Probab. Theory Relat. Fields 133, 43–56 (2005)
Digital Object Identifier (DOI) 10.1007/s00440-004-0412-5

J.M.C. Clark · D. Crisan

On a robust version of the integral representation formula of nonlinear filtering

Received: 14 August 2004 / Revised version: 16 November 2004 /
Published online: 10 February 2005 –

Abstract. The paper is concerned with completing "unfinished business" on a robust representation formula for the conditional expectation operator of nonlinear filtering. Such a formula, robust in the sense that its dependence on the process of observations is continuous, was stated in [2] without proof. The main purpose of this paper is to repair this deficiency.

The formula is "almost obvious" as it can be derived at a formal level by a process of integration-by-parts applied to the stochastic integrals that appear in the integral representation formula. However, the rigorous justification of the formula is quite subtle, as it hinges on a measurability argument the necessity of which is easy to miss at first glance. The continuity of the representation (but not its validity) was proved by Kushner [9] for a class of diffusions.

1. Introduction

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let (X, Y) be a system of partially observed random processes, where X is the unobserved component and Y is the observable one. Let $(\mathcal{X}_t)_{t\geq 0}$ and $(\mathcal{Y}_t)_{t\geq 0}$ be the filtrations generated by X and respectively Y,

$$\mathcal{X}_t = \sigma\left(X_s, 0 \leq s \leq t\right), \quad \mathcal{Y}_t = \sigma\left(Y_s, 0 \leq s \leq t\right).$$

The problem of stochastic filtering for the partially observed system (X, Y) comprises the construction of $\pi_t(F)$, the optimal mean square estimate of any $\mathcal{X}_t \vee \mathcal{Y}_t$-measurable random variable F, on the basis of the data collected by observing Y in the time interval $[0, t]$.[1] If we assume that F is square integrable, then

$$\pi_t(F) = E\left[F|\mathcal{Y}_t\right], P- \text{ almost surely,}$$

where $E\left[F|\mathcal{Y}_t\right]$ is the conditional expectation of F given $\mathcal{Y}_t$.

At such a level of generality this is all that can be said about the solution of the stochastic filtering problem. However, by adding additional assumptions on X, Y and F one can obtain further, more amenable, representations. To fix the

J.M.C. Clark: Department of Electrical and Electronic Engineering, Imperial College London, Exhibition Road London SW7 2BT, United Kingdom
e-mail: j.m.c.clark@imperial.ac.uk

D. Crisan: Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7 2BZ, United Kingdom. e-mail: d.crisan@imperial.ac.uk

[1] Here we follow the definition given in [11].

ideas we will assume that X and Y are continuous $\mathbb{R}^d$-valued, respectively $\mathbb{R}^m$-valued, processes. Further, throughout the paper, we will fix the parameter t and denote by $y.$ and $Y.$ the respective continuous path and path-valued random variables $(y(s); 0 \leq s \leq t)$ and $(Y_s; 0 \leq s \leq t)$.

The purpose of this paper is to obtain a *robust* representation of $\pi_t(F)$. Using the above notation, a robust representation of $\pi_t(F)$ (following [2]) is a representation of the form

$$\pi_t(F) = \hat{f}(Y.), \tag{1}$$

where $\hat{f} : C_{\mathbb{R}^m}[0,t] \to \mathbb{R}$ is a *continuous* function (with respect to the supremum norm on $C_{\mathbb{R}^m}[0,t]$) and (1) holds P-almost surely.

The need for this type of representation arises when the filtering framework is used to model and solve "real-life" problems. As explained in a substantial number of papers (cf [2], [3], [4], [5], [6], [9], [10]) the model Y chosen for the "real-life" observation process $\bar{Y}$ may not be a perfect one. However, as long as the distribution of $\bar{Y}.$ is close in a weak sense to that of $Y.$(and some integrability assumptions hold), the estimate $\hat{f}\left(\bar{Y}.\right)$ computed on the actual observation will still be reasonable, as $E[(F - \hat{f}(\bar{Y}.))^2]$ is well approximated by the idealized error $E[(F - \hat{f}(Y.))^2]$.

Even when Y and $\bar{Y}$ coincide, one is never able to obtain and exploit a continuous stream of data as modelled by the *continuous* path $Y.(\omega)$. Instead the observation arrives and is processed at discrete moments in time

$$0 = t_0 < t_1 < t_2 < \cdots < t_n = t.$$

However the continuous path $\hat{Y}.(\omega)$ obtained from the discrete observations $(Y_{t_i}(\omega))_{i=1}^n$ by linear interpolation is close to $Y.(\omega)$ (with respect to the supremum norm on $C_{\mathbb{R}^m}[0,t]$); hence, by the same argument, $\hat{f}(\hat{Y}.)$ will be a sensible approximation of $\pi_t(F)$.

Finally, if P_Y positively charges all open sets in $C_{\mathbb{R}^m}[0,t]$ (as is the case in what follows), a continuous $\hat{f}$ has the virtue of uniqueness: Though there are many measurable functions ϱ on $C_{\mathbb{R}^m}[0,t]$ for which $\varrho(Y.)$ is a version of $E[F|\mathcal{Y}_t]$, there is only one which is also continuous.

Our main results are presented in Theorems 1 and 2. A version of Theorem 1 has previously been proved by Kushner [9] in its application to diffusion processes. His argument relied on the weak approximation of diffusions by continuous time Markov chains. We present an alternative argument, which, being more direct, allows for greater generality.

In the next section we introduce the main results of the paper together with the required additional notation and assumptions.

