The Annals of Applied Probability
2002, Vol. 12, No. 3, 1096–1113

ON INVARIANT MEASURES OF DISCRETE TIME FILTERS IN THE CORRELATED SIGNAL–NOISE CASE[1]

BY A. BUDHIRAJA

University of North Carolina–Chapel Hill

The classical results on the ergodic properties of the nonlinear filter previously have been proved under the crucial assumption that the signal process and the observation noise are independent. This assumption is quite restrictive and many important problems in engineering and stochastic control correspond to filtering models with correlated signal and noise. Unlike the case of independent signal and noise, the filter process in the general correlated case may not be Markov even if the signal is a Markov process. In this work a broad class of discrete time filtering problems with signal–noise correlation is studied. It is shown that the pair process $(Y_j, \pi_j)_{j\in\mathbb{N}_0}$ is a Feller–Markov process, where $(Y_j)_{j\in\mathbb{N}_0}$ is the observation process and π_j is the filter, that is, the conditional distribution of the signal: X_j given past and current observations. It is shown that if the signal process (X_j) has an invariant measure, then so does (Y_j, π_j). Finally, it is proved that if (X_j) has a unique invariant measure and the stationary flow corresponding to the signal process is purely nondeterministic, then the pair (Y_j, π_j) has a unique invariant measure.

1. Introduction. Stochastic nonlinear filtering is one of the central areas of application of stochastic processes. The basic object of the study is a pair of stochastic processes $(X_j, Y_j)_{j\in\mathbb{N}_0}$, where (X_j) is called the signal process and (Y_j) the observation process. The central problem in nonlinear filtering is the study of the measure valued process (π_j) which is the conditional distribution of X_j given $\sigma\{Y_k : k \in \mathbb{N}_0;\ k \le j\}$. This measure valued process is called the nonlinear filter. In the classical setting of nonlinear filtering, the signal is taken to be a Markov process with values in some Polish space E and the observations are given via the relation

$$Y_j = h(X_j) + \eta_j, \tag{1.1}$$

where (η_j) is an i.i.d. sequence of $\mathbb{R}^d$ valued random variables, referred to as the observation noise sequence, and h is the observation function which is a map from $E \to \mathbb{R}^d$. The study of ergodic properties of the nonlinear filter has generated significant research in recent years [5, 9, 6, 10, 8, 1, 7, 2]. The pioneering work in

Received January 2000; revised October 2000.

[1]Supported in part by NSF Grant DMI-98-12857 and the University of Notre Dame Faculty Research Program.

AMS 2000 subject classifications. 60G35, 60J05, 60H15.

Key words and phrases. Nonlinear filtering, invariant measures, asymptotic stability, measure valued processes.

this direction is by Kunita [5]. In this classic paper Kunita used the uniqueness of the solution of the Kushner–Stratonovich equation, in the classical filtering model with independent signal and noise, to study the Markov properties of the filter. It was shown that if the signal is Feller–Markov with a compact, separable Hausdorff state space E, then the optimal filter is also a Feller–Markov process with state space $\mathcal{P}(E)$, where $\mathcal{P}(E)$ is the space of all probability measures on E. Furthermore, [5] shows that if the signal in addition has a unique invariant measure μ for which (4.6) holds, then the filter has a unique invariant measure. In subsequent papers Kunita [6] and Stettner [9] extended the above results to the case where the state space is a locally compact Polish space. In all the above papers [5, 6, 9] the observation function h is assumed to be bounded. In a recent paper [2] we extend the results of Kunita and Stettner to the case of unbounded h and signals with state space an arbitrary Polish space. The proofs in [2] are of independent interest since unlike the arguments in [5, 6, 9] they do not rely on the uniqueness of the solution to the Kushner–Stratonovich equation.

The analysis in the above-stated works is greatly simplified by the assumption that the signal process and the observation noise are independent. In general, however, the assumption of signal–noise independence is quite restrictive and many important problems in engineering and stochastic control correspond to filtering models with correlated signal and noise. In this work we show that the techniques developed in [2] can be used to study Markov and ergodicity properties for the nonlinear filter for quite general models with correlations as well. For the sake of exposition we restrict ourselves to signals and observations evolving in discrete time; however, similar techniques can be used to study the continuous time problem.

In the classical setup [5, 9, 6, 2] if the signal is a Feller–Markov process, then so is the filter. The first obstacle in the study of the correlated case is that, in general, even if the signal is a Markov process, the filter need not be Markov. To see this problem consider the following elementary filtering model. Suppose that the signal $(X_n)_{n\in\mathbb{N}_0}$ and the observations $(Y_n)_{n\in\mathbb{N}_0}$ are given as follows:

$$X_n = X_{n-1} + Y_{n-1} + \xi_n, \qquad n \in \mathbb{N},$$
$$Y_n = X_n + \eta_n, \qquad n \in \mathbb{N}_0,$$

where $(\xi_n)_{n\in\mathbb{N}}$ and $(\eta_n)_{n\in\mathbb{N}_0}$ are i.i.d. standard scalar normal random variables. Suppose that X_0 has a density with respect to the Lebesgue measure. Denote by ρ_n the filtering density, that is, the conditional density of X_n given $\mathbf{Y}_n \doteq (Y_0, \dots, Y_n)$. For random vectors $\mathbf{Z}_1$, $\mathbf{Z}_2$ let $f_{\mathbf{Z}_1|\mathbf{Z}_2}$ denote the conditional density of $\mathbf{Z}_1$ given $\mathbf{Z}_2$. Then

$$\rho_n(x) = c f_{X_n|\mathbf{Y}_{n-1}}(x) f_{Y_n|(\mathbf{Y}_{n-1}, X_n=x)}(Y_n),$$

where c is the normalizing constant. Next, denoting the standard normal density by ϕ, we have $f_{Y_n|(\mathbf{Y}_{n-1}, X_n=x)}(Y_n) = \phi(Y_n - x)$ and

$$f_{X_n|\mathbf{Y}_{n-1}}(x) = \int_{\mathbb{R}} f_{X_n|(\mathbf{Y}_{n-1}, X_{n-1}=y)}(x)\rho_{n-1}(y)\,dy.$$

Next note that $f_{X_n|(\mathbf{Y}_{n-1}, X_{n-1}=y)}(x) = \phi(x - y - Y_{n-1})$. Combining the above observations we have that

$$\rho_n(x) = \phi(Y_n - x) \int_{\mathbb{R}} \phi(x - y - Y_{n-1}) \rho_{n-1}(y)\, dy. \tag{1.2}$$

Thus the filter update formula for ρ_n, unlike the case of independent signal and noise, in addition to using Y_n and ρ_{n-1}, involves Y_{n-1}. This destroys the Markov property for $(\rho_n)_{n \in \mathbb{N}_0}$, since from (1.2) it follows that $E(\int_{\mathbb{R}} g(x)\rho_n(x)\, dx \mid \mathbf{Y}_{n-1})$ equals $E(\int_{\mathbb{R}} g(x)\rho_n(x)\, dx \mid \rho_{n-1}, Y_{n-1})$, where g is an appropriate test function.

