Electronic Journal of Probability

Vol. 13 (2008), Paper no. 72, pages 2190–2216.

Journal URL
http://www.math.washington.edu/~ejpecp/

On the Innovations Conjecture of Nonlinear Filtering with Dependent Data*

Andrew J. Heunis
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, Ontario N2L 3G1, Canada
e-mail: heunis@kingcong.uwaterloo.ca

Vladimir M. Lucic
Barclays Capital
5 The North Colonnade, Canary Warf
London E14 4BB, United Kingdom
e-mail: vladimir.lucic@btopenworld.com

Abstract

We establish the innovations conjecture for a nonlinear filtering problem in which the signal to be estimated is conditioned by the observations. The approach uses only elementary stochastic analysis, together with a variant due to J.M.C. Clark of a theorem of Yamada and Watanabe on pathwise-uniqueness and strong solutions of stochastic differential equations.

Key words: nonlinear filter, innovations conjecture, pathwise-uniqueness.

AMS 2000 Subject Classification: Primary 60G35; Secondary: 60H10, 60G44, 60G57.

Submitted to EJP on November 11, 2007, final version accepted November 5, 2008.

*Supported by NSERC of Canada

1 Introduction

A continuing challenge in the theory of nonlinear filtering is the identification of natural conditions under which the so-called *innovations conjecture* is valid. The essence of the problem is as follows: on a filtered probability space $(\Omega, \mathscr{F}, P; \{\mathscr{F}_t\})$ one is given an $\mathbb{R}^D$-valued $\mathscr{F}_t$-standard Wiener process $\{W_t\}$, together with an $\mathbb{R}^D$-valued $\mathscr{F}_t$-progressively measurable signal process $\{\beta_t\}$ subject to the "finite energy" condition

$$\mathrm{E}\int_0^T |\beta_t|^2 \,\mathrm{d}t < \infty, \tag{1.1}$$

for some finite "horizon" $T \in (0, \infty)$. Define the observation process $\{Y_t\}$ and its associated filtration $\{\mathscr{F}_t^Y\}$ in the usual way, namely

$$Y_t := W_t + \int_0^t \beta_s \;\mathrm{d}s, \qquad \mathscr{F}_t^Y := \sigma\{Y_s, s \in [0,t]\}, \qquad t \in [0,T]. \tag{1.2}$$

Put $\hat{\beta}_t := \mathrm{E}\left[\beta_t \,\middle|\, \mathscr{F}_t^Y\right]$, and define the innovation process $\{I_t\}$ and filtration $\{\mathscr{F}_t^I\}$ as

$$I_t := Y_t - \int_0^t \hat{\beta}_s \;\mathrm{d}s, \qquad \mathscr{F}_t^I := \sigma\{I_s, s \in [0,t]\}, \qquad t \in [0,T]. \tag{1.3}$$

In this rather informal introduction we disregard such measure-theoretic technicalities as the need to deal with the usual augmentations of the filtrations $\{\mathscr{F}_t^Y\}$ and $\{\mathscr{F}_t^I\}$, and for $\{\hat{\beta}_t\}$ to be interpreted as the optional projection of $\{\beta_t\}$ onto the usual augmentation of the filtration $\{\mathscr{F}_t^Y\}$; all these refined technicalities will be given due consideration from Section 2 onwards. One sees from (1.3) that $\{I_t\}$ is $\mathscr{F}_t^Y$-adapted, hence $\mathscr{F}_t^I \subset \mathscr{F}_t^Y$. The innovations conjecture states that one actually has equality of the σ-algebras, namely $\mathscr{F}_t^I = \mathscr{F}_t^Y$ for each $t \in [0,T]$.

The innovations conjecture is in fact false under the "minimal" conditions above, as follows from results of Benes ([2], Sect. 8), which are based upon a subtle example due to Tsirel'son [25] of a functional stochastic differential equation (SDE) which fails to have a strong solution. It is therefore necessary to formulate supplementary conditions under which the innovations conjecture can be shown to hold. One of the earliest results of this kind is due to Clark [5], who established the conjecture in the case where the signal $\{\beta_t\}$ is uniformly bounded and independent of $\{W_t\}$, using an argument based on the Kallianpur-Striebel representation for the conditional expectation $\hat{\beta}_t$ (the main result of [5] appears as Theorem 11.4.1 of Kallianpur [[11], p.284], as well as in Meyer [[19], pp.244-246]).

A powerful approach for dealing with the innovations problem was introduced in the late 1970's with the recognition by Allinger, Benes, Clark, Erzhov and Mitter that a theorem of Yamada and Watanabe [27] on pathwise uniqueness and strong solutions of SDEs was key to an alternative and very elegant approach for establishing the innovations conjecture. The central idea is the following: when $\{\beta_t\}$ and $\{W_t\}$ are independent then the Kallianpur-Striebel formula yields an $\mathbb{R}^D$-valued non-anticipative function $\{\gamma_t\}$ on $C[0, T : \mathbb{R}^D]$ (the space of continuous functions from $[0,T]$ into $\mathbb{R}^D$, with the raw canonical filtration) such that $\hat{\beta}(t,\omega) = \gamma_t(Y(\omega))$ $(P \otimes \lambda)$ a.e., and therefore the defining relation for the innovation process in (1.3) can be written in the differential form

$$\mathrm{d}Y_t = \mathrm{d}I_t + \gamma_t(Y)\,\mathrm{d}t, \qquad Y_0 = 0. \tag{1.4}$$

Since $\{I_t\}$ is a standard $\mathscr{F}_t^Y$-Wiener process (see e.g. Rogers and Williams [22], Theorem VI(8.4)(i), p.323) it follows that the pair $\{(Y_t, I_t)\}$ on the filtered probability space $(\Omega, \mathscr{F}, P; \{\mathscr{F}_t^Y\})$ is a solution of (1.4) in the "weak" or "usual" sense (see e.g. Revuz and Yor [20], Definition IX(1.2), p.350). Consequently, if it can be verified that (1.4) has the property of *pathwise uniqueness*, then the theorem of Yamada and Watanabe (see e.g. Theorem IX(1.7)(ii) of [20], p.352) asserts that $\{(Y_t, I_t)\}$ is a *strong solution* of (1.4), that is $\{Y_t\}$ is $\mathscr{F}_t^I$-adapted, and hence $\mathscr{F}_t^Y \subset \mathscr{F}_t^I$, as required to get the innovations conjecture. This is the approach used in [1], in which the innovations conjecture is established when $\{\beta_t\}$ is independent of $\{W_t\}$ and satisfies the finite energy condition (1.1).

The preceding results rely completely on the basic condition that the given processes $\{\beta_t\}$ and $\{W_t\}$ in (1.2) are *independent*. When these processes are correlated it is not possible to establish the innovations conjecture unless specific models which define the structure of the correlation are postulated. The simplest of these is the linear Gaussian model, for which the innovations conjecture may be established by functional-analytic methods which depend on the underlying linearity (see e.g. Kallianpur [11], Section 10.2). In the genuinely nonlinear and non-Gaussian case the problem of establishing the innovations conjecture with correlated data becomes much more challenging. The most significant results on this problem are due to Krylov [15], who considers the following model defined on the finite interval $t \in [0, T]$:

$$\begin{cases} (1') \quad \mathrm{d}X_t = b_1(t, X_t, Y_t)\,\mathrm{d}t + \sigma_{11}(t, X_t, Y_t)\,\mathrm{d}W_t^1 + \sigma_{12}(t, X_t, Y_t)\,\mathrm{d}W_t^2, \qquad X_0 = \xi_0, \\ (2') \quad \mathrm{d}Y_t = b_2(t, X_t, Y_t)\,\mathrm{d}t + \mathrm{d}W_t^2, \qquad Y_0 = 0; \end{cases} \tag{1.5}$$

(in fact, Krylov considers a rather more general model - see (1) of [15], p.773 - which we have simplified to (1.5) in order to focus on just the main essentials). In this model $\{X_t\}$ is an $\mathbb{R}^d$-valued "state" process (typically not available for exact measurement), $\{Y_t\}$ is an $\mathbb{R}^D$-valued observation process, $\{(W_t^1, W_t^2)\}$ is a standard Wiener process independent of the initial state ξ_0, and b_1, b_2, σ_{11}, σ_{12} are vector or matrix-valued Borel measurable functions (of appropriate dimension) defined and uniformly bounded on $[0, T] \times \mathbb{R}^d \times \mathbb{R}^D$, and globally Lipschitz-continuous on $\mathbb{R}^d \times \mathbb{R}^D$ with Lipschitz-constant independent of $t \in [0, T]$. For the filtration $\{\mathscr{F}_t\}$ given by $\mathscr{F}_t := \sigma\{\xi_0, W_s^1, W_s^2, s \in [0, t]\}$ and $\{\beta_t\}$ defined by $\beta_t := b_2(t, X_t, Y_t)$ it follows that $\{W_t^2\}$ is a standard $\mathscr{F}_t$-Wiener process, and $\{\beta_t\}$ is a uniformly bounded and $\mathscr{F}_t$-adapted process. Thus (1.5)(2′) corresponds to the observation equation in (1.2) (with W^2 interpreted as W), and we are therefore in the setting of the first paragraph of the present section, and can pose the question of the validity of the innovations conjecture for the model (1.5). This was established by Krylov [15], using an approach which relies on the normalized (Kushner-Stratonovich) filter equation written as a "forward equation" in density form (due to Krylov and Rozovskii [16], Theorem 1.2, pp.341-342), together with an ingenious recursive approximation of the optimal nonlinear filter, each approximation being adapted to the innovations filtration; this effectively yields that the optimal nonlinear filter itself is adapted to the innovations filtration, from which the innovations conjecture follows. The associated hypotheses are enumerated in ([15], pp.773-774), as well as in Kallianpur ([11], pp.302-303), and in particular include (1) the law of ξ_0 has a continuous density which belongs to an appropriate Sobolev space of functions on $\mathbb{R}^d$; (2) the least eigenvalue of the symmetric matrix $\sigma_{11}\sigma_{11}'(t, x, y)$ is lower-bounded by some $\mu > 0$ uniformly in (t, x, y); (3) $\sigma_{11}(t, x, y)$ and $\sigma_{12}(t, x, y)$ are third-order smooth with respect to x, and $b_1(t, x, y)$ and $b_2(t, x, y)$ are second-order smooth with respect to x; and (4) the "sensor function" $b_2(t, x, y)$ is square-integrable in x with respect to Lebesgue measure in the sense that there is some Borel-measurable $g : \mathbb{R}^d \to [0, \infty)$ such that $\int_{\mathbb{R}^d} g^2(x)\,\mathrm{d}x < \infty$ and

$$\left| b_2(t, x, y) \right| \le g(x) \quad \text{for all} \quad (t, x, y) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^D, \tag{1.6}$$

(see eqn. (4) of [15], p.774, or (v) of [11], p.302). This latter condition is fairly strong, since it supplements the previous postulate of uniform boundedness of the sensor function $b_2(t,x,y)$ with respect to $(t,x,y) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^D$ with the further condition that $b_2(t,x,y)$ be a member of $L^2(\mathbb{R}^d)$, when regarded as a function of the "state" parameter $x \in \mathbb{R}^d$. In addition, the requirement that the initial state ξ_0 have a continuous density is also quite strong, since it excludes for example the simple case where the initial state is non-random.

In view of the preceding discussion, an obvious goal is to try to extend the basic approach used for resolving the innovations conjecture when $\{\beta_t\}$ and $\{W_t\}$ in the observation equation (1.2) are independent, to models such as (1.5) for which this independence no longer holds. It is this problem that we address here. The advantages of this approach are that it is conceptually quite simple, and, as will be seen, it also enables one to remove several of the stronger hypotheses that appear to be necessary when resolving the innovations conjecture for the model (1.5) by methods based on the density form of the normalized filter equation. In fact we shall study a model which is a non-Markov variant of (1.5) with path-dependent coefficients and hypotheses which do not entail uniform boundedness, smoothness, or square-integrability of the coefficients, or require the initial state distribution ξ_0 to have a density. The approach relies only on elementary stochastic analysis, together with a modification due to Clark [6] of the classical theorem of Yamada and Watanabe [27] on the relation between strong solutions of SDEs and pathwise uniqueness.

In Section 2 we formulate the model and the basic hypotheses, and state the main result (Proposition 2.6) on equality of the observation and innovation σ-algebras. In Section 3 we summarize a number of preliminaries that we shall need, in particular a "pathwise" Bayes formula and the afore-mentioned result of Clark [6], and in Section 4 we establish the innovation conjecture for the model of Section 2. Finally, in Section 5, we specialize the nonlinear filtering problem of Section 2, and obtain as a by-product of the innovations conjecture that the corresponding measure-valued normalized filter equation has the property of pathwise-uniqueness; this supplements some earlier results of this kind of Kurtz and Ocone [17] and the authors [18]. Throughout the exposition we isolate in numbered Remarks various facts and observations which we shall need for later reference.

2 Conditions, Model and Main Result

To facilitate easy reference we summarize all of our notation as follows:

Notation 2.1. (I) For a positive integer q write $\mathbb{R}^q$ for the space of all real q-dimensional column vectors, write x^i or $[x]^i$ for the i-th scalar entry and $|x| := [\sum_{i=1}^q (x^i)^2]^{1/2}$ for the Euclidean norm of $x \in \mathbb{R}^q$. Likewise, for positive integers q and r, write $\mathbb{R}^{q\times r}$ for the space of all q by r matrices with real entries, write A^{ij} or $[A]^{ij}$ for the (i,j)-th scalar entry, A' for the transpose, and $\|A\| := \max_{x\in\mathbb{R}^r, |x|=1} |Ax|$ for the operator norm of $A \in \mathbb{R}^{q\times r}$.
(II) For a fixed $T \in (0,\infty)$ let $C[0,T;\mathbb{R}^q]$ denote the space of all continuous mappings from $[0,T]$ into $\mathbb{R}^q$, with the usual supremum norm. For notational brevity we also denote this space by C_q when there is no possibility of confusion, and put $\mathscr{B}_t(C_q) = \sigma\{\psi(s), s \in [0,t]\}$, $t \in [0,T]$, for the canonical filtration on C_q (with ψ a generic member of C_q).
(III) $\mathscr{B}(S)$ denotes the Borel σ-algebra of a separable metric space (S,ρ), and, for a $\mathscr{F}/\mathscr{B}(S)$-measurable mapping ξ from a probability space $(\Omega,\mathscr{F},P)$ into S, let $P\xi^{-1}$ denote the probability measure on $\mathscr{B}(S)$ given by $P\xi^{-1}(A) := P[\xi \in A]$ for each $A \in \mathscr{B}(S)$.
(IV) For an $\mathbb{R}^q$-valued process $\{\eta_t\}$ put $\mathscr{F}_t^\eta := \sigma\{\eta_s, s \in [0,t]\}$ for its raw filtration.

(V) For a continuous semimartingale $\{M_t\}$ put $\mathscr{E}(M)_t := \exp\{M_t - (1/2)\langle M\rangle_t\}$.
(VI) For a measure space $(E, \mathscr{S}, \mu)$ (not necessarily complete) write $\mathbf{Z}^{\mu}[\mathscr{S}]$ for the collection $\{N \subset E : N \subset H$ for some $H \in \mathscr{S}$ with $\mu(H) = 0\}$. For a σ-algebra $\mathscr{M} \subset \mathscr{S}$ write $\mathscr{M} \vee \mathbf{Z}^{\mu}[\mathscr{S}]$ for the minimal σ-algebra on E which includes all members of $\mathscr{M}$ and $\mathbf{Z}^{\mu}[\mathscr{S}]$.
(VII) Let $C^{\infty}(\mathbb{R}^q)$ denote the set of all infinitely smooth $\mathbb{R}$-valued mappings on $\mathbb{R}^q$, and let $C_c^{\infty}(\mathbb{R}^q)$ denote the set of all members of $C^{\infty}(\mathbb{R}^q)$ with compact support.