2. Assumptions and statement of the main results

The rôle of the unobserved component X is essential in the statement of the filtering problem. However, for the purpose of the robustness results, it only enters the framework and the analysis as the generator of the filtration $(\mathcal{X}_t)_{t\geq 0}$.[2] As such,

[2] This observation was pointed out to us by an anonymous referee, to whom we are grateful.

unless otherwise stated, we will assume from now on that $(\mathcal{X}_t)_{t\geq 0}$ is an arbitrary filtration, not necessarily generated by a continuous process X. We will also assume that Y is given by

$$Y_s = \int_0^s A_r dr + W_s, \quad s \geq 0, \tag{2}$$

where W is a Brownian motion independent of $\bigvee_{s\geq 0} \mathcal{X}_s$ and $A = \left(A^i\right)_{i=1}^m$ is an $(\mathcal{X}_s)_{s\geq 0}$-adapted continuous semimartingale. We will denote by $A^m = \left(A^{m,i}\right)_{i=1}^d$ the martingale part of A (with $A_0^m = 0$) and $A^{fv} = \left(A^{fv,i}\right)_{i=1}^d$ is the finite variation part. Both A_s^m and A_s^{fv} are assumed to be continuous $(\mathcal{X}_s)_{s\geq 0}$-adapted processes satisfying, for all positive $k > 0$,

$$E\left[\exp\left(k\sum_{i=1}^d \int_0^t d\left\langle A^{m,i}\right\rangle_s\right)\right] < \infty, E\left[\exp\left(k\sum_{i=1}^d \int_0^t \left|dA_s^{fv,i}\right|\right)\right] < \infty. \tag{3}$$

Here $\langle A^m\rangle$ is the quadratic variation of A^m.

In order to obtain a representation of the form (1) we introduce a new probability measure $\tilde{P}$ absolutely continuous with respect to P (and vice versa) such that the Radon-Nikodym derivative of $\tilde{P}$ with respect to P is given by ($A_t^\top$ represents the row vector $\left(A_t^1, \ldots, A_t^m\right)$)

$$\left.\frac{d\tilde{P}}{dP}\right|_{\mathcal{X}_t \vee \mathcal{Y}_t} = \exp\left(-\int_0^t A_s^\top dW_s - \frac{1}{2}\int_0^t A_s^\top A_s ds\right). \tag{4}$$

Condition (3) implies that A satisfies the Novikov condition

$$E\left[\exp\left(\frac{1}{2}\int_0^t A_s^\top A_s dt\right)\right] < \infty;$$

hence, by Girsanov's theorem, Y is a Brownian motion under $\tilde{P}$, independent of X.

As stated in the introduction, F is an $\mathcal{X}_t \vee \mathcal{Y}_t$-measurable random variable. Hence $F(\omega) = \varphi(\omega, Y_\cdot(\omega))$ for some $\mathcal{X}_t \otimes \mathcal{B}(C_{\mathbb{R}^m}[0,t])$ -measurable function $\varphi : \Omega \times C_{\mathbb{R}^m}[0,t] \to \mathbb{R}$ ($\mathcal{B}(C_{\mathbb{R}^m}[0,t])$ is the Borel σ-field on $C_{\mathbb{R}^m}[0,t]$). We will assume that φ is locally Lipschitz in the second component. Further, we will assume that for any $R > 0$

$$K_R^{F,1}(\omega) \triangleq \sup_{\|y_\cdot^1\|,\|y_\cdot^2\|\leq R,\, y_\cdot^1\neq y_\cdot^2} \frac{\left|\varphi(\omega, y_\cdot^1) - \varphi(\omega, y_\cdot^2)\right|}{\left\|y_\cdot^1 - y_\cdot^2\right\|} < \infty, P\text{-a.s.} \tag{5}$$

and

$$K_R^{F,2}(\omega) \triangleq \sup_{\|y_\cdot\|\leq R} |\varphi(\omega, y_\cdot)| < \infty, P\text{-a.s.}. \tag{6}$$

In (5) and (6) $K_R^{F,1}$ and $K_R^{F,2}$ are $\mathcal{X}_t$-measurable random variables. Further we assume that they are square integrable and we denote by $M_R^{F,1}$, respectively, $M_R^{F,2}$, their L^2-norms:

$$M_R^{F,1} \triangleq E\left[\left(K_R^{F,1}\right)^2\right]^{\frac{1}{2}} < \infty, M_R^{F,2} \triangleq E\left[\left(K_R^{F,2}\right)^2\right]^{\frac{1}{2}} < \infty. \tag{7}$$

We also assume that $F$is square-integrable under the modified measure $\tilde{P}$:

$$\tilde{E}\left[F^2\right] < \infty. \tag{8}$$

However, if φ has a square-integrable *global* Lipschitz constant; that is,

$$M^{F,1} = \left(E\left[\left(\sup_{R>0} K_R^{F,1}\right)^2\right]\right)^{\frac{1}{2}} < \infty,$$

then (8) is automatically satisfied as

$$\begin{aligned}|\varphi(\omega, y.)| &\leq |\varphi(\omega, y.) - \varphi(\omega, \mathbf{0}.)| + |\varphi(\omega, \mathbf{0}.)| \\ &\leq \sup_{R>0} K_R^{F,1}(\omega) \left\|y_\cdot^1\right\| + K_1^{F,2}(\omega),\end{aligned}$$

where $\mathbf{0} : [0, t] \rightarrow \mathbb{R}$ is the constant path $\mathbf{0}(s) = 0$ for all $s \in [0, t]$ and

$$\tilde{E}\left[F^2\right]^{\frac{1}{2}} \leq M^{F,1} \tilde{E}\left[\left(\max_{s \in [0,t]} Y_s\right)^2\right]^{\frac{1}{2}} + M_1^{F,2} < \infty. \tag{9}$$

In (9), $\tilde{E}\left[\left(\max_{s\in[0,t]} Y_s\right)^2\right]$ is finite, since the running maximum of the Brownian motion Y has finite second moment and, as the change of measure does not affect the law of any $\mathcal{X}_t$-measurable random variable, in particular the laws of $\sup_{R>0} K_R^{F,1}$ and $K_1^{F,2}$,

$$\tilde{E}\left[\left(\sup_{R>0} K_R^{F,1}\right)^2\right] = E\left[\left(\sup_{R>0} K_R^{F,1}\right)^2\right] = \left(M^{F,1}\right)^2$$
$$\tilde{E}\left[\left(K_1^{F,2}\right)^2\right] = \tilde{E}\left[\left(K_1^{F,2}\right)^2\right] = \left(M_1^{F,2}\right)^2.$$

Note that if F is independent of Y, that is, F is just an $\mathcal{X}_t$-measurable random variable, then conditions (5) and (6) are trivially satisfied with $K_R^{F,1} \equiv 0$ and $K_R^{F,2} \equiv F$. Also, since $\tilde{E}\left[F^2\right] = E\left[F^2\right]$, condition (7) is satisfied as well (we assumed in the introduction that F is square integrable with respect to the original probability measure P). In other words, in the language of filtering theory, if F is a function of the unobservable component only, *no additional conditions are required.*