In view of the above problem it is natural to consider instead the process (Y_n, π_n), where π_n is the conditional distribution of X_n given $\mathbf{Y}_n$. We show that for a quite general class of discrete time filtering models (see Section 2 for the precise setup) (Y_n, π_n) is Feller–Markov. The proof, as in [2] for the independent signal–noise case, uses a change of measure technique. The change of measure is such that, under the new measure X_0, (Y_n) and (ξ_n) are mutually independent and the observation sequence has the same distribution as that of the observation noise sequence under the original measure. The key step in the proof of the Markov property is the filter update formula analogous to (1.2). This is obtained in Proposition 3.1. The Markov property for (Y_j, π_j) (Corollary 3.4) is then a consequence of Theorem 3.3. We next show that the above Markov process has the Feller property. This is done in Theorem 3.5.

In Section 4 we study the problem of existence and uniqueness of invariant measures for (Y_j, π_j). We show, in Theorem 4.4, that if the signal process has an invariant measure, then so does the above pair process. Our final result is that if the signal has a unique invariant measure and it satisfies Assumption 4.7, then the pair (Y_j, π_j) has a unique invariant measure. The key steps in the proof are Theorem 4.4 and Proposition 4.8. Once these are proved the result follows immediately upon taking limits as $m \to \infty$ in the inequality (4.7) and noting that the two extreme terms in the limit inequality are identical in view of Assumption 4.7.

We now list the common notation used in this paper. For a complete separable metric space S, let $BM(S)$ be the class of real valued bounded measurable functions on S, let $C_b(S)$ be the subclass of $BM(S)$ of continuous functions on S, let $\mathcal{B}(S)$ be the Borel σ-field on S, let $\mathcal{P}(S)$ be the space of probability measures on $(S, \mathcal{B}(S))$ endowed with the weak convergence topology and let $\mathcal{M}(S)$ be the class of positive finite measures on $(S, \mathcal{B}(S))$ with the weak convergence topology. For $f \in BM(S)$ and $\nu \in \mathcal{P}(S)$ we denote $\int_S f(x)\, d\nu(x)$ by $\nu(f)$. The probability measure on S which is concentrated at the single point $x \in S$ is denoted by δ_x. The indicator function of a set A is denoted by χ_A. If Z is an S valued random variable on some probability space $(\Omega, \mathcal{F}, P)$, then the law of Z will be written as PoZ^{-1}.

2. The filtering model. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let the filtering model be given as

$$X_n = \mathcal{A}(X_{n-1}, Y_{n-1}, \xi_n), \qquad n \in \mathbb{N},$$
$$Y_n = h(X_n) + \eta_n, \qquad n \in \mathbb{N}_0,$$

where X_n takes values in a Polish space E and Y_n takes values in $\mathbb{R}^d$. In the above model, $(\xi_n)_{n\geq 1}$ is an i.i.d. sequence of E_0 valued random variables with law μ_1, where E_0 is another Polish space and $(\eta_n)_{n\geq 0}$ is an i.i.d. sequence of $\mathbb{R}^d$ valued random variables with a continuous and bounded density function $g(\cdot)$. We assume that $(X_0, (\xi_n)_{n\in\mathbb{N}}, (\eta_n)_{n\in\mathbb{N}_0})$ are mutually independent. We denote the distribution of X_0 by p_0. The maps $h : E \to \mathbb{R}^d$ and $\mathcal{A} : E \times \mathbb{R}^d \times E_0 \to E$ are taken to be continuous.

The basic object of interest is the $\mathcal{P}(E)$ valued stochastic process

$$\pi_n(A) \doteq P(X_n \in A \mid \mathcal{F}^Y_{0,n}), \qquad A \in \mathcal{B}(E),$$

where for a sequence of random variables $\{Z_n\}$ we denote by $\mathcal{F}^Z_{m,k}$ the σ-field generated by $\{Z_m, Z_{m+1}, \dots, Z_k\}$ for $m \leq k$. It will be convenient to work with the following canonical spaces. Denote by $(\mathbb{R}^d)^{\mathbb{N}_0}$ the space of all sequences $\gamma^1 \equiv (\gamma^1_n)_{n\in\mathbb{N}_0}$ in $\mathbb{R}^d$. Also, denote by $(E_0)^{\mathbb{N}}$ the space of all sequences $\gamma^2 \equiv (\gamma^2_n)_{n\in\mathbb{N}}$ in E_0. Endow the above spaces with the Borel σ-fields, corresponding to the pointwise convergence topology, $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively. Denote by Q_1 the probability measure on $((\mathbb{R}^d)^{\mathbb{N}_0}, \mathcal{B}_1)$ under which the canonical coordinate sequence is i.i.d. with probability density function $g(\cdot)$. Also, denote by Q_2 the probability measure on $((E_0)^{\mathbb{N}}, \mathcal{B}_2)$ under which the canonical coordinate sequence is i.i.d. with law μ_1. Now for fixed $\nu \in \mathcal{P}(E)$ consider the probability space

$$(\Omega', \mathcal{F}', R_\nu) \doteq \big(E \times (\mathbb{R}^d)^{\mathbb{N}_0} \times (E_0)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}_1 \otimes \mathcal{B}_2, \nu \otimes Q_1 \otimes Q_2\big).$$

A typical element of Ω' will be denoted by $\gamma \equiv (\gamma^0, \gamma^1, \gamma^2)$, and (β^i_n) for $i = 1, 2$ denotes the canonical processes on $(\Omega', \mathcal{F}')$ given as $\beta^i_n(\gamma) \doteq \gamma^i_n$. With an abuse of notation, β^1_n also denotes the canonical coordinate process on $((\mathbb{R}^d)^{\mathbb{N}_0}, \mathcal{B}_1)$. Now we define a sequence (θ_n) of E valued random variables on the above probability space as follows:

$$\theta_n(\gamma) \doteq \mathcal{A}(\theta_{n-1}, \beta^1_{n-1}, \beta^2_n)(\gamma), \qquad n \in \mathbb{N},$$
$$\theta_0(\gamma) \doteq \gamma^0.$$

For $n \in \mathbb{N}_0$, define $\mathcal{G}_n \doteq \sigma\{\theta_0, \beta^i_j;\ j \leq n;\ i = 1, 2\}$. One of the key representation formulas in nonlinear filtering is the so-called Kallianpur–Striebel formula (see [4]), which we now present in our notation. Define, for $0 \leq m \leq n < \infty$,

$$L_{m,n}(\gamma) \doteq \prod_{i=m}^{n} L^{(i)}(\gamma) \tag{2.1}$$

and for $j \geq 0$, $L_j(\gamma) \doteq L_{0,j}(\gamma)$, where

$$L^{(i)}(\gamma) \doteq \frac{g(\beta_i^1 - h(\theta_i))}{g(\beta_i^1)}.$$

Note that with respect to the filtration $\{\mathcal{F}_{0,j}^{\beta^1} \vee \mathcal{F}_{0,j}^{\beta^2} \vee \mathcal{F}_{0,0}^{\theta}\}$, L_j is an R_ν-martingale with mean 1. Furthermore, if the probability measure $\tilde{R}_\nu$ on $(\Omega', \mathcal{F}')$ is defined as

$$\frac{d\tilde{R}_\nu}{dR_\nu} \doteq L_n \qquad \text{on } \mathcal{G}_n,\ n \in \mathbb{N}_0,$$

then $\tilde{R}_\nu o(\{\theta_n, \beta_n^1\}_{n\in\mathbb{N}_0})^{-1} = Po(\{X_n, Y_n\}_{n\in\mathbb{N}_0})^{-1}$.