To specify a model for the filtering problem, fix a constant $T \in (0, \infty)$, which determines a finite "time-horizon" $t \in [0, T]$ of interest, together with an $\mathbb{R}^d$-valued "state" process $\{X_t, t \in [0, T]\}$ and an $\mathbb{R}^D$-valued observation process $\{Y_t, t \in [0, T]\}$ determined by the coupled SDE (2.7), which is subject to the following Conditions 2.2 and 2.3:

$$\begin{cases} (1') \quad \mathrm{d}X_t = a_t(X,Y)\,\mathrm{d}t + b_t(X,Y)\,\mathrm{d}W_t^1 + c_t(X,Y)\,\mathrm{d}Y_t, \qquad X_0 = \xi_0, \\ (2') \quad \mathrm{d}Y_t = h_t(X,Y)\,\mathrm{d}t + \mathrm{d}W_t^2, \qquad Y_0 = 0. \end{cases} \tag{2.7}$$

Condition 2.2. The process $\{(W_t^1, W_t^2), t \in [0, T]\}$ in (2.7) is an $\mathbb{R}^{q+D}$-valued standard Wiener process with respect to the filtration $\{\mathscr{F}_t\}$ on a given filtered probability space $(\Omega, \mathscr{F}, P; \{\mathscr{F}_t\})$ satisfying the "usual conditions" (see e.g. [21], Definition II(67.1), p.172), and ξ_0 is an $\mathbb{R}^d$-valued $\mathscr{F}_0$-measurable random variable such that $\mathrm{E}\left|\xi_0\right|^4 < \infty$.

Condition 2.3. For $\Xi := [0, T] \times C[0, T; \mathbb{R}^d] \times C[0, T; \mathbb{R}^D]$ the mappings $a : \Xi \to \mathbb{R}^d$, $b : \Xi \to \mathbb{R}^{d\times q}$, $c : \Xi \to \mathbb{R}^{d\times D}$, and $h : \Xi \to \mathbb{R}^D$ in (2.7) have the following properties: (i) if α denotes a generic scalar entry from a, b, c or h, then the mapping $(t, x, y) \mapsto \alpha_t(x, y) : \Xi \mapsto \mathbb{R}$ is continuous, and the the mapping $(x, y) \mapsto \alpha_t(x, y) : C_d \times C_D \mapsto \mathbb{R}$ is $\mathscr{B}_t(C_d \times C_D)$-measurable for each $t \in [0, T]$; (ii) there is a constant $K \in [0, \infty)$ such that, if α denotes a generic scalar entry from a, b, c, h, or ch, then

$$\left|\alpha_t(x_1, y_1) - \alpha_t(x_2, y_2)\right| \le K \sup_{s\in[0,t]} \{\left|x_1(s) - x_2(s)\right| + \left|y_1(s) - y_2(s)\right|\},$$

for all $t \in [0, T]$, $(x_i, y_i) \in C_d \times C_D$; (iii) h is uniformly bounded on Ξ.

Remark 2.4. The third term on the right of (2.7)(1′) represents "feedback" of the observation $\{Y_t\}$ to the dynamics of the state process $\{X_t\}$. When (2.7)(2′) is used to remove "$\mathrm{d}Y_t$" from (2.7)(1′) then we get state-dynamics in a form similar to (1.5)(1′) (but for path-dependent coefficients). From Condition 2.3 and standard results on SDE's with Lipschitz-continuous "functional" coefficients (e.g. Kallianpur [11], Theorem 5.1.1, p.97) we know that the processes $\{X_t\}$ and $\{Y_t\}$ are uniquely determined to within indistinguishability, and adapted to the filtration $\sigma\{\xi_0, W_s^1, W_s^2, s \in [0, t]\} \vee \mathbf{Z}^P[\mathscr{F}] \subset \mathscr{F}_t$, hence are continuous $\mathscr{F}_t$-semimartingales.

Remark 2.5. Let $\{\mathscr{Y}_t\}$ be the "usual augmentation" (Rogers and Williams [21], Definition II(67.3) and Lemma II(67.4), p.172) of the raw filtration $\mathscr{F}_t^Y = \sigma\{Y_s, s \in [0, t]\}$ of the observation process $\{Y_t\}$ given by (2.7)(2′) (recall Notation 2.1(IV)). Put

$$\beta_t := h_t(X, Y), \qquad t \in [0, T], \tag{2.8}$$

(for $\{(X_t, Y_t)\}$ given by (2.7)), and let $\{\hat{\beta}_t, t \in [0, T]\}$ denote the *optional projection* of $\{\beta_t, t \in [0, T]\}$ onto the filtration $\{\mathscr{Y}_t\}$ (see e.g. Definition VI(7.1) of [22], p.319). With $\hat{\beta}_t$ thus understood, define the innovations process $\{I_t\}$ by

$$I_t := Y_t - \int_0^t \hat{\beta}_s\,\mathrm{d}s, \qquad t \in [0, T], \tag{2.9}$$

and let $\{\mathscr{I}_t\}$ be the usual augmentation of the raw filtration $\{\mathscr{F}_t^I\}$ of the process given by (2.9). It follows that $\{I_t\}$ is $\mathscr{Y}_t$-adapted, hence $\mathscr{I}_t \subset \mathscr{Y}_t$. Our main result asserts that the opposite set inclusion holds, which establishes the innovations conjecture for the model (2.7):

Proposition 2.6. Suppose Conditions 2.2 and 2.3 for the coupled SDE (2.7). Then, with reference to Remark 2.5, we have that $\mathscr{Y}_t = \mathscr{I}_t$ for each $t \in [0, T]$.

Comparing this result with that summarized in Section 1, we see that (2.7) admits coefficients with "functional" dependence (in contrast to the "diffusion-type" coefficients in (1.5)), without any postulated smoothness, allows unbounded coefficients in the "state" equation (2.7)(1′), and does not require that the initial state ξ_0 in (2.7) have a density function. We do need uniform boundedness of the "sensor function" h_t in the observation equation (2.7)(2′), but do not postulate that this function enjoy any square-integrability properties comparable to (1.6) (which would of course not make sense in the setting of path-dependent coefficients).

3 A Pathwise Bayes Formula and Other Preliminaries

Essential to the proof of Proposition 2.6 is a "pathwise" Bayes representation for the process $\{\hat{\beta}_t\}$ (recall Remark 2.5) which is given in the present section. We begin by recalling a Bayes formula for the conditional expectation $\mathrm{E}\left[\beta_t \,\middle|\, \mathscr{Y}_t\right]$ which is established in Elliott ([7], (18.24), p.290) in the case of Markov dynamics for the state/observation pair (X, Y). We briefly summarize the derivation of the formula, since our setting is somewhat different from that of [7], and because later on we shall need some of the ideas which arise in the derivation. The process $\{\beta_t\}$ is continuous, $\mathscr{F}_t$-adapted and uniformly bounded (see (2.8), Remark 2.4, and Condition 2.3(i)(iii)), and hence (see Notation 2.1(V)) the process given by

$$\Gamma_t := \exp\left[-\int_0^t \beta_\tau' \,\mathrm{d}W_\tau^2 - \frac{1}{2}\int_0^t \left|\beta_\tau\right|^2 \,\mathrm{d}\tau\right] = \mathscr{E}(-[0 \;\; \beta]' \bullet [W^1 \;\; W^2])_t, \tag{3.10}$$

defines a continuous $\mathscr{F}_t$-martingale on $(\Omega, \mathscr{F}, P)$. It follows that

$$P_0(A) := \mathrm{E}[\Gamma_T;\, A], \qquad A \in \mathscr{F}, \tag{3.11}$$

defines a *probability measure* on $(\Omega, \mathscr{F})$ equivalent to P, and for $A \in \mathscr{F}_0$, we have $P_0(A) = \mathrm{E}[\mathrm{E}\left[\Gamma_T \,\middle|\, \mathscr{F}_0\right];\, A] = \mathrm{E}[\Gamma_0;\, A] = P(A)$. Since ξ_0 is $\mathscr{F}_0$-measurable (Condition 2.2) it follows in particular that ξ_0 has the same law relative to both P and P_0, namely

$$\pi_0(B) := P[\xi_0 \in B] = P_0[\xi_0 \in B], \quad B \in \mathscr{B}(\mathbb{R}^d). \tag{3.12}$$

Moreover, from (2.8) and (2.7)(2′), we have

$$\begin{bmatrix} W_t^1 \\ Y_t \end{bmatrix} = \begin{bmatrix} W_t^1 \\ W_t^2 \end{bmatrix} + \int_0^t \begin{bmatrix} 0 \\ \beta_s \end{bmatrix} \mathrm{d}s, \tag{3.13}$$

and it follows from (3.10), (3.11), (3.13), and the Girsanov theorem that

$$\{(W_t^1, Y_t),\ t \in [0, T]\} \text{ is an } \mathbb{R}^{q+D}\text{-valued standard } \mathscr{F}_t\text{-Wiener process on } (\Omega, \mathscr{F}, P_0). \tag{3.14}$$

Now put

$$\Lambda_t := 1/\Gamma_t = \exp\left[\left[\int_0^t \beta'_\tau \, \mathrm{d}Y_\tau - \frac{1}{2}\int_0^t \left|\beta_\tau\right|^2 \, \mathrm{d}\tau\right]\right], \quad t \in [0,T], \tag{3.15}$$

(where we have used (3.13) at the second equality). From (3.11), for each $t \in [0,T]$ we have $P(A) = \mathrm{E}_0[\Lambda_t;\ A]$ for all $A \in \mathscr{F}_t$ (E_0 denotes P_0-expectation). Since $\mathscr{Y}_t \subset \mathscr{F}_t$, the change-of-variables formula for Radon-Nikodym derivatives (e.g. Wong and Hajek [26], Lemma 7.1, p.243) then establishes the desired Bayes representation, namely for each $t \in [0,T]$ we have

$$\mathrm{E}\left[\beta_t \,\middle|\, \mathscr{Y}_t\right] = \frac{\mathrm{E}_0\left[\beta_t\Lambda_t \,\middle|\, \mathscr{Y}_t\right]}{\mathrm{E}_0\left[\Lambda_t \,\middle|\, \mathscr{Y}_t\right]}, \qquad P-\text{a.s.} \tag{3.16}$$

In the sequel it will be essential to have at hand a "pathwise" representation of the Bayes formula (3.16) which is due to Bhatt and Karandikar [3]. As preparation for stating this representation we summarize in the following Theorem 3.1 two fundamental results of Karandikar ([13], Theorem 3) and ([12], Theorem 4.3) on pathwise representations for stochastic integrals and for solutions of SDEs with Lipschitz-continuous coefficients. These results are needed to state the "pathwise" representation of the Bayes formula (3.16) and will also be used for the proof of Proposition 2.6 in Section 4.

Theorem 3.1. The following "pathwise" representations hold:

(I) There exists a universal "stochastic integral" mapping $\mathfrak{J} : C_1 \times C_1 \to C_1$ (where $C_1 := C[0,T:\mathbb{R}]$ - recall Notation 2.1(II)) with the following properties:
(a) $(\psi^1,\psi^2) \mapsto \mathfrak{J}(\psi^1,\psi^2) : C_1 \times C_1 \mapsto C_1$ is $\mathscr{B}(C_1 \times C_1)/\mathscr{B}(C_1)$-measurable;
(b) if $\{\rho_t\}$ is an $\mathbb{R}$-valued continuous $\mathscr{A}_t$-adapted process and $\{\eta_t\}$ is an $\mathbb{R}$-valued continuous semi-martingale on a filtered probability space $(E,\mathscr{A},\mu;\{\mathscr{A}_t\})$ satisfying the usual conditions, then the process $\{\mathfrak{J}_t(\rho,\eta)\}$ (defined pathwise) is indistinguishable from the ($\mathscr{A}_t$-adapted) stochastic integral process

$$(\rho \bullet \eta)_t := \int_0^t \rho_\tau \, \mathrm{d}\eta_\tau, \qquad t \in [0,T].$$

(II) Suppose that Condition 2.3 holds for the coefficients a, b and c in (2.7)(1′). Then there exists a universal "solution" mapping $\mathfrak{e} : \mathbb{R}^d \times C_q \times C_D \to C_d$ with the following properties:
(a) $(\xi,w,y) \mapsto \mathfrak{e}(\xi,w,y) : \mathbb{R}^d \times C_q \times C_D \mapsto C_d$ is $\mathscr{B}(\mathbb{R}^d \times C_q \times C_D)/\mathscr{B}(C_d)$-measurable;
(b) if $\{(\eta_t^1,\eta_t^2)\}$ is an $\mathbb{R}^{q+D}$-valued continuous semimartingale on a filtered probability space $(E,\mathscr{A},\mu;\{\mathscr{A}_t\})$ satisfying the usual conditions, and χ_0 is some $\mathbb{R}^d$-valued $\mathscr{A}_0$-measurable random vector, then the process $\{\mathfrak{e}_t(\chi_0,\eta^1,\eta^2)\}$ (defined pathwise) is indistinguishable from the ($\mathbb{R}^d$-valued, continuous, and $\mathscr{A}_t$-adapted) process $\{\zeta_t\}$ defined by

$$\mathrm{d}\zeta_t = a_t(\zeta,\eta^2)\,\mathrm{d}t + b_t(\zeta,\eta^2)\,\mathrm{d}\eta_t^1 + c_t(\zeta,\eta^2)\,\mathrm{d}\eta_t^2, \qquad \zeta_0 = \chi_0, \tag{3.17}$$

(existence and uniqueness - to within indistinguishability - of solutions for (3.17) follows from Picard iterations and Condition 2.3)

The mapping $\mathfrak{J}$ in Theorem 3.1(I) is "universal" in the sense that it does not depend in any way on the laws of the continuous process $\{\rho_t\}$ and the continuous semimartingale $\{\eta_t\}$. Similarly, the solution mapping $\mathfrak{e}$ in Theorem 3.1(II) is "universal" in that it is determined by the coefficients a, b

and c only, and does not depend on the laws of χ_0 or the semimartingale $\{(\eta_t^1, \eta_t^2)\}$. It will be useful to define a "canonical set-up" as follows: recalling Notation 2.1(II)(III) put

$$Q_1 := \text{Wiener measure on } \mathscr{B}(C_q), \qquad Q_2 := \text{Wiener measure on } \mathscr{B}(C_D), \tag{3.18}$$

$$E := \mathbb{R}^d \times C_q \times C_D, \qquad \mu(A) := (\pi_0 \times Q_1 \times Q_2)(A), \quad A \in \mathscr{B}(E), \tag{3.19}$$

in which π_0 is the law of the initial state ξ_0 for the model (2.7) (see (3.12)).

Remark 3.2. For later use we note the following simple observation: By (3.14) and (3.18) we have $P_0 Y^{-1} = Q_2$ on $\mathscr{B}(C_D)$; since P and P_0 are equivalent on $\mathscr{F}$ (by (3.11)), we then see that PY^{-1} and Q_2 are equivalent probability measures on $\mathscr{B}(C_D)$.

Denote by $\mathscr{A} := \mathscr{B}(E) \vee \mathbf{Z}^\mu[\mathscr{B}(E)]$ the μ-completion of $\mathscr{B}(E)$ (see Notation 2.1(VI)), and put

$$\mathscr{A}_t := (\mathscr{B}(\mathbb{R}^d) \otimes \mathscr{B}_t(C_q) \otimes \mathscr{B}_t(C_D)) \vee \mathbf{Z}^\mu[\mathscr{B}(E)]. \tag{3.20}$$

Then the filtered probability space $(E, \mathscr{A}, \mu; \{\mathscr{A}_t\})$ satisfies the usual conditions (see e.g. [21], Theorem (II)(68.4), p.175). If (ξ, w, y) is a generic member of E then $\{\mathfrak{e}_t(\xi, w, y)\}$ is an $\mathbb{R}^d$-valued, continuous and $\mathscr{A}_t$-adapted process on $(E, \mathscr{A}, \mu; \{\mathscr{A}_t\})$, and then, for

$$\vartheta_t(\xi, w, y) := h_t(\mathfrak{e}(\xi, w, y), y), \qquad (t, \xi, w, y) \in [0, T] \times \mathbb{R}^d \times C_q \times C_D, \tag{3.21}$$

we see from Condition 2.3(i) that $\{\vartheta_t(\xi, w, y)\}$ is an $\mathbb{R}^D$-valued, continuous and $\mathscr{A}_t$-adapted process on $(E, \mathscr{A}, \mu; \{\mathscr{A}_t\})$. From this it follows that $\{\mathfrak{I}_t(\vartheta^k(\xi, w, y), y^k)\}$ is $\mathbb{R}$-valued, continuous and $\mathscr{A}_t$-adapted for each $k = 1, 2, \ldots, D$, and hence the $(0, \infty)$-valued process $\{\rho_t(\xi, w, y)\}$ defined on $(E, \mathscr{A}, \mu; \{\mathscr{A}_t\})$ by

$$\rho_t(\xi, w, y) := \exp\left[\sum_{k=1}^D \mathfrak{I}_t(\vartheta^k(\xi, w, y), y^k) - \frac{1}{2}\int_0^t \left|\vartheta_\tau(\xi, w, y)\right|^2 \, \mathrm{d}\tau\right], \tag{3.22}$$

is likewise continuous and $\mathscr{A}_t$-adapted.

