Pathwise Approximations for the Solution of the Non-Linear Filtering Problem

Dan Crisan, Alexander Lobbe, and Salvador Ortiz-Latorre

Abstract We consider high order approximations of the solution of the stochastic filtering problem, derive their pathwise representation in the spirit of the earlier work of Clark [2] and Davis [10, 11] and prove their robustness property. In particular, we show that the high order discretised filtering functionals can be represented by Lipschitz continuous functions defined on the observation path space. This property is important from the practical point of view as it is in fact the pathwise version of the filtering functional that is sought in numerical applications. Moreover, the pathwise viewpoint will be a stepping stone into the rigorous development of machine learning methods for the filtering problem. This work is a cotinuation of [5] where a discretisation of the solution of the filtering problem of arbitrary order has been established. We expand the work in [5] by showing that robust approximations can be derived from the discretisations therein.

1 Introduction

With the present article on non-linear filtering we wish to honor the work of Mark H. A. Davis in particular to commemorate our great colleague. The topic of filtering is an area that has seen many excellent contributions by Mark. It is remarkable that

Dan Crisan
Department of Mathematics, Imperial College London, Huxley's Building, 180 Queen's Gate, London SW7 2AZ, UK,
e-mail: dcrisan@imperial.ac.uk

Alexander Lobbe
Department of Mathematics, University of Oslo, P.O. Box 1053, Blindern, 0316 Oslo, Norway,
e-mail: alexalob@math.uio.no

Salvador Ortiz-Latorre
Department of Mathematics, University of Oslo, P.O. Box 1053, Blindern, 0316 Oslo, Norway,
e-mail: salvadoo@math.uio.no

G. Yin, T. Zariphopoulou (eds.), *Stochastic Analysis, Filtering, and Stochastic Optimization*,
https://doi.org/10.1007/978-3-030-98519-6_4

he was able to advance the understanding of non-linear filtering from a variety of angles. He considered many aspects of the field in his work, spanning the full range from the theory of the filtering equations to the numerical solution of the filtering problem via Monte-Carlo methods.

Mark Davis' work on filtering can be traced back to his doctoral thesis where he treats stochastic control of partially observable processes. The first article specifically on the topic of filtering that was co-authored by Mark appeared back in 1975 and considered a filtering problem with discontinuous observation process [12]. There, they used the so-called innovations method to compute the evolution of the conditional density of a process that is used to modulate the rate of a counting process. This method is nowadays well-known and is a standard way also to compute the linear (Kalman) filter explicitly. Early on in his career, Mark also contributed to the dissemination of filtering in the mathematics community with his monograph *Linear Estimation and Stochastic Control* [7], published in 1977, which deals with filtering to a significant degree. Moreover, his paper *An Introduction to Nonlinear Filtering* [9], written together with S. I. Marcus in 1981, has gained the status of a standard reference in the field.

Importantly, and in connection to the theme of the present paper, Mark has worked on computation and the robust filter already in 1980 [8]. Directly after the conception of the robust filter by Clark in 1978 [2], Mark took up the role of a driving figure in the subsequent development of robust, also known as pathwise, filtering theory [10, 11]. Here, he was instrumental in the development of the pathwise solution to the filtering equations with one-dimensional observation processes. Additionally, also correlated noise was already analysed in this work.

Robust filtering remains a highly relevant and challenging problem today. Some more recent work on this topic includes the article [6] which can be seen as an extension of the work by Mark, where correlated noise and a multidimensional observation process are considered. The work [4] is also worth mentioning in this context, as it establishes the validity of the robust filter rigorously.

Non-linear filtering is an important area within stochastic analysis and has numerous applications in a variety of different fields. For example, numerical weather prediction requires the solution of a high dimensional, non-linear filtering problem. Therefore, accurate and fast numerical algorithms for the approximate solution of the filtering problem are essential. In this contribution we analyse a recently developed high order time discretisation of the solution of the filtering problem from the literature [5] and prove that the so discretised solution possesses a property known as *robustness*. Thus, the present paper is a continuation of the previous work [5] by two of the authors which gives a new high-order time discretisation for the filtering functional. We extend this result to produce the robust version, of any order, of the discretisation from [5]. The implementation of the resulting numerical method remains open and is subject of future research. In subsequent work, the authors plan to deal with suitable extensions, notably a machine learning approach to pathwise filtering.

Robustness is a property that is especially important for the numerical approximation of the filtering problem in continuous time, since numerical observations can only be made in a discrete way. Here, the robustness property ensures that despite the discrete approximation, the solution obtained from it will still be a reasonable approximation of the true, continuous filter.

The present paper is organised as follows: In Section 2 we discuss the established theory leading up to the contribution of this paper. We introduce the stochastic filtering problem in sufficient generality in Subsection 2.1 whereafter the high order discretisation from the recent paper [5] is presented in Subsection 2.2 together with all the necessary notations. The Subsection 2.2 is concluded with the Theorem 1, taken from [5], which shows the validity of the high order discretisation and is the starting point for our contribution. Then, Section 3 serves to concisely present the main result of this work, which is Theorem 2 below. Our Theorem is a general result applying to corresponding discretisations of arbitrary order and shows that all of these discretisations do indeed assume a robust version. In Section 4 we present the proof of the main result in detail. The argument proceeds along the following lines. First, we establish the robust version of the discretisations for any order by means of a *formal* application of the integration by parts formula. In Lemma 1 we then show that the new robust approximation is locally bounded over the set of observation paths. Thereafter, Lemma 2 shows that the robustly discretised filtering functionals are locally Lipschitz continuous over the set of observation paths. Based on the elementary but important auxilliary Lemma 3 we use the path properties of the typical observation in Lemma 4 to get a version of the stochastic integral appearing in the robust approximation which is product measurable on the Borel sigma-algebra of the path space and the chosen filtration. Finally, after simplifying the arguments by lifting some of the random variables to an auxilliary copy of the probability space, we can show in Lemma 5 that, up to a null-set, the lifted stochastic integral appearing in the robust approximation is a random variable on the correct space. And subsequently, in Lemma 6 that the pathwise integral almost surely coincides with the standard stochastic integral of the observation process. The argument is concluded with Theorem 3 where we show that the robustly discretised filtering functional is a version of the high-order discretisation of the filtering functional as derived in the recent paper [5].

Our result in Theorem 2 can be interpreted as a remedy for some of the shortcomings of the earlier work [5] where the discretisation of the filter is viewed as a random variable and the dependence on the observation path is not made explicit. Here, we are correcting this in the sense that we give an interpretation of said random variable as a continuous function on path space. Our approach has two main advantages. Firstly, from a practitioner's point of view, it is exactly the path dependent version of the discretised solution that we are computing in numerical applications. Thus it is natural to consider it explicitly. The second advantage lies in the fact that here we are building a foundation for the theoretical development of machine learning approaches to the filtering problem which rely on the simulation of observation paths. With Theorem 2 we offer a first theoretical justification for this approach.

2 Preliminaries

Here, we begin by introducing the theory leading up to the main part of the paper which is presented in Sections 3 and 4.

2.1 The filtering problem

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a complete and right-continuous filtration $(\mathcal{F}_t)_{t\geq 0}$. We consider a $d_X \times d_Y$-dimensional partially observed system (X, Y) satisfying the system of stochastic integral equations

$$\begin{cases} X_t = X_0 + \int_0^t f(X_s)\,ds + \int_0^t \sigma(X_s)\,dV_s, \\ Y_t = \int_0^t h(X_s)\,ds + W_t, \end{cases} \tag{1}$$

where V and W are independent $(\mathcal{F}_t)_{t\geq 0}$-adapted d_V- and d_Y-dimensional standard Brownian motions, respectively. Further, X_0 is a random variable, independent of V and W, with distribution denoted by π_0. We assume that the coefficients

$$f = (f_i)_{i=1,\ldots,d_X} : \mathrm{R}^{d_X} \to \mathrm{R}^{d_X} \text{ and } \sigma = \left(\sigma_{i,j}\right)_{i=1,\ldots,d_X, j=1,\ldots,d_V} : \mathrm{R}^{d_X} \to \mathrm{R}^{d_X \times d_V}$$

of the *signal process* X are globally Lipschitz continuous and that the *sensor function*

$$h = (h_i)_{i=1,\ldots,d_Y} : \mathrm{R}^{d_X} \to \mathrm{R}^{d_Y}$$

is Borel-measurable and has linear growth. These conditions ensure that strong solutions to the system (1) exist and are almost surely unique. A central object in filtering theory is the *observation filtration* $\{\mathcal{Y}_t\}_{t\geq 0}$ that is defined as the augmentation of the filtration generated by the *observation process* Y, so that $\mathcal{Y}_t = \sigma(Y_s, s \in [0,t]) \vee \mathcal{N}$, where $\mathcal{N}$ are all P-null sets of $\mathcal{F}$.