The following representation is due, in its full generality, to Kallianpur and Striebel [7], though special versions occur in the papers of Bucy [1], Kushner [8], Wonham [13] and Zakai [14]:

$$\pi_t(F) = \frac{\rho_t(F)}{\rho_t(1)}, \tag{10}$$

where

$$\rho_t(F) = \tilde{E}\left[F \exp\left(\int_0^t A_s^\top dY_s - \frac{1}{2}\int_0^t A_s^\top A_s ds \right) \middle| \mathcal{Y}_t \right].$$

Now let $\Theta(y.)$ be the following random variable

$$\Theta(y.) \triangleq \exp\left(A_t^\top y_t - I(y.) - \frac{1}{2}\int_0^t A_s^\top A_s ds \right), \tag{11}$$

where $I(y.)$ is a "well behaved" version of the stochastic integral $\int_0^t y_s^\top dA_s$, the definition of which will be specified at the end of the section. The exponent of $\Theta(y.)$ will be recognized as a formal integration-by-parts of the exponent in the version of the Kallianpur-Striebel formula given above.

Finally, let $\hat{g}^\varphi, \hat{g}^{\mathbf{1}}, \hat{f} : C_{\mathbb{R}^m}[0,t] \to \mathbb{R}$ be the following functions

$$\hat{g}^\varphi(y.) = \tilde{E}[\varphi(\iota, y.)\Theta(y.)], \hat{g}^{\mathbf{1}}(y.) = \tilde{E}[\Theta(y.)], \hat{f}(y.) = \frac{\hat{g}^\varphi(y.)}{\hat{g}^{\mathbf{1}}(y.)},$$

where ι is the identity function on Ω, $\iota(\omega) = \omega$. The function $\hat{f}$ is our candidate for the robust form of $\pi_t(F)$. More precisely:

Theorem 1. *The function $\hat{f}$ is locally Lipschitz. In other words, for any $R > 0$, there exists a constant K_R such that*

$$\left| \hat{f}\left(y_\cdot^1\right) - \hat{f}\left(y_\cdot^2\right) \right| \leq K_R \left|\left| y_\cdot^1 - y_\cdot^2 \right|\right|$$

for any two paths $y_\cdot^1, y_\cdot^2$ such that $||y_\cdot^1||, ||y_\cdot^2|| \leq R$.

Theorem 2. *The random variable $\hat{f}(Y.)$ is a version of $\pi_t(F)$; that is $\pi_t(F) = \hat{f}(Y.)$, P-almost surely. Hence $\hat{f}(Y.)$ is the unique robust representation of $\pi_t(F)$.*

The choice of the version of $\int_0^t y_s^\top dA_s$ used in defining $\hat{f}$ is irrelevant in the proof of the continuity of $\hat{f}$, but it is important in justifying that $\hat{f}(Y.)$ is a version of $\pi_t(F)$.

Some form of condition on the continuity of the function $(\omega, y.) \to \varphi(\omega, y.)$ in the second component appears to be required for Theorem 1 to hold. For instance, if φ is the function $\varphi(\omega, y.) = 1_{\{y_t \geq 0\}}$, then there is no continuous $\hat{f}$ for which $\hat{f}$ $(Y.)$ is a version of $E[F|\mathcal{Y}_t]$.

However, there are many tantalizing examples that show that the continuity is not strictly necessary. For example, take X to be a constant process $X_s = X_0$, for all $s \geq 0$, where X_0 is a Gaussian variable of zero mean and unit variance and let

$A_s = X_0$. Then $Y_s = sX_0 + W_s$. It is easily shown that the law of X_0 conditioned on $\mathcal{Y}_1$ can be expressed as a Gaussian law with mean $\frac{1}{2}Y_1$ and variance $\frac{1}{2}$. The probability $\rho(a, y_1) = P(X_0 > a | Y_1 = y_1)$ is jointly continuous in (a, y_1). Then $\rho(Y_1, Y_1)$ is a robust version of $P(X_0 > Y_1 | \mathcal{Y}_1)$.

The theorems can be extended to $(\mathcal{X}_s \vee \mathcal{Y}_s)_{s\geq 0}$-adapted semimartingales A. For example, let X be a semimartingale and assume that there exists a continuously differentiable function $\Phi : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ and a Borel measurable function $\beta = \left(\beta^i\right)_{i=1}^m : \mathbb{R}^m \to \mathbb{R}^m$ such that $A = \left(A^i\right)_{i=1}^m$ and

$$A_s^i = \alpha^i(X_s, Y_s) + \beta^i(Y_s),$$

where

$$\alpha^i(x_1, \ldots, x_d, y_1, \ldots, y_m) = \frac{d\Phi}{dy_i}(x_1, \ldots, x_d, y_1, \ldots, y_m).$$

Let F be an $\mathcal{X}_t \vee \mathcal{Y}_t$-measurable random variable ($(\mathcal{X}_s)_{s\geq 0}$ is the filtration generated by X), hence $F = \varphi(X_\cdot, Y_\cdot)$ for some Borel measurable function $\varphi : C_{\mathbb{R}^d}[0,t] \times C_{\mathbb{R}^m}[0,t] \to \mathbb{R}$. In this case the unique robust representation of $\pi_t(F)$ has the form

$$\hat{f}(y_\cdot) = \frac{\tilde{E}\left[\varphi(X_\cdot, y_\cdot)\,\Theta(y_\cdot)\right]}{\tilde{E}\left[\Theta(y_\cdot)\right]},$$

where

$$\Theta(y_\cdot) \triangleq \exp\left(\Phi(X_t, y_t) - \sum_{i=1}^d \int_0^t \frac{d\Phi}{dx_i}(X_s, y_s)\, dX_s^i - \frac{1}{2}\int_0^t A_s^\top A_s ds\right).$$

Note that the representation is independent of β as the terms corresponding to β cancel out due to the normalization.