From now on $\overline{\mathscr{B}(C_D)}^{Q_2}$ denotes the Q_2-completion of $\mathscr{B}(C_D)$ and $\overline{\mathscr{B}_t(C_D)}^{Q_2}$ denotes the completion of $\mathscr{B}_t(C_D)$ with the Q_2-null sets of $\overline{\mathscr{B}(C_D)}^{Q_2}$, namely (see Notation 2.1(VI))

$$\overline{\mathscr{B}(C_D)}^{Q_2} := \mathscr{B}(C_D) \vee \mathbf{Z}^{Q_2}[\mathscr{B}(C_D)], \qquad \overline{\mathscr{B}_t(C_D)}^{Q_2} := \mathscr{B}_t(C_D) \vee \mathbf{Z}^{Q_2}[\mathscr{B}(C_D)]. \tag{3.23}$$

Then the filtered probability space $(C_D, \overline{\mathscr{B}(C_D)}^{Q_2}, Q_2; \{\overline{\mathscr{B}_t(C_D)}^{Q_2}\})$ is just the usual augmentation of $(C_D, \mathscr{B}(C_D), Q_2; \{\mathscr{B}_t(C_D)\})$ (since Q_2 is Wiener measure on $\mathscr{B}(C_D)$).

With these preliminaries in place, we can state the following representation formulae of Bhatt and Karandikar ([3], p.45, p.46): The $(0, \infty)$-valued process $\{G_t(y)\}$ defined by

$$G_t(y) := \int_{\mathbb{R}^d \times C_q} \rho_t(\xi, w, y) \, \mathrm{d}(\pi_0 \times Q_1)(\xi, w), \qquad (t, y) \in [0, T] \times C_D, \tag{3.24}$$

is $\overline{\mathscr{B}_t(C_D)}^{Q_2}$-predictable (or previsible), the $\mathbb{R}^D$-valued process $\{F_t(y)\}$ defined by

$$F_t(y) := \int_{\mathbb{R}^d \times C_q} \vartheta_t(\xi, w, y) \rho_t(\xi, w, y) \, \mathrm{d}(\pi_0 \times Q_1)(\xi, w), \qquad (t, y) \in [0, T] \times C_D, \tag{3.25}$$

is $\overline{\mathscr{B}_t(C_D)}^{Q_2}$-predictable, and for each $t \in [0,T]$ we have

$$\mathrm{E}_0\left[\Lambda_t \,\middle|\, \mathscr{Y}_t\right] = G_t(Y) \qquad \text{and} \qquad \mathrm{E}_0\left[\beta_t\Lambda_t \,\middle|\, \mathscr{Y}_t\right] = F_t(Y), \qquad P-\text{a.s.} \tag{3.26}$$

Thus, from (3.26) and (3.16), for each $t \in [0,T]$ one has

$$\mathrm{E}\left[\beta_t \,\middle|\, \mathscr{Y}_t\right] = \gamma_t(Y), \qquad P-\text{a.s.} \tag{3.27}$$

for the $\mathbb{R}^D$-valued and $\overline{\mathscr{B}_t(C_D)}^{Q_2}$-predictable function $\{\gamma_t(y)\}$ given by

$$\gamma_t(y) := \frac{F_t(y)}{G_t(y)}, \qquad (t,y) \in [0,T] \times C_D. \tag{3.28}$$

Remark 3.3. The "pathwise" representation (3.27) will be essential to establishing Proposition 2.6. In fact, it will be useful to have the representation in the form of a functional which is predictable with respect to the *raw* canonical filtration $\{\mathscr{B}_t(C_D)\}$ rather than its Q_2-augmentation $\{\overline{\mathscr{B}_t(C_D)}^{Q_2}\}$, in order to facilitate later application of results on SDE's with coefficients having functional dependence (see Rogers and Williams [22], Definition V(8.3) and V(8.7), p.122-123). To this end, observe from (3.23) and ([22], V(10.12), p.128) that there exists some $\mathbb{R}^D$-valued $\mathscr{B}_t(C_D)$-predictable process $\{\gamma_t^0\}$ such that

$$Q_2\{y \in C_D : \gamma_t(y) = \gamma_t^0(y),\ t \in [0,T]\} = 1. \tag{3.29}$$

Moreover, from (3.21), (3.24), (3.25), and the uniform bound on h (Condition 2.3(iii)), we see that $|F_t(y)| \le \|h\|\, G_t(y)$ (for $\|h\| := \sup_{(t,x,y)\in\Xi} |h_t(x,y)| < \infty$), thus $|\gamma_t(y)| \le \|h\|$ for all $(t,y) \in [0,T] \times C_D$, and it is easily seen from the monotone-class argument leading to (3.29) that γ^0 is subject to the same uniform bound, namely

$$|\gamma_t^0(y)| \le \|h\|, \qquad (t,y) \in [0,T] \times C_D. \tag{3.30}$$

Since $\{\gamma_t^0\}$ is $\mathscr{B}_t(C_D)$-predictable, it follows that $\{\gamma_t^0(Y)\}$ is $\mathscr{Y}_t$-predictable (see [22], V(8.6), p.123) hence $\mathscr{Y}_t$-optional. Moreover, from (3.29), (3.27), Remark 3.2, and $\hat{\beta}_t = \mathrm{E}\left[\beta_t \,\middle|\, \mathscr{Y}_t\right]$, $P-$a.s. (recall Remark 2.5), we get $\hat{\beta}_t = \gamma_t^0(Y)$, $P-$a.s. for each $t \in [0,T]$. From this relation, along with (2.9) and Fubini's theorem, we see that $\{I_t\}$ and $\{Y_t\}$ satisfy the identity

$$I_t = Y_t - \int_0^t \gamma_\tau^0(Y)\,\mathrm{d}\tau, \qquad t \in [0,T], \qquad P-\text{a.s.} \tag{3.31}$$

Remark 3.4. For later reference we next recall the notions of solution and strong solution in the context of the particular SDE arising from (3.31). Although these ideas are quite standard in the theory of SDE's, there is some variation in terminology and formulation from one reference to another. In view of their considerable importance it seems appropriate to recall the definitions clearly at this point:

(I) A pair $\{(\bar{\Omega},\bar{\mathscr{F}},\bar{P};\{\bar{\mathscr{F}}_t\}),(\bar{Y}_t,\bar{I}_t)\}$ is a solution of the *SDE with drift function* $\{\gamma_t^0\}$ *and unit covariance* when $(\bar{\Omega},\bar{\mathscr{F}},\bar{P};\{\bar{\mathscr{F}}_t\})$ is a filtered probability space satisfying the usual conditions, $\{\bar{Y}_t\}$ is an $\mathbb{R}^D$-valued continuous $\bar{\mathscr{F}}_t$-adapted process, and $\{\bar{I}_t\}$ is an $\mathbb{R}^D$-valued standard $\bar{\mathscr{F}}_t$-Wiener process on $(\bar{\Omega},\bar{\mathscr{F}},\bar{P})$, such that

$$\bar{Y}_t = \bar{I}_t + \int_0^t \gamma_\tau^0(\bar{Y})\,\mathrm{d}\tau, \qquad t \in [0,T], \qquad \bar{P}-\text{a.s.}$$

Furthermore, a solution $\{(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P}; \{\bar{\mathscr{F}}_t\}), (\bar{Y}_t, \bar{I}_t)\}$ is said to be a *strong solution* when $\{\bar{Y}_t\}$ is adapted to the filtration $\{\bar{\mathscr{I}}_t\}$ (which is the usual augmentation of the raw filtration $\{\mathscr{F}_t^{\bar{I}}\}$ of the Wiener process $\{\bar{I}_t\}$), or equivalently when $\bar{\mathscr{Y}}_t \subset \bar{\mathscr{I}}_t$, $t \in [0,T]$, where $\{\bar{\mathscr{Y}}_t\}$ is the usual augmentation of the raw filtration $\{\mathscr{F}_t^{\bar{Y}}\}$ of the process $\{\bar{Y}_t\}$.

(II) For the innovation process $\{I_t\}$ and observation filtration $\{\mathscr{Y}_t\}$ defined at Remark 2.5 it is a standard result that $\{I_t\}$ is an $\mathbb{R}^D$-valued $\mathscr{Y}_t$-Wiener process on $(\Omega, \mathscr{F}, P)$ (see e.g. Rogers and Williams [22], Theorem VI(8.4)(i), p.323). It then follows from (3.31) that $\{(\Omega, \mathscr{F}, P; \{\mathscr{Y}_t\}), (Y_t, I_t)\}$ is a solution of the SDE with drift function $\{\gamma_t^0\}$ and unit covariance (in the sense of (I)). It remains to show that it is a *strong solution*, for then we have that $\mathscr{Y}_t \subset \mathscr{I}_t$, which gives Proposition 2.6. To this end, we shall use a variant of the theorem of Yamada and Watanabe [27] due to Clark ([6], Proposition and Corollary p.157), which, in the context of the SDE with drift function $\{\gamma_t^0\}$ and unit covariance, states the following:

Theorem 3.5. Suppose that the SDE with drift function $\{\gamma_t^0\}$ and unit covariance has the property of pathwise-uniqueness in the following restricted sense: given any pair of solutions $\{(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P}; \{\bar{\mathscr{F}}_t\}), (\bar{Y}_t^i, \bar{I}_t)\}$, $i = 1,2$, each having joint-law identical to that of the solution $\{(\Omega, \mathscr{F}, P; \{\mathscr{Y}_t\}), (Y_t, I_t)\}$ in Remark 3.4(II) (that is $P(I,Y)^{-1} = \bar{P}(\bar{I}, \bar{Y}^1)^{-1} = \bar{P}(\bar{I}, \bar{Y}^2)^{-1}$), it necessarily follows that $\{\bar{Y}_t^1\}$ and $\{\bar{Y}_t^2\}$ are $\bar{P}$-indistinguishable. Then $\{(\Omega, \mathscr{F}, P; \{\mathscr{Y}_t\}), (Y_t, I_t)\}$ is a strong solution, that is $\mathscr{Y}_t \subset \mathscr{I}_t$ for each $t \in [0,T]$.

Remark 3.6. Some remarks on the relationship between the classical theorem of Yamada and Watanabe [27] and Clark's modification of this theorem are in order. The theorem of Yamada and Watanabe asserts (among other things) that if pathwise-uniqueness holds among *all* pairs of postulated solutions of an SDE (on an arbitrary common filtered probability space, with common initial value and common "driving" Wiener process) then *each and every* solution of the SDE is a strong solution (see e.g. Theorem IX(1.7)(ii) of [20], p.352, for a very nice rendition of this result). In contrast, Clark's modification [6] of this theorem postulates the weaker hypothesis of pathwise-uniqueness among all pairs of postulated solutions of the SDE *having the same joint law as some designated solution*, and in return gives the weaker conclusion that just the designated solution is strong. For many applications this conclusion is all that is wanted (the present one being a case in point), and the weaker hypothesis is often easier to verify, since pathwise-uniqueness need be established only among putative solutions with the same law as the solution whose strength must be demonstrated, rather than among all postulated solutions. Although Theorem 3.5 is a statement of Clark's result only for the specific case of the SDE with drift function $\{\gamma_t^0\}$ and unit covariance, the result in fact pertains to completely general SDE's with functional coefficients (see Clark [6], Proposition and Corollary on p.157).

Remark 3.7. From now on fix a pair of solutions $\{(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P}; \{\bar{\mathscr{F}}_t\}), (\bar{Y}_t^i, \bar{I}_t)\}$, $i = 1,2$, of the SDE with drift function $\{\gamma_t^0\}$ and unit covariance, having the same joint-law as the solution $\{(\Omega, \mathscr{F}, P; \{\mathscr{Y}_t\}), (Y_t, I_t)\}$ of Remark 3.4(II), that is $P(I,Y)^{-1} = \bar{P}(\bar{I}, \bar{Y}^1)^{-1} = \bar{P}(\bar{I}, \bar{Y}^2)^{-1}$ (as in Theorem 3.5). In Section 4 we shall use the structure of the drift functional $\{\gamma_t^0\}$ to establish

$$\bar{P}\{\bar{Y}_t^1 = \bar{Y}_t^2, \quad t \in [0,T]\} = 1. \tag{3.32}$$

Once this has been shown, Proposition 2.6 follows immediately from Theorem 3.5. To establish (3.32) it is necessary to "expand" the filtered probability space $(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P}; \{\bar{\mathscr{F}}_t\})$ on which the solutions are defined. Fix a standard $\mathbb{R}^q$-valued $\hat{\mathscr{F}}_t$-Wiener process $\{\hat{W}_t,\ t \in [0,T]\}$ and an $\mathbb{R}^d$-valued $\hat{\mathscr{F}}_0$-measurable random vector $\hat{\xi}_0$ on a filtered probability space $(\hat{\Omega}, \hat{\mathscr{F}}, \hat{P}; \{\hat{\mathscr{F}}_t\})$, such that

$\hat{P}\hat{\xi}_0^{-1} = P\xi_0^{-1} := \pi_0$ (recall Condition 2.2 and definition of π_0 at (3.12)). Next, put

$$\tilde{\Omega} := \bar{\Omega} \times \hat{\Omega}, \qquad \tilde{\mathscr{G}} := \bar{\mathscr{F}} \otimes \hat{\mathscr{F}}, \qquad \tilde{\mathscr{G}}_t := \bar{\mathscr{F}}_t \otimes \hat{\mathscr{F}}_t, \qquad \tilde{P} := \bar{P} \times \hat{P}, \tag{3.33}$$

and let $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P}; \{\tilde{\mathscr{F}}_t\})$ be the usual augmentation of the filtered probability space $(\tilde{\Omega}, \tilde{\mathscr{G}}, \tilde{P}; \{\tilde{\mathscr{G}}_t\})$ (see [21], Lemma II(67.4), p.172). Next, for $\tilde{\omega} := (\bar{\omega}, \hat{\omega})$, put

$$\tilde{I}_t(\tilde{\omega}) := \bar{I}_t(\bar{\omega}), \qquad \tilde{Y}_t^i(\tilde{\omega}) := \bar{Y}_t^i(\bar{\omega}), \qquad \tilde{W}_t(\tilde{\omega}) := \hat{W}_t(\hat{\omega}), \qquad \tilde{\xi}_0(\tilde{\omega}) := \hat{\xi}_0(\hat{\omega}). \tag{3.34}$$

Since a Wiener process remains so when its filtration is augmented ([21], Lemma II(72.2), p.180), we have

$$\{\tilde{W}_t\} \text{ and } \{\tilde{I}_t\} \text{ are } \mathbb{R}^q \text{ and } \mathbb{R}^D\text{-valued } \tilde{\mathscr{F}}_t\text{-standard Wiener processes on } (\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P}), \tag{3.35}$$

$$\sigma\{\tilde{\xi}_0,\, \tilde{W}_t,\, t \in [0,T]\} \text{ and } \sigma\{\tilde{I}_t, \tilde{Y}_t^1, \tilde{Y}_t^2,\, t \in [0,T]\} \text{ are } \tilde{P}\text{-independent}, \tag{3.36}$$

$$\tilde{P}(\tilde{\xi}_0, \tilde{W})^{-1} = \pi_0 \times Q_1, \qquad P(I,Y)^{-1} = \bar{P}(\bar{I}, \bar{Y}^i)^{-1} = \tilde{P}(\tilde{I}, \tilde{Y}^i)^{-1}, \qquad i = 1,2, \tag{3.37}$$

and $\{\tilde{Y}_t^1\}$ and $\{\tilde{Y}_t^2\}$ are $\mathbb{R}^D$-valued continuous $\tilde{\mathscr{F}}_t$-semimartingales such that

$$\tilde{Y}_t^i = \tilde{I}_t + \int_0^t \gamma_\tau^0(\tilde{Y}^i)\, \mathrm{d}\tau, \qquad t \in [0,T], \qquad \tilde{P} - \text{a.s.} \tag{3.38}$$

Upon defining

$$\tilde{l}_t := \gamma_t^0(\tilde{Y}^1) - \gamma_t^0(\tilde{Y}^2), \tag{3.39}$$

we see from (3.38) that

$$\tilde{Y}_t^1 - \tilde{Y}_t^2 = \int_0^t \tilde{l}_\tau\, \mathrm{d}\tau. \tag{3.40}$$

It therefore remains to show that

$$\tilde{l}_t(\tilde{\omega}) = 0, \qquad \lambda \times \tilde{P} - \text{a.e.} \tag{3.41}$$

for then (3.40) and the Fubini theorem establish that $\{\tilde{Y}_t^1\}$ and $\{\tilde{Y}_t^2\}$ are $\tilde{P}$-indistinguishable, which, in view of (3.33) and (3.34), gives (3.32).