In this context, non-linear filtering means that we are interested in determining, for all $t > 0$, the conditional law, called *filter* and denoted by π_t, of the signal X at time t given the information accumulated from observing Y on the interval $[0,t]$. Furthermore, this is equivalent to knowing for every bounded and Borel measurable function φ and every $t > 0$, the value of

$$\pi_t(\varphi) = \mathrm{E}[\varphi(X_t) \mid \mathcal{Y}_t].$$

A common approach to the non-linear filtering problem introduced above is via a change of probability measure. This approach is explained in detail in the monograph [1]. In summary, a probability measure $\tilde{P}$ is constructed that is absolutely continuous with respect to P and such that Y becomes a $\tilde{P}$-Brownian motion independent of X. Additionally, the law of X remains unchanged under $\tilde{P}$. The

Radon-Nikodym derivative of $\tilde{P}$ with respect to P is further given by the process Z that is given, for all $t \geq 0$, by

$$Z_t = \exp\left(\sum_{i=1}^{d_Y} \int_0^t h_i(X_s)\, dY_s^i - \frac{1}{2}\sum_{i=1}^{d_Y} \int_0^t h_i^2(X_s)\, ds\right).$$

Note that Z is an $(\mathcal{F}_t)_{t\geq 0}$-adapted martingale under $\tilde{P}$. This process is used in the definition of another, measure-valued process ρ that is given, for all bounded and Borel measurable functions φ and all $t \geq 0$, by

$$\rho_t(\varphi) = \tilde{\mathrm{E}}[\varphi(X_t)Z_t \mid \mathcal{Y}_t], \tag{2}$$

where we denote by $\tilde{\mathrm{E}}$ the expectation with respect to $\tilde{P}$. We call ρ the *unnormalised filter*, because it is related to the probability measure-valued process π through the Kallianpur-Striebel formula establishing that for all bounded Borel measurable functions φ and all $t \geq 0$ we have P-almost surely that

$$\pi_t(\varphi) = \frac{\rho_t(\varphi)}{\rho_t(\mathbf{1})} = \frac{\tilde{\mathrm{E}}\left[\varphi(X_t)Z_t|\mathcal{Y}_t\right]}{\tilde{\mathrm{E}}\left[Z_t|\mathcal{Y}_t\right]} \tag{3}$$

where $\mathbf{1}$ is the constant function. Hence, the denominator $\rho_t(\mathbf{1})$ can be viewed as the normalising factor for π_t.

2.2 High order time discretisation of the filter

As shown by the Kallianpur-Striebel formula (3), $\pi_t(\varphi)$ is a ratio of two conditional expectations. In the recent paper [5] a high order time discretisation of these conditional expectations was introduced which leads further to a high order time discretisation of $\pi_t(\varphi)$. The idea behind this discretisation is summarised as follows.

First, for the sake of compactness, we augment the observation process as $\hat{Y}_t = (\hat{Y}_t^i)_{i=0}^{d_Y} = (t, Y_t^1, \ldots, Y_t^{d_Y})$ for all $t \geq 0$ and write

$$\hat{h} = \left(-\frac{1}{2}\sum_{i=1}^{d_Y} h_i^2, h_1, \ldots, h_{d_Y}\right).$$

Then, consider the *log-likelihood* process

$$\xi_t = \log(Z_t) = \sum_{i=0}^{d_Y} \int_0^t \hat{h}_i(X_s)\, \mathrm{d}\hat{Y}_s^i, \qquad t \geq 0. \tag{4}$$

Now, given a positive integer m, the order m time discretisation is achieved by a stochastic Taylor expansion up to order m of the processes $(\hat{h}_i(X_t))_{t\geq 0}$, $i = 0, \ldots, d_Y$ in (4). Finally, we substitute the discretised log-likelihood back into the original relationships (2) and the Kallianpur-Striebel formula (3) to obtain a discretisation of the filtering functionals. However, it is important to note that for the orders $m > 2$ an additional truncation procedure is needed, which we will make precise shortly, after introducing the necessary notation for the stochastic Taylor expansion.

2.2.1 Stochastic Taylor expansions

Let $\mathcal{M} = \{\alpha \in \{0, \ldots, d_V\}^l : l = 0, 1, \ldots\}$ be the set of all multi-indices with range $\{0, \ldots, d_V\}$, where $\emptyset$ denotes the multi-index of length zero. For $\alpha = (\alpha_1, ..., \alpha_k) \in \mathcal{M}$ we adopt the notation $|\alpha| = k$ for its length, $|\alpha|_0 = \#\{j : \alpha_j = 0\}$ for the number of zeros in α, and $\alpha- = (\alpha_1, ..., \alpha_{k-1})$ and $-\alpha = (\alpha_2, ..., \alpha_k)$, for the right and left truncations, respectively. By convention $|\emptyset| = 0$ and $-\emptyset = \emptyset- = \emptyset$. Given two multi-indices $\alpha, \beta \in \mathcal{M}$ we denote their concatenation by $\alpha * \beta$. For positive and non-zero integers n and m, we will also consider the subsets of multi-indices

$$\mathcal{M}_{n,m} = \{\alpha \in \mathcal{M} : n \leq |\alpha| \leq m\}, \text{ and}$$
$$\mathcal{M}_m = \mathcal{M}_{m,m} = \{\alpha \in \mathcal{M} : |\alpha| = m\}.$$

For brevity, and by slight abuse of notation, we augment the Brownian motion V and now write $V = \left(V^i\right)_{i=0}^{d_V} = (t, V_t^1, \ldots, V_t^{d_V})$ for all $t \geq 0$. We will consider the filtration $\{\mathcal{F}_t^{0,V}\}_{t\geq 0}$ defined to be the usual augmentation of the filtration generated by the process V and initially enlarged with the random variable X_0. Moreover, for fixed $t \geq 0$, we will also consider the filtration $\{\mathcal{H}_s^t = \mathcal{F}_s^{0,V} \vee \mathcal{Y}_t\}_{s\leq t}$. For all $\alpha \in \mathcal{M}$ and all suitably integrable $\mathcal{H}_s^t$-adapted processes $\gamma = \{\gamma_s\}_{s\leq t}$ denote by $I_\alpha(\gamma_\cdot)_{s,t}$ the ItÃť iterated integral given for all $s \leq t$ by

$$I_\alpha(\gamma_\cdot)_{s,t} = \begin{cases} \gamma_t, & \text{if } |\alpha| = 0 \\ \displaystyle\int_s^t I_{\alpha-}(\gamma_\cdot)_{s,u}\, dV_u^{\alpha_{|\alpha|}}, & \text{if } |\alpha| \geq 1. \end{cases}$$

Based on the coefficient functions of the signal X, we introduce the differential operators L^0 and L^r, $r = 1, ..., d_V$, defined for all twice continuously differentiable functions $g : \mathrm{R}^{d_X} \to \mathrm{R}$ by

$$L^0 g = \sum_{k=1}^{d_X} f_k \frac{\partial g}{\partial x^k} + \frac{1}{2} \sum_{k,l=1}^{d_X} \sum_{r=1}^{d_V} \sigma_{k,r}\sigma_{l,r} \frac{\partial^2 g}{\partial x^k \partial x^l} \quad \text{and}$$
$$L^r g = \sum_{k=1}^{d_X} \sigma_{k,r} \frac{\partial g}{\partial x^k}, \qquad r = 1, ..., d_V.$$

Lastly, for $\alpha = (\alpha_1, ..., \alpha_k) \in \mathcal{M}$, the differential operator L^α is defined to be the composition $L^\alpha = L^{\alpha_1} \circ \cdots \circ L^{\alpha_k}$, where, by convention, $L^\emptyset g = g$.

2.2.2 Discretisation of the log-likelihood process

With the stochastic Taylor expansion at hand, we can now describe the discretisation of the log-likelihood in (4). To this end, let for all $t > 0$,

$$\Pi(t) = \{\{t_0, \ldots, t_n\} \subset [0,t]^{n+1} : 0 = t_0 < t_1 < \cdots < t_n = t,\ n = 1,2,\ldots\}$$

be the set of all partitions of the interval $[0,t]$. For a given partition we call the quantity $\delta = \max\{t_{j+1} - t_j : j = 0,\ldots,n-1\}$ the *meshsize* of τ. Then we discretise the log-likelihood as follows. For all $t > 0$, $\tau \in \Pi(t)$ and all positive integers m we consider

$$\begin{aligned}\xi_t^{\tau,m} = \sum_{j=0}^{n-1} \xi_t^{\tau,m}(j) &= \sum_{j=0}^{n-1} \sum_{i=0}^{d_Y} \sum_{\alpha \in \mathcal{M}_{0,m-1}} L^\alpha \hat{h}_i(X_{t_j}) \int_{t_j}^{t_{j+1}} I_\alpha(\mathbf{1})_{t_j,s} d\hat{Y}_s^i \\ &= \sum_{j=0}^{n-1} \{\kappa_j^{0,m} + \int_{t_j}^{t_{j+1}} \langle \eta_j^{0,m}(s), \mathrm{d}Y_s \rangle\},\end{aligned}$$

where we define for all integers $l \leq m-1$ and $j = 0,\ldots,n-1$ the quantities

$$\kappa^{l,m} = \sum_{j=0}^{n-1} \kappa_j^{l,m} = \sum_{j=0}^{n-1} \Big\{ -\frac{1}{2} \sum_{\alpha \in \mathcal{M}_{l,m-1}} L^\alpha \langle h(\cdot), h(\cdot) \rangle (X_{t_j}) \int_{t_j}^{t_{j+1}} I_\alpha(\mathbf{1})_{t_j,s}\, \mathrm{d}s \Big\}$$