For $B \in \mathcal{B}(E)$, $\nu \in \mathcal{M}(E)$ and $j \in \mathbb{N}_0$ define

$$\Gamma_j(\nu, B)(\gamma^1) \doteq \int_E \int_{E_0^{\mathbb{N}}} \chi_B(\theta_j(\gamma)) L_j(\gamma)\, dQ_2(\gamma^2)\, d\nu(\gamma^0).$$

Finally, define for $\nu \in \mathcal{M}(E)$, $B \in \mathcal{B}(E)$, $\Lambda_j(\nu, B) \doteq \Gamma_j(\nu, B)/\Gamma_j(\nu, E)$. For notational convenience we sometimes write $\Gamma_j(\nu, \cdot)$ and $\Lambda_j(\nu, \cdot)$ as $\Gamma_j(\nu)$ and $\Lambda_j(\nu)$, respectively. Then the Kallianpur–Striebel formula in this notation states that

$$\Lambda_j(p_0)(Y_\cdot(\omega)) = \pi_j \qquad \text{a.e. } \omega\ [P]. \tag{2.2}$$

3. Feller–Markov property of the filter. In this section we prove that $((Y_j, \pi_j), \mathcal{F}_{0,j}^Y)_{j\in\mathbb{N}_0}$ is a Markov chain with a Feller semigroup. The proof of the Markov property for this pair process is similar to the proof of the Markov property of the filter process, in the uncorrelated case, presented in [2] and so some details are omitted. We also refer the reader to [3], where a different proof for the Markov property is given. The key step in the proof is establishing the semigroup relation in Proposition 3.1 below.

We begin with the following notation. Define for $l \in \mathbb{N}_0$ the map $\beta_{l+\cdot}^1 : (\mathbb{R}^d)^{\mathbb{N}_0} \to (\mathbb{R}^d)^{\mathbb{N}_0}$ by

$$[\beta_{l+\cdot}^1(\gamma^1)](j) \doteq \gamma_{l+j}^1, \qquad j \geq 0,\ \gamma^1 \in (\mathbb{R}^d)^{\mathbb{N}_0}.$$

For $B \in \mathcal{B}(E)$, $\nu \in \mathcal{M}(E)$ and $j \in \mathbb{N}$ define

$$\tilde{\Gamma}_j(\nu, B)(\gamma^1) \doteq \int_E \int_{E_0^{\mathbb{N}}} \chi_B(\theta_j(\gamma)) L_{1,j}(\gamma)\, dQ_2(\gamma^2)\, d\nu(\gamma^0).$$

Set $\tilde{\Gamma}_0(\nu, \cdot) = \nu$. Also, for $j, k \in \mathbb{N}_0$, $k \leq j$, let

$$\tilde{\Gamma}_{k,j}(\nu, B)(\gamma^1) \doteq \tilde{\Gamma}_{j-k}(\nu, B)(\beta_{k+\cdot}^1(\gamma^1)).$$

Finally, define for $\nu \in \mathcal{M}(E)$, $B \in \mathcal{B}(E)$,

$$\tilde{\Lambda}_{k,j}(\nu, B) \doteq \frac{\tilde{\Gamma}_{k,j}(\nu, B)}{\tilde{\Gamma}_{k,j}(\nu, E)}.$$

PROPOSITION 3.1. *For* $\nu \in \mathcal{M}(E)$, $B \in \mathcal{B}(E)$ *and* $j, k \in \mathbb{N}_0$, $k \leq j$,

$$\Gamma_j(\nu, B)(\gamma^1) = \tilde{\Gamma}_{k,j}\big(\Gamma_k(\nu)(\gamma^1), B\big)(\gamma^1) \qquad a.e.\ \gamma^1\ [Q_1]. \tag{3.1}$$

PROOF. Equation (3.1) holds trivially if $j = k$ so henceforth we assume that $0 \leq k < j$. Note that both the left- and right-hand sides in (3.1) are $\mathcal{F}_{0,j}^{\beta^1}$-measurable. Define, for $A \in \mathcal{F}_{0,j}^{\beta^1}$,

$$G(A, B) \doteq \int_{(\mathbb{R}^d)^{\mathbb{N}_0}} \Gamma_j(\nu, B)(\gamma^1)\chi_A(\gamma^1)\, dQ_1(\gamma^1).$$

We need to show that

$$G(A, B) = \int_{(\mathbb{R}^d)^{\mathbb{N}_0}} \tilde{\Gamma}_{k,j}\big(\Gamma_k(\nu)(\gamma^1), B\big)(\gamma^1)\chi_A(\gamma^1)\, dQ_1(\gamma^1). \tag{3.2}$$

From (2.1) we have that $L_j = L_k L_{k+1,j}$ and thus using the definition of $\Gamma_j(\nu, B)$ we have that

$$G(A, B) = \int_{(\mathbb{R}^d)^{\mathbb{N}_0}} \int_E \int_{E_0^{\mathbb{N}}} L_k(\gamma)\chi_A(\gamma^1)\big(F(\theta_k, \beta_k^1, \dots, \beta_j^1)\big)\, dR_\nu(\gamma),$$

where, for $(x, \gamma_0^1, \dots, \gamma_{j-k}^1) \in E \times (\mathbb{R}^d)^k$,

$$F(x, \gamma_0^1, \dots, \gamma_{j-k}^1) \doteq \int_E \int_{E_0^{\mathbb{N}}} \chi_B(\theta_{j-k}) L_{1,j-k}(\gamma)\, dQ_2(\gamma^2)\, d\delta_x(\gamma^0)$$

and the last step follows on observing that, under R_ν, $\{\beta_{k+1+s}^2\}_{s\in\mathbb{N}_0}$ is independent of $\{\mathcal{F}_{0,k}^{\theta} \vee \mathcal{F}_{0,j}^{\beta^1}\}$ and $L_k(\gamma)\chi_A(\gamma^1)$ is $\{\mathcal{F}_{0,k}^{\theta} \vee \mathcal{F}_{0,j}^{\beta^1}\}$-measurable. Also observe that, for $\nu \in \mathcal{P}(E)$,

$$\int_E F(\gamma^0, \gamma_0^1, \dots, \gamma_{j-k}^1)\, d\nu(\gamma^0) = \tilde{\Gamma}_{j-k}(\nu, B)(\gamma^1). \tag{3.3}$$

Using the definition of $\Gamma_k(\nu)$ once more we have that

$$\begin{aligned} G(A, B) &= \int_{(\mathbb{R}^d)^{\mathbb{N}_0}} \chi_A(\gamma^1)\bigg(\int_E F(x, \beta_k^1, \dots, \beta_j^1)\Gamma_k(\nu, dx)\bigg)\, dQ_1(\gamma^1) \\ &= \int_{(\mathbb{R}^d)^{\mathbb{N}_0}} \chi_A(\gamma^1)\tilde{\Gamma}_{k,j}\big(\Gamma_k(\nu)(\gamma^1), B\big)(\gamma^1)\, dQ_1(\gamma^1). \end{aligned}$$

□

Using the above semigroup property we have the following result, the proof of which is similar to that of Theorem 4.3 of [2] and thus is omitted.