4 Proof of Proposition 2.6

In the present section our goal is to show (3.41); as noted at the end of the previous section, this establishes (3.32) and hence Proposition 2.6. To this end, and recalling Remark 3.7 and Theorem 3.1(II), define

$$\tilde{X}_t^i := \mathfrak{e}_t(\tilde{\xi}_0, \tilde{W}, \tilde{Y}^i), \qquad i = 1,2, \qquad\qquad \Delta\tilde{X}_t := \sup_{s \le t} |\tilde{X}_s^1 - \tilde{X}_s^2|, \tag{4.42}$$

$$\mathscr{H}_t := \sigma\{\tilde{Y}_s^1,\, \tilde{Y}_s^2,\, s \in [0,t]\}, \qquad\qquad \tilde{\mathscr{Y}}_t := \mathscr{H}_{t+} \vee \mathbf{Z}^{\tilde{P}}[\tilde{\mathscr{F}}] \subset \tilde{\mathscr{F}}_t, \tag{4.43}$$

(i.e. $\{\tilde{\mathscr{Y}}_t\}$ is the usual augmentation of the filtration $\{\mathscr{H}_t\}$; the set-inclusion at (4.43) follows because the $\{\tilde{Y}^i\}$ are $\tilde{\mathscr{F}}_t$-adapted and $\{\tilde{\mathscr{F}}_t\}$ satisfies the usual conditions). To establish (3.41) we shall need the following result, the proof of which is deferred to later in this section:

Proposition 4.1. Suppose that Conditions 2.2 and 2.3 hold. Then there exists a sequence of $\tilde{\mathscr{Y}}_t$-stopping times $\{T_n,\ n = 1,2,\ldots\}$ on $(\tilde{\Omega},\tilde{\mathscr{F}},\tilde{P})$, together with a sequence of constants $\{K(n), n = 1,2,\ldots\} \subset [0,\infty)$, such that

(i) $0 \le T_n(\tilde{\omega}) \le T_{n+1}(\tilde{\omega}) \le T$ for all $\tilde{\omega} \in \tilde{\Omega}$ and $n = 1,2,\ldots$;

(ii) $\lim_{n\to\infty} T_n(\tilde{\omega}) = T$ for $\tilde{P}$-almost all $\tilde{\omega} \in \tilde{\Omega}$;

(iii) for each $n = 1,2,\ldots$, the following inequalities (4.44) and (4.45) hold for all $t \in [0,T]$:

$$\begin{aligned}\tilde{\mathrm{E}}[|\tilde{l}_t|^2 I[0,T_n)(t)] &\le K(n)\Bigg\{\int_0^t \tilde{\mathrm{E}}[I[0,T_n)(\tau)|\Delta\tilde{X}_\tau|^2]\,\mathrm{d}\tau \\ &+ \int_0^t \tilde{\mathrm{E}}[I[0,T_n)(\tau)\,|\tilde{l}_\tau|^2]\,\mathrm{d}\tau + \tilde{\mathrm{E}}[I[0,T_n)(t)|\Delta\tilde{X}_t|^2]\Bigg\},\end{aligned} \tag{4.44}$$

$$\begin{aligned}\tilde{\mathrm{E}}[I[0,T_n)(t)|\Delta\tilde{X}_t|^2] &\le K(n)\Bigg\{\int_0^t \tilde{\mathrm{E}}[I[0,T_n)(\tau)\,|\Delta\tilde{X}_\tau|^2]\,\mathrm{d}\tau \\ &+ \int_0^t \tilde{\mathrm{E}}[I[0,T_n)(\tau)\,|\tilde{l}_\tau|^2]\,\mathrm{d}\tau\Bigg\}.\end{aligned} \tag{4.45}$$

Remark 4.2. From (3.35) and (3.38), we see that $\{(\tilde{W}_t,\tilde{Y}_t^i)\}$ is an R^{q+D}-valued continuous $\tilde{\mathscr{F}}_t$-semimartingale, hence, for $\{\tilde{X}_t^i\}$ defined by (4.42), it follows from Theorem 3.1(II)(b) that

$$\tilde{X}_t^i = \tilde{\xi}_0 + \int_0^t a_\tau(\tilde{X}^i,\tilde{Y}^i)\,\mathrm{d}\tau + \int_0^t b_\tau(\tilde{X}^i,\tilde{Y}^i)\,\mathrm{d}\tilde{W}_\tau + \int_0^t c_\tau(\tilde{X}^i,\tilde{Y}^i)\,\mathrm{d}\tilde{Y}_\tau^i, \qquad t \in [0,T], \tag{4.46}$$

and in particular $\{\tilde{X}_t^i\}$ is an $\mathbb{R}^d$-valued continuous $\tilde{\mathscr{F}}_t$-semimartingale. Now $\tilde{\mathrm{E}}|\tilde{\xi}_0|^4 < \infty$ (by Condition 2.2 and Remark 3.7), the coefficients in (4.46) are globally Lipschitz-continuous hence linearly bounded (by Condition 2.3), and $\{\gamma_t^0(\tilde{Y}^i)\}$ is uniformly bounded (by (3.30)). From this, together with (4.46), (3.38), (3.35), and a proof identical to that for no. 5.3.15 of Karatzas and Shreve ([14], p.306), we get the following bound which will often be used:

$$\tilde{\mathrm{E}}[\max_{t\in[0,T]} |\tilde{X}_t^i|^4] < \infty, \qquad i = 1,2. \tag{4.47}$$

Proof of (3.41): For each $t \in [0,T]$ and $n = 1,2,\ldots$, put

$$\alpha_t^n := \tilde{\mathrm{E}}[I[0,T_n)(t)|\Delta\tilde{X}_t|^2], \qquad \eta_t^n := \tilde{\mathrm{E}}[I[0,T_n)(t)|\tilde{l}_t|^2], \qquad \beta_t^n := K(n)\int_0^t \eta_\tau^n\,\mathrm{d}\tau, \tag{4.48}$$

where T_n and $K(n)$ are as asserted in Proposition 4.1. Then (4.45) gives

$$\alpha_t^n \le \beta_t^n + K(n)\int_0^t \alpha_\tau^n\,\mathrm{d}\tau, \qquad t \in [0,T], \tag{4.49}$$

for each $n = 1,2,\ldots$ From (4.47) it follows in particular that $t \mapsto \alpha_t^n$ is integrable on $t \in [0,T]$, and $t \mapsto \beta_t^n$ is of course integrable on $t \in [0,T]$ (in view of (3.30) and (3.39)). Hence (4.49) and Gronwall's inequality (Kallianpur [11], Proposition 5.1.1, p.94-95) give, for all $t \in [0,T]$,

$$\alpha_t^n \le \beta_t^n + K(n)\int_0^t e^{(t-\tau)K(n)}\beta_\tau^n\,\mathrm{d}\tau \le \beta_t^n + T\,K(n)\,e^{TK(n)}\beta_t^n = K_1(n)\beta_t^n, \tag{4.50}$$

where $K_1(n) := 1 + TK(n)\exp\{TK(n)\} \in [0,\infty)$, and we have used the fact that the mapping $t \mapsto \beta_t^n$ is nondecreasing at the second inequality of (4.50). In light of (4.50) and (4.48), we obtain the following integral inequalities: for each $n = 1, 2, \ldots$,

$$\alpha_t^n \leq K_2(n) \int_0^t \eta_\tau^n \, \mathrm{d}\tau, \qquad t \in [0,T], \tag{4.51}$$

with $K_2(n) := K_1(n)K(n) \in [0,\infty)$. From (4.51), (4.48) and (4.44), for each $t \in [0,T]$ and $n = 1, 2, \ldots$,

$$\eta_t^n \leq K(n) \left\{ (1 + K_2(n)) \int_0^t \eta_\tau^n \, \mathrm{d}\tau + K_2(n) \int_0^t \left[\int_0^\tau \eta_s^n \, \mathrm{d}s \right] \, \mathrm{d}\tau \right\}. \tag{4.52}$$

Since $\eta_s^n \geq 0$, $s \in [0,T]$, we have

$$\int_0^t [\int_0^\tau \eta_s^n \, \mathrm{d}s] \, \mathrm{d}\tau \leq t \int_0^t \eta_s^n \, \mathrm{d}s \leq T \int_0^t \eta_s^n \, \mathrm{d}s, \qquad t \in [0,T];$$

it follows from this inequality and (4.52) that there is a constant $K_3(n) \in [0,\infty)$ such that

$$\eta_t^n \leq K_3(n) \int_0^t \eta_\tau^n \, \mathrm{d}\tau, \qquad t \in [0,T], \tag{4.53}$$

and therefore, from (4.53), (4.48) and the Gronwall inequality, for each $t \in [0,T]$ we have

$$\tilde{\mathrm{E}}[I[0,T_n)(t) \, |\tilde{l}_t|^2] = 0, \qquad n = 1, 2, 3, \ldots \tag{4.54}$$

Now $\lim_{n\to\infty} T_n = T$ ($\tilde{P}$ − a.s.) and monotonically (by Proposition 4.1), and therefore, in view of (4.54) and the monotone convergence theorem, we obtain $\tilde{\mathrm{E}}|\tilde{l}_t|^2 = 0$ for each $t \in [0,T)$; now (3.41) follows from this and the Fubini theorem. □

Proof of Proposition 4.1: The proof involves constructing a sequence of $\tilde{\mathscr{Y}}_t$-stopping times T_n and constants $K(n)$ with the stated properties, and relies of course on the structure of the drift functional $\{\gamma_t^0\}$ (defined by (3.21), (3.22), (3.24), (3.25), (3.28) and (3.29)). Since the proof divides quite naturally into five distinct steps, we choose to present it as such (rather than breaking it up into a number of independent lemmas, propositions, etc).

Step 1: In this step we begin construction of the stopping times T_n asserted in Proposition 4.1. From Remark 3.2 and (3.37),

$$\tilde{P}(\tilde{Y}^i)^{-1} \text{ and } Q_2 \text{ are equivalent probability measures on } \mathscr{B}(C_D) \text{ for } i = 1, 2, \tag{4.55}$$

and hence, from this, together with (3.29) and (3.28),

$$\tilde{P}\left\{ \tilde{\omega} : \gamma_t^0(\tilde{Y}^i(\tilde{\omega})) = \frac{F_t(\tilde{Y}^i(\tilde{\omega}))}{G_t(\tilde{Y}^i(\tilde{\omega}))}, \qquad t \in [0,T] \right\} = 1, \qquad i = 1, 2. \tag{4.56}$$

Next, we need the following result, which follows immediately from Bhatt and Karandikar (see Theorem 4.1 and eqn. (4.4) on p.47 of [3]):

Theorem 4.3. There exists some $N \in \overline{\mathscr{B}(C_D)}^{Q_2}$ such that $Q_2(N) = 0$, and, for each $y \notin N$, the mappings $t \mapsto F_t(y)$ and $t \mapsto G_t(y)$ are continuous on $[0,T]$ with $\inf_{t\in[0,T]} G_t(y) > 0$.

Now use the set N from Theorem 4.3 to define processes $\{\tilde{F}_t(y)\}$ and $\{\tilde{G}_t(y)\}$ on the probability space $(C_D, \overline{\mathscr{B}(C_D)}^{Q_2}, Q_2)$ by

$$\begin{cases} \tilde{F}_t(y) := F_t(y) & \& \quad \tilde{G}_t(y) := G_t(y), \quad t \in [0,T], \text{ when } y \notin N, \\ \tilde{F}_t(y) := 0 & \& \quad \tilde{G}_t(y) := 1, \quad t \in [0,T], \text{ when } y \in N. \end{cases} \tag{4.57}$$

Since $Q_2(N) = 0$, and we have already seen that $\{F_t(y)\}$ and $\{G_t(y)\}$ are $\overline{\mathscr{B}_t(C_D)}^{Q_2}$-adapted (at (3.24), (3.25)), it follows that $\{\tilde{F}_t(y)\}$ and $\{\tilde{G}_t(y)\}$ are continuous and $\overline{\mathscr{B}_t(C_D)}^{Q_2}$-adapted processes on $(C_D, \overline{\mathscr{B}(C_D)}^{Q_2}, Q_2)$. Moreover, $\overline{\mathscr{B}_t(C_D)}^{Q_2} = \mathscr{B}_t(C_D) \vee \mathbf{Z}^{\tilde{P}(\tilde{Y}^i)^{-1}}[\mathscr{B}(C_D)]$ for $i = 1,2$, (as follows from (4.55) and (3.23)), thus each set $A \in \overline{\mathscr{B}_t(C_D)}^{Q_2}$ has the form $A = B \triangle C$ for some $B \in \mathscr{B}_t(C_D)$ and $C \in \mathbf{Z}^{\tilde{P}(\tilde{Y}^i)^{-1}}[\mathscr{B}(C_D)]$, and from this together with (4.43) it is seen that $\tilde{F}_t(\tilde{Y}^i)$ and $\tilde{G}_t(\tilde{Y}^i)$ are $\tilde{\mathscr{Y}}_t$-measurable, that is

$$\begin{cases} \{\tilde{F}_t(\tilde{Y}^i)\} \text{ and } \{\tilde{G}_t(\tilde{Y}^i)\} \text{ are continuous and } \tilde{\mathscr{Y}}_t\text{-adapted processes,} \\ \inf\limits_{t\in[0,T]} \tilde{G}_t(\tilde{Y}^i(\tilde{\omega})) > 0, \qquad \text{for each } \tilde{\omega} \in \tilde{\Omega}, \end{cases} \tag{4.58}$$

$$\tilde{P}\{\tilde{F}_t(\tilde{Y}^i) = F_t(\tilde{Y}^i),\ t \in [0,T]\} = \tilde{P}\{\tilde{G}_t(\tilde{Y}^i) = G_t(\tilde{Y}^i),\ t \in [0,T]\} = 1, \tag{4.59}$$

(the latter following from (4.55), (4.57) and $Q_2(N) = 0$). Then, from (4.59), (4.56), (3.39),

$$\tilde{P}\left\{\tilde{l}_t = \frac{\tilde{F}_t(\tilde{Y}^1) - \tilde{F}_t(\tilde{Y}^2)}{\tilde{G}_t(\tilde{Y}^1)} + \frac{\tilde{F}_t(\tilde{Y}^2)[\tilde{G}_t(\tilde{Y}^2) - \tilde{G}_t(\tilde{Y}^1)]}{\tilde{G}_t(\tilde{Y}^1)\tilde{G}_t(\tilde{Y}^2)},\ t \in [0,T]\right\} = 1. \tag{4.60}$$

In view of (4.58) we see that

$$\begin{cases} T_n^i(\tilde{\omega}) := \inf\{t \in [0,T] :\ \tilde{G}_t(\tilde{Y}^i(\tilde{\omega})) \le 1/n\} \wedge T, \qquad i = 1,2, \\ T_n^3(\tilde{\omega}) := \inf\{t \in [0,T] :\ |\tilde{F}_t(\tilde{Y}^2(\tilde{\omega}))| \ge n\} \wedge T, \end{cases} \tag{4.61}$$

are $\tilde{\mathscr{Y}}_t$-stopping times for all $n = 1,2,\ldots$, and

$$\lim_{n\to\infty} T_n^i(\tilde{\omega}) = T, \qquad \text{for each } \tilde{\omega} \in \tilde{\Omega},\ i = 1,2,3. \tag{4.62}$$

Now put

$$\tilde{l}_t^1 := \tilde{G}_t(\tilde{Y}^1) - \tilde{G}_t(\tilde{Y}^2), \qquad \tilde{l}_t^2 := \tilde{F}_t(\tilde{Y}^1) - \tilde{F}_t(\tilde{Y}^2). \tag{4.63}$$

Then it follows from (4.60), (4.61) and (4.63) that, for $\tilde{P}$-almost all $\tilde{\omega}$,

$$|\tilde{l}_t(\tilde{\omega})| \le n|\tilde{l}_t^2(\tilde{\omega})| + n^3|\tilde{l}_t^1(\tilde{\omega})|, \qquad 0 \le t \le (T_n^1 \wedge T_n^2 \wedge T_n^3)(\tilde{\omega}). \tag{4.64}$$

Step 2: In this step we work out upper-bounds for the quantities $|\tilde{l}_t^1|$ and $|\tilde{l}_t^2|$ defined at (4.63) (these upper-bounds are given by (4.79) and (4.82)).