$$\eta_j^{l,m}(s) = \Big(\sum_{\alpha \in \mathcal{M}_{l,m-1}} L^\alpha h_i(X_{t_j}) I_\alpha(\mathbf{1})_{t_j,s} \Big)_{i=1,\ldots,d_Y}.$$

and $\langle \cdot, \cdot \rangle$ denotes the euclidean inner product. Note that by setting, in the case of $m > 2$,

$$\begin{aligned}\mu^{\tau,m}(j) &= \sum_{i=0}^{d_Y} \sum_{\alpha \in \mathcal{M}_{2,m-1}} L^\alpha \hat{h}_i(X_{t_j}) \int_{t_j}^{t_{j+1}} I_\alpha(\mathbf{1})_{t_j,s} d\hat{Y}_s^i \\ &= \kappa_j^{2,m} + \int_{t_j}^{t_{j+1}} \langle \eta_j^{2,m}(s), \mathrm{d}Y_s \rangle,\end{aligned}$$

we may write the above as

$$\xi_t^{\tau,m} = \xi_t^{\tau,2} + \sum_{j=0}^{n-1} \mu^{\tau,m}(j).$$

As outlined before, the discretisations $\xi^{\tau,m}$ are obtained by replacing the processes $(\hat{h}_i(X_t))_{t \geq 0}$, $i = 0,\ldots,d_Y$ in (4) with the truncation of degree $m-1$ of the corresponding stochastic Taylor expansion of $\hat{h}_i(X_t)$. These discretisations are subsequently used to obtain discretisation schemes of first and second order for the filter

$\pi_t(\varphi)$. However, they cannot be used directly to produce discretisation schemes of any order $m > 2$ because they do not have finite exponential moments (required to define the discretisation schemes). More precisely, the quantities $\mu^{\tau,m}(j)$ do not have finite exponential moments because of the high order iterated integral involved. For this, we need to introduce a truncation of $\mu^{\tau,m}(j)$ resulting in a (partial) taming procedure to the stochastic Taylor expansion of $(\hat{h}_i(X_t))_{t\geq 0}$. To achieve this, we introduce for every positive integer q and all $\delta > 0$ the truncation functions $\Gamma_{q,\delta}\colon \mathrm{R} \to \mathrm{R}$ such that

$$\Gamma_{q,\delta}(z) = \frac{z}{1+(z/\delta)^{2q}} \tag{5}$$

and set, for all $j = 0, \ldots, n-1$,

$$\bar{\xi}_t^{\tau,m}(j) = \begin{cases} \xi_t^{\tau,m}(j), & \text{if } m = 1,2 \\ \xi_t^{\tau,2}(j) + \Gamma_{m,(t_{j+1}-t_j)}\left(\mu^{\tau,m}(j)\right), & \text{if } m > 2 \end{cases}.$$

Utilising the above, the truncated discretisations of the log-likelihood finally read

$$\bar{\xi}_t^{\tau,m} = \sum_{j=0}^{n-1} \bar{\xi}_t^{\tau,m}(j). \tag{6}$$

We end this section with a remark about the properties of the truncation function before we go on to discretising the filter.

Remark 1 The following two properties of the truncation function Γ, defined in (5), are readily checked. For all positive integers q and all $\delta > 0$ we have that

i) the truncation function is bounded, specifically, for all $z \in \mathrm{R}$,

$$|\Gamma_{q,\delta}(z)| \leq \frac{\delta}{(2q-1)^{1/2q}},$$

ii) and that its derivative is bounded for all $z \in \mathrm{R}$ as

$$\frac{q(1-q)-1}{2q} \leq \frac{\mathrm{d}}{\mathrm{d}z}\Gamma_{q,\delta}(z) \leq 1.$$

In particular, the truncation function is Lipschitz continuous.

2.2.3 Discretisation of the filter

Since $\bar{\xi}_t^{\tau,m}$ in (6) is a discretisation of the log-likelihood we will now consider, for all $t > 0$, $\tau \in \Pi(t)$ and all positive integers m, the discretised likelihood

$$Z_t^{\tau,m} = \exp\left(\bar{\xi}_t^{\tau,m}\right).$$

The filter is now discretised, under the condition that the Borel measurable function φ satisfies $\tilde{\mathrm{E}}[|\varphi(X_t)Z_t^{\tau,m}|] < \infty$, to the m-th order by

$$\rho_t^{\tau,m}(\varphi) = \tilde{\mathrm{E}}\left[\varphi(X_t)Z_t^{\tau,m} \mid \mathcal{Y}_t\right]$$

and

$$\pi_t^{\tau,m}(\varphi) = \frac{\rho_t^{\tau,m}(\varphi)}{\rho_t^{\tau,m}(\mathbf{1})}. \tag{7}$$

It remains to show that the achieved discretisation is indeed of order m.

2.2.4 Order of approximation for the filtering functionals

In the framework developed thus far, we can state the main result of [5] which justifies the construction and proves the high order approximation. To this end, we consider the L^p-norms $\|\cdot\|_{L^p} = \tilde{\mathrm{E}}[|\cdot|^p]^{1/p}$, $p \geq 1$.

Theorem 1 (Theorem 2.3 in [5])
Let m be a positive integer, let $t > 0$, let φ be an $(m+1)$-times continuously differentiable function with at most polynomial growth and assume further that the coefficients of the partially observed system (X,Y) in (1) satisfy that

- *f is bounded and $\max\{2, 2m-1\}$-times continuously differentiable with bounded derivatives,*
- *σ is bounded and $2m$-times continuously differentiable with bounded derivatives,*
- *h is bounded and $(2m+1)$-times continuously differentiable with bounded derivatives, and that*
- *X_0 has moments of all orders.*

Then there exist positive constants δ_0 and C, such that for all partitions $\tau \in \Pi(t)$ with meshsize $\delta < \delta_0$ we have that

$$\|\rho_t(\varphi) - \rho_t^{\tau,m}(\varphi)\|_{L^2} \leq C\delta^m.$$

Moreover, there exist positive constants $\bar{\delta}_0$ and $\bar{C}$, such that for all partitions $\tau \in \Pi(t)$ with meshsize $\delta < \bar{\delta}_0$,

$$\mathrm{E}\left[|\pi_t(\varphi) - \pi_t^{\tau,m}(\varphi)|\right] \leq \bar{C}\delta^m.$$

Remark 2 Under the above assumption that h is bounded and φ has at most polynomial growth, the required condition from Theorem 2.4 in [5] that there exists $\varepsilon > 0$ such that $\sup_{\{\tau\in\Pi(t):\delta<\delta_0\}} \|\pi_t^{\tau,m}(\varphi)\|_{L^{2+\varepsilon}} < \infty$ holds.

3 Robustness of the approximation

The classical robustness of the filter as in Theorem 5.12 in [1] states that for every $t > 0$ and bounded Borel measurable function φ the filter $\pi_t(\varphi)$ can be represented as a function of the observation *path*

$$Y_{[0,t]}(\omega) = \{Y_s(\omega) \colon s \in [0,t]\}, \qquad \omega \in \Omega.$$

In particular, $Y_{[0,t]}$ is here a path-valued random variable. The precise meaning of robustness is then that there exists a unique bounded Borel measurable function $F^{t,\varphi}$ on the path space $C([0,t];\mathrm{R}^{d_Y})$, that is the space of continuous R^{d_Y}-valued functions on $[0,t]$, with the properties that

i) P-almost surely,

$$\pi_t(\varphi) = F^{t,\varphi}(Y_{[0,t]})$$

and

ii) $F^{t,\varphi}$ is continuous with respect to the supremum norm[1].

The volume [1] contains further details on the robust representation. In the present paper, we establish the analogous result for the discretised filter $\pi_t^{\tau,m}(\varphi)$ from (7). It is formulated as follows.

Theorem 2 *Let $t > 0$, $\tau = \{t_0, \ldots, t_n\} \in \Pi(t)$, let m be a positive integer and let φ be a bounded Borel measurable function. Then there exists a function $F_\varphi^{\tau,m} \colon C([0,t];\mathrm{R}^{d_Y}) \to \mathrm{R}$ with the properties that*

i) P-almost surely,

$$\pi_t^{\tau,m}(\varphi) = F_\varphi^{\tau,m}(Y_{[0,t]})$$

and

ii) for every two bounded paths $y_1, y_2 \in C([0,t];\mathrm{R}^{d_Y})$ there exists a positive constant C such that

$$|F_\varphi^{\tau,m}(y_1) - F_\varphi^{\tau,m}(y_2)| \le C\|\varphi\|_\infty \|y_1 - y_2\|_\infty.$$

Note that Theorem 2 implies the following statement in the total variation norm.

Corollary 1 *Let $t > 0$, $\tau = \{t_0, \ldots, t_n\} \in \Pi(t)$, and let m be a positive integer. Then, for every two bounded paths $y_1, y_2 \in C([0,t];\mathrm{R}^{d_Y})$ there exists a positive constant C such that*

$$\|\pi_t^{\tau,m,y_1} - \pi_t^{\tau,m,y_2}\|_{TV} = \sup_{\varphi \in B_b,\, \|\varphi\|_\infty \le 1} |F_\varphi^{\tau,m}(y_1) - F_\varphi^{\tau,m}(y_2)| \le C\|y_1 - y_2\|_\infty,$$

where B_b is the set of bounded and Borel measurable functions.