Finally let us remark that, following the rough paths theory of Lyons [12], there will be no robustness result for general $\mathcal{X}_s \vee \mathcal{Y}_s$-adapted semimartingales A. We illustrate this by means of the following simple example: Let X be the constant process

$$X_s = X_0 = \begin{cases} 1 \text{ with probability} \frac{1}{2} \\ 0 \text{ with probability} \frac{1}{2} \end{cases}.$$

Again, let $(\mathcal{X}_s)_{s\geq 0}$ be the filtration generated by X and let $Y = \left(Y^1, Y^2\right)^\top$ be the solution of the equation

$$\begin{aligned} dY_s^1 &= X_s Y_s^2 ds + dW_s^1 \\ dY_s^2 &= dW_s^2 \end{aligned}$$

and choose $F = X_t$. Then

$$\pi_t(F) = \left(1 + e^{-\int_0^t Y_s^2 dY_s^1 + \frac{1}{2}\int_0^t \left(Y_s^2\right)^2 ds}\right)^{-1},$$

which is not a continuous functional over $C_{\mathbb{R}^2}[0,t]$ due to the discontinuity of the stochastic integral $\int_0^t Y_s^2 dY_s^1$.

We specify now the version of the stochastic integral $\int_0^t y_s^\top dA_s$ that we choose in order to define $I(y.)$. Denote by $I^{fv}(y.)$ the Stieltjes integral with respect to A^{fv}. $I^{fv}(y.)$ is defined unambiguously pathwise. More precisely for arbitrary $y. \in C_{\mathbb{R}^m}[0,t]$ and all $\omega \in \Omega$

$$I^{fv}(y.)(\omega) = \lim_{n\to\infty}\left(\sum_{i=0}^{n-1} y_{\frac{it}{n}}^\top\left(A^{fv}_{\frac{(i+1)t}{n}}(\omega) - A^{fv}_{\frac{it}{n}}(\omega)\right)\right).$$

Hence defining $I(y.)$ only depends on selecting the version of $\int_0^t y_s^\top dA_s^m$, the stochastic integral with respect to the martingale part of A, which we will denote by $I^m(y.)$. Let $H_{\frac{1}{3}}$ be the following subset of $C_{\mathbb{R}^m}[0,t]$

$$H_{\frac{1}{3}} = \left\{ y. \in C_{\mathbb{R}^m}[0,t] \,\middle|\, K(y.) \triangleq \sup_{s_1,s_2\in[0,t]} \frac{\left|\left|y_{s_1} - y_{s_2}\right|\right|}{|s_1 - s_2|^{\frac{1}{3}}} < \infty\,, \right\},$$

where $\left|\left|y_{s_1} - y_{s_2}\right|\right|$ is the Euclidean distance between the two points $y_{s_1}, y_{s_2} \in \mathbb{R}^m$ and let

$$I_n^m(y.) \triangleq \sum_{i=0}^{n-1} y_{\frac{it}{n}}^\top\left(A^m_{\frac{(i+1)t}{n}}(\hat{\omega}) - A^m_{\frac{it}{n}}(\hat{\omega})\right).$$

Since, for $y. \in H_{\frac{1}{3}}$,

$$\begin{aligned}\tilde{E}\left[\left(I^m_{2^k}(y.) - \int_0^t y_s^\top dA_s^m\right)^2\right] &= \tilde{E}\left[\left(\sum_{i=1}^d \int_0^t \left(y^i(s) - y^i\left(\left\lceil\frac{s2^k}{t}\right\rceil\frac{t}{2^k}\right)\right) dA_s^{m,i}\right)^2\right] \\ &\le d\sum_{i=1}^d \tilde{E}\left[\int_0^t \left(y^i(s) - y^i\left(\left\lceil\frac{s2^k}{t}\right\rceil\frac{t}{2^k}\right)\right)^2 d\left\langle A^{m,i}\right\rangle_s\right] \\ &\le \frac{dK(y.)^2}{2^{\frac{2k}{3}}}\sum_{i=1}^d \tilde{E}\left[\left(A_t^{m,i}\right)^2\right],\end{aligned}$$

it follows that, for $y \in H_{\frac{1}{3}}$, $I^m_{2^k}(y.)$ converges to $\int_0^t y_s^\top dA_s^m$, $\tilde{P}$ -almost surely. We define $I^m(y.)$ to be the limit

$$I^m(y.)(\omega) = \limsup_{k\to\infty} I^m_{2^k}(y.)(\omega)$$

on the set $H_{\frac{1}{3}} \times \Omega$ and any version of $\int_0^t y_s^\top dA_s^m$ on $C_{\mathbb{R}^m}[0,t]\backslash H_{\frac{1}{3}} \times \Omega$. Though the resulting map is generally non-measurable with respect to $\mathcal{B}(C_{\mathbb{R}^m}[0,t]) \bigotimes \mathcal{F}$, where $\mathcal{B}(C_{\mathbb{R}^m}[0,t])$ is the Borel σ -field on $C_{\mathbb{R}^m}[0,t]$, it is equal on $H_{\frac{1}{3}} \times \Omega$ to the following jointly measurable function

$$J^m(y.) \triangleq \limsup_{k\to\infty} I^m_{2^k}(y.) \tag{12}$$

defined on the whole of $C_{\mathbb{R}^m}[0,t] \times \Omega$.

3. Proofs of the main results

The proof of Theorem 1 uses the following two lemmas

Lemma 3. *For any $R > 0$ and $p \geq 1$ there exists a positive constant $M^{\Theta}_{R,p}$ such that*

$$\sup_{||y.|| \leq R} E\left[\Theta\left(y.\right)^p\right]^{\frac{1}{p}} \leq M^{\Theta}_{R,p}. \tag{13}$$

Also, for any $R > 0$ there exists a constant M^{Θ}_R such that

$$\sqrt{E\left[\left(\Theta\left(y^1_.\right) - \Theta\left(y^2_.\right)\right)^2\right]} \leq M^{\Theta}_R \left|\left|y^1_. - y^2_.\right|\right| \tag{14}$$

for any two paths $y^1_., y^2_.$ such that $\left|\left|y^1_.\right|\right|, \left|\left|y^2_.\right|\right| \leq R$. In particular, (14) implies that $\hat{g}^{\mathbf{1}}$ is locally Lipschitz; more precisely

$$\left|\hat{g}^{\mathbf{1}}\left(y^1_.\right) - \hat{g}^{\mathbf{1}}\left(y^2_.\right)\right| \leq M^{\Theta}_R \left|\left|y^1_. - y^2_.\right|\right|$$

for any two paths $y^1_., y^2_.$ such that $\left|\left|y^1_.\right|\right|, \left|\left|y^2_.\right|\right| \leq R$.

