. We generate 2000 approximate realizations of paths of the solution using Milstein’s method with discretization step 0.001. We use these paths to approximate the iterated integrals over the interval \([0, \frac{1}{4}]\). We use the values \(a = 1\) and \(b = 2\). Then, we get an approximation to the empirical moments at \(T = \frac{1}{4}\) by averaging the different realizations of the iterated integrals of \(Y_{[0,\frac{1}{4}]}\).

Finally, by equating the empirical and theoretical approximations to the moments for \(t = \frac{1}{4}\), we get a system of polynomials of \((a, b)\) of degree 14. We get two exact real solutions to this sytem: (.996353, −2.12892) and (.996353, 2.12892). As expected, the sign of \(b\) cannot be identified. Apart from that, the estimates are very close to the true values.

We repeat this process 100 times and get 100 different estimates of \((a, b)\). In figure 4.1 we show the positive solutions. We also check asymptotic normality: we normalize the 100 realizations by the asymptotic variance (25), where \(V_i(\theta)_{i,\tau}\) and \(\Sigma_V(\theta_0)_{\tau,\tau'}\) in (22) are computed, the first using approximation of the theoretical moments from Picard iterations and the second is computed from the data by Monte Carlo. The normalized estimates are shown in figure 4.1. Their covariance matrix is

\[
\begin{pmatrix}
0.97172 & 0.0243445 \\
0.0243445 & 0.954654
\end{pmatrix}
\]

which is very close to the identity.

4.2 Fractional Diffusions

We now apply the ESME to estimate the parameters of the differential equation driven by fractional diffusion with Hurst parameter \(h > 1/4\). We choose the same
Figure 1: 100 realizations of the expected signature matching estimator. True parameters are $a = 1$ and $b = 2$.

Figure 2: 100 realizations of the expected signature matching estimator, after centering and normalizing by the asymptotic variance.
vector field as before. Let

\[ dY_t = a(1 - Y_t) dX_t^{(1)} + bY_t^2 dX_t^{(2)}, \quad Y_0^{(1)} = 0, \] (29)

where \( X_t^{(1)} = t \) and \( X_t^{(2)} = B_t^h \), where \( B_t^h \) is fractional Brownian motion with Hurst parameter \( h \). Fractional Brownian motion generalizes Brownian motion, in the sense that it is a self-similar Gaussian process. It is defined as the Gaussian process with correlation given by

\[ E(B_s^h B_t^h) = \frac{1}{2} \left( |s|^{2h} + |t|^{2h} - |t-s|^{2h} \right). \]

Clearly, for \( h = \frac{1}{2} \) we get independent intervals and Brownian motion. For \( h > \frac{1}{2} \) the intervals are positively correlated and “smoother” than Brownian motion while for \( h < \frac{1}{2} \) they are negatively correlated and they get more and more “rough” as \( h \) gets smaller. In particular, the paths of fractional Brownian motion possess finite \( p \)-variation for every \( p > \frac{1}{h} \).

Defining integration with respect to fractional Brownian motion is necessary in order for (29) to make sense. This is non-trivial and it is a very active area of research. One of the most successful approaches is given by rough paths - but it is limited to \( h > \frac{1}{4} \) (see [17] or [25] for a more recent approach), i.e. to paths of finite \( p \)-variation for \( p < 4 \).

Having defined (29) as a differential equation driven by the rough path \((t, B_t^h)\), we can proceed to estimate the parameters \( a \) and \( b \). As in the diffusion case, we first construct an approximation to the theoretical moments, using Picard iterations. One difference is that up to this moment, an analytic expression for the expected signature is not known. Instead, we get a numerical approximation by simulating many paths of fractional Brownian motion, computing their iterated integral and then averaging.

We need to set some parameters: we choose \( T = \frac{1}{4} \) as before and \( h = \frac{11}{24} \). The Hurst parameter \( h \) is chosen so that the paths are more rough than Brownian motion but not too much – we will see later that the smaller the \( h \), the smaller the discretization step needs to be in order for the simulation of the paths to be good. We use 1000 paths of fractional Brownian motion with Hurst parameter \( h = \frac{11}{24} \) – these are exact simulations with discretization step \( 10^{-3} \) – to compute the iterated integrals appearing in the Picard iteration and then average to approximate their expectations. We get the following formulas for the theoretical approximation of the first two moments of the response \( Y \):
Figure 3: 100 realizations of the expected signature matching estimator. True parameters are $a = 1$ and $b = 2$.

\[
\mathbb{E}(Y(3)_{0,\frac{1}{4}}^{(1)}) = 0.25a - 0.03125a^2 + 0.00260417a^3 + 0.00044726a^2b - 0.000111815a^3b + 4.97138 \times 10^{-6}a^4b + 0.00116494a^3b^2 - 0.000115953a^4b^2 + 2.53676 \times 10^{-6}a^4b^3
\]

\[
\mathbb{E}(2Y(3)_{0,\frac{1}{4}}^{(1,1)}) = 0.0625a^2 - 0.015625a^3 + 0.00227865a^4 - 0.00016276a^5 + 6.78168 \times 10^{-6}a^6 + 0.00022363a^3b - 0.0000838612a^4b + 0.0000118036a^5b - 8.93081 \times 10^{-7}a^6b
\]

We create the data by numerically simulating 2000 paths of the solution of (29) for $h = \frac{11}{24}$, $a = 1$ and $b = 2$ and discretization step $\delta = 10^{-3}$. We use a method proposed by Davie that is the equivalent of Milstein’s method for differential equations driven by fractional Brownian motion (see [27] and references within). The error is of order $\delta^{3h-1}$, which for our choices of discretization step $\delta$ and Hurst parameter $h$ is 0.075.