Remark 3 A natural question that arises in this context is to seek the rate of pathwise convergence of $F_\varphi^{\tau,m}$ to F_φ (defined as the limit of $F_\varphi^{\tau,m}$ when the meshsize goes to zero) as functions on the path space. The rate of pathwise convergence is expected to be dependent on the Hölder constant of the observation path. Therefore, it is expected to be not better than $\frac{1}{2} - \epsilon$ for a semimartingale observation. The absence of high order iterated integrals of the observation process in the construction of $F_\varphi^{\tau,m}$ means

[1] For a subset $D \subseteq \mathrm{R}^l$ and a function $\psi \colon D \to \mathrm{R}^d$ we set $\|\psi\|_\infty = \max_{i=1,\ldots,d} \|\psi_i\|_\infty = \max_{i=1,\ldots,d} \sup_{x \in D} |\psi_i(x)|$

that one cannot obtain *pathwise* high order approximations based on the work in [5]. Such approximations will no longer be continuous in the supremum norm. Thus we need to consider rough path norms in this context. In a different setting, Clark showed in the earlier paper [3] that one cannot construct pathwise approximations of solutions of SDEs by using only increments of the driving Brownian motion.

In the following and final part of the paper, we exhibit the proof of Theorem 2.

4 Proof of the robustness of the approximation

We begin by constructing what will be the robust representation. Consider, for all $y \in C([0,t];\mathbb{R}^{d_Y})$,

$$\begin{aligned}
\Xi_t^{\tau,m}(y) &= \sum_{j=0}^{n-1} \{\kappa_j^{0,m} + \langle \eta_j^{0,m}(t_{j+1}), y_{t_{j+1}}\rangle - \langle \eta_j^{0,m}(t_j), y_{t_j}\rangle - \int_{t_j}^{t_{j+1}} \langle y_s, \mathrm{d}\eta_j^{0,m}(s)\rangle\} \\
&= \sum_{j=0}^{n-1} \{\kappa_j^{0,m} + \langle \eta_j^{0,m}(t_{j+1}), y_{t_{j+1}}\rangle - \langle h(X_{t_j}), y_{t_j}\rangle - \int_{t_j}^{t_{j+1}} \langle y_s, \mathrm{d}\eta_j^{0,m}(s)\rangle\} \\
&= \langle h(X_{t_n}), y_{t_n}\rangle - \langle h(X_{t_0}), y_{t_0}\rangle \\
&+ \sum_{j=0}^{n-1} \{\kappa_j^{0,m} + \langle \eta_j^{0,m}(t_{j+1}) - h(X_{t_{j+1}}), y_{t_{j+1}}\rangle - \int_{t_j}^{t_{j+1}} \langle y_s, \mathrm{d}\eta_j^{0,m}(s)\rangle\}
\end{aligned}$$

and further, for $m > 2$,

$$M_j^{\tau,m}(y) = \kappa_j^{2,m} + \langle \eta_j^{2,m}(t_{j+1}), y_{t_{j+1}}\rangle - \int_{t_j}^{t_{j+1}} \langle y_s, d\eta_j^{2,m}(s)\rangle$$

so that we can define

$$\bar{\Xi}_t^{\tau,m}(y) = \begin{cases} \Xi_t^{\tau,m}(y), & \text{if } m = 1,2 \\ \Xi_t^{\tau,2}(y) + \sum_{j=0}^{n-1} \Gamma_{m,(t_{j+1}-t_j)}(M_j^{\tau,m}(y)), & \text{if } m > 2 \end{cases} .$$

Furthermore, set

$$\mathcal{Z}_t^{\tau,m}(y) = \exp(\bar{\Xi}_t^{\tau,m}(y)).$$

Example 1 The robust approximation for $m = 1$ and $m = 2$ are given as follows. First, if $m = 1$, then

$$\begin{aligned}\Xi_t^{\tau,1}(y) &= \sum_{j=0}^{n-1}\{\kappa_j^{0,1} + \langle \eta_j^{0,1}(t_{j+1}), y_{t_{j+1}}\rangle - \langle h(X_{t_j}), y_{t_j}\rangle - \int_{t_j}^{t_{j+1}} \langle y_s, \mathrm{d}\eta_j^{0,1}(s)\rangle\}\\ &= \sum_{j=0}^{n-1}\{-\frac{1}{2}\langle h,h\rangle(X_{t_j})(t_{j+1}-t_j) + \langle h(X_{t_j}), y_{t_{j+1}} - y_{t_j}\rangle\}\end{aligned}$$

and also $\bar{\Xi}_t^{\tau,1}(y) = \Xi_t^{\tau,1}(y)$ so that $\mathcal{Z}_t^{\tau,1}(y) = \exp(\Xi_t^{\tau,1}(y))$. If $m = 2$, then

$$\begin{aligned}\Xi_t^{\tau,2}(y) &= \Xi_t^{\tau,1}(y) + \sum_{j=0}^{n-1}\{\kappa_j^{1,2} + \langle \eta_j^{1,2}(t_{j+1}), y_{t_{j+1}}\rangle - \int_{t_j}^{t_{j+1}} \langle y_s, \mathrm{d}\eta_j^{1,2}(s)\rangle\}\\ &= \Xi_t^{\tau,1}(y) - \sum_{\alpha\in\mathcal{M}_1}\sum_{j=0}^{n-1}\frac{1}{2}L^\alpha\langle h,h\rangle(X_{t_j})\int_{t_j}^{t_{j+1}} V_s^\alpha - V_{t_j}^\alpha\,\mathrm{d}s\\ &\quad + \sum_{\alpha\in\mathcal{M}_1}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\langle L^\alpha h(X_{t_j}), y_{t_{j+1}} - y_s\rangle\,\mathrm{d}V_s^\alpha.\end{aligned}$$

Therefore, also $\bar{\Xi}_t^{\tau,2}(y) = \Xi_t^{\tau,2}(y)$ so that $\mathcal{Z}_t^{\tau,2}(y) = \exp(\Xi_t^{\tau,2}(y))$. □

First, we show that the newly constructed $\mathcal{Z}_t^{\tau,m}$ is locally bounded.

Lemma 1 *Let $t > 0$, let $\tau = \{t_0,\dots,t_n\} \in \Pi(t)$ be a partition with mesh size δ and let m be a positive integer. Then, for all $R > 0$, $p \geq 1$ there exists a positive constant $B_{p,R}$ such that*

$$\sup_{\|y\|_\infty \leq R} \|\mathcal{Z}_t^{\tau,m}(y)\|_{L^p} \leq B_{p,R}.$$

Proof Notice that, by Remark 1, in the case $m \geq 2$, we have for all $y \in C([0,t];\mathrm{R}^{d_Y})$ that

$$\bar{\Xi}_t^{\tau,m}(y) \leq \Xi_t^{\tau,2}(y) + \frac{n\delta}{(2m-1)^{1/2m}}.$$

This implies that for all $y \in C([0,t];\mathrm{R}^{d_Y})$,

$$\mathcal{Z}_t^{\tau,m}(y) = \exp(\bar{\Xi}_t^{\tau,m}(y)) \leq \exp(\Xi_t^{\tau,2}(y))\exp\left(\frac{n\delta}{(2m-1)^{1/2m}}\right).$$

For $m = 1$, we clearly have $\mathcal{Z}_t^{\tau,1}(y) = \exp(\Xi_t^{\tau,1}(y))$. Hence, it suffices to show the result for $m = 1, 2$ only. We have

$$\Xi_t^{\tau,2}(y) = \Xi_t^{\tau,1}(y) + \sum_{j=0}^{n-1}\{\kappa_j^{1,2} + \langle \eta_j^{1,2}(t_{j+1}), y_{t_{j+1}}\rangle - \int_{t_j}^{t_{j+1}} \langle y_s, \mathrm{d}\eta_j^{1,2}(s)\rangle\}.$$

Now, by the triangle inequality, boundedness of y, and boundedness of h, we get

$$|\Xi_t^{\tau,1}(y)| = \left|\sum_{j=0}^{n-1}\{\kappa_j^{0,1} + \langle \eta_j^{0,1}(t_{j+1}), y_{t_{j+1}}\rangle - \langle h(X_{t_j}), y_{t_j}\rangle - \int_{t_j}^{t_{j+1}} \langle y_s, \mathrm{d}\eta_j^{0,1}(s)\rangle\}\right|$$

$$
\begin{aligned}
&= \Big|\sum_{j=0}^{n-1}\{\kappa_j^{0,1} + \langle h(X_{t_j}), y_{t_{j+1}} - y_{t_j}\rangle\}\Big| \\
&= \Big|\sum_{j=0}^{n-1}\{-\frac{1}{2}\langle h(X_{t_j}), h(X_{t_j})\rangle(t_{j+1}-t_j) + \langle h(X_{t_j}), y_{t_{j+1}} - y_{t_j}\rangle\}\Big| \\
&\leq \frac{td_Y\|h\|_\infty^2}{2} + 2R\|h\|_\infty = C_0,
\end{aligned}
$$

where we denote the final constant by C_0. Furthermore, by the triangle inequality, boundedness of y, and boundedness of h and its derivatives,