By (4.42) and (3.21) we have $\{\vartheta_t(\tilde{\xi}_0, \tilde{W}, \tilde{Y}^i)\} = \{h_t(\tilde{X}^i, \tilde{Y}^i)\}$, and this process is continuous $\tilde{\mathscr{F}}_t$-adapted (by Condition 2.2 and the fact that $\{(\tilde{X}_t^i, \tilde{Y}_t^i)\}$ is continuous $\tilde{\mathscr{F}}_t$-adapted - see Remark 4.2

and Remark 3.7). Since $\{\tilde{Y}_t^i\}$ is also a $\tilde{\mathscr{F}}_t$-semimartingale, we get from Theorem 3.1(I)(b) that the process $\{\mathfrak{J}_t(\vartheta^k(\tilde{\xi}_0,\tilde{W},\tilde{Y}^i),(\tilde{Y}^i)^k)\}$ and stochastic integral $h^k(\tilde{X}^i,\tilde{Y}^i)\bullet(\tilde{Y}^i)^k$ are indistinguishable for each $k=1,2,\dots,D$. From these observations and (3.22), we find

$$\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^i)=\exp\left[\int_0^t h'_\tau(\tilde{X}^i,\tilde{Y}^i)\,\mathrm{d}\tilde{Y}^i_\tau-\frac{1}{2}\int_0^t |h_\tau(\tilde{X}^i,\tilde{Y}^i)|^2\,\mathrm{d}\tau\right]. \tag{4.65}$$

To simplify the notation, put

$$h_t^i:=h'_t(\tilde{X}^i,\tilde{Y}^i),\qquad i=1,2, \tag{4.66}$$

(thus h_t^i is D-dimensional row-vector), and observe from (4.65) and (3.38), together with the uniform boundedness of h and γ^0 (see Condition 2.3(iii) and (3.30)), that there is a constant $C\in[0,\infty)$ such that

$$\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^i)=\mathscr{E}(h^i\bullet\tilde{I})_t\,\exp\left[\int_0^t h_\tau^i\,\gamma_\tau^0(\tilde{Y}^i)\,\mathrm{d}\tau\right]\le C\mathscr{E}(h^i\bullet\tilde{I})_t,\qquad t\in[0,T], \tag{4.67}$$

(recall Notation 2.1(V), and the fact that $\{\tilde{I}_t\}$ is an $\mathbb{R}^D$-valued $\tilde{\mathscr{F}}_t$-Wiener process - see (3.35)). Since $\{h_t^i\}$ is uniformly bounded, continuous, and $\tilde{\mathscr{F}}_t$-adapted, we see that $\{\mathscr{E}(h^i\bullet\tilde{I})_t\}$ is a continuous L_p-bounded $\tilde{\mathscr{F}}_t$-martingale for each $p\in[1,\infty)$, hence Doob's maximal inequality and (4.67) establish

$$\tilde{\mathrm{E}}[\sup_{t\in[0,T]}\{\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^i)\}^\alpha]<\infty\quad\text{for each}\quad\alpha\in[1,\infty). \tag{4.68}$$

Now $\tilde{P}(\tilde{\xi}_0,\tilde{W})^{-1}=\pi_0\times Q_1$ (by (3.37)), while $\sigma\{\tilde{\xi}_0,\tilde{W}_t,\ t\in[0,T]\}$ and $\sigma\{\tilde{Y}^1,\tilde{Y}^2,\ t\in[0,T]\}$ are $\tilde{P}$-independent (by (3.36)). It then follows from the Fubini theorem for conditional expectations (Ethier and Kurtz [8], Proposition 4.5, p.498), (3.24) and (4.59), that, for each $t\in[0,T]$,

$$\tilde{\mathrm{E}}\left[\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^i)\,\middle|\,\tilde{\mathscr{Y}}_t\right]=G_t(\tilde{Y}^i)=\tilde{G}_t(\tilde{Y}^i),\qquad\tilde{P}-\text{a.s.} \tag{4.69}$$

and thus, from (4.69), (4.63) and Jensen's inequality, for each $t\in[0,T]$,

$$|\tilde{l}_t^1|\le\tilde{\mathrm{E}}\left[|\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^1)-\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^2)|\,\middle|\,\tilde{\mathscr{Y}}_t\right],\qquad\tilde{P}-\text{a.s.} \tag{4.70}$$

We next upper-bound the quantity on the right side of (4.70). To this end, put

$$\chi_t:=(1/2)[\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^1)+\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^2)]. \tag{4.71}$$

From the elementary upper-bound $|e^x-e^y|\le(1/2)(e^x+e^y)|x-y|$, $x,y\in\mathbb{R}$, together with (4.71), (4.66), (4.65), and (3.40), we find

$$\begin{aligned}|\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^1)-\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^2)|\quad&\le\quad\chi_t\left\{\left|\int_0^t[h_\tau^1-h_\tau^2]\,\mathrm{d}\tilde{Y}_\tau^1+\int_0^t h_\tau^2\,\tilde{l}_\tau\,\mathrm{d}\tau\right.\right.\\&+\quad\left.\left.\frac{1}{2}\int_0^t[|h_\tau^2|^2-|h_\tau^1|^2]\,\mathrm{d}\tau\right|\right\}.\end{aligned} \tag{4.72}$$

Since $|h^2_\tau|^2 - |h^1_\tau|^2 = [h^2_\tau + h^1_\tau][h^2_\tau - h^1_\tau]'$, we can use (3.38) and the Cauchy-Schwarz inequality for the $\mathrm{d}\tau$-integrals on the right side of (4.72) to get

$$\begin{aligned}
&|\rho_t(\tilde{\xi}_0, \tilde{W}, \tilde{Y}^1) - \rho_t(\tilde{\xi}_0, \tilde{W}, \tilde{Y}^2)| \\
\leq\ & \chi_t \left\{ \left| \int_0^t [h^1_\tau - h^2_\tau]\ \mathrm{d}\tilde{I}_\tau \right| + \left(\int_0^t |h^1_\tau - h^2_\tau|^2\ \mathrm{d}\tau \right)^{1/2} \left(\int_0^t |\gamma^0_\tau(\tilde{Y}^1)|^2\ \mathrm{d}\tau \right)^{1/2} \right. \\
+\ & \left(\int_0^t |h^2_\tau|^2\ \mathrm{d}\tau \right)^{1/2} \left(\int_0^t |\tilde{l}_\tau|^2\ \mathrm{d}\tau \right)^{1/2} \\
+\ & \left. \frac{1}{2} \left(\int_0^t |h^1_\tau - h^2_\tau|^2\ \mathrm{d}\tau \right)^{1/2} \left(\int_0^t |h^1_\tau + h^2_\tau|^2\ \mathrm{d}\tau \right)^{1/2} \right\}.
\end{aligned} \tag{4.73}$$

Since $\{\gamma^0_t(y)\}$ is $\mathscr{B}_t(C_D)$-predictable and $\{\tilde{Y}^i_t\}$ are continuous $\tilde{\mathscr{Y}}_t$-adapted, we see from (3.39) that $\{\tilde{l}_t\}$ is $\tilde{\mathscr{Y}}_t$-predictable, and hence $\int_0^t |\tilde{l}_\tau|^2\ \mathrm{d}\tau$ is $\tilde{\mathscr{Y}}_t$-measurable. Then, upon taking $\tilde{\mathscr{Y}}_t$-conditional expectations on each side of (4.73) and using the uniform-boundedness of $\{h^i_t\}$ and $\{\gamma^0_t(\tilde{Y}^1)\}$, for each $t \in [0, T]$ we get

$$\begin{aligned}
&\tilde{\mathrm{E}}\left[|\rho_t(\tilde{\xi}_0, \tilde{W}, \tilde{Y}^1) - \rho_t(\tilde{\xi}_0, \tilde{W}, \tilde{Y}^2)| \,\middle|\, \tilde{\mathscr{Y}}_t \right] \leq \tilde{\mathrm{E}}\left[\chi_t \left| \int_0^t [h^1_\tau - h^2_\tau]\ \mathrm{d}\tilde{I}_\tau \right| \,\middle|\, \tilde{\mathscr{Y}}_t \right] \\
&+ K_1 \tilde{\mathrm{E}}\left[\chi_t \left(\int_0^t |h^1_\tau - h^2_\tau|^2\ \mathrm{d}\tau \right)^{1/2} \,\middle|\, \tilde{\mathscr{Y}}_t \right] + K_1 \tilde{\mathrm{E}}\left[\chi_t \,\middle|\, \tilde{\mathscr{Y}}_t \right] \left(\int_0^t |\tilde{l}_\tau|^2\ \mathrm{d}\tau \right)^{1/2},\ \tilde{P} - \text{a.s.}
\end{aligned} \tag{4.74}$$

in which $K_1 \in [0, \infty)$ is a constant (depending only on the uniform bound $\|h\|$ on $\{h^i_t\}$ and $\{\gamma^0_t(\tilde{Y}^1)\}$ - see (3.30)). Next, consider each term on the right of (4.74). Put

$$\psi^1_t := \tilde{\mathrm{E}}\left[\sup_{t \in [0,T]} (\chi_t)^2 \,\middle|\, \tilde{\mathscr{Y}}_t \right], \tag{4.75}$$

and observe, from (4.68) and (4.71), that

$$\{\psi^1_t\} \text{ is a } \tilde{\mathscr{Y}}_t\text{-martingale such that } \sup_{t \in [0,T]} \tilde{\mathrm{E}}[(\psi^1_t)^\alpha] < \infty, \text{ for each } \alpha \in [1, \infty). \tag{4.76}$$

From (4.75) and the Cauchy-Schwarz inequality for conditional expectations (Chow and Teicher [4], Theorem 7.2.4, p.219) we obtain

$$\tilde{\mathrm{E}}\left[\chi_t \left| \int_0^t [h^1_\tau - h^2_\tau]\ \mathrm{d}\tilde{I}_\tau \right| \,\middle|\, \tilde{\mathscr{Y}}_t \right] \leq [\psi^1_t]^{1/2}\, \tilde{\mathrm{E}}^{1/2}\left[\left| \int_0^t [h^1_\tau - h^2_\tau]\ \mathrm{d}\tilde{I}_\tau \right|^2 \,\middle|\, \tilde{\mathscr{Y}}_t \right],\ \tilde{P} - \text{a.s.} \tag{4.77}$$

$$\tilde{\mathrm{E}}\left[\chi_t \left(\int_0^t |h^1_\tau - h^2_\tau|^2\ \mathrm{d}\tau \right)^{1/2} \,\middle|\, \tilde{\mathscr{Y}}_t \right] \leq [\psi^1_t]^{1/2}\, \tilde{\mathrm{E}}^{1/2}\left[\int_0^t |h^1_\tau - h^2_\tau|^2\ \mathrm{d}\tau \,\middle|\, \tilde{\mathscr{Y}}_t \right],\ \tilde{P} - \text{a.s.} \tag{4.78}$$

Now substitute (4.77) - (4.78) into (4.74) and square both sides. From this, together with the upper-bound $[\sum_{i=1}^m a_i]^2 \le m^2 \sum_{i=1}^m a_i^2$ (for $a_i \in \mathbb{R}$) and (4.70), for each $t \in [0,T]$ we get

$$
\begin{aligned}
|\tilde{l}_t^1|^2 &\le \left(\tilde{\mathrm{E}}\left[|\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^1)-\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^2)|\,\Big|\,\tilde{\mathscr{Y}}_t\right]\right)^2 \qquad (4.79)\\
&\le K_2\psi_t^1\left\{\tilde{\mathrm{E}}\left[\left|\int_0^t [h_\tau^1-h_\tau^2]\,\mathrm{d}\tilde{I}_\tau\right|^2\,\Bigg|\,\tilde{\mathscr{Y}}_t\right]\right.\\
&+ \left.\tilde{\mathrm{E}}\left[\int_0^t |h_\tau^1-h_\tau^2|^2\,\mathrm{d}\tau\,\Bigg|\,\tilde{\mathscr{Y}}_t\right]+\int_0^t |\tilde{l}_\tau|^2\,\mathrm{d}\tau\right\}, \qquad \tilde{P}-\text{a.s.}
\end{aligned}
$$

the constant $K_2 \in [0,\infty)$ depending only on the uniform bound $\|h\|$ on $\{h_t^i\}$ and $\{\gamma_t^0(\tilde{Y}^1)\}$.

We next establish a similar upper-bound on $|\tilde{l}_t^2|$. From (4.42) and an argument identical to that which led to (4.69) we obtain

$$
\tilde{\mathrm{E}}\left[h_t(\tilde{X}^i,\tilde{Y}^i)\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^i)\,\Big|\,\tilde{\mathscr{Y}}_t\right]=\tilde{F}_t(\tilde{Y}^i), \qquad \tilde{P}-\text{a.s.}
$$

Then it follows from (4.63) that

$$
\begin{aligned}
\tilde{l}_t^2 &= \tilde{\mathrm{E}}\left[h_t(\tilde{X}^1,\tilde{Y}^1)[\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^1)-\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^2)]\,\Big|\,\tilde{\mathscr{Y}}_t\right]\\
&+ \tilde{\mathrm{E}}\left[[h_t(\tilde{X}^1,\tilde{Y}^1)-h_t(\tilde{X}^2,\tilde{Y}^2)]\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^2)\,\Big|\,\tilde{\mathscr{Y}}_t\right], \qquad \tilde{P}-\text{a.s.}
\end{aligned}
$$

and then, from Jensen's inequality, for each $t \in [0,T]$ we get

$$
\begin{aligned}
|\tilde{l}_t^2| &\le K_3\tilde{\mathrm{E}}\left[|\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^1)-\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^2)|\,\Big|\,\tilde{\mathscr{Y}}_t\right] \qquad (4.80)\\
&+ \tilde{\mathrm{E}}\left[|h_t(\tilde{X}^1,\tilde{Y}^1)-h_t(\tilde{X}^2,\tilde{Y}^2)|\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^2)\,\Big|\,\tilde{\mathscr{Y}}_t\right], \qquad \tilde{P}-\text{a.s.}
\end{aligned}
$$

for a constant $K_3 \in [0,\infty)$ depending only on the uniform bound on h (Condition 2.3(iii)). From the Cauchy-Schwarz inequality for conditional expectations, together with (4.71), (4.66),

$$
\tilde{\mathrm{E}}\left[|h_t(\tilde{X}^1,\tilde{Y}^1)-h_t(\tilde{X}^2,\tilde{Y}^2)|\rho_t(\tilde{\xi}_0,\tilde{W},\tilde{Y}^2)\,\Big|\,\tilde{\mathscr{Y}}_t\right] \le 2[\psi_t^1]^{1/2}\,\tilde{\mathrm{E}}^{1/2}\left[|h_t^1-h_t^2|^2\,\Big|\,\tilde{\mathscr{Y}}_t\right]. \qquad (4.81)
$$

From (4.81), (4.80), and (the second inequality of) (4.79), for each $t \in [0,T]$ we have

$$
\begin{aligned}
|\tilde{l}_t^2|^2 &\le K_4\psi_t^1\left\{\tilde{\mathrm{E}}\left[|h_t^1-h_t^2|^2\,\Big|\,\tilde{\mathscr{Y}}_t\right]+\tilde{\mathrm{E}}\left[\left|\int_0^t [h_\tau^1-h_\tau^2]\,\mathrm{d}\tilde{I}_\tau\right|^2\,\Bigg|\,\tilde{\mathscr{Y}}_t\right]\right. \qquad (4.82)\\
&+ \left.\tilde{\mathrm{E}}\left[\int_0^t |h_\tau^1-h_\tau^2|^2\,\mathrm{d}\tau\,\Bigg|\,\tilde{\mathscr{Y}}_t\right]+\int_0^t |\tilde{l}_\tau|^2\,\mathrm{d}\tau\right\}, \qquad \tilde{P}-\text{a.s.}
\end{aligned}
$$

for a constant $K_4 \in [0,\infty)$ depending only on the uniform bound $\|h\|$.