Proof. We have first that

$$\Theta\left(y.\right)^p \leq \exp\left(2pR\sum_{i=1}^{d}\int_0^t \left|dA^{fv,i}_s\right| + 4R^2p^2\sum_{i,j=1}^{d}\int_0^t d\left|\left\langle A^{m,i}, A^{m,j}\right\rangle_s\right|\right)\Theta'_t\left(y.\right),$$

where

$$\begin{aligned}\Theta'_r\left(y.\right) \triangleq \exp\Bigg(&\int_0^r p\left(y_t - y_s\right)^{\top} dA^m_s \\ &- \sum_{i,j=1}^{d}\int_0^r p^2\left(y^i_t - y^i_s\right)\left(y^j_t - y^j_s\right) d\left\langle A^{m,i}, A^{m,j}\right\rangle_s\Bigg).\end{aligned}$$

We note now that the process $r \longrightarrow \left(\Theta'_r\left(y.\right)\right)^2$ is a martingale; so

$$E\left[\left(\Theta'_r\left(y.\right)\right)^2\right] = 1.$$

Hence, by the Cauchy-Schwarz inequality and the fact that

$$\int_0^t d\left|\left\langle A^{m,i}, A^{m,j}\right\rangle_s\right| \leq \frac{1}{2}\int_0^t d\left\langle A^{m,i}\right\rangle_s + \frac{1}{2}\int_0^t d\left\langle A^{m,j}\right\rangle_s,$$

(13) follows with

$$M^{\Theta}_{R,p} \triangleq E\left[\exp\left(2pR\sum_{i=1}^{d}\int_0^t \left|dA^{fv,i}_s\right| + 4R^2p^2d\sum_{i=1}^{d}\int_0^t d\left\langle A^{m,i}\right\rangle_s\right)\right]^{\frac{1}{2p}}.$$

Note that $M^{\Theta}_{R,p}$is finite by virtue of condition (3).

Now if two paths $y^1_\cdot$, $y^2_\cdot$ are such that $||y^1_\cdot||, ||y^2_\cdot|| \leq R$, then

$$\left|\Theta\left(y^1_\cdot\right) - \Theta\left(y^2_\cdot\right)\right| \leq \Theta\left(y^1_\cdot\right)\left|\int_0^t \left(y^{12}_t - y^{12}_s\right)^\top dA_s\right|,$$

where $y^{12}_\cdot$ is the difference path $y^{12}_\cdot \triangleq y^1_\cdot - y^2_\cdot$. Then, using again the Cauchy-Schwarz inequality

$$\sqrt{E\left[\left(\Theta\left(y^1_\cdot\right) - \Theta\left(y^2_\cdot\right)\right)^2\right]} \leq \left(M^{\Theta}_{R,4}\right)\sqrt{E\left[\left(\int_0^t \left(y^{12}_t - y^{12}_s\right)^\top dA_s\right)^4\right]}. \quad (15)$$

Finally, since $||y^{12}_\cdot|| = ||y^1_\cdot - y^2_\cdot|| \leq 2R$, a standard argument based on Doob's maximal inequality shows that the expectation on the right hand side of (15) is bounded by a constant M'_R. Hence (14) holds true with $M^{\Theta}_R = M^{\Theta}_{R,4}\sqrt{M'_R}$. □

Lemma 4. *The function $\hat{g}^{\varphi}$ is locally Lipschitz and locally bounded.*

Proof. We have first

$$\tilde{E}\left[\left|\varphi\left(\iota, y^1_\cdot\right) - \varphi\left(\iota, y^2_\cdot\right)\right|\Theta\left(y^1_\cdot\right)\right] \leq M^F_R\left|\left|y^1_\cdot - y^2_\cdot\right|\right|, \quad (16)$$

where

$$M^F_R = \tilde{E}\left[K^{F,1}_R\Theta\left(y^1_\cdot\right)\right] \leq M^{F,1}_R M^{\Theta}_{R,2}$$

and M^F_R is finite from (7) and (13). Then, using (7) and (14), we have that

$$\tilde{E}\left[\left|\varphi\left(\iota, y^2_\cdot\right)\right|\left|\Theta\left(y^1_\cdot\right) - \Theta\left(y^2_\cdot\right)\right|\right] \leq M^{F,2}_R M^{\Theta}_R\left|\left|y^1_\cdot - y^2_\cdot\right|\right|. \quad (17)$$

The Lipschitz property of $\hat{g}^{\varphi}$ then follows from (16), (17) and the identity

$$\begin{aligned}\hat{g}^{\varphi}\left(y^1_\cdot\right) - \hat{g}^{\varphi}\left(y^2_\cdot\right) = \tilde{E}&\left[\left(\varphi\left(\iota, y^1_\cdot\right) - \varphi\left(\iota, y^2_\cdot\right)\right)\Theta\left(y^1_\cdot\right)\right] \\ &+\tilde{E}\left[\varphi\left(\iota, y^2_\cdot\right)\left(\Theta\left(y^1_\cdot\right) - \Theta\left(y^2_\cdot\right)\right)\right].\end{aligned}$$

Again, from (7) and (13)

$$\sup_{||y_\cdot|| \leq R}\left|\hat{g}^{\varphi}\left(y_\cdot\right)\right| = \sup_{||y_\cdot|| \leq R}\left|\tilde{E}\left[\varphi\left(\iota, y_\cdot\right)\Theta\left(y^1_\cdot\right)\right]\right| \leq M^{F,2}_R M^{\Theta}_{R,2}.$$