$$
\begin{aligned}
&\Big|\sum_{j=0}^{n-1}\{\kappa_j^{1,2} + \langle \eta_j^{1,2}(t_{j+1}), y_{t_{j+1}}\rangle - \int_{t_j}^{t_{j+1}} \langle y_s, \mathrm{d}\eta_j^{1,2}(s)\rangle\}\Big| \\
&= \sum_{\alpha\in\mathcal{M}_1}\Big\{\Big|\sum_{j=0}^{n-1}\frac{1}{2}L^\alpha\langle h,h\rangle(X_{t_j})\int_{t_j}^{t_{j+1}} V_s^\alpha - V_{t_j}^\alpha\,\mathrm{d}s \\
&+\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\langle L^\alpha h(X_{t_j}), y_{t_{j+1}} - y_s\rangle\,\mathrm{d}V_s^\alpha\Big|\Big\} \\
&\leq \sum_{\alpha\in\mathcal{M}_1\setminus\{0\}}\Big\{\Big|\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\frac{1}{2}L^\alpha\langle h,h\rangle(X_{t_j})(t_{j+1}-s) + \langle L^\alpha h(X_{t_j}), y_{t_{j+1}} - y_s\rangle\,\mathrm{d}V_s^\alpha\Big|\Big\} \\
&+\Big|\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\frac{1}{2}L^0\langle h,h\rangle(X_{t_j})(s-t_j) + \langle L^0 h(X_{t_j}), y_{t_{j+1}} - y_s\rangle\,\mathrm{d}s\Big| \\
&\leq \sum_{\alpha\in\mathcal{M}_1\setminus\{0\}}\Big\{\Big|\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\frac{1}{2}L^\alpha\langle h,h\rangle(X_{t_j})(t_{j+1}-s) + \langle L^\alpha h(X_{t_j}), y_{t_{j+1}} - y_s\rangle\,\mathrm{d}V_s^\alpha\Big|\Big\} \\
&+\frac{1}{2}\delta t\|L^0\langle h,h\rangle\|_\infty + 2d_Y Rt\|L^0 h\|_\infty \\
&= C_1 + \sum_{\alpha\in\mathcal{M}_1\setminus\{0\}}\Big\{\Big|\int_0^t\frac{1}{2}L^\alpha\langle h,h\rangle(X_{\lfloor s\rfloor})(\lceil s\rceil - s) + \langle L^\alpha h(X_{\lfloor s\rfloor}), y_{\lceil s\rceil} - y_s\rangle\,\mathrm{d}V_s^\alpha\Big|\Big\}.
\end{aligned}
$$

Here, C_1 is a constant introduced for conciseness. Then,

$$
\begin{aligned}
&\|\mathcal{Z}_t^{\tau,2}(y)\|_{L^p} \\
&= \Big\|\mathcal{Z}_t^{\tau,1}(y)\exp(\sum_{j=0}^{n-1}\{\kappa_j^{1,2} - \langle\eta_j^{1,2}(t_{j+1}), y_{t_{j+1}}\rangle - \int_{t_j}^{t_{j+1}}\langle y_s, \mathrm{d}\eta_j^{1,2}(s)\rangle\})\Big\|_{L^p} \\
&\leq \exp(C_0 + C_1) \\
&\Big\|\exp\Big(\sum_{\alpha\in\mathcal{M}_1\setminus\{0\}}\Big\{\Big|\int_0^t\frac{1}{2}L^\alpha\langle h,h\rangle(X_{\lfloor s\rfloor})(\lceil s\rceil - s) + \langle L^\alpha h(X_{\lfloor s\rfloor}), y_{\lceil s\rceil} - y_s\rangle\,\mathrm{d}V_s^\alpha\Big|\Big\}\Big)\Big\|_{L^p}
\end{aligned}
$$

$< \infty.$

The lemma is thus proved. □

In analogy to the filter, we define the functions

$$G_\varphi^{\tau,m}(y) = \tilde{\mathrm{E}}[\varphi(X_t)\mathcal{Z}_t^{\tau,m}(y)]$$

and

$$F_\varphi^{\tau,m}(y) = \frac{G_\varphi^{\tau,m}(y)}{G_{\mathbf{1}}^{\tau,m}(y)} = \frac{\tilde{\mathrm{E}}[\varphi(X_t)\mathcal{Z}_t^{\tau,m}(y)]}{\tilde{\mathrm{E}}[\mathcal{Z}_t^{\tau,m}(y)]}.$$

Lemma 2 *Let $\tau \in \Pi(t)$ be a partition, let m be a positive integer and let φ be a bounded Borel measurable function. Then the functions $G_\varphi^{\tau,m}: C([0,t];\mathrm{R}^{d_Y}) \to \mathrm{R}$ and $F_\varphi^{\tau,m}: C([0,t];\mathrm{R}^{d_Y}) \to \mathrm{R}$ are locally Lipschitz continuous and locally bounded. Specifically, for every two paths $y_1, y_2 \in C([0,t];\mathrm{R}^{d_Y})$ such that there exists a real number $R > 0$ with $\|y_1\|_\infty \le R$ and $\|y_2\|_\infty \le R$, there exist constants L_G, M_G, L_F, and M_F such that*

$$|G_\varphi^{\tau,m}(y_1) - G_\varphi^{\tau,m}(y_2)| \le L_G\|\varphi\|_\infty\|y_1 - y_2\|_\infty \qquad \text{and} \qquad |G_\varphi^{\tau,m}(y_1)| \le M_G\|\varphi\|_\infty$$

and

$$|F_\varphi^{\tau,m}(y_1) - F_\varphi^{\tau,m}(y_2)| \le L_F\|\varphi\|_\infty\|y_1 - y_2\|_\infty \qquad \text{and} \qquad |F_\varphi^{\tau,m}(y_1)| \le M_F\|\varphi\|_\infty.$$

Proof We first show the results for $G_\varphi^{\tau,m}$. Note that

$$|\mathcal{Z}_t^{\tau,m}(y_1) - \mathcal{Z}_t^{\tau,m}(y_2)| \le (\mathcal{Z}_t^{\tau,m}(y_1) + \mathcal{Z}_t^{\tau,m}(y_2))|\bar{\Xi}_t^{\tau,m}(y_1) - \bar{\Xi}_t^{\tau,m}(y_2)|.$$

Then, by the Cauchy-Schwarz inequality, for all $p \ge 1$ we have

$$\|\varphi(X_t)\mathcal{Z}_t^{\tau,m}(y_1) - \varphi(X_t)\mathcal{Z}_t^{\tau,m}(y_2)\|_{L^p} \le 2B_{2p,R}\|\varphi\|_\infty\|\bar{\Xi}_t^{\tau,m}(y_1) - \bar{\Xi}_t^{\tau,m}(y_2)\|_{L^{2p}}. \tag{8}$$

Thus, for $m > 2$, we can exploit the effect of the truncation function and, similarly to the proof of Lemma 1, it suffices to show the result for $m = 1,2$. To this end, consider for all $q \ge 1$,

$$\begin{aligned}\|\Xi_t^{\tau,2}(y_1) - \Xi_t^{\tau,2}(y_2)\|_{L^q} &\le \|\Xi_t^{\tau,1}(y_1) - \Xi_t^{\tau,1}(y_2)\|_{L^q} \\ &+ \left\|\sum_{j=0}^{n-1}\{\langle\eta_j^{1,2}(t_{j+1}), y_1(t_{j+1}) - y_2(t_{j+1})\rangle - \int_{t_j}^{t_{j+1}}\langle y_1(s) - y_2(s), \mathrm{d}\eta_j^{1,2}(s)\rangle\}\right\|_{L^q}.\end{aligned}$$

First, we obtain for all $q \ge 1$,

$$\|\Xi_t^{\tau,1}(y_1)-\Xi_t^{\tau,1}(y_2)\|_{L^q} = \left\|\sum_{j=0}^{n-1}\langle h(X_{t_j}),(y_1(t_{j+1})-y_2(t_{j+1}))-(y_1(t_j)-y_2(t_j))\rangle\right\|_{L^q}$$
$$\leq 2d_Y\|h\|_\infty\|y_1-y_2\|_\infty.$$

And second we have for all $q \geq 1$ that

$$\left\|\sum_{j=0}^{n-1}\{\int_{t_j}^{t_{j+1}}\langle(y_1(t_{j+1})-y_1(s))-(y_2(t_{j+1})-y_2(s)),\mathrm{d}\eta_j^{1,2}(s)\rangle\}\right\|_{L^q}$$
$$\leq \sum_{j=0}^{n-1}\Bigg\|\left|\langle L^0 h(X_{t_j}),y_1(t_{j+1})-y_2(t_{j+1})\rangle(t_{j+1}-t_j)\right|$$
$$+\left|\int_{t_j}^{t_{j+1}}\langle L^0 h(X_{t_j}),y_1(s)-y_2(s)\rangle\,\mathrm{d}s\right|$$
$$+\sum_{\alpha\in\mathcal{M}_1\setminus\{0\}}\left|\langle L^\alpha h(X_{t_j}),y_1(t_{j+1})-y_2(t_{j+1})\rangle(V_{t_{j+1}}^\alpha-V_{t_j}^\alpha)\right|$$
$$+\left|\int_{t_j}^{t_{j+1}}\langle L^\alpha h(X_{t_j}),y_1(s)-y_2(s)\rangle\,\mathrm{d}V_s^\alpha\right|\Bigg\|_{L^q}$$
$$\leq\left[\bar{C}_1+\bar{C}_2\sum_{j=0}^{n-1}\sum_{\alpha\in\mathcal{M}_1\setminus\{0\}}\|V_{t_{j+1}}^\alpha-V_{t_j}^\alpha\|_{L^q}\right]\|y_1-y_2\|_\infty$$
$$\leq C\|y_1-y_2\|_\infty$$