PROPOSITION 3.2. *Let* $\nu \in \mathcal{M}(E)$ *be arbitrary and let* $k, j \in \mathbb{N}_0$, $k < j$. *Let* $\psi : \mathbb{R}^d \times \mathcal{M}(E) \to \mathbb{R}$ *be such that*

$$\mathbb{E}_{Q_1}\big|\psi\big(\beta_j^1, \Gamma_j(\nu)\big)\big| < \infty.$$

Then

$$\mathbb{E}_{Q_1}\Big(\psi(\beta^1_j, \Gamma_j(\nu)) \,|\, \mathcal{F}^{\beta^1}_{0,k}\Big) = \psi_1(\beta^1_k, \Gamma_k(\nu)),$$

where, for $z \in \mathbb{R}^d$ *and* $\nu \in \mathcal{M}(E)$,

$$\psi_1(z, \nu) \doteq \mathbb{E}_{Q_1}[\psi(\beta^1_{j-k}, \tilde{\Gamma}_{j-k}(\nu)(z, \beta^1_{1+\cdot}))]. \tag{3.4}$$

We now introduce the probability measure on $(\mathbb{R}^d)^{\mathbb{N}_0}$ under which the canonical sequence $\{\beta^1_j\}$ has the same law as the observation sequence. Given $\lambda \in \mathcal{P}(E)$ define $\tilde{Q}_\lambda \in \mathcal{P}((\mathbb{R}^d)^{\mathbb{N}_0})$ by the relation

$$\frac{d\tilde{Q}_\lambda}{dQ_1} = \Gamma_j(\lambda, E) \qquad \text{on } \mathcal{F}^{\beta^1}_{0,j}, \ j \in \mathbb{N}_0.$$

It can be easily verified that $\tilde{Q}_{p_0}$ is the probability measure induced by $\{Y_j\}$ on $(\mathbb{R}^d)^{\mathbb{N}_0}$, that is, $\tilde{Q}_{p_0} = Po(\{Y_j\}_{j\in\mathbb{N}_0})^{-1}$. Thus in view of (2.2) we have that

$$\tilde{Q}_{p_0} o\Big(\big\{\beta^1_j, \Lambda_j(p_0)\big\}_{j\in\mathbb{N}_0}\Big)^{-1} = Po\Big(\{Y^1_j, \pi_j\}_{j\in\mathbb{N}_0}\Big)^{-1}. \tag{3.5}$$

The Markov property of (Y_j, π_j) is a consequence of relation (3.5) and Theorem 3.3 below.

THEOREM 3.3. *Let* $\lambda \in \mathcal{P}(E)$. *Then* $((\beta^1_j, \Lambda_j(\lambda)), \mathcal{F}^{\beta^1}_{0,j})$ *is a Markov process on* $((\mathbb{R}^d)^{\mathbb{N}_0}, \mathcal{B}_1, \tilde{Q}_\lambda)$. *Furthermore, for* $\phi \in BM(\mathbb{R}^d \times \mathcal{P}(E))$ *and* $j, k \in \mathbb{N}_0$, $k < j$, *we have that*

$$\mathbb{E}_{\tilde{Q}_\lambda}\Big[\phi(\beta^1_j, \Lambda_j(\lambda)) \,|\, \mathcal{F}^{\beta^1}_{0,k}\Big] = \phi_1(\beta^1_k, \Lambda_k(\lambda)),$$

where $\phi_1 : \mathbb{R}^d \times \mathcal{P}(E) \to \mathbb{R}$ *is defined as follows. For* $(z, \lambda_1) \in \mathbb{R}^d \times \mathcal{P}(E)$,

$$\phi_1(z, \lambda_1) \doteq \mathbb{E}_{Q_1}\Big[\phi(\beta^1_{j-k}, \tilde{\Lambda}_{j-k}(\lambda_1)(z, \beta^1_{1+\cdot}))\tilde{\Gamma}_{j-k}(\lambda_1, E)(z, \beta^1_{1+\cdot})\Big].$$

PROOF. Fix $j, k \in \mathbb{N}_0$, $k < j$, and let $A \in \mathcal{F}^{\beta^1}_{0,k}$. Let ϕ be as in the statement of the theorem. Then we have that

$$\int_A \phi(\beta^1_j, \Lambda_j(\lambda))\, d\tilde{Q}_\lambda = \int_A \psi(\beta^1_j, \Gamma_j(\lambda))\, dQ_1, \tag{3.6}$$

where $\psi : \mathbb{R}^d \times \mathcal{M}(E) \to \mathbb{R}$ is defined as follows. For $(z, \nu_1) \in \mathbb{R}^d \times \mathcal{M}(E)$,

$$\psi(z, \nu_1) \doteq \phi\left(z, \frac{\nu_1}{\nu_1(E)}\right)\nu_1(E).$$

Applying Proposition 3.2 we have that

$$\int_A \psi(\beta^1_j, \Gamma_j(\lambda))\, dQ_1 = \int_A \psi_1(\beta^1_k, \Gamma_k(\lambda))\, dQ_1, \tag{3.7}$$

where ψ_1 is as in (3.4). Next, as in the proof of Theorem 4.4 of [2], it follows that

$$\int_A \psi_1(\beta_k^1, \Gamma_k(\lambda))\, dQ = \int_A f_1(\beta_k^1, \Gamma_k(\lambda))\, d\tilde{Q}_\lambda, \tag{3.8}$$

where for $(z,\nu) \in E \times \mathcal{M}(E)$, $f_1(z,\nu) \doteq \frac{\psi_1(z,\nu)}{\nu(E)}$. Finally, using the definition of ψ_1 and ψ we have that

$$\begin{aligned} f_1(z,\nu) &= \mathbb{E}_{Q_1}\Big[\phi(\beta_{j-k}^1, \tilde{\Lambda}_{j-k}(\hat{\nu})(z, \beta_{1+\cdot}^1))\tilde{\Gamma}_{j-k}(\hat{\nu}, E)(z, \beta_{1+\cdot}^1)\Big] \\ &= \phi_1(z, \hat{\nu}), \end{aligned} \tag{3.9}$$

where $\hat{\nu} \doteq \frac{\nu}{\nu(E)}$ and the equalities in the above display follow upon noting that, for $m \in \mathbb{N}_0$, $\tilde{\Gamma}_m(\hat{\nu}, \cdot) = \tilde{\Gamma}_m(\nu, \cdot)/\nu(E)$ and $\tilde{\Lambda}_m(\hat{\nu}, \cdot) = \tilde{\Lambda}_m(\nu, \cdot)$. The result now follows on combining (3.6)–(3.9) and noting that $\hat{\Gamma}_k(\lambda) = \Lambda_k(\lambda)$. □

As an immediate consequence of the above theorem and (3.5) we have the Markov property of (Y_j, π_j).

COROLLARY 3.4. *$\{(Y_j, \pi_j), \mathcal{F}_{0,j}^Y\}_{j\geq 0}$ is an $\mathbb{R}^d \times \mathcal{P}(E)$ valued Markov process on $(\Omega, \mathcal{F}, P)$ with an associated semigroup $\{T_m\}$ given as follows. For $\phi \in BM(\mathbb{R}^d \times \mathcal{P}(E))$ and $(z,\lambda) \in (\mathbb{R}^d \times \mathcal{P}(E))$,*

$$(T_m\phi)(z,\lambda) = \mathbb{E}_{Q_1}\Big[\phi(\beta_m^1, \tilde{\Lambda}_m(\lambda)(z, \beta_{1+\cdot}^1))\tilde{\Gamma}_m(\lambda, E)(z, \beta_{1+\cdot}^1)\Big].$$

We now prove the Feller property of the above Markov chain.