Step 3: In this step we complete construction of the stopping times T_n in Proposition 4.1 (by introducing stopping times for the martingale $\{\psi_t^1\}$ at (4.75) and for the martingale $\{\psi_t^2\}$ to be defined at (4.86)).

In view of (4.43), the filtered probability space $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P}; \{\tilde{\mathscr{Y}}_t\})$ satisfies the usual conditions (i.e. is an R-filtered probability space in the sense of [21], Definition II(67.1), p.172), and thus, from (4.76) and ([21], Theorem II(67.7), p.173), $\{\psi_t^1\}$ has a modification which is a cadlag process (or R-process in the sense of [21], Definitions II(62.1) and II(63.5), p. 163 and p.167). Since (4.79) and (4.82) hold $\tilde{P}$ − a.s. for each $t \in [0, T]$, with no loss of generality we shall always suppose that $\{\psi_t^1\}$ is a $\tilde{\mathscr{Y}}_t$-adapted cadlag process. Now define

$$T_n^4 := \inf\{t \in [0, T] : \psi_t^1 > n\} \wedge T. \tag{4.83}$$

Since $\{\tilde{\mathscr{Y}}_t\}$ is a right-continuous filtration we then see that T_n^4 is a $\tilde{\mathscr{Y}}_t$-stopping time ([21], Lemma II(74.3), p.184). Moreover, from (4.76) and Doob's maximal inequality, we obtain $\tilde{\mathrm{E}}[\sup_{t\in[0,T]}(\psi_t^1)^\alpha] < \infty$ for each $\alpha > 1$, thus $\sup_{t\in[0,T]} \psi_t^1 < \infty$, $\tilde{P}$ − a.s., and hence

$$\lim_{n\to\infty} T_n^4 = T, \qquad \tilde{P} - \text{a.s.} \tag{4.84}$$

Next, put

$$\zeta := \int_0^T |c_\tau(\tilde{X}^2, \tilde{Y}^2)|^2 \, \mathrm{d}\tau. \tag{4.85}$$

Now $\tilde{\mathrm{E}}[\max_{t\in[0,T]} |\tilde{Y}_t^2|^4] < \infty$ (see (3.38), (3.35) and (3.30)); from this, together with (4.47) and the fact that coefficient c is linearly bounded (being globally Lipschitz-continuous - see Condition 2.3), we obtain $\mathrm{E}|\zeta|^2 < \infty$. For each $t \in [0, T]$ put

$$\psi_t^2 := \tilde{\mathrm{E}}\left[\zeta \mid \tilde{\mathscr{Y}}_t\right]. \tag{4.86}$$

Then $\{\psi_t^2\}$ is a square-integrable $\tilde{\mathscr{Y}}_t$-martingale, which, without loss in generality, we shall suppose is a cadlag process. Then

$$T_n^5 := \inf\{t \in [0, T] : \psi_t^2 > n\} \wedge T, \tag{4.87}$$

defines a $\tilde{\mathscr{Y}}_t$-stopping time, and, from the square-integrability of ζ together with Doob's maximal inequality, we find that

$$\lim_{n\to\infty} T_n^5 = T, \qquad \tilde{P} - \text{a.s.} \tag{4.88}$$

(exactly as at (4.84)). Recalling (4.61) and (4.83), put

$$T_n := \bigwedge_{k=1}^{5} T_n^k. \tag{4.89}$$

Then T_n is a $\tilde{\mathscr{Y}}_t$-stopping time (since the T_n^k are $\tilde{\mathscr{Y}}_t$-stopping times), and, from (4.88), (4.84) and (4.62) we see that $\lim_{n\to\infty} T_n = T$, $\tilde{P}$ − a.s., and monotonically, as required.

Step 4: In this step we use the inequalities (4.79) and (4.82) (that were obtained in Step 2) to establish (4.44) for some constants $K(n) \in [0, \infty)$ (and stopping times T_n given by (4.89)). From (4.64) and (4.89) we have

$$|\tilde{l}_t|^2 I[0, T_n)(t) \le 2n^6 |\tilde{l}_t^1|^2 I[0, T_n)(t) + 2n^2 |\tilde{l}_t^2|^2 I[0, T_n)(t), \qquad \tilde{P} - \text{a.s.} \tag{4.90}$$

for each $n = 1, 2, \ldots$ and $t \in [0, T]$. We next use the inequalities (4.79) and (4.82) obtained in Step 2 to upper-bound each term on the right of (4.90). From (4.89) and (4.83) we have the upper-bound

$\psi^1_t(\tilde\omega)I[0,T_n(\tilde\omega))(t)\le n$ for all $(t,\tilde\omega)\in[0,T]\times\tilde\Omega$, thus, from (4.79) and the $\tilde{\mathscr{Y}}_t$-measurability of $I[0,T_n)(t)$, for each $n=1,2,\dots$ and $t\in[0,T]$,

$$
\begin{aligned}
|\tilde l^1_t|^2 I[0,T_n)(t) \quad &\le \quad K_2 n\left\{\tilde{\mathrm{E}}\left[I[0,T_n)(t)\left|\int_0^t [h^1_\tau-h^2_\tau]\,\mathrm{d}\tilde I_\tau\right|^2\,\middle|\,\tilde{\mathscr{Y}}_t\right]\right. \\
&+\quad \left.\tilde{\mathrm{E}}\left[I[0,T_n)(t)\int_0^t |h^1_\tau-h^2_\tau|^2\,\mathrm{d}\tau\,\middle|\,\tilde{\mathscr{Y}}_t\right]+I[0,T_n)(t)\int_0^t|\tilde l_\tau|^2\,\mathrm{d}\tau\right\}.
\end{aligned}
\tag{4.91}
$$

Since $\tilde{\mathscr{Y}}_t\subset\tilde{\mathscr{F}}_t$ (from (4.43)) we see that T_n is a $\tilde{\mathscr{F}}_t$-stopping time. But $\{h^1_t-h^2_t\}$ is $\tilde{\mathscr{F}}_t$-progressively measurable (recall (4.66)), thus it follows from (3.35) and Karatzas and Shreve ([14], (2.24) of Proposition 3.2.10, p.139-140) that

$$
\begin{aligned}
I[0,T_n)(t)\left|\int_0^t [h^1_\tau-h^2_\tau]\,\mathrm{d}\tilde I_\tau\right|^2 \quad &= \quad \left|I[0,T_n)(t)\int_0^{t\wedge T_n}[h^1_\tau-h^2_\tau]\,\mathrm{d}\tilde I_\tau\right|^2 \\
&\le \quad \left|\int_0^t I[0,T_n](\tau)\,[h^1_\tau-h^2_\tau]\,\mathrm{d}\tilde I_\tau\right|^2,
\end{aligned}
\tag{4.92}
$$

hence, upon taking expectations in (4.92), and using (3.35) together with the Itô isometry and Fubini theorem, for each $t\in[0,T]$ we have

$$
\tilde{\mathrm{E}}\left[I[0,T_n)(t)\left|\int_0^t [h^1_\tau-h^2_\tau]\,\mathrm{d}\tilde I_\tau\right|^2\right]\le\int_0^t\tilde{\mathrm{E}}[I[0,T_n)(\tau)\,|h^1_\tau-h^2_\tau|^2]\,\mathrm{d}\tau. \tag{4.93}
$$

Moreover, since $I[0,T_n)(t)\le I[0,T_n)(\tau)$ when $0\le\tau\le t$, it is clear that

$$
\tilde{\mathrm{E}}\left[I[0,T_n)(t)\int_0^t|h^1_\tau-h^2_\tau|^2\,\mathrm{d}\tau\right]\le\int_0^t\tilde{\mathrm{E}}[I[0,T_n)(\tau)\,|h^1_\tau-h^2_\tau|^2]\,\mathrm{d}\tau, \tag{4.94}
$$

and from (4.94), (4.93) and (4.91), for each $t\in[0,T]$ and $n=1,2,\dots$,

$$
\begin{aligned}
\tilde{\mathrm{E}}[|\tilde l^1_t|^2 I[0,T_n)(t)] \quad &\le \quad 2K_2 n\left\{\int_0^t\tilde{\mathrm{E}}[I[0,T_n)(\tau)|h^1_\tau-h^2_\tau|^2]\,\mathrm{d}\tau\right. \\
&+ \quad \left.\int_0^t\tilde{\mathrm{E}}[I[0,T_n)(\tau)\,|\tilde l_\tau|^2]\,\mathrm{d}\tau\right\}.
\end{aligned}
\tag{4.95}
$$

Similarly, from (4.82), (4.93), (4.94) and the $\tilde{\mathscr{Y}}_t$-measurability of $I[0,T_n)(t)$, we obtain

$$
\begin{aligned}
\tilde{\mathrm{E}}[|\tilde l^2_t|^2 I[0,T_n)(t)] \quad &\le \quad 2K_4 n\left\{\int_0^t\tilde{\mathrm{E}}[I[0,T_n)(\tau)|h^1_\tau-h^2_\tau|^2]\,\mathrm{d}\tau\right. \\
&+ \quad \left.\int_0^t\tilde{\mathrm{E}}[I[0,T_n)(\tau)\,|\tilde l_\tau|^2]\,\mathrm{d}\tau+\tilde{\mathrm{E}}[I[0,T_n)(t)|h^1_t-h^2_t|^2]\right\},
\end{aligned}
\tag{4.96}
$$

and, upon combining (4.96), (4.95) and (4.90), for each $t \in [0, T]$ and $n = 1, 2, \ldots$, we have

$$\begin{aligned}\tilde{\mathrm{E}}[|\tilde{l}_t|^2 I[0,T_n)(t)] \quad &\leq \quad K_5 n^7 \left\{ \int_0^t \tilde{\mathrm{E}}[I[0,T_n)(\tau)|h^1_\tau - h^2_\tau|^2]\, \mathrm{d}\tau \right. \\ &+ \quad \left. \int_0^t \tilde{\mathrm{E}}[I[0,T_n)(\tau)\,|\tilde{l}_\tau|^2]\, \mathrm{d}\tau + \tilde{\mathrm{E}}[I[0,T_n)(t)|h^1_t - h^2_t|^2] \right\},\end{aligned} \tag{4.97}$$

in which $K_5 \in [0, \infty)$ is a constant. Next consider the first and third terms on the right of (4.97). From Condition 2.3, (3.40), (4.42), (4.66) and the Cauchy-Schwarz inequality,

$$\begin{aligned}|h^1_t - h^2_t|^2 I[0,T_n)(t) \quad &\leq \quad 2K^2 \left\{ \sup_{s \leq t} |\tilde{X}^1_s - \tilde{X}^2_s|^2 + \sup_{s \leq t} |\tilde{Y}^1_s - \tilde{Y}^2_s|^2 \right\} I[0,T_n)(t) \\ &\leq \quad 2K^2(T+1) \left\{ |\Delta\tilde{X}_t|^2 I[0,T_n)(t) + \int_0^t I[0,T_n)(\tau)\,|\tilde{l}_\tau|^2\, \mathrm{d}\tau \right\},\end{aligned} \tag{4.98}$$

(since $I[0,T_n)(t) \leq I[0,T_n)(\tau)$ when $0 \leq \tau \leq t$). Now it is clear that

$$\int_0^t \left[\int_0^\tau I[0,T_n)(s)\,|\tilde{l}_s|^2\, \mathrm{d}s \right] \mathrm{d}\tau \leq t \int_0^t I[0,T_n)(s)|\tilde{l}_s|^2\, \mathrm{d}s, \qquad t \in [0, T], \tag{4.99}$$

and upon inserting (4.98) into (4.97) and then using (4.99), we see that there is a constant $K_6 \in [0, \infty)$ such that (4.44) holds for each $t \in [0, T]$ and $n = 1, 2, \ldots$, with $K(n) := K_6 n^7$.

Step 5: In this step we establish (4.45) for some constants $K(n) \in [0, \infty)$ (with the stopping times T_n given by (4.89)). From (4.46), (3.40) and (3.38),

$$\begin{aligned}\tilde{X}^1_t - \tilde{X}^2_t \quad &= \quad \int_0^t [a_\tau(\tilde{X}^1, \tilde{Y}^1) - a_\tau(\tilde{X}^2, \tilde{Y}^2)]\, \mathrm{d}\tau + \int_0^t [b_\tau(\tilde{X}^1, \tilde{Y}^1) - b_\tau(\tilde{X}^2, \tilde{Y}^2)]\, \mathrm{d}\tilde{W}_\tau \\ &+ \quad \int_0^t [c_\tau(\tilde{X}^1, \tilde{Y}^1) - c_\tau(\tilde{X}^2, \tilde{Y}^2)]\, \mathrm{d}\tilde{Y}^1_\tau + \int_0^t c_\tau(\tilde{X}^2, \tilde{Y}^2)\, \mathrm{d}(\tilde{Y}^1 - \tilde{Y}^2)_\tau \\ &= \quad \int_0^t [a_\tau(\tilde{X}^1, \tilde{Y}^1) - a_\tau(\tilde{X}^2, \tilde{Y}^2)]\, \mathrm{d}\tau + \int_0^t [b_\tau(\tilde{X}^1, \tilde{Y}^1) - b_\tau(\tilde{X}^2, \tilde{Y}^2)]\, \mathrm{d}\tilde{W}_\tau \qquad (4.100) \\ &+ \quad \int_0^t [c_\tau(\tilde{X}^1, \tilde{Y}^1) - c_\tau(\tilde{X}^2, \tilde{Y}^2)]\, \mathrm{d}\tilde{I}_\tau + \int_0^t [c_\tau(\tilde{X}^1, \tilde{Y}^1) - c_\tau(\tilde{X}^2, \tilde{Y}^2)]\, \gamma^0_\tau(\tilde{Y}^1)\, \mathrm{d}\tau \\ &+ \quad \int_0^t c_\tau(\tilde{X}^2, \tilde{Y}^2)\, \tilde{l}_\tau\, \mathrm{d}\tau.\end{aligned}$$

In view of this, together with (4.42), (4.85), and the Cauchy-Schwarz inequality, we find a constant

$K_7 \in [0,\infty)$ (depending only on T) such that, for each $t \in [0,T]$, $n = 1,2,\ldots$,

$$\begin{aligned} |\Delta\tilde{X}_t|^2 I[0,T_n)(t) \quad \leq \quad & K_7 \Bigg\{ I[0,T_n)(t) \int_0^t |a_\tau(\tilde{X}^1,\tilde{Y}^1) - a_\tau(\tilde{X}^2,\tilde{Y}^2)|^2 \, \mathrm{d}\tau \\ & + \; I[0,T_n)(t) \sup_{s\leq t} \left| \int_0^s [b_\tau(\tilde{X}^1,\tilde{Y}^1) - b_\tau(\tilde{X}^2,\tilde{Y}^2)] \, \mathrm{d}\tilde{W}_\tau \right|^2 \\ & + \; I[0,T_n)(t) \sup_{s\leq t} \left| \int_0^s [c_\tau(\tilde{X}^1,\tilde{Y}^1) - c_\tau(\tilde{X}^2,\tilde{Y}^2)] \, \mathrm{d}\tilde{I}_\tau \right|^2 \\ & + \; I[0,T_n)(t) \left(\int_0^t |c_\tau(\tilde{X}^1,\tilde{Y}^1) - c_\tau(\tilde{X}^2,\tilde{Y}^2)|^2 \, \mathrm{d}\tau \right) \left(\int_0^t |\gamma^0_\tau(\tilde{Y}^1)|^2 \, \mathrm{d}\tau \right) \\ & + \; I[0,T_n)(t) \, \zeta \left(\int_0^t |\tilde{l}_\tau|^2 \, \mathrm{d}\tau \right) \Bigg\}. \end{aligned} \tag{4.101}$$