Hence $\hat{g}^{\varphi}$ is locally bounded. □

Proof of Theorem 1. The ratio $\frac{\hat{g}^{\varphi}}{\hat{g}^{\mathbf{1}}}$ of the two locally Lipschitz functions $\hat{g}^{\varphi}$ and $\hat{g}^{\mathbf{1}}$ (Lemma 3 and Lemma 4) is locally Lipschitz provided both $\hat{g}^{\varphi}$ and $\frac{1}{\hat{g}_t^{\mathbf{1}}}$ are locally bounded. The local boundedness property of $\hat{g}^{\varphi}$ is shown in Lemma 4 and that of $\frac{1}{\hat{g}_t^{\mathbf{1}}}$ follows from the following simple argument: If $||y.|| \leq R$, Jensen's inequality implies that

$$\begin{aligned}\tilde{E}\left[\Theta\left(y.\right)\right] &\geq \exp\left(E\left[\int_0^t \left(y_t - y_s\right)^\top dA_s^m + \int_0^t \left(y_t - y_s\right)^\top dA_s^{fv}\right.\right.\\ &\qquad\left.\left. -\frac{1}{2}\int_0^t A_s^\top A_s ds\right]\right)\\ &\geq \exp\left(-2R\sum_{i=1}^d E\left[\int_0^t \left|dA_s^{fv,i}\right|\right] - \frac{1}{2}E\left[\int_0^t A_s^\top A_s ds\right]\right). \quad (18)\end{aligned}$$

Note that both expectations in (18) are finite, again by virtue of condition (3). □

We turn now to the proof of Theorem 2. First let us remark that it suffices to prove that, P-almost surely (or, equivalently, $\tilde{P}$ -almost surely),

$$\rho_t\left(F\right) = \hat{g}^{\varphi}\left(Y.\right) \quad \text{and} \quad \rho_t\left(\mathbf{1}\right) = \hat{g}^{\mathbf{1}}\left(Y.\right).$$

We need only prove the first identity as the second is just a special case. For this, it is useful to 'decouple' the two filtrations $(\mathcal{X}_s)_{s\geq 0}$ and $(\mathcal{Y}_s)_{s\geq 0}$. Let $(\hat{\Omega}, \mathcal{F}, \tilde{P})$ be an identical copy of $(\Omega, \mathcal{F}, \tilde{P})$ and let $(\hat{\mathcal{X}}_s)_{s\geq 0}$ be the copy of $(\mathcal{X}_s)_{s\geq 0}$ within the new space $(\hat{\Omega}, \mathcal{F}, \tilde{P})$. Let $\hat{A}$, $\hat{A}^m$ and $\hat{A}^{fv}$ be the processes within the new space $(\hat{\Omega}, \mathcal{F}, \tilde{P})$ corresponding to the original A, A^m and A^{fv}. Then the function $\hat{g}^{\varphi}$ has the following representation

$$\hat{g}^{\varphi}\left(y.\right) = \hat{E}\left[\varphi\left(\hat{\iota}, y.\right)\exp\left(\hat{A}_t^\top y_t - \hat{I}\left(y.\right) - \frac{1}{2}\int_0^t \hat{A}_s^\top \hat{A}_s ds\right)\right], \quad (19)$$

where $\hat{E}\left[\cdot\right]$ denotes integration on $(\hat{\Omega}, \mathcal{F}, \tilde{P})$, $\hat{\iota}$ is the identity function on $\hat{\Omega}$ and $\hat{I}\left(y.\right)$ is the version of the stochastic integral $\int_0^t y_s^\top d\hat{A}_s$ corresponding to $I\left(y.\right)$. Denote by $\hat{I}^m\left(y.\right)$ the respective version of the stochastic integral with respect to the martingale $\hat{A}^m$and by $\hat{I}^{fv}\left(y.\right)$ the Stieltjes integral with respect to $\hat{A}^{fv}$. Let $\hat{J}^m\left(y.\right)$ be the function corresponding to $J^m\left(y.\right)$ as defined in (12). Then, for $y. \in H_{\frac{1}{3}}$, (19) can be written as

$$\hat{g}^{\varphi}\left(y.\right) = \hat{E}\left[\varphi\left(\hat{\iota}, y.\right)\exp\left(\hat{A}_t^\top y_t - \hat{I}^{fv}\left(y.\right) - \hat{J}^m\left(y.\right) - \frac{1}{2}\int_0^t \hat{A}_s^\top \hat{A}_s ds\right)\right]. \quad (20)$$

Finally, let $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ be the product space

$$(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P}) = (\Omega \times \hat{\Omega}, \mathcal{F}\bigotimes\mathcal{F}, \tilde{P}\bigotimes\tilde{P})$$

on which we 'lift' the processes $\hat{A}$ and Y from the component spaces. In other words, $Y\left(\omega, \hat{\omega}\right) = Y\left(\omega\right)$ and $\hat{A}\left(\omega, \hat{\omega}\right) = \hat{A}\left(\hat{\omega}\right)$for all $\left(\omega, \hat{\omega}\right) \in \Omega \times \hat{\Omega}$.

Lemma 5. *There exists a null set* $\mathcal{N} \in \mathcal{F}$ *such that mapping* $(\omega, \hat{\omega}) \in \bar{\Omega} \longrightarrow \hat{I}(Y(\omega))(\hat{\omega})$ *coincides on* $(\Omega \backslash \mathcal{N}) \times \hat{\Omega}$ *with an* $\bar{\mathcal{F}}$*-measurable mapping.*

Proof. First let us remark that $(\omega, \hat{\omega}) \longrightarrow \hat{I}^{fv}(Y(\omega))(\hat{\omega})$ is equal to

$$\hat{I}^{fv}(Y_{\cdot}(\omega))(\hat{\omega}) = \lim_{n\to\infty}\left(\sum_{i=0}^{n-1} Y_{\frac{it}{n}}^{\top}(\omega)\left(\hat{A}_{\frac{it}{n}}^{fv}(\hat{\omega}) - \hat{A}_{\frac{(i-1)t}{n}}^{fv}(\hat{\omega})\right)\right) \tag{21}$$

and since

$$(\omega, \hat{\omega}) \in \bar{\Omega} \longrightarrow \sum_{i=0}^{n-1} Y_{\frac{it}{n}}^{\top}(\omega)\left(\hat{A}_{\frac{it}{n}}^{fv}(\hat{\omega}) - \hat{A}_{\frac{(i-1)t}{n}}^{fv}(\hat{\omega})\right)$$