THEOREM 3.5. *(T_m) is a Feller semigroup.*

PROOF. Let $\phi \in C_b(\mathbb{R}^d \times \mathcal{P}(E))$. We need to show that $T_1\phi \in C_b(\mathbb{R}^d \times \mathcal{P}(E))$. Let (z_m, λ_m) be a sequence in $\mathbb{R}^d \times \mathcal{P}(E)$ converging to (z_0, λ_0) as $m \to \infty$. Then, for $m \in \mathbb{N}_0$,

$$(T_1\phi)(z_m, \lambda_m) = \mathbb{E}_{Q_1}\Big[\phi(\beta_1^1, \tilde{\Lambda}_1(\lambda_m)(z_m, \beta_{1+\cdot}^1))\tilde{\Gamma}_1(\lambda_m, E)(z_m, \beta_{1+\cdot}^1)\Big].$$

To study the convergence of the above expression as $m \to \infty$ we write it as an expectation over a more convenient probability space.

Let $(\Omega_1, \mathcal{F}_1, P_1)$ be a probability space which supports E valued random variables $\{X^{(m)}\}_{m\in\mathbb{N}_0}$ such that the law of $X^{(m)}$ is λ_m and $X^{(m)}$ converges a.s. to $X^{(0)}$ as $m \to \infty$. Let $(\Omega_2, \mathcal{F}_2, P_2)$ be another probability space which supports independent random variables ξ_1 and η_1 which are mutually independent and where ξ_1 is E_0 valued with law μ_1 and η_1 is $\mathbb{R}^d$ valued with density g. Define the probability space

$$(\Omega^*, \mathcal{F}^*, P^*) \doteq (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P_1 \otimes P_2)$$

and random variables on this space:

$$X_1^{(m)} \doteq \mathcal{A}\big(X_0^{(m)}, z_m, \xi_1\big);$$
$$X_0^{(m)} \doteq X^{(m)}.$$

Also define the random variable

$$L^{(m)} \doteq \frac{g(\eta_1 - h(X_1^{(m)}))}{g(\eta_1)}.$$

Finally define $\mathcal{M}(E)$ valued random variable $\tilde{\Gamma}_1^{(m)}$ as follows. For $B \in \mathcal{B}(E)$,

$$\tilde{\Gamma}_1^{(m)}(B) \doteq \mathbb{E}_{P^*}\big[\chi_B\big(X_1^{(m)}\big)L^{(m)} \mid \eta_1\big],$$

where $\mathbb{E}_{P^*}$ denotes the expectation with respect to the probability measure P^*, and define $\tilde{\Lambda}_1^{(m)}(\cdot)$ to be the normalized measure. Then in this notation,

$$(T_1\phi)(z_m, \lambda_m) = \mathbb{E}_{P^*}\big[\phi\big(\eta_1, \tilde{\Lambda}_1^{(m)}\big)\tilde{\Gamma}_1^{(m)}(E)\big]. \tag{3.10}$$

Next note that by continuity of $\mathcal{A}$ we have that $X_i^{(m)}$ converges a.s. to $X_i^{(0)}$ as $m \to \infty$ for $i = 0, 1$. Thus using the continuity of g we have that $L^{(m)}$ converges a.s. to $L^{(0)}$ as $m \to \infty$. This, along with the fact that g is bounded, implies that, for all $f \in C_b(E)$, $\tilde{\Gamma}_1^{(m)}(f)$ converges almost surely to $\tilde{\Gamma}_1^{(0)}(f)$ as $m \to \infty$. Thus in particular, $\tilde{\Gamma}_1^{(m)}(E)$ converges almost surely to $\tilde{\Gamma}_1^{(0)}(E)$ and $\tilde{\Lambda}_1^{(m)}$ converges almost surely to $\tilde{\Lambda}_1^{(0)}$ as $m \to \infty$.

Also note that $\mathbb{E}_{P^*}(\tilde{\Gamma}_1^{(m)}(E)) = 1$ and so $\tilde{\Gamma}_1^{(m)}(E)$ is a sequence of nonnegative random variables with mean 1 which converge a.s. to $\tilde{\Gamma}_1^{(0)}(E)$ as $m \to \infty$ and therefore also in L^1, that is,

$$\tilde{\Gamma}_1^{(m)}(E) \overset{L^1(P^*)}{\to} \tilde{\Gamma}_1^{(0)}(E) \qquad \text{as } m \to \infty. \tag{3.11}$$

Combining (3.10) with the above observations we have that

$$(T_1\phi)(z_m, \lambda_m) \to (T_1\phi)(z_0, \lambda_0)$$

as $m \to \infty$. This proves the theorem. □

4. Ergodicity properties of the nonlinear filter. In this section we obtain conditions for existence and uniqueness of (T_m) invariant measures. As in the uncorrelated case the ergodicity properties of the filter process depend crucially on that of the signal process. We begin by noting that, for $n \in \mathbb{N}$,

$$X_n = \mathcal{A}\big(X_{n-1}, h(X_{n-1}) + \eta_{n-1}, \xi_n\big)$$
$$\doteq \mathcal{A}^*(X_{n-1}, \eta_{n-1}, \xi_n),$$

where $\mathcal{A}^*$ is a continuous map from $E \times \mathbb{R}^d \times E_0$ to E. Hence $\{X_n\}_{n\geq 0}$ is a time homogeneous Markov chain. Denote the semigroup of the Markov chain (X_n) by (S_n); that is, for $f \in BM(E)$, $x \in E$ and $n \in \mathbb{N}$,

$$(S_n f)(x) \doteq \mathbb{E}\big(f(X_n)|X_0 = x\big).$$

One basic assumption in many results of this section is the following.

ASSUMPTION 4.1. There exists a unique (S_m) invariant measure μ.

We begin by showing that if there is an (S_m) invariant measure, then (T_m) also admits an invariant measure. The proof will use the following lemma, whose proof is standard and thus is omitted.

LEMMA 4.2. *Let S be a Polish space and let $(\zeta_n)_{n\geq 0}$ be an S valued time homogeneous Markov chain with a Feller semigroup $(\mathcal{T}_n)$. Suppose that for some $\nu \in \mathcal{P}(S)$ the measure $\nu\mathcal{T}_n$ defined as*

$$\nu\mathcal{T}_n(\phi) \doteq \int_S (\mathcal{T}_n\phi)(x)\nu(dx).$$

converges weakly to ν_0 as $n \to \infty$. Then ν_0 is $(\mathcal{T}_n)$ invariant.

Now let μ be an (S_m) invariant measure. To show that there exists an (T_m) invariant measure we show that there exist probability measures $m_1^{(i)}$ on $\mathbb{R}^d \times \mathcal{P}(E)$, $i = 1, 2$, such that $m_1^{(i)} T_n$ converge weakly as $n \to \infty$. To introduce these measures we find it convenient to work with a different probability space.