This and Lemma 1 imply that $G_\varphi^{\tau,m}$ is locally Lipschitz and locally bounded. To show the result for $F_\varphi^{\tau,m}$ we need to establish that $1/G_{\mathbf{1}}^{\tau,m}$ is locally bounded. We have, using Jensen's inequality, that for $m \geq 2$

$$G_{\mathbf{1}}^{\tau,m}=\tilde{\mathrm{E}}[\mathcal{Z}_t^{\tau,m}]\geq\exp\left(\tilde{\mathrm{E}}[\bar{\Xi}_t^{\tau,m}]\right)\geq\exp\left(\tilde{\mathrm{E}}[\Xi_t^{\tau,2}]\right)\exp\left(-\frac{n\delta}{(2m-1)^{1/2m}}\right)$$

and for $m = 1$ clearly

$$G_{\mathbf{1}}^{\tau,1}=\tilde{\mathrm{E}}[\mathcal{Z}_t^{\tau,1}]\geq\exp\left(\tilde{\mathrm{E}}[\Xi_t^{\tau,1}]\right).$$

Since the quantities $\tilde{\mathrm{E}}[\Xi_t^{\tau,1}]$ and $\tilde{\mathrm{E}}[\Xi_t^{\tau,2}]$ are finite, the lemma is proved. □

In the following, given $t > 0$, we set for every $\gamma \in (0, 1/2)$,

$$\mathcal{H}_\gamma=\left\{y\in C([0,t];\mathrm{R}^{d_Y}):\sup_{s_1,s_2\in[0,t]}\frac{\|y_{s_1}-y_{s_2}\|_\infty}{|s_1-s_2|^\gamma}<\infty\right\}\subseteq C([0,t];\mathrm{R}^{d_Y})$$

and recall that $Y_{[0,t]}:\Omega\to C([0,t];\mathrm{R}^{d_Y})$ denotes the random variable in path space corresponding to the observation process Y.

Lemma 3 *For all $t > 0$ and $\gamma \in (0, 1/2)$, we have $\tilde{P}$-almost surely that $Y_{[0,t]} \in \mathcal{H}_\gamma$.*

Proof Recall that, under $\tilde{P}$, the observation process Y is a Brownian motion and, by the Brownian scaling property, it suffices to show the result for $t = 1$. Therefore, let $\gamma \in (0, 1/2)$ and note that for all $\delta \in (0, 1]$ we have

$$\sup_{s_1,s_2\in[0,1]} \frac{\|Y_{s_1} - Y_{s_2}\|_\infty}{|s_1 - s_2|^\gamma} = \max\left\{ \sup_{\substack{s_1,s_2\in[0,1]\\ |s_1-s_2|\le\delta}} \frac{\|Y_{s_1} - Y_{s_2}\|_\infty}{|s_1 - s_2|^\gamma}, \sup_{\substack{s_1,s_2\in[0,1]\\ |s_1-s_2|\ge\delta}} \frac{\|Y_{s_1} - Y_{s_2}\|_\infty}{|s_1 - s_2|^\gamma} \right\}.$$

The second element of the maximum above is easily bounded, $\tilde{P}$-almost surely, by the sample path continuity. For the first element, note that there exists $\delta_0 \in (0, 1)$ such that for all $\delta \in (0, \delta_0]$,

$$\delta^\gamma \ge \sqrt{2\delta \log(1/\delta)}.$$

Therefore, it follows that $\tilde{P}$-almost surely,

$$\sup_{\substack{s_1,s_2\in[0,1]\\ |s_1-s_2|\le\delta_0}} \frac{\|Y_{s_1} - Y_{s_2}\|_\infty}{|s_1 - s_2|^\gamma} \le \sup_{\substack{s_1,s_2\in[0,1]\\ |s_1-s_2|\le\delta_0}} \frac{\|Y_{s_1} - Y_{s_2}\|_\infty}{\sqrt{2|s_1 - s_2| \log(1/|s_1 - s_2|)}}.$$

The Lévy modulus of continuity of Brownian motion further ensures that $\tilde{P}$-almost surely,

$$\limsup_{\delta\downarrow 0} \sup_{\substack{s_1,s_2\in[0,1]\\ |s_1-s_2|\le\delta}} \frac{\|Y_{s_1} - Y_{s_2}\|_\infty}{\sqrt{2\delta \log(1/\delta)}} = 1.$$

The Lemma 3 thus follows. □

Lemma 4 *Let $\tau = \{0 = t_1 < \ldots < t_n = t\} \in \Pi(t)$ be a partition, let $j \in \{0, \ldots, n-1\}$ and let c be a positive integer. Then, there exists a version of the stochastic integral*

$$C([0,t]; \mathrm{R}^{d_Y}) \times \Omega \ni (y, \omega) \mapsto \int_{t_j}^{t_{j+1}} \langle y_s, \mathrm{d}\eta_j^{c,c+1}(s, \omega)\rangle \in \mathrm{R}$$

such that it is equal on $\mathcal{H}_\gamma \times \Omega$ to a $\mathcal{B}(C([0,t]; \mathrm{R}^{d_Y})) \times \mathcal{F}$-measurable mapping.

Proof For k a positive integer, define for $y \in C([0,t]; \mathrm{R}^{d_Y})$,

$$\mathcal{J}_j^{c,k}(y) = \sum_{i=0}^{k-1} \left\langle y_{s_{i,j}}, \left(\eta_j^{c,c+1}(s_{i+1,j}) - \eta_j^{c,c+1}(s_{i,j})\right)\right\rangle,$$

where $s_{i,j} = \frac{i(t_{j+1}-t_j)}{k} + t_j$, $i = 0, \ldots, k$. Furthermore, we set $\lfloor s \rfloor = s_{i,j}$ for $s \in [\frac{i(t_{j+1}-t_j)}{k} + t_j, \frac{(i+1)(t_{j+1}-t_j)}{k} + t_j)$. Then, for $y \in \mathcal{H}_\gamma$, we have

$$\tilde{\mathrm{E}}\left[\left(\mathcal{J}_j^{c,2^l}(y) - \int_{t_j}^{t_{j+1}} \langle y_s, \mathrm{d}\eta_j^{c,c+1}(s)\rangle\right)^2\right]$$

$$= \tilde{\mathrm{E}}\Big[\Big(\int_{t_j}^{t_{j+1}} \langle y_{\lfloor s\rfloor} - y_s, \mathrm{d}\eta_j^{c,c+1}(s)\rangle\Big)^2\Big]$$

$$= \tilde{\mathrm{E}}\Big[\Big(\sum_{\alpha\in\mathcal{M}_c}\sum_{i=0}^{d_Y}\int_{t_j}^{t_{j+1}} (y^i_{\lfloor s\rfloor} - y^i_s) L^\alpha h_i(X_{t_j})\,\mathrm{d}I_\alpha(\mathbf{1})_{t_j,s}\Big)^2\Big]$$

$$\le (d_V+1)d_Y\sum_{i=0}^{d_Y}\sum_{\alpha\in\mathcal{M}_c}\tilde{\mathrm{E}}\Big[\Big(\int_{t_j}^{t_{j+1}} (y^i_{\lfloor s\rfloor} - y^i_s) L^\alpha h_i(X_{t_j})\,\mathrm{d}I_\alpha(\mathbf{1})_{t_j,s}\Big)^2\Big]$$

$$= (d_V+1)d_Y\sum_{i=0}^{d_Y}\sum_{\substack{\alpha\in\mathcal{M}_c\\ \alpha_{|\alpha|}\neq 0}}\tilde{\mathrm{E}}\Big[\int_{t_j}^{t_{j+1}} ((y^i_{\lfloor s\rfloor} - y^i_s) L^\alpha h_i(X_{t_j}))^2\,\mathrm{d}\langle I_\alpha(\mathbf{1})_{t_j,\cdot}\rangle_s\Big]$$

$$+ (d_V+1)d_Y\sum_{i=0}^{d_Y}\sum_{\substack{\alpha\in\mathcal{M}_c\\ \alpha_{|\alpha|}=0}}\tilde{\mathrm{E}}\Big[\Big(\int_{t_j}^{t_{j+1}} (y^i_{\lfloor s\rfloor} - y^i_s) L^\alpha h_i(X_{t_j})\,\mathrm{d}\big[\int_{t_j}^{s} I_{\alpha-}(\mathbf{1})_{t_j,r}\mathrm{d}r\big]\Big)^2\Big]$$

$$\le (d_V+1)d_Y\frac{K(t_{j+1}-t_j)^{2\gamma}}{2^{2l\gamma}}\max_{\alpha\in\mathcal{M}_c}\|L^\alpha h(X_{t_j})\|_\infty\Big\{\sum_{i=0}^{d_Y}\sum_{\substack{\alpha\in\mathcal{M}_c\\ \alpha_{|\alpha|}\neq 0}}\tilde{\mathrm{E}}\Big[\int_{t_j}^{t_{j+1}} (I_{\alpha-}(\mathbf{1})_{t_j,s})^2\,\mathrm{d}s\Big]$$

$$+\sum_{i=0}^{d_Y}\sum_{\substack{\alpha\in\mathcal{M}_c\\ \alpha_{|\alpha|}=0}}\tilde{\mathrm{E}}\Big[\Big(\int_{t_j}^{t_{j+1}} I_{\alpha-}(\mathbf{1})_{t_j,s}\,\mathrm{d}s\Big)^2\Big]\Big\}$$

$$\le \frac{(d_V+1)d_Y CK(t_{j+1}-t_j)^{2\gamma}}{2^{2l\gamma}},$$

Where the constant C is independent of l. Thus, by Chebyshev's inequality, we get for all $\epsilon > 0$ that

$$\tilde{P}\Big(\Big|\mathcal{J}_j^{c,2^l}(y) - \int_{t_j}^{t_{j+1}}\langle y_s, \mathrm{d}\eta_j^{c,c+1}(s)\rangle\Big| > \epsilon\Big) \le \frac{1}{\epsilon^2}\frac{(d_V+1)d_Y CK(t_{j+1}-t_j)^{2\gamma}}{2^{2l\gamma}}.$$