is $\bar{\mathcal{F}}$-measurable so is its limit. Define $\mathcal{N} \triangleq \left\{\omega \in \Omega | Y_{\cdot}(\omega) \notin H_{\frac{1}{3}}\right\}$. Then $\mathcal{N} \in \mathcal{F}$ and $\tilde{P}(\mathcal{N}) = 0$. Following the definition of $I^m(y_{\cdot})$, the mapping $(\omega, \hat{\omega}) \longrightarrow \hat{I}^m(Y(\omega))(\hat{\omega})$ coincides with the mapping $(\omega, \hat{\omega}) \longrightarrow \hat{J}^m(Y(\omega))(\hat{\omega})$ on $(\Omega \backslash \mathcal{N}) \times \hat{\Omega}$. $\hat{J}^m$ is an $\bar{\mathcal{F}}$ -measurable random variable, since

$$\hat{J}^m(Y(\omega))(\hat{\omega}) = \limsup_{k\to\infty}\left(\sum_{i=0}^{n_k-1} Y_{\frac{i}{n_k}t}^{\top}(\omega)\left(\hat{A}_{\frac{(i+1)t}{n_k}}^{m}(\hat{\omega}) - \hat{A}_{\frac{it}{n_k}}^{m}(\hat{\omega})\right)\right) \tag{22}$$

Combining this with the measurability of $\hat{I}^{fv}(Y_{\cdot})$ gives us the lemma. □

Lemma 6. $\bar{P}$ –almost surely,

$$\int_0^t Y_s^{\top} d\hat{A}_s = \hat{I}^{fv}(Y_{\cdot}) + \hat{J}^m(Y_{\cdot}). \tag{23}$$

Proof. We have

$$\int_0^t Y_s^{\top} d\hat{A}_s = \int_0^t Y_s^{\top} d\hat{A}_t^m + \int_0^t Y_s^{\top} d\hat{A}_t^{fv}.$$

Following (21) it is obvious that $\int_0^t Y_s^{\top} d\hat{A}_t^{fv} = \hat{I}^{fv}(Y_{\cdot})$. Hence, following the proof of the previous lemma, it suffices to prove that, $\bar{P}$ –almost surely, $\int_0^t Y_s^{\top} d\hat{A}_t^m = \hat{J}^m(Y_{\cdot})$ where $\hat{J}^m(Y_{\cdot})$ is the function defined in (22). Without loss of generality we will assume that $m = 1$ (the general case follows by treating each of the m components in turn) and we note that we only need to prove that, for arbitrary $K > 0$, $\bar{P}$ –almost surely,

$$\int_0^t Y_s^K d\hat{A}_t^m = \hat{J}^m\left(Y_{\cdot}^K\right), \tag{24}$$

where $Y_s^K = \operatorname{sgn}(Y_s)\min(|Y_s|, K), s \geq 0$. In turn, (24) follows from the proof of

$$\lim_{n\to\infty} \bar{E}\left[\left(\sum_{i=0}^{n-1}\left(Y_{\frac{it}{n}}^K\right)^{\top}\left(\hat{A}_{\frac{(i+1)t}{n}}^m - \hat{A}_{\frac{it}{n}}^m\right) - \hat{J}^m\left(Y_{\cdot}^K\right)\right)^2\right] = 0.$$

By Fubini, we have that (using the $\bar{\mathcal{F}}$-measurability of $\hat{J}^m\left(Y_\cdot^K\right)$ and the fact that $\hat{I}^m\left(Y_\cdot^K\right)$ coincide with $\hat{J}^m\left(Y_\cdot^K\right)$ on $(\Omega\backslash\mathcal{N})\times\hat{\Omega}$)

$$\begin{aligned}\bar{E}&\left[\left(\sum_{i=0}^{n-1}\left(Y_{\frac{it}{n}}^K\right)^\top\left(\hat{A}^m_{\frac{(i+1)t}{n}}-\hat{A}^m_{\frac{it}{n}}\right)-\hat{J}^m\left(Y_\cdot^K\right)\right)^2\right]\\ &=\int_{\Omega\backslash\mathcal{N}}\hat{E}\left[\left(\hat{I}_n^m\left(Y_\cdot^K(\omega)\right)-\hat{J}^m\left(Y_\cdot^K\right)\right)^2\right]d\tilde{P}(\omega)\\ &=\int_{\Omega\backslash\mathcal{N}}\hat{E}\left[\left(\hat{I}_n^m\left(Y_\cdot^K(\omega)\right)-\hat{I}^m\left(Y_\cdot^K\right)\right)^2\right]d\tilde{P}(\omega)\end{aligned}$$

Now since $s\to Y_s^K(\omega)$ is a continuous function and $\hat{I}^m\left(Y_\cdot^K(\omega)\right)$ is a version of the stochastic integral $\int_0^t\left(Y_s^K\right)^\top(\omega)\,d\hat{A}_s^m$, it follows that

$$\lim_{n\to\infty}\hat{E}\left[\left(\hat{I}_n^m\left(Y_\cdot^K(\omega)\right)-\hat{I}^m\left(Y_\cdot^K(\omega)\right)\right)^2\right]=0$$

for all $\omega\in\Omega\backslash\mathcal{N}$. Also, we have the following upper bound

$$\hat{E}\left[\left(\hat{I}_n^m\left(Y_\cdot^K(\omega)\right)-\hat{I}^m\left(Y_\cdot^K(\omega)\right)\right)^2\right]\le 4K^2\hat{E}\left[\left(\hat{A}_t^m\right)^2\right]<\infty.$$

Hence, by the Dominated Convergence Theorem,

$$\begin{aligned}\lim_{n\to\infty}\bar{E}&\left[\left(\sum_{i=0}^{n-1}\left(Y_{\frac{it}{n}}^K\right)^\top\left(\hat{A}^m_{\frac{(i+1)t}{n}}-\hat{A}^m_{\frac{it}{n}}\right)-\hat{I}^m\left(Y_\cdot^K\right)\right)^2\right]\\ &=\int_{\Omega\backslash\mathcal{N}}\lim_{n\to\infty}\hat{E}\left[\left(\hat{I}_n^m\left(Y_\cdot^K(\omega)\right)-\hat{I}^m\left(Y_\cdot^K(\omega)\right)\right)^2\right]d\tilde{P}(\omega)=0.\end{aligned}$$