Note initially that $(X_n, \eta_n, \xi_{n+1})_{n\in\mathbb{N}_0}$ is an $E \times \mathbb{R}^d \times E_0$ valued Markov chain and $\mu \otimes \mu_g \otimes \mu_1$ is an invariant probability measure for this chain, where $\mu_g \in \mathcal{P}(\mathbb{R}^d)$ is defined as follows. For $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\mu_g(A) \doteq \int_A g(z)\, dz.$$

Let $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ be a probability space which supports the stationary flow corresponding to the above invariant measure; that is, there exist sequences which we denote as $(\tilde{X}_n)_{n\in\mathbb{Z}}$, $(\tilde{\eta}_n)_{n\in\mathbb{Z}}$, $(\tilde{\xi}_n)_{n\in\mathbb{Z}}$ such that

$$(\Psi_n)_{n\in\mathbb{Z}} \doteq (\tilde{X}_n, \tilde{\eta}_n, \tilde{\xi}_{n+1})_{n\in\mathbb{Z}}$$

is a stationary Markov chain with the invariant law $\mu \otimes \mu_g \otimes \mu_1$ and the same transition probability function as that of $(X_n, \eta_n, \xi_{n+1})_{n\in\mathbb{N}}$.

Define

$$\tilde{Y}_n \doteq h(\tilde{X}_n) + \tilde{\eta}_n, \qquad n \in \mathbb{Z}.$$

Clearly $(\tilde{X}_n, \tilde{Y}_n)_{n=-\infty}^{\infty}$ is a stationary Markov chain. Define for $m, n \in \mathbb{Z},\ m < n$, the probability measure valued processes $\pi_{m,n}^{(i)}$, $i = 1, 2$, as follows. For $A \in \mathcal{B}(E)$,

$$\pi_{m,n}^{(1)}(A) \doteq \tilde{P}\big[\tilde{X}_n \in A \mid \mathcal{F}_{m,n}^{\tilde{Y}}\big]$$

and

$$\pi_{m,n}^{(2)}(A) \doteq \tilde{P}\big[\tilde{X}_n \in A \mid \mathcal{F}_{m,n}^{\tilde{Y}} \vee \sigma(\tilde{X}_m)\big].$$

An application of the martingale convergence theorem shows that, for each fixed n as $m \to -\infty$, $\pi_{m,n}^{(1)}$ converges almost surely to $\pi_{-\infty,n}^{(1)}$, where

$$\pi_{-\infty,n}^{(1)}(A) \doteq \tilde{P}\big(\tilde{X}_n \in A \mid \mathcal{F}_{-\infty,n}^{\tilde{Y}}\big)$$

and $\mathcal{F}_{-\infty,n}^{\tilde{Y}} \doteq \sigma(\bigcup_{m=-\infty}^{n} \mathcal{F}_{m,n}^{\tilde{Y}})$. Also, using the Markov property of $\tilde{X}$ and an application of the reverse martingale convergence theorem (cf. [2]) we have that, for each fixed n, $\pi_{m,n}^{(2)}$ converges almost surely to $\pi_{-\infty,n}^{(2)}$, as $m \to -\infty$, where

$$\pi_{-\infty,n}^{(2)}(A) \doteq \tilde{P}\big(\tilde{X}_n \in A \mid \mathcal{F}_{-\infty,n}^{\tilde{Y}} \vee \mathcal{F}_{-\infty,-\infty}^{\tilde{X}}\big)$$

and $\mathcal{F}_{-\infty,-\infty}^{\tilde{X}} \doteq \bigcap_{n=-\infty}^{\infty} \mathcal{F}_{-\infty,n}^{\tilde{X}}$. Furthermore since, for all k,

$$\tilde{P} o(\tilde{Y}_n, \pi_{m,n}^{(i)})^{-1} = \tilde{P} o(\tilde{Y}_{n+k}, \pi_{m+k,n+k}^{(i)})^{-1} \doteq M_{n-m}^{(i)}$$

we have that the joint law of $(\tilde{Y}_n, \pi_{-\infty,n}^{(i)})$ does not depend on n. Denote this law by $M^{(i)}$. We show in Theorem 4.4 that this probability measure, for $i = 1, 2$, is (T_m) invariant. To show this it suffices to show, in view of Lemma 4.2, that $M_k^{(i)} = m_1^{(i)} T_k$ for some $m_1^{(i)} \in \mathcal{P}(\mathbb{R}^d \times \mathcal{P}(E))$. This is done in Proposition 4.3 below. Henceforth, denote the law of $\tilde{Y}_n$ by μ_Y, the law of $(\tilde{X}_n, \tilde{Y}_n)$ by μ_{XY} and the law of $(\tilde{Y}_n, \pi_{n,n}^{(1)})$ by μ_1.

PROPOSITION 4.3. *Define, for $F \in C_b(\mathbb{R}^d \times \mathcal{P}(E))$ and $k \in \mathbb{N}$,*

$$\overline{M}_k^{(1)}(F) \doteq \int_{\mathbb{R}^d \times \mathcal{P}(E)} (T_k F)(z, \pi) \mu_1(dz, d\pi)$$

and

$$\overline{M}_k^{(2)}(F) \doteq \int_{E \times \mathbb{R}^d} (T_k F)(z, \delta_x) \mu_{XY}(dx, dz).$$

Then $\overline{M}_k^{(i)} = M_k^{(i)}$ for $i = 1, 2$.

PROOF. Denote the expectation with respect to the probability measure $\tilde{P}$ by $\tilde{\mathbb{E}}$. Then, for $F \in C_b(\mathbb{R}^d \times \mathcal{P}(E))$,

$$\tilde{\mathbb{E}}\Big[F(\tilde{Y}_k, \pi^{(1)}_{0,k}) \,\big|\, \tilde{Y}_0 = z, \pi^{(1)}_{0,0} = \pi\Big] = \mathbb{E}[\phi(Y_k, \pi_k) \mid Y_0 = z, \pi_0 = \pi] = (T_k F)(z, \pi).$$

Hence

$$\begin{aligned} M^{(1)}_k(F) &= \tilde{\mathbb{E}}\big[F(\tilde{Y}_k, \pi^{(1)}_{0,k})\big] \\ &= \tilde{\mathbb{E}}\big(\tilde{\mathbb{E}}\big(F(\tilde{Y}_k, \pi^{(1)}_{0,k}) \mid \tilde{Y}_0, \pi^{(1)}_{0,0}\big)\big) \\ &= \int_{\mathbb{R}^d} T_k F(z, \pi) \mu_1(dz, d\pi) \\ &= \overline{M}^{(1)}_k(F). \end{aligned}$$

Next note that

$$\begin{aligned} \pi^{(2)}_{0,k}(\phi) &= \tilde{\mathbb{E}}\big(\phi(\tilde{X}_k) \mid \mathcal{F}^{\tilde{Y}}_{0,k} \vee \sigma(\tilde{X}_0)\big) \\ &= G(\tilde{X}_0, \tilde{Y}_0, \dots, \tilde{Y}_k), \end{aligned}$$

where $G : E \times (\mathbb{R}^d)^k \to \mathbb{R}$ is given as follows. For $(\gamma^0, \gamma^1_0, \dots, \gamma^1_k) \in E \times (\mathbb{R}^d)^k$,

$$\begin{aligned} G(\gamma^0, \gamma^1_0, \dots, \gamma^1_k) &= \frac{\mathbb{E}_{R_{\delta_{\gamma^0}}}[\phi(\theta_k) L_k(\gamma) \mid \mathcal{F}^{\beta^1}_{0,k}]}{\mathbb{E}_{R_{\delta_{\gamma^0}}}[L_k(\gamma) \mid \mathcal{F}^{\beta^1}_{0,k}]} \\ &= \Lambda_k(\delta_{\gamma^0})(\phi). \end{aligned}$$