Finally, we match the theoretical moments that are polynomials of $(a, b)$ with the empirical moments and solve the system. As in the diffusion case, we get two solutions corresponding to $b$ positive or negative. Since fractional Brownian motion is mean zero Gaussian process, we cannot expect to identify the sign of $b$.

We repeat the process 100 times to get 100 realizations of the estimates. These are shown in figure 3. Also, figure 4 shows the estimates after centering and normalizing by the asymptotic variance.
Figure 4: 100 realizations of the expected signature matching estimator, after centering and normalizing by the asymptotic variance.

References

[1] J.P.N. Bishwal. Parameter Estimation in Stochastic Differential Equations. Lecture Notes in Mathematics vol. 1923, Springer, Berlin, 2008.

[2] C.P. Calderon. Fitting Effective Diffusion Models to Data Associated with a “Glassy Potential”: Estimation, Classical Inference Procedures, and Some Heuristics. Multiscale Model. Simul. 6(2): 656–687, 2007.

[3] L. P. Hansen and J. A. Scheinkman. Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes. Econometrica 63(4):767-804, 1995.

[4] Y. Ait-Sahalia and P. A. Mykland. An Analysis of Hansen-Scheinkman Moment Estimators for Discretely and Randomly Sampled Diffusions. Journal of Econometrics 144:1–26, 2008.

[5] A. Beskos, O. Papaspiliopoulos and G. Roberts. Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. Ann. Statist. 37(1): 223–245, 2009.

[6] P.K. Friz and N. Victoir. Multidimensional Stochastic Processes as Rough Paths. Cambridge University Press, 2010.
[7] E. Gobet, M. Hoffmann and M. Reiss. Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Stat.* 32(5): 2223–2253, 2004.

[8] L.P. Hansen and J.A. Scheinkman. Back to the future: generating moment implications for continuous-time Markov processes. *Econometrica* 63(4): 767–804, 1995.

[9] Y. Hu and D. Nualart. Parameter estimation for fractional Ornstein-Uhlenbeck processes. *Preprint*: arXiv:0901.4925v1.

[10] H. Hult. Approximating some Volterra type stochastic integrals with applications to parameter estimation. *Stoch. Process. Appl.* 105: 1–32, 2003.

[11] J. Li, P.G. Kevrekidis, C.W. Gear and I.G. Kevrekidis. Deciding the nature of the coarse equation through microscopic simulations: the baby-bathwater scheme. *SIAM Review* 49(3): pp. 469–487, 2007.

[12] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Application of Mathematics 23. Springer-Verlag, Berlin 1999.

[13] U. Küchler and M. Sørensen. *Exponential families of stochastic processes*. Springer Series in Statistics, Springer-Verlag, New York, 1997.

[14] Y.A. Kutoyants. *Statistical inference for ergodic diffusion processes*. Springer-Verlag, London, 2004.

[15] W. E, D. Liu and E. Vanden-Eijnden. Analysis of multiscale methods for stochastic differential equations. *Comm. Pure Appl. Math.* 58(11): 1544-1585, 2005.

[16] T.J. Lyons, M.J. Caruna and T. Lévy. *Differential Equations Driven by Rough Paths*. In *Ecole d’été de Probabilités de Saint-Flour XXXIV* (Editor: J. Picard), Springer, Berlin 2007.

[17] T. Lyons and Z. Qian. *System control and rough paths*. Oxford University Press, Oxford, 2002.

[18] A. Neuenkirch, I. Nourdin and S. Tindel. Delay equations driven by rough paths. *Electronic Journal of Probability* 13, 2031–2068, 2008.

[19] A. Papavasiliou. Coarse-grained modeling of multiscale diffusions: the p-variation estimates. arXiv:1002.3241

[20] B.L.S. Prakasa Rao. *Statistical Inference for Diffusion Type Processes*. In *Kendall’s Library of Statistics* 8, Oxford University Press, New York, 1999.

[21] M. Reiss. Adaptive estimation for affine stochastic delay equations. *Bernoulli* 11(1): 67–102, 2005.

[22] C. A. Tudor and F. G. Viens. Statistical aspects of the fractional stochastic calculus. *Ann. Statist.* 35(3): 1183–1212, 2007.
[23] M. Hairer and N.S. Pillai. Ergodicity of hypoelliptic SDEs driven by fractional Brownian motion. arXiv:0909.4505.

[24] W. Willinger, M. S. Taqqu and V. Teverovsky. Stock market prices and long-range dependence. Finance and Stochastics 3(1): 1432–1122, 1999.

[25] J. Unterberger. Stochastic Calculus for fractional Brownian motion with Hurst exponent $H > \frac{1}{4}$: a rough path method by analytic extension. Ann. Prob. 37(2): 565–614, 2009.

[26] A. Tocino. Multiple stochastic integrals with Mathematica. Mathematics and Computers in Simulation 79(5):1658–1667, 2009.

[27] A. Deya, A. Neuenkirch and S. Tindel. A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. arXiv:1001.3344.

[28] M. Rosenbaum. Integrated volatility and round-off error. Bernoulli 15(3): 687–720, 2009.

[29] L. Zhang, P.A. Mykland and Y. At-Sahalia. A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc. 100(472) 1394–1411, 2005.

[30] P. Cheredito. Regularizing Fractional Brownian Motion with a View towards Stock Price Modelling, PhD thesis, ETH Zurich, 2001.