However, the bound on the right-hand side is summable over l so that we conclude using the first Borel-Cantelli Lemma that, for all $\epsilon > 0$,

$$\tilde{P}\Big(\limsup_{l\to\infty}\Big|\mathcal{J}_j^{c,2^l}(y) - \int_{t_j}^{t_{j+1}}\langle y_s, \mathrm{d}\eta_j^{c,c+1}(s)\rangle\Big| > \epsilon\Big) = 0.$$

Thus, for all $y\in\mathcal{H}_\gamma$, the integral $\mathcal{J}_j^{c,k}(y)$ converges $\tilde{P}$-almost surely to the integral $\int_{t_j}^{t_{j+1}}\langle y_s, \mathrm{d}\eta_j^{c,c+1}(s)\rangle$. Hence, we can define the limit on $\mathcal{H}_\gamma\times\Omega$ to be

$$\mathcal{J}_j^{c}(y)(\omega) = \limsup_{l\to\infty}\mathcal{J}_j^{c,l}(y)(\omega);\qquad (y,\omega)\in\mathcal{H}_\gamma\times\Omega.$$

Since the mapping

$$C([0,T];\mathbb{R}^{d_Y})\times\Omega \ni (y,\omega)\mapsto \limsup_{l\to\infty}\mathcal{J}_j^{c,l}(y)(\omega)\in\mathbb{R}$$

is jointly $\mathcal{B}(C([0,T];\mathbb{R}^{d_Y}))\otimes\mathcal{F}$ measurable the lemma is proved. □

It turns out that proving the robustness result is simplified by first decoupling the processes X and Y in the following manner. Let $(\mathring{\Omega},\mathring{\mathcal{F}},\mathring{P})$ be an indentical copy of the probability space $(\Omega,\mathcal{F},\tilde{P})$. Then

$$\mathring{G}_{\varphi}^{\tau,m}(y)=\mathring{\mathrm{E}}[\varphi(\mathring{X}_t)\mathring{\mathcal{Z}}_t^{\tau,m}(y)]$$

is the corresponding representation of $G_{\varphi}^{\tau,m}(y)$ in the new space, where $\mathring{\mathcal{Z}}_t^{\tau,m}(y)=\exp(\mathring{\bar{\Xi}}_t^{\tau,m}(y))$ with

$$\mathring{\Xi}_t^{\tau,m}(y)=\sum_{j=0}^{n-1}\mathring{\kappa}_j^{0,m}+\langle\mathring{\eta}_j^{0,m}(t_{j+1}),y_{t_{j+1}}\rangle-\langle h(\mathring{X}_{t_j}),y_{t_j}\rangle-\int_{t_j}^{t_{j+1}}\langle y_s,\mathrm{d}\mathring{\eta}_j^{0,m}(s)\rangle$$

and, for $m>2$,

$$\mathring{M}_j^{\tau,m}(y)=\mathring{\kappa}_j^{2,m}-\langle\mathring{\eta}_j^{2,m}(t_{j+1}),y_{t_{j+1}}\rangle-\int_{t_j}^{t_{j+1}}\langle y_s,d\mathring{\eta}_j^{2,m}(s)\rangle,$$

so that, finally,

$$\mathring{\bar{\Xi}}_t^{\tau,m}(y)=\begin{cases}\mathring{\Xi}_t^{\tau,m}(y), & \text{if } m=1,2\\ \mathring{\Xi}_t^{\tau,2}(y)+\sum_{j=0}^{n-1}\Gamma_{m,(t_{j+1}-t_j)}(\mathring{M}_j^{\tau,m}(y)), & \text{if } m>2.\end{cases}$$

Moreover, with $\mathring{\mathcal{J}}_j^{c}(y)$ corresponding to Lemma 4 we can write for $y\in\mathcal{H}_\gamma$,

$$\begin{aligned}\mathring{\Xi}_t^{\tau,m}(y)=&\sum_{j=0}^{n-1}\mathring{\kappa}_j^{0,m}+\langle\mathring{\eta}_j^{0,m}(t_{j+1}),y_{t_{j+1}}\rangle-\langle h(\mathring{X}_{t_j}),y_{t_j}\rangle\\&-\sum_{c=0}^{m-1}\sum_{j=0}^{n-1}\mathring{\mathcal{J}}_j^{c}(y).\end{aligned}$$

In the same way we get, *mutatis mutandis*, the expression for $\mathring{\bar{\Xi}}_t^{\tau,m}(y)$ on $\mathcal{H}_\gamma$. Now, we denote by

$$(\check{\Omega},\check{\mathcal{F}},\check{P})=(\Omega\times\mathring{\Omega},\mathcal{F}\otimes\mathring{\mathcal{F}},\tilde{P}\otimes\mathring{P})$$

the product probability space. In the following we lift the processes $\mathring{\eta}$ and Y from the component spaces to the product space by writing $Y(\omega,\mathring{\omega})=Y(\omega)$ and $\mathring{\eta}_j^{c,c+1}(\omega,\mathring{\omega})=\mathring{\eta}_j^{c,c+1}(\mathring{\omega})$ for all $(\omega,\mathring{\omega})\in\check{\Omega}$.

Lemma 5 *Let c be a positive integer and let $j \in \{0,\dots,n\}$. Then there exists a nullset $N_0 \in \mathcal{F}$ such that the mapping $(\omega,\mathring{\omega}) \mapsto \mathring{\mathcal{J}}_j^c(Y_{[0,t]}(\omega))(\mathring{\omega})$ coincides on $(\Omega \setminus N_0)\times\mathring{\Omega}$ with an $\check{\mathcal{F}}$-measurable map.*

Proof Notice first that the set

$$N_0 = \{\omega \in \Omega\colon\ Y_{[0,t]}(\omega) \notin \mathcal{H}_\gamma\}$$

is clearly a member of $\mathcal{F}$ and we have that $\tilde{P}(N_0) = 0$. With N_0 so defined, the lemma follows from the definition and measurability of $(\omega,\mathring{\omega}) \mapsto \mathring{\mathcal{J}}_j^c(Y_{[0,t]}(\omega))(\mathring{\omega})$. □

Lemma 6 *Let c be a positive integer and $j \in \{0,\dots,n\}$. Then we have $\check{P}$-almost surely that*

$$\int_{t_j}^{t_{j+1}} \langle Y_s, \mathrm{d}\mathring{\eta}_j^{c,c+1}(s)\rangle = \mathring{\mathcal{J}}_j^c(Y_{[0,t]}).$$

Proof Note that we can assume without loss of generality that $d_Y = 1$ because the result follows componentwise. Then, let $K > 0$ and $T = \inf\{s \in [0,t]\colon |Y_s| \le K\}$ to define

$$Y_s^K = Y_s\mathbb{I}_{s\le T} + Y_T\mathbb{I}_{s>T}\ ; \qquad s \in [0,t].$$

Then Fubini's theorem and Lemma 5 imply that

$$\check{\mathrm{E}}\left[\left(\sum_{i=0}^{k-1} Y_{s_{i,j}}^K\left(\mathring{\eta}_j^{c,c+1}(s_{i+1,j}) - \mathring{\eta}_j^{c,c+1}(s_{i,j})\right) - \mathring{\mathcal{J}}_j^c(Y_{[0,t]}^K(\omega))\right)^2\right]$$
$$= \int_{\Omega\setminus N_0} \mathring{\mathrm{E}}[(\mathring{\mathcal{J}}_j^{c,k}(Y_{[0,t]}^K(\omega)) - \mathring{\mathcal{J}}_j^c(Y_{[0,t]}^K(\omega)))^2]\,\mathrm{d}\tilde{P}(\omega)$$

Now, since the function $s \mapsto Y_s^K(\omega)$ is continuous and $\mathring{\mathcal{J}}_j^c(Y_{[0,t]}^K(\omega))$ is a version of the integral $\int_{t_j}^{t_{j+1}} Y_s^K(\omega)\,\mathrm{d}\mathring{\eta}_j^{c,c+1}(s)$ we have for every $\omega \in \Omega \setminus N_0$ that

$$\lim_{k\to\infty} \mathring{\mathrm{E}}[(\mathring{\mathcal{J}}_j^{c,k}(Y_{[0,t]}^K(\omega)) - \mathring{\mathcal{J}}_j^c(Y_{[0,t]}^K(\omega)))^2] = 0.$$