Hence

$$\pi^{(2)}_{0,k}(\phi) = \Lambda_k(\delta_{\tilde{X}_0})(\tilde{Y}_{0+\cdot})(\phi).$$

Now

$$\begin{aligned} &\tilde{\mathbb{E}}\big[F(\tilde{Y}_k, \pi^{(2)}_{0,k}) \mid \tilde{Y}_0 = z, \tilde{X}_0 = x\big] \\ &\quad = \tilde{\mathbb{E}}\Big[F\big(\tilde{Y}_k, \Lambda_k(\delta_{\tilde{X}_0})(\tilde{Y}_{0+\cdot})\big) \mid \tilde{Y}_0 = z, \tilde{X}_0 = x\Big] \\ &\quad = \tilde{\mathbb{E}}\Big[\tilde{\mathbb{E}}\Big[F\big(\tilde{Y}_k, \Lambda_k(\delta_{\tilde{X}_0})(\tilde{Y}_{0+\cdot})\big) \mid \tilde{Y}_0 = z,\ \tilde{X}_0 = x,\ \Lambda_0(\delta_{\tilde{X}_0})\Big] \Big| \tilde{Y}_0 = z, \tilde{X}_0 = x\Big] \\ &\quad = \tilde{\mathbb{E}}\Big[T_k F\big(z, \Lambda_0(\delta_x)\big) \,\big|\, \tilde{Y}_0 = z, \tilde{X}_0 = x\Big] \\ &\quad = T_k F(z, \delta_x), \end{aligned}$$

where the last equality follows on observing that $\Lambda_0(\delta_x) = \delta_x$. Hence

$$M_k^{(2)}(F) = \int (T_k F)(z, \delta_x)\mu_{XY}(dx, dz) \ = \overline{M}_k^{(2)}(F). \qquad \square$$

THEOREM 4.4. *Let μ be an (S_n) invariant measure. Then, for $i = 1, 2$, $M_k^{(i)}$ converges weakly to $M^{(i)}$ as $k \to \infty$. Furthermore, both $M^{(1)}$ and $M^{(2)}$ are (T_m) invariant.*

PROOF. Note that $M_k^{(i)}$ is the law of $(\tilde{Y}_n, \pi_{n-k,n}^{(i)})$. Also recall that, as $k \to \infty$, $(\tilde{Y}_n, \pi_{n-k,n}^{(i)}) \to (\tilde{Y}_n, \pi_{-\infty,n}^{(i)})$ a.s., and hence in particular the law of $(\tilde{Y}_n, \pi_{n-k,n}^{(i)})$, namely $M_k^{(i)}$, converges to the law of $(\tilde{Y}_n, \pi_{-\infty,n}^{(i)})$, which is $M^{(i)}$. Next, from Proposition 4.3, $M_k^{(i)}$ equals $\overline{M}_k^{(i)}$ and so $\overline{M}_k^{(i)}$ converges to $M^{(i)}$ as $k \to \infty$. Finally from Lemma 4.2 we have that both $M^{(1)}$ and $M^{(2)}$ are (T_m) invariant. $\square$

We now present the following consistency property of (T_m) invariant measures, which is used in Proposition 4.8, which in turn enables us to establish the uniqueness of the (T_m) invariant measure.

PROPOSITION 4.5. *Let Φ be a (T_m) invariant measure. Suppose that Assumption* 4.1 *holds. Then, for all $f \in BM(E)$ and $\phi \in BM(\mathbb{R}^d)$,*

$$\int_{\mathbb{R}^d \times \mathcal{P}(E)} \nu(f)\phi(y)\Phi(dy, d\nu) = \int_{E \times \mathbb{R}^d} f(x)\phi(y)\mu_{XY}(dx, dy). \tag{4.1}$$

PROOF. Denote the semigroup corresponding to the Markov chain (X_n, Y_n) by $(\mathcal{T}_m)$; that is, for $F \in BM(E \times \mathbb{R}^d)$,

$$(\mathcal{T}_m F)(x, y) = \mathbb{E}\big[F(X_m, Y_m) \mid X_0 = x, Y_0 = y\big].$$

From Assumption 4.1 we have that μ_{XY} is the unique $(\mathcal{T}_m)$ invariant measure. Define $\tilde{\mu}_{XY} \in \mathcal{P}(E \times \mathbb{R}^d)$ as follows. For $A \in \mathcal{B}(E)$ and $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\tilde{\mu}_{XY}(A \times B) \doteq \int_{B \times \mathcal{P}(E)} \nu(A)\Phi(dy, d\nu).$$

Note that with this notation the left-hand side of (4.1) equals $\tilde{\mu}_{XY}(f \otimes \phi)$. We now show that $\tilde{\mu}_{XY}$ is $(\mathcal{T}_m)$ invariant and thus equals μ_{XY}. This will clearly prove the proposition. To show that $\tilde{\mu}_{XY}$ is $(\mathcal{T}_m)$ invariant it suffices to show that, for arbitrary $f \in C_b(E)$ and $\phi \in C_b(\mathbb{R}^d)$,

$$\tilde{\mu}_{XY}(f \otimes \phi) = \tilde{\mu}_{XY}\big(\mathcal{T}_1(f \otimes \phi)\big). \tag{4.2}$$

Now note that

$$\begin{aligned}\mathcal{T}_1(f\otimes\phi)(x,y) &= \mathbb{E}\big(f(X_1)\phi(Y_1)\mid X_0=x, Y_0=y\big)\\ &= \int_{\mathbb{R}^d\times E_0} f\big(\mathcal{A}(x,y,u)\big)\phi\big(z_1+h\big(\mathcal{A}(x,y,u)\big)\big)g(z_1)\,dz\,d\mu_1(u).\end{aligned}$$

Hence $\tilde{\mu}_{XY}(\mathcal{T}_1(f\otimes\phi)) = \int_{\mathbb{R}^d\times\mathcal{P}(E)}\Psi(y,\nu)\,d\Phi(y,\nu)$, where

$$\begin{aligned}\Psi(y,\nu) \doteq \int_{\mathbb{R}^d\times E_0\times E} & f\big(\mathcal{A}(x,y,u)\big)\phi\big(z_1+h\big(\mathcal{A}(x,y,u)\big)\big)\\ &\times g(z_1)\,dz_1\,d\mu_1(u)\,d\nu(x).\end{aligned}\tag{4.3}$$

Next, recalling that Φ is a (T_m) invariant measure, we have that

$$\begin{aligned}&\tilde{\mu}_{XY}(f\otimes\phi)\\ &\quad= \int_{\mathbb{R}^d\times\mathcal{P}(E)}\nu(f)\phi(y)\Phi oT(dy,d\nu)\\ &\quad= \int_{\mathbb{R}^d\times\mathcal{P}(E)}\mathbb{E}_{Q_1}\big[\phi(\beta_1^1)\tilde{\Lambda}_1(\nu)(y,\beta_1^1)(f)\tilde{\Gamma}_1(\nu,E)(y,\beta_1^1)\big]\Phi(dy,d\nu).\end{aligned}\tag{4.4}$$