Now upper-bound each term on the right of (4.101). It follows from Condition 2.3, reasoning identical to that used at (4.98), and (4.99), that

$$\begin{aligned} & I[0,T_n)(t) \int_0^t |a_\tau(\tilde{X}^1,\tilde{Y}^1) - a_\tau(\tilde{X}^2,\tilde{Y}^2)|^2 \, \mathrm{d}\tau \\ & \leq 2K^2(T+1)^2 \left\{ \int_0^t |\Delta\tilde{X}_\tau|^2 I[0,T_n)(\tau) \, \mathrm{d}\tau + \int_0^t I[0,T_n)(\tau)|\tilde{l}_\tau|^2 \, \mathrm{d}\tau \right\}. \end{aligned} \tag{4.102}$$

As for the second and third terms on the right of (4.101), put

$$\Delta b_t := b_t(\tilde{X}^1,\tilde{Y}^1) - b_t(\tilde{X}^2,\tilde{Y}^2), \qquad \Delta c_t := c_t(\tilde{X}^1,\tilde{Y}^1) - c_t(\tilde{X}^2,\tilde{Y}^2). \tag{4.103}$$

Then, again since $I[0,T_n)(t) \leq I[0,T_n)(s)$ when $0 \leq s \leq t$, we have

$$\begin{aligned} I[0,T_n)(t) \sup_{s\leq t} \left| \int_0^s \Delta b_\tau \, \mathrm{d}\tilde{W}_\tau \right|^2 \quad & \leq \quad \sup_{s\leq t} I[0,T_n)(s) \left| \int_0^s \Delta b_\tau \, \mathrm{d}\tilde{W}_\tau \right|^2 \\ & \leq \quad \sup_{s\leq t} \left| \int_0^s I[0,T_n](\tau) \Delta b_\tau \, \mathrm{d}\tilde{W}_\tau \right|^2, \end{aligned} \tag{4.104}$$

where the second inequality of (4.104) follows exactly as at (4.92). For the third term on the right of (4.101) we clearly have a bound identical to that of (4.104) but with Δc_τ and $\tilde{I}$ in place of Δb_τ and $\tilde{W}$ respectively. For the fourth term on the right of (4.101), it follows from the uniform-boundedness of $\{\gamma^0_t(\tilde{Y}^1)\}$ and the same reasoning which led to (4.102), that

$$\begin{aligned} & I[0,T_n)(t) \left(\int_0^t |\Delta c_\tau|^2 \, \mathrm{d}\tau \right) \left(\int_0^t |\gamma^0_\tau(\tilde{Y}^1)|^2 \, \mathrm{d}\tau \right) \\ & \leq CK^2(T+1)^2 \left\{ \int_0^t |\Delta\tilde{X}_\tau|^2 I[0,T_n)(\tau) \, \mathrm{d}\tau + \int_0^t I[0,T_n)(\tau)|\tilde{l}_\tau|^2 \, \mathrm{d}\tau \right\} \end{aligned} \tag{4.105}$$

for some constant $C \in [0, \infty)$ (depending only on T and the uniform bound on γ^0). Upon combining (4.101), (4.102), (4.104), and (4.105), we find a constant $K_8 \in [0, \infty)$ such that, for each $t \in [0, T]$ and $n = 1, 2, \ldots$,

$$\begin{aligned} |\Delta\tilde{X}_t|^2 I[0,T_n)(t) \quad &\leq \quad K_8 \left\{ \int_0^t |\Delta\tilde{X}_\tau|^2 I[0,T_n)(\tau)\, \mathrm{d}\tau + \int_0^t |\tilde{l}_\tau|^2 I[0,T_n)(\tau)\, \mathrm{d}\tau \right. \\ &+ \quad \sup_{s\leq t} \left| \int_0^s I[0,T_n](\tau)\Delta b_\tau \, \mathrm{d}\tilde{W}_\tau \right|^2 + \sup_{s\leq t} \left| \int_0^s I[0,T_n](\tau)\Delta c_\tau \, \mathrm{d}\tilde{I}_\tau \right|^2 \\ &+ \quad \left. \zeta\, I[0,T_n)(t) \int_0^t I[0,T_n)(\tau) |\tilde{l}_\tau|^2 \, \mathrm{d}\tau \right\}. \end{aligned} \tag{4.106}$$

We next take expectations on each side of (4.106). To this end, from Doob's L^2-maximal inequality, (3.35), and the Itô isometry, we get

$$\tilde{\mathrm{E}} \left[\sup_{s\leq t} \left| \int_0^s I[0,T_n](\tau)\Delta b_\tau \, \mathrm{d}\tilde{W}_\tau \right|^2 \right] \leq 4 \int_0^t \tilde{\mathrm{E}}[I[0,T_n)(\tau)\, |\Delta b_\tau|^2]\, \mathrm{d}\tau, \tag{4.107}$$

and, exactly as at (4.98) together with Fubini's theorem, we obtain

$$\tilde{\mathrm{E}}[|\Delta b_\tau|^2 I[0,T_n)(\tau)] \leq 2K^2(T+1) \left\{ \tilde{\mathrm{E}}[|\Delta\tilde{X}_\tau|^2 I[0,T_n)(\tau)] + \int_0^\tau \tilde{\mathrm{E}}[|\tilde{l}_s|^2 I[0,T_n)(s)]\, \mathrm{d}s \right\}.$$

Combining this with (4.107) then establishes

$$\begin{aligned} \tilde{\mathrm{E}} \left[\sup_{s\leq t} \left| \int_0^s I[0,T_n](\tau)\Delta b_\tau \, \mathrm{d}\tilde{W}_\tau \right|^2 \right] \quad &\leq \quad 8K^2(T+1) \left\{ \int_0^t \tilde{\mathrm{E}}[I[0,T_n)(\tau)|\Delta\tilde{X}_\tau|^2]\, \mathrm{d}\tau \right. \\ &+ \quad \left. T \int_0^t \tilde{\mathrm{E}}[I[0,T_n)(\tau)|\tilde{l}_\tau|^2]\, \mathrm{d}\tau \right\}, \end{aligned} \tag{4.108}$$

and an identical upper-bound holds for the expectation of the fourth term on the right side of (4.106). As for the fifth term on the right of (4.106), for each $t \in [0, T]$ we have

$$\begin{aligned} &\tilde{\mathrm{E}} \left[\zeta\, I[0,T_n)(t) \int_0^t I[0,T_n)(\tau)|\tilde{l}_\tau|^2 \, \mathrm{d}\tau \,\middle|\, \tilde{\mathscr{Y}}_t \right] \\ &= \psi_t^2 I[0,T_n)(t) \int_0^t I[0,T_n)(\tau)|\tilde{l}_\tau|^2 \, \mathrm{d}\tau \leq n \int_0^t I[0,T_n)(\tau)\, |\tilde{l}_\tau|^2 \, \mathrm{d}\tau, \qquad \tilde{P} - \text{a.s.} \end{aligned} \tag{4.109}$$

where we have used the $\tilde{\mathscr{Y}}_t$-measurability of $I[0,T_n)(t)\int_0^t I[0,T_n)(\tau)|\tilde{l}_\tau|^2\, \mathrm{d}\tau$ and (4.86) at the equality in (4.109), and (4.89) and (4.87) at the inequality in (4.109). Taking expectations in (4.109) then gives, for each $t \in [0, T]$ and $n = 1, 2, \ldots$,

$$\tilde{\mathrm{E}} \left[\zeta I[0,T_n)(t) \int_0^t I[0,T_n)(\tau)|\tilde{l}_\tau|^2 \, \mathrm{d}\tau \right] \leq n \int_0^t \tilde{\mathrm{E}}[I[0,T_n)(\tau)|\tilde{l}_\tau|^2]\, \mathrm{d}\tau. \tag{4.110}$$

Combining (4.110), (4.108), and (4.106), we see that there is a constant $K(n) \in [0, \infty)$ such that (4.45) holds for each $t \in [0, T]$ and $n = 1, 2, \ldots$ □

Remark 4.4. As we have noted at Remark 3.4(II), establishing the innovations conjecture is really a matter of showing that $\{(\Omega, \mathscr{F}, P; \{\mathscr{Y}_t\}), (Y_t, I_t)\}$ is a strong solution of the SDE with drift function $\{\gamma_t^0\}$ and unit covariance (see Remark 3.4(I)), and this in turn follows from Theorem 3.5 once we have shown that this SDE has the property of pathwise uniqueness. Although global Lipschitz-continuity is postulated for the coefficients of the state/observation equation (2.7) (see Condition 2.3) it is well to note that the complexity of the drift term $\{\gamma_t^0\}$ means that standard results on pathwise-uniqueness for classical Itô SDEs (see e.g. Theorem 5.1.1 of [Kallianpur [11], p.97]) do not apply directly to the SDE with unit covariance and drift $\{\gamma_t^0\}$. In fact, the global Lipschitz-continuity of Condition 2.3 is essential for securing the representation of the drift $\{\gamma_t^0\}$ due to Bhatt and Karandikar [3], but is otherwise used only rather indirectly in the proof of pathwise-uniqueness (at (4.98), (4.102), (4.105) and (4.108)).

5 Pathwise-uniqueness for the Normalized Filter Equation

In this section we shall use Proposition 2.6 to establish *pathwise-uniqueness* for the measure-valued SDE which gives the nonlinear filter for a simplified version of the model (2.7). To this end we shall suppose from now on that $\{X_t,\ t \in [0,T]\}$ is an $\mathbb{R}^d$-valued "state" process and $\{Y_t,\ t \in [0,T]\}$ is an $\mathbb{R}^D$-valued observation process given by

$$\begin{cases} (1') \quad \mathrm{d}X_t = a(X_t)\,\mathrm{d}t + b(X_t)\,\mathrm{d}W_t^1 + c(X_t)\,\mathrm{d}Y_t, \qquad X_0 = \xi_0, \\ (2') \quad \mathrm{d}Y_t = h(X_t)\,\mathrm{d}t + \mathrm{d}W_t^2, \qquad Y_0 = 0. \end{cases} \tag{5.111}$$

subject to Condition 2.2 as well as

Condition 5.1. The mappings $a : \mathbb{R}^d \to \mathbb{R}^d$, $b : \mathbb{R}^d \to \mathbb{R}^{d\times q}$, $c : \mathbb{R}^d \to \mathbb{R}^{d\times D}$ and $h : \mathbb{R}^d \to \mathbb{R}^D$ are globally Lipschitz-continuous, h is uniformly bounded, and the $d \times d$-matrix $b(x)b'(x)$ is strictly positive definite for each $x \in \mathbb{R}^d$.

Remark 5.2. The model (5.111) is just a special case of (2.7), and therefore Remark 2.4 continues to hold for (5.111). We define the observation filtration $\{\mathscr{Y}_t\}$, the innovations process $\{I_t\}$, and the innovations filtration $\{\mathscr{I}_t\}$, exactly as in Remark 2.5, except that $\{(X_t, Y_t)\}$ is now given by (5.111) (in place of (2.7)), and we put

$$\beta_t := h(X_t), \qquad t \in [0,T], \tag{5.112}$$

(in place of (2.8)). From Proposition 2.6 we then have $\mathscr{Y}_t = \mathscr{I}_t$, $t \in [0,T]$.

Our goal is to establish pathwise-uniqueness for the measure-valued SDE giving the optimal nonlinear filter for the model (5.111). For completeness we briefly recall this equation together with the related ideas of a solution, pathwise-uniqueness, and uniqueness-in-law. From (5.111) we see that $\{X_t\}$ is a diffusion process given by

$$\mathrm{d}X_t = (a + ch)(X_t)\,\mathrm{d}t + c(X_t)\,\mathrm{d}W_t^2 + b(X_t)\,\mathrm{d}W_t^1, \qquad X_0 = \xi_0, \tag{5.113}$$

for which the corresponding linear second-order differential operator is

$$\mathcal{A}\phi(x) = \sum_{i=1}^{d}[a + ch]^i(x)\partial_i\phi(x) + \frac{1}{2}\sum_{i,j=1}^{d}[m(x)m'(x)]^{ij}\partial_i\partial_j\phi(x), \tag{5.114}$$

for all $x \in \mathbb{R}^d$ and $\phi \in C^\infty(\mathbb{R}^d)$, with $m : \mathbb{R}^d \to \mathbb{R}^{d\times(D+q)}$ given by $m(x) := \begin{bmatrix} c(x) & b(x) \end{bmatrix}$, $x \in \mathbb{R}^d$. To account for the dependence of the state-process $\{X_t\}$ on the Wiener process $\{W_t^2\}$ in the observation equation (5.111)(2′) define the linear first-order differential operator

$$\mathcal{B}_k\phi(x) := \sum_{i=1}^{d} c^{ik}(x)\partial_i\phi(x), \qquad k = 1, \dots, D, \tag{5.115}$$

for all $x \in \mathbb{R}^d$ and $\phi \in C^\infty(\mathbb{R}^d)$. From Kurtz and Ocone ([17], Lemma 1.1, p.82) there exists some $\mathscr{P}(\mathbb{R}^d)$-valued and $\mathscr{Y}_t$-optional process $\{\pi_t,\ t \in [0,T]\}$ (called the *optimal nonlinear filter*), defined on $(\Omega, \mathscr{F}, P)$ and unique to within indistinguishability, such that

$$\pi_t\phi = \mathrm{E}[\phi(X_t)|\mathscr{Y}_t], \quad P-\text{a.s.} \quad \text{for each } t \in [0,T] \text{ and } \phi \in B(\mathbb{R}^d). \tag{5.116}$$

From Fujisaki, Kallianpur and Kunita ([9], Theorem 5.1, p.34) (see also Rogers and Williams [22], VI(8.11), p.325) we then have the following

Theorem 5.3. *Suppose that $\{\pi_t,\ t \in [0,T]\}$ is the optimal nonlinear filter for the model (5.111). Then, for each $\phi \in C_c^\infty(\mathbb{R}^d)$, the following holds to within indistinguishability:*

$$\pi_t\phi = \pi_0\phi + \int_0^t \pi_s(\mathcal{A}\phi)\, ds + \int_0^t \sum_{k=1}^{D} [\pi_s(h^k\phi + \mathcal{B}_k\phi) - (\pi_s h^k)(\pi_s\phi)]\, dI_s^k, \qquad t \in [0,T]. \tag{5.117}$$

Remark 5.4. Theorem 5.3 shows that $\{\pi_t\phi,\ t \in [0,T]\}$ is continuous for each $\phi \in C_c^\infty(\mathbb{R}^d)$; since $C_c^\infty(\mathbb{R}^d)$ is convergence-determining (Ethier and Kurtz [8], Problem 3.11.11, p.151), it follows that $\{\pi_t,\ t \in [0,T]\}$ is a *continuous* $\mathscr{P}(\mathbb{R}^d)$-valued and $\mathscr{Y}_t$-adapted process.