Moreover, clearly,

$$\mathring{\mathrm{E}}[(\mathring{\mathcal{J}}_j^{c,k}(Y_{[0,t]}^K(\omega)) - \mathring{\mathcal{J}}_j^c(Y_{[0,t]}^K(\omega)))^2] \le 4K^2\mathring{\mathrm{E}}[\mathring{\eta}_t^2] < \infty$$

So that we can conclude by the dominated convergence theorem that

$$\lim_{k\to\infty} \check{\mathrm{E}}\left[\left(\sum_{i=0}^{k-1} Y_{s_{i,j}}^K\left(\eta_j^{c,c+1}(s_{i+1,j}) - \eta_j^{c,c+1}(s_{i,j})\right) - \mathring{\mathcal{J}}_j^c(Y_{[0,t]}^K(\omega))\right)^2\right]$$
$$= \int_{\Omega\setminus N_0} \lim_{k\to\infty} \mathring{\mathrm{E}}[(\mathring{\mathcal{J}}_j^{c,k}(Y_{[0,t]}^K(\omega)) - \mathring{\mathcal{J}}_j^c(Y_{[0,t]}^K(\omega)))^2]\,\mathrm{d}\tilde{P}(\omega) = 0$$

As K is arbitrary, the lemma is proved. □

Finally, we are ready to show the main result, Theorem 2. We restate it here again, in a slightly different manner which reflects the current line of argument.

Theorem 3 *The random variable* $F_\varphi^{\tau,m}(Y_{[0,t]})$ *is a version of* $\pi_t^{\tau,m}(\varphi)$.

Proof By the Kallianpur-Striebel formula it suffices to show that for all bounded and Borel measurable functions φ we have $\tilde{P}$-almost surely

$$\rho_t^{\tau,m}(\varphi) = G_\varphi^{\tau,m}(Y_{[0,t]}).$$

Furthermore, this is equivalent to showing that for all continuous and bounded functions $b\colon C([0,t];\mathrm{R}^{d_Y}) \to \mathrm{R}$ the equality

$$\tilde{\mathrm{E}}[\rho_t^{\tau,m}(\varphi)b(Y_{[0,t]})] = \tilde{\mathrm{E}}[G_\varphi^{\tau,m}(Y_{[0,t]})b(Y_{[0,t]})].$$

holds. As for the left-hand side we can write

$$\begin{aligned}
&\tilde{\mathrm{E}}[\rho_t^{\tau,m}(\varphi)b(Y_{[0,t]})] \\
&= \tilde{\mathrm{E}}[\varphi(X_t)Z_t^{\tau,m}b(Y_{[0,t]})] \\
&= \tilde{\mathrm{E}}[\varphi(X_t)\exp(\bar{\xi}_t^{\tau,m})b(Y_{[0,t]})] \\
&= \check{\mathrm{E}}[\varphi(\mathring{X}_t)\exp(\mathring{\bar{\xi}}_t^{\tau,m})b(Y_{[0,t]})] \\
&= \check{\mathrm{E}}[\varphi(\mathring{X}_t)\exp(\mathtt{IBP}(\mathring{\bar{\xi}}_t^{\tau,m}))b(Y_{[0,t]})]
\end{aligned}$$

where $\mathtt{IBP}(\mathring{\bar{\xi}}_t^{\tau,m})$ is given by the application of the integration by parts formula for semimartingales as

$$\begin{aligned}
\mathtt{IBP}(\mathring{\xi}_t^{\tau,m}) &= \sum_{j=0}^{n-1}\mathtt{IBP}(\mathring{\xi}_t^{\tau,m})(j) \\
&= \sum_{j=0}^{n-1}\{\mathring{\kappa}_j^{0,m} + \langle\mathring{\eta}_j^{0,m}(t_{j+1}), Y_{t_{j+1}}\rangle - \langle h(\mathring{X}_{t_j}), Y_{t_j}\rangle - \int_{t_j}^{t_{j+1}}\langle Y_s, \mathrm{d}\mathring{\eta}_j^{0,m}(s)\rangle\}
\end{aligned}$$

$$\mathtt{IBP}(\mathring{\mu}^{\tau,m})(j) = \mathring{\kappa}_j^{2,m} + \langle\mathring{\eta}_j^{2,m}(t_{j+1}), Y_{t_{j+1}}\rangle - \int_{t_j}^{t_{j+1}}\langle Y_s, \mathrm{d}\mathring{\eta}_j^{2,m}(s)\rangle$$

$$\mathtt{IBP}(\mathring{\bar{\xi}}_t^{\tau,m})(j) = \begin{cases}\mathtt{IBP}(\mathring{\xi}_t^{\tau,m})(j), & \text{if } m = 1,2 \\ \mathtt{IBP}(\mathring{\xi}_t^{\tau,m})(j) + \Gamma_{m,(t_{j+1}-t_j)}\left(\mathtt{IBP}(\mathring{\mu}^{\tau,m})(j)\right), & \text{if } m > 2\end{cases}.$$

And, on the other hand, the right-hand side is

$$
\begin{aligned}
&\tilde{\mathrm{E}}[G_{\varphi}^{\tau,m}(Y_{[0,t]})b(Y_{[0,t]})] \\
&= \tilde{\mathrm{E}}[\varphi(X_t)\mathcal{Z}_t^{\tau,m}(Y_{[0,t]})b(Y_{[0,t]})] \\
&= \tilde{\mathrm{E}}[\varphi(X_t)\exp(\bar{\Xi}_t^{\tau,m}(Y_{[0,t]}))b(Y_{[0,t]})] \\
&= \tilde{\mathrm{E}}[\mathring{\mathrm{E}}[\varphi(\mathring{X}_t)\exp(\mathring{\bar{\Xi}}_t^{\tau,m}(Y_{[0,t]}))]b(Y_{[0,t]})] \\
&= \check{\mathrm{E}}[\varphi(\mathring{X}_t)\exp(\mathring{\bar{\Xi}}_t^{\tau,m}(Y_{[0,t]}))b(Y_{[0,t]})],
\end{aligned}
$$

where the last equality follows from Fubini's theorem. As the representations coincide, the theorem is thus proved. □

Acknowledgements Part of this research was funded within the project STORM: Stochastics for Time-Space Risk Models, from the Research Council of Norway (RCN). Project number: 274410.

References

1. Bain, A., Crisan, D.: Fundamentals of stochastic filtering. Springer, Stochastic modelling and applied probability 60 (2009)
2. Clark, J. M. C.: The design of robust approximations to the stochastic differential equations of nonlinear filtering. In J.K. Skwirzynski, editor, Communication systems and random processes theory, volume 25 of Proc. 2nd NATO Advanced Study Inst. Ser. E, Appl. Sci., pages 721-734. Sijthoff & Noordhoff, Alphen aan den Rijn (1978)
3. Clark, J.M.C., Cameron, R.J.: The maximum rate of convergence of discrete approximations for stochastic differential equations. In Stochastic differential systems, Lecture Notes in Control and Information Science 25, pages 162–171, Springer, Berlin-New York (1980)
4. Clark, J., Crisan, D.: On a robust version of the integral representation formula of nonlinear filtering. In Probab. Theory Relat. Fields 133, 43âĂŞ56 (2005). https://doi.org/10.1007/s00440-004-0412-5
5. Crisan, D., Ortiz-Latorre, S.: A high order time discretization of the solution of the non-linear filtering problem. Stoch PDE: Anal Comp (2019) doi: 10.1007/s40072-019-00157-3
6. Crisan, D., Diehl, J., Friz, P. K., Oberhauser, H.: Robust filtering: Correlated noise and multidimensional observation. In Ann. Appl. Probab. 23 (2013), no. 5, 2139–2160. doi:10.1214/12-AAP896.
7. Davis, M.H.A.: Linear estimation and stochastic control. Chapman and Hall, London (1977)
8. Davis, M.H.A., Wellings, P.H.: Computational problems in nonlinear filtering. In Analysis and Optimization of Systems, Springer (1980)
9. Davis, M.H.A., Marcus S.I.: An Introduction to Nonlinear Filtering. In Hazewinkel M., Willems J.C. (eds) Stochastic Systems - The Mathematics of Filtering and Identification and Applications. NATO Advanced Study Institutes Series (Series C - Mathematical and Physical Sciences), vol 78. Springer, Dordrecht. doi: 10.1007/978-94-009-8546-9_4 (1981)
10. Davis, M.H.A.: A Pathwise Solution of the Equations of Nonlinear Filtering. In Theory of Probability and its Applications, 27:1, pages 167-175 (1982)
11. Davis, M.H.A., Spathopoulos, M. P.: Pathwise Nonlinear Filtering for Nondegenerate Diffusions with Noise Correlation. In SIAM Journal on Control and Optimization, 25:2, pages 260-278 (1987)
12. Segall, A., Davis, M.H.A., Kailath, T.: Nonlinear filtering with counting observations. In IEEE Transactions on Information Theory, volume 21, no. 2, pages 143-149. doi: 10.1109/TIT.1975.1055360. (1975)