□

Proof of Theorem 2. To prove Theorem 2 it suffices to show that

$$\tilde{E}\left[\rho_t(F)\,\Upsilon(Y_\cdot)\right]=\tilde{E}\left[\hat{g}^\varphi(Y_\cdot)\,\Upsilon(Y_\cdot)\right],\tag{25}$$

where Υ is an arbitrary continuous bounded function $\Upsilon : C_{\mathbb{R}^m}[0,t]\to\mathbb{R}$. Since A and Y are independent under $\tilde{P}$, it follows that the pair processes (A,Y) (under $\tilde{P}$) and $(\hat{A},Y)$ (under $\bar{P}$) have the same distribution. Hence, the left hand side of (25) has the following representation

$$\begin{aligned}\tilde{E}\left[\rho_t(F)\,\Upsilon(Y_\cdot)\right]&=\tilde{E}\left[\varphi(\iota,Y_\cdot)\exp\left(\int_0^t A_s^\top dY_s-\frac{1}{2}\int_0^t A_s^\top A_s ds\right)\Upsilon(Y_\cdot)\right]\\ &=\bar{E}\left[\varphi\left(\hat{\iota},Y_\cdot\right)\exp\left(\int_0^t \hat{A}_s^\top dY_s-\frac{1}{2}\int_0^t \hat{A}_s^\top\hat{A}_s ds\right)\Upsilon(Y_\cdot)\right]\\ &=\bar{E}\left[\varphi\left(\hat{\iota}_\cdot,Y_\cdot\right)\exp\left(\hat{A}_t^\top Y_t-\int_0^t Y_s^\top d\hat{A}_s-\frac{1}{2}\int_0^t \hat{A}_s^\top\hat{A}_s ds\right)\Upsilon(Y_\cdot)\right]\end{aligned}$$

On the other hand, using (20), the right hand side of (25) has the representation

$$
\begin{aligned}
&\tilde{E}\left[\hat{g}^{\varphi}(Y.)\Upsilon(Y.)\right] \\
&= \tilde{E}\left[\hat{E}\left[\varphi\left(\hat{\iota}, Y.\right)\exp\left(\hat{A}_t^{\top}Y_t - \hat{I}^{fv}(Y.) - \hat{I}^{m}(Y.) - \frac{1}{2}\int_0^t \hat{A}_s^{\top}\hat{A}_s ds\right)\right]\Upsilon(Y.)\right] \\
&= \tilde{E}\left[\hat{E}\left[\varphi\left(\hat{\iota}., Y.\right)\exp\left(\hat{A}_t^{\top}Y_t - \hat{I}^{fv}(Y.) - \hat{J}^{m}(Y.) - \frac{1}{2}\int_0^t \hat{A}_s^{\top}\hat{A}_s ds\right)\right]\Upsilon(Y.)\right].
\end{aligned}
$$

Hence by Fubini's theorem (using, again, the $\bar{\mathcal{F}}$ -measurability of $\hat{J}^m(Y.)$)

$$
\begin{aligned}
&\tilde{E}\left[\hat{g}^{\varphi}(Y.)\Upsilon(Y.)\right] \\
&\quad = \bar{E}\left[\varphi\left(\hat{\iota}., Y.\right)\exp\left(\hat{A}_t^{\top}Y_t - \hat{I}^{fv}(Y.) - \hat{J}^{m}(Y.) - \frac{1}{2}\int_0^t \hat{A}_s^{\top}\hat{A}_s ds\right)\Upsilon(Y.)\right].
\end{aligned}
$$

Finally, from Lemma (6), the two representations coincide. □

References

1. Bucy, R.C.: Nonlinear filtering theory. IEEE Trans. Automatic Control **10**, 198–199 (1965)
2. Clark, J.M.C.: The design of robust approximations to the stochastic differential equations of nonlinear filtering. Communication systems and random process theory (Proc. 2nd NATO Advanced Study Inst., Darlington, 1977), pp. 721–734. NATO Advanced Study Inst. Ser., Ser. E: Appl. Sci., No. 25, Sijthoff & Noordhoff, Alphen aan den Rijn, (1978)
3. Davis, M.H.A.: Pathwise nonlinear filtering. Stochastic systems: the mathematics of filtering and identification and applications. pp. 505–528. Proceedings of the Nato Advanced Study Institute held at Les Arcs, June 22–July 5, 1980. Edited by Michiel Hazewinkel and Jan C. Willems. NATO Advanced Study Institute Series C: Mathematical and Physical Sciences, 78. D. Reidel Publishing Co., Dordrecht-Boston, Mass., (1981)
4. Davis, M.H.A.: New approach to filtering for nonlinear systems. Proc. IEE-D **128** (5), 166–172 (1981)
5. Davis, M.H.A.: A pathwise solution of the equations of nonlinear filtering. Teor. Veroyatnost. i Primenen. **27** (1), 160–167 (1982)
6. Davis, M.H.A., Spathopoulos, M.P.: Pathwise nonlinear filtering for nondegenerate diffusions with noise correlation. SIAM J. Control Optim. **25** (2), 260–278 (1987)
7. Kallianpur, G., Striebel, C.: Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors. Ann. Math. Statist. **39**, 785–801 (1968)
8. Kushner, H.J.: Dynamical equations for optimal nonlinear filtering. J. Differential Equations **3**, 179–190 (1967)
9. Kushner, H.J.: A robust discrete state approximation to the optimal nonlinear filter for a diffusion. Stochastics **3** (2), 75–83 (1979)
10. Kushner, H.J. Robustness and convergence of approximations to nonlinear filters for jump-diffusions. Matemática Aplicada e Computacional, **16** (2), 153–183 (1997)
11. Liptser, R.S., Shiryaev, A.N.: Statistics of random processes. I. General theory. Translated from the 1974 Russian original by A. B. Aries. Second, revised and expanded edition. Applications of Mathematics (New York), 5. Stochastic Modelling and Applied Probability. Berlin: Springer-Verlag, (2001)

12. Lyons, T.J.: Differential equations driven by rough signals. Rev. Mat. Iberoamericana **14** (2), 215–310 (1998)
13. Wonham, W.M.: Some applications of stochastic differential equations to optimal nonlinear filtering. J. Soc. Indust. Appl. Math. Ser. A Control **2**, 347–369 (1965)
14. Zakai, M.: On the optimal filtering of diffusion processes. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **11**, 230–243 (1969)