Remark 5.5. The family of relations (5.117) indexed by $\phi \in C_c^\infty(\mathbb{R}^d)$, constitutes the Fujisaki-Kallianpur-Kunita-Kushner-Stratonovich equation or *normalized filter equation*, and effectively defines a "probability measure-valued SDE" in which $\{I_t\}$ is the "driving" Wiener process. A basic viewpoint introduced by Szpirglas [24] and used in [18], is to regard the normalized filter equation as a probability measure-valued SDE determined completely by the functions h^k and linear operators $\mathcal{A}$ and $\mathcal{B}_k$ (given by (5.114) and (5.115)), but otherwise quite separate from the nonlinear filtering problem, for which one can formulate the notions of solution, pathwise-uniqueness, and uniqueness-in-law by direct analogy with these concepts for "ordinary" SDEs (see e.g. Revuz and Yor [20], Definitions IX(1.2)-(1.3), pp. 350-351). For completeness we repeat the basic definitions from [18]:

Definition 5.6. A pair $\{(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P}; \{\bar{\mathscr{F}}_t\}), (\bar{\pi}_t, \bar{I}_t)\}$ is a *solution* of the normalized filter equation when:

1. $(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P}; \{\bar{\mathscr{F}}_t\})$ is an R-filtered probability space (Definition II(67.1) on p.172 of [21]);
2. $\{\bar{I}_t,\ t \in [0,T]\}$ is an $\mathbb{R}^D$-valued standard $\bar{\mathscr{F}}_t$-Wiener process on $(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P})$;
3. $\{\bar{\pi}_t,\ t \in [0,T]\}$ is a $\mathscr{P}(\mathbb{R}^d)$-valued, continuous, $\bar{\mathscr{F}}_t$-adapted process such that, for each $\phi \in C_c^\infty(\mathbb{R}^d)$, the following holds to within $\bar{P}$-indistinguishability

$$\bar{\pi}_t\phi = \bar{\pi}_0\phi + \int_0^t \bar{\pi}_s(\mathcal{A}\phi)\, \mathrm{d}s + \int_0^t \sum_{k=1}^{D} [\bar{\pi}_s(h^k\phi + \mathcal{B}_k\phi) - (\bar{\pi}_s h^k)(\bar{\pi}_s\phi)]\, \mathrm{d}\bar{I}_s^k, \quad t \in [0,T].$$

Remark 5.7. Since the innovation $\{I_t\}$ is an $\mathbb{R}^D$-valued $\mathscr{Y}_t$-Wiener process on $(\Omega, \mathscr{F}, P)$ (by Remark 5.2 and ([22], Theorem VI(8.4), p.323)), it follows from Remark 5.4 and Theorem 5.3 that $\{(\Omega, \mathscr{F}, P; \{\mathscr{Y}_t\}), (\pi_t, I_t)\}$ is a solution of the normalized filter equation (5.117) in the sense of Definition 5.6 (where $\{\pi_t\}$ is the optimal nonlinear filter defined by (5.116)).

Definition 5.8. (1) The normalized filter equation has the property of *pathwise-uniqueness* when the following holds: If $\{(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P}; \{\bar{\mathscr{F}}_t\}), (\bar{\pi}_t^i, \bar{I}_t)\}$, $i = 1, 2$, are solutions of the normalized filter equation such that $\bar{P}(\bar{\pi}_0^1 = \bar{\pi}_0^2) = 1$, then $\{\bar{\pi}_t^1,\ t \in [0, T]\}$ and $\{\bar{\pi}_t^2,\ t \in [0, T]\}$ are $\bar{P}$-indistinguishable.
(2) The normalized filter equation has the property of *uniqueness-in-joint law* when the following holds: If $\{(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P}; \{\tilde{\mathscr{F}}_t\}), (\tilde{\pi}_t, \tilde{I}_t)\}$ and $\{(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P}; \{\bar{\mathscr{F}}_t\}), (\bar{\pi}_t, \bar{I}_t)\}$ are solutions of the normalized filter equation and $\tilde{P}(\tilde{\pi}_0)^{-1} = \bar{P}(\bar{\pi}_0)^{-1}$, then the $\mathscr{P}(\mathbb{R}^d) \times \mathbb{R}^D$-valued processes $\{(\tilde{\pi}_t, \tilde{I}_t),\ t \in [0, T]\}$ and $\{(\bar{\pi}_t, \bar{I}_t),\ t \in [0, T]\}$ have identical finite-dimensional distributions.

In ([18], Theorem 2.21(ii), p.190) it is established that the normalized filter equation has the property of *uniqueness-in-joint law*, but it is *not* shown that this equation has the property of *pathwise-uniqueness* (see [18], Remark 3.8, p.197). Using the innovations conjecture asserted in Remark 5.2 we can establish a restricted form of pathwise-uniqueness:

Proposition 5.9. *Suppose Condition 5.1 and fix a pair of solutions $\{(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P}; \{\bar{\mathscr{F}}_t\}), (\bar{\pi}_t^i, \bar{I}_t)\}$, $i = 1, 2$, of the normalized filter equation such that $\bar{P}(\bar{\pi}_0^1 = \bar{\pi}_0^2) = 1$. If $\bar{\pi}_0^1$ and $\bar{\pi}_0^2$ are such that $\bar{E}[\bar{\pi}_0^i \psi] < \infty$ for $\psi(x) := |x|^4$, $x \in \mathbb{R}^d$, then $\{\bar{\pi}_t^1,\ t \in [0, T]\}$ and $\{\bar{\pi}_t^2,\ t \in [0, T]\}$ are $\bar{P}$-indistinguishable.*

Remark 5.10. Proposition 5.9 postulates fourth-power integrability of the initial-value probability measures $\bar{\pi}_0^i$, and for this reason gives a restricted form of pathwise-uniqueness when compared with Definition 5.8(1), in which there is no comparable integrability condition on the initial-value measures. Proposition 5.9 should be compared with Theorem 4.5 of Kurtz and Ocone ([17], p.99), which establishes a partial pathwise-uniqueness for the normalized filter equation in the following sense: if $\{\mu_t,\ t \in [0, T]\}$ is a $\mathscr{P}(\mathbb{R}^d)$-valued continuous process on the probability space $(\Omega, \mathscr{F}, P)$ *on which the random data of the nonlinear filtering problem is specified*, adapted to the observation filtration $\{\mathscr{Y}_t\}$ and satisfying a relation exactly analogous to the normalized filter equation, then $\{\mu_t\}$ and the optimal nonlinear filter $\{\pi_t\}$ are P-indistinguishable. The hypotheses for this result are quite general (see (4.19)(i)-(iv) in [17], p.98), and in particular do not entail the uniform-boundedness and Lipschitz-continuity of the sensor function $h(\cdot)$ or fourth-power integrability of the initial laws that are needed for Proposition 5.9. On the other hand, in return for these stronger hypotheses, Proposition 5.9 establishes pathwise uniqueness in the genuine "Yamada-Watanabe" sense, that is among pairs of candidate solutions on an arbitrary common filtered probability space $(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P}; \{\bar{\mathscr{F}}_t\})$ and "driven" by an arbitrary $\bar{\mathscr{F}}_t$-Wiener Process $\{\bar{I}_t\}$, rather than for candidate solutions $\{\mu_t\}$ on the particular filtered probability space $(\Omega, \mathscr{F}, P; \{\mathscr{Y}_t\})$, adapted specifically to the observation filtration, and "driven" specifically by the innovations process.

Proof of Proposition 5.9: From the hypotheses we have

$$\bar{\pi}_0^1 \psi = \bar{\pi}_0^2 \psi < \infty, \qquad \bar{P} - \text{a.s.} \qquad \text{for} \qquad \psi(x) := |x|^4,\ x \in \mathbb{R}^d. \tag{5.118}$$

By a simple conditioning argument (see Ikeda and Watanabe [10], Remark IV.1.4, p.149) it is enough to establish pathwise-uniqueness in the special case where the $\bar{\pi}_0^i$ are non-random, and thus, without loss of generality, we shall suppose that

$$\bar{\pi}_0^1 = \bar{\pi}_0^2 \ \text{ for some } \bar{\pi}_0^i \in \mathscr{P}(\mathbb{R}^d) \text{ such that } \bar{\pi}_0^i \psi < \infty \text{ (for } \psi \text{ given by (5.118)).} \tag{5.119}$$

Fix some $\mathbb{R}^{q+D}$-valued Wiener process $\{(W_t^1, W_t^2),\ t \in [0,T]\}$ and some $\mathbb{R}^d$-valued random vector ξ_0 on a common complete probability space $(\Omega, \mathscr{F}, P)$, with $\sigma\{W_t^1, W_t^2,\ t \in [0,T]\}$ and ξ_0 being independent, and ξ_0 having the common distribution $\bar{\pi}_0^i$ at (5.119), and let $\{\mathscr{F}_t\}$ be the usual augmentation of the filtration $\{\mathscr{G}_t\}$ given by $\mathscr{G}_t := \sigma\{\xi_0, W_s^1, W_s^2,\ s \in [0,t]\}$. Then the filtered probability space $(\Omega, \mathscr{F}, P; \{\mathscr{F}_t\})$, Wiener process $\{(W_t^1, W_t^2),\ t \in [0,T]\}$, and random vector ξ_0 satisfy Condition 2.2, thus, for $\{(X_t, Y_t)\}$ given by (5.111), it follows from Remark 5.2 that $\mathscr{Y}_t = \mathscr{I}_t$ for each $t \in [0,T]$. Denote by $\{\pi_t,\ t \in [0,T]\}$ the optimal nonlinear filter corresponding to the model (5.111) (that is, $\{\pi_t\}$ is characterized by (5.116)). From Condition 5.1 one sees immediately that Conditions 2.18, 2.19 and 2.20 of [18] hold, and hence Theorem 2.21(ii) of ([18], p.190) establishes that the normalized filter equation (5.117) has the property of uniqueness-in-joint law (see Definition 5.8(2)). Accordingly, since $\{(\Omega, \mathscr{F}, P; \{\mathscr{Y}_t\}), (\pi_t, I_t)\}$ is a solution of the normalized filter equation (recall Remark 5.7), and $\pi_0 = P\xi_0^{-1} = \bar{\pi}_0^i$, $i = 1,2$ (the first equality following from (5.116) together with $X_0 = \xi_0$, and the second equality following from the definition of ξ_0), we find that

$$\{(\pi_t, I_t),\ t \in [0,T]\} \text{ and } \{(\bar{\pi}_t^i, \bar{I}_t),\ t \in [0,T]\} \text{ are identically distributed processes.} \tag{5.120}$$

Fix an arbitrary $t \in [0,T]$. We know that π_t is $\mathscr{Y}_t$-measurable and therefore $\mathscr{I}_t$-measurable (since we have seen that $\mathscr{Y}_t = \mathscr{I}_t$), and, since $\{I_t\}$ is Wiener-distributed, we also have $\mathscr{I}_t = \sigma\{I_s,\ s \in [0,t]\} \vee \mathbf{Z}^{\mathbf{P}}[\mathscr{F}]$ (by Rogers and Williams [21], II(68.4), p.175). But $\mathscr{P}(\mathbb{R}^d)$ is a Polish space (when it carries the Prohorov metric - see Ethier and Kurtz [8], Theorem 3.1.7, p.101), and therefore the $\sigma\{I_s,\ s \in [0,t]\} \vee \mathbf{Z}^{\mathbf{P}}[\mathscr{F}]$-measurability of π_t implies that there is a sequence $\{t_n, n = 1,2,\dots\} \subset [0,t]$ and a Borel-measurable mapping $\Psi_t : \Pi_{n=1}^{\infty}\mathbb{R}^D \to \mathscr{P}(\mathbb{R}^d)$ such that

$$\pi_t = \Psi_t(I_{t_1}, I_{t_2}, I_{t_3}, \dots), \qquad P-\text{a.s.} \tag{5.121}$$

(see e.g. Stroock and Varadhan [23], Ex.1.5.6, p.44). In light of (5.121) and (5.120) we find

$$\bar{\pi}_t^i = \Psi_t(\bar{I}_{t_1}, \bar{I}_{t_2}, \bar{I}_{t_3}, \dots), \qquad \bar{P}-\text{a.s.},$$

for $i = 1,2$. Thus $\bar{\pi}_t^1 = \bar{\pi}_t^2$, $\bar{P}-$a.s., for arbitrary $t \in [0,T]$, and the processes $\{\bar{\pi}_t^1,\ t \in [0,T]\}$ and $\{\bar{\pi}_t^2,\ t \in [0,T]\}$, being continuous, are therefore indistinguishable. □

References

[1] D.F. Allinger and S.K. Mitter, New results on the innovations problem of non-linear filtering, *Stochastics*, pp. 339–348, v.4, (1981). MR0609692

[2] V.E. Benes, Nonexistence of strong nonanticipating solutions to stochastic DEs: implications for functional DEs, filtering and control, *Stoch. Proc. Appl.* pp. 243–263, v.5, (1977). MR0458588

[3] A.G. Bhatt and R.L. Karandikar, Robustness of the nonlinear filter: the correlated case. *Stoch. Proc. Appl.* pp. 41–58, v.97, (2002). MR1870719

[4] Y.S.Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*, 2nd ed., Springer-Verlag, New York, (1988). MR0953964

[5] J.M.C. Clark, *Conditions for the one-to-one correspondence between an observation process and its innovation*, Technical Report, Centre for Computing and Automation, Imperial College, London, (1969).

[6] J.M.C. Clark, A remark on a strong solution corollary of Watanabe and Yamada. In *Recent Advances in Communication and Control Theory* (volume in honor of A.V. Balakrishnan; R.E. Kalman, G.I. Marchuk, A. Ruperti and A.J. Viterbi, eds.) Optimization Software Inc., New York, pp. 155–158, (1987).

[7] R.J. Elliott, *Stochastic Calculus and Applications*, Springer-Verlag, New York, (1982). MR0678919

[8] S.N. Ethier and T.G. Kurtz, *Markov Processes: Characterization and Convergence*. J. Wiley & Sons, New York, (1986). MR0838085

[9] M. Fujisaki, G. Kallianpur and H. Kunita, Stochastic differential equations for the non-linear filtering problem. *Osaka J. Math.*, pp. 19–40, v.9, (1972). MR0336801

[10] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland, Amsterdam (Kodansha Ltd. Tokyo), (1981). MR0637061

[11] G. Kallianpur, *Stochastic Filtering Theory*, Springer-Verlag, New York, (1980). MR0583435

[12] R.L. Karandikar, On the Métivier-Pellaumail inequality, Emery topology and pathwise formulae in stochastic calculus. *Sankhyā Ser. A*, pp. 121–143, v.51, (1989). MR1065565

[13] R.L. Karandikar, On pathwise stochastic integration, *Stoch. Proc. Appl.* pp. 11–18, v.57, (1995). MR1327950

[14] I. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York (1991). MR1121940

[15] N.V. Krylov, On the equivalence of σ-algebras in the filtering problem of diffusion processes, *Theor. Probab. Appl.* pp. 772–781, v.24, (1979). MR0550532

[16] N.V. Krylov and B.L. Rozovskii, On conditional distributions of diffusion processes, *Math. USSR Izvestia*, pp.336–356, v. 12, (1978). MR0517867

[17] T.G. Kurtz and D.L. Ocone, Unique characterization of conditional distributions in nonlinear filtering. *Ann. Prob.*, pp. 80–107, v.16, (1988). MR0920257

[18] V.M. Lucic and A.J. Heunis, On uniqueness of solutions for the stochastic differential equations of nonlinear filtering, *Ann. Appl. Prob.*, pp. 182–209, v.11, (2001). MR1825463

[19] P.A. Meyer, Sur un problème de filtration, *Séminaire de Probabilités VII*, Lecture Notes in Mathematics no. 321, pp. 223–247, Springer-Verlag, Berlin, (1973). MR0378084

[20] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, 2nd ed., Springer-Verlag, Berlin, (1994). MR1303781

[21] L.C.G Rogers and D. Williams, *Diffusions, Markov Processes and Martingales: Volume 1 Foundations*, 2nd Ed. Cambridge University Press, Cambridge, (2000). MR1796539

[22] L.C.G Rogers and D. Williams, *Diffusions, Markov Processes and Martingales: Volume 2 Itô Calculus*, 2nd Ed. Cambridge University Press, Cambridge, (2000). MR1780932

[23] D.W. Stroock and S.R.S. Varadhan, *Multidimensional Diffusion Processes*, Springer-Verlag, New York, (1979). MR0532498

[24] J. Szpirglas, Sur l'équivalence d'équations différentielles stochastiques á valeurs mesures intervenant dans le filtrage markovien non linéaire, *Ann. Inst. Henri Poincaré*, pp. 33–59, v.XIV, (1978). MR0495059

[25] B.S. Tsirel'son, An example of a stochastic differential equation having no strong solution, *Theor. Probab. Appl.* pp. 416–418, v.20, (1975).

[26] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*, Springer-Verlag, New York, (1985). MR0787046

[27] T. Yamada and S. Watanabe, On the uniqueness of solutions of stochastic differential equations, *J. Math. Kyoto Univ.* pp. 155–167, v.11, (1971). MR0278420