Finally note that

$$\begin{aligned}&\mathbb{E}_{Q_1}\big[\phi(\beta_1^1)\tilde{\Lambda}_1(\nu)(y,\beta_1^1)(f)\tilde{\Gamma}_1(\nu,E)(y,\beta_1^1)\big]\\ &\quad= \mathbb{E}_{R_\nu}\left[\phi(\beta_1^1)f(\theta_1)\frac{g(\beta_1^1-h(\theta_1))}{g(\beta_1^1)}\,\middle|\,\beta_0^1=y\right]\\ &\quad= \int_{\mathbb{R}^d\times E_0\times E}\phi(z_1)f\big(\mathcal{A}(\theta_0,y,u)\big)\frac{g(z_1-h(\mathcal{A}(\theta_0,y,u)))}{g(z_1)}\\ &\qquad\qquad \times g(z_1)\,dz_1\,d\nu(\theta_0)\,d\mu_1(u)\\ &\quad= \int_{\mathbb{R}^d\times E_0\times E}\phi\big(z_1+h\big(\mathcal{A}(\theta_0,y,u)\big)\big)\\ &\qquad\qquad\times f\big(\mathcal{A}(\theta_0,y,u)\big)g(z_1)\,dz_1\,d\nu(\theta_0)\,d\mu_1(u)\\ &\quad= \Psi(y,\nu).\end{aligned}\tag{4.5}$$

Combining (4.3)–(4.5) we have (4.2). □

Let $\mathcal{G}$ be the class of all $G\in C_b(\mathbb{R}^d\times\mathcal{P}(E))$ which are bounded from below and are such that, for all $x\in\mathbb{R}^d$, $G(x,\cdot)$ is a convex function on $\mathcal{P}(E)$. It can be shown that $\mathcal{G}$ is a probability measure determining class (cf. [6]).

The following extension of Jensen's inequality, proved in [6], will be used in the proof of Proposition 4.8.

LEMMA 4.6 [6]. *Let π be a $\mathcal{P}(E)$ valued random variable on some probability space $(\Omega_1,\mathcal{F}_1,P_1)$ and let $\mathcal{F}_2$ be a sub-σ-field of $\mathcal{F}_1$. The conditional expectation of π with respect to $\mathcal{F}_2$, denoted by $\mathbb{E}[\pi\mid\mathcal{F}_2]$ is defined as a $\mathcal{P}(E)$ valued random variable π' such that $\mathbb{E}[F(\pi)\mid\mathcal{F}_2]=F(\pi')$ holds for any continuous affine function F on $\mathcal{P}(E)$. Let $G\in\mathcal{G}$ and let X be an $\mathcal{F}_2$ measurable, $\mathbb{R}^d$ valued random variable. Then $G(X,\mathbb{E}[\pi\mid\mathcal{F}_2])\le E[G(X,\pi)\mid\mathcal{F}_2]$.*

Our basic condition for the uniqueness of a (T_m) invariant measure is the following.

ASSUMPTION 4.7. For all $f \in C_b(E)$,

$$\lim_{n\to\infty} \int_E |S_n f(x) - \mu(f)| \, d\mu(x) = 0. \tag{4.6}$$

It is well known that this condition is equivalent to the statement that the σ-field $\mathcal{F}^{\tilde{X}}_{-\infty,-\infty}$ is trivial. Note that if the above condition holds, then $\pi^{(1)}_{-\infty,n} = \pi^{(2)}_{-\infty,n}$ a.s. and so $M^{(1)}$ equals $M^{(2)}$. The key step in showing that the above condition in fact implies that there is exactly one (T_m) invariant measure is the following result.

PROPOSITION 4.8. *Suppose that Assumption* 4.1 *holds. Let* $G \in \mathcal{G}$ *and let* Φ *be a* (T_m) *invariant measure. Then, for all* $m \geq 1$,

$$M^{(1)}_m(G) \leq \int_{\mathbb{R}^d \times \mathcal{P}(E)} G(z,\nu)\Phi(dz,d\nu) \leq M^{(2)}_m(G). \tag{4.7}$$

PROOF. Define $P_* \in \mathcal{P}(\mathbb{R}^d \times \mathcal{P}(E) \times E)$ as follows. For $A \in \mathcal{B}(\mathbb{R}^d)$, $B \in \mathcal{B}(\mathcal{P}(E))$ and $C \in \mathcal{B}(E)$,

$$P_*(A \times B \times C) \doteq \int_{A\times B} \nu(C)\Phi(dz,d\nu). \tag{4.8}$$

Note that, from Proposition 4.5,

$$P_*\big(\mathbb{R}^d \times \mathcal{P}(E) \times C\big) = \int_{\mathcal{P}(E)} \nu(C)\Phi(\mathbb{R}^d, d\nu) = \mu(C).$$

Next consider the space $\overline{\Omega} \doteq \mathbb{R}^d \times \mathcal{P}(E) \times E \times (\mathbb{R}^d)^{\mathbb{N}} \times (E_0)^{\mathbb{N}}$ and endow it with the natural product σ-field denoted by $\overline{\mathcal{F}}$. Let $\overline{P}$ be the probability measure on $(\overline{\Omega}, \overline{\mathcal{F}})$ defined as

$$\overline{P} \doteq P_* \otimes \mu_g^{\otimes\mathbb{N}} \otimes \mu_1^{\otimes\mathbb{N}}.$$

The expectation with respect to the probability measure $\overline{P}$ is denoted by $\overline{\mathbb{E}}$. Similarly as in Section 2 a typical element of $\overline{\Omega}$ is denoted by $\overline{\omega} = (\overline{y}, \overline{\nu}_0, \overline{\gamma}^0, \overline{\gamma}^1, \overline{\gamma}^2)$. Define the canonical sequences $\overline{\beta}^i_n(\overline{\omega}) \doteq \overline{\gamma}^i_n$, $i = 1,2$, $n \in \mathbb{N}$. Finally define

$$\overline{X}_n \doteq \mathcal{A}(\overline{X}_{n-1}, \overline{Y}_{n-1}, \overline{\beta}^2_n), \qquad n \in \mathbb{N},\ \overline{X}_0 \doteq \overline{\gamma}^0,$$
$$\overline{Y}_n \doteq h(\overline{X}_n) + \overline{\beta}^1_n, \qquad n \in \mathbb{N},\ \overline{Y}_0 \doteq \overline{y}.$$

From (4.8) we have that, for $f \in C_b(E)$ and $\phi \in C_b(\mathbb{R}^d)$,

$$\overline{\mathbb{E}}[f(\overline{X}_0)\phi(\overline{Y}_0)] = \int_{\mathbb{R}^d\times\mathcal{P}(E)} \nu(f)\phi(z)\,d\Phi(z,\nu)$$
$$= \int_{E\times\mathbb{R}^d} f(x)\phi(y)\mu_{XY}(dx,dy),$$

where the last step follows from Proposition 4.5. Thus we have shown that $\overline{P}o(\overline{X}_0, \overline{Y}_0)^{-1} = \tilde{P}o(\tilde{X}_0, \tilde{Y}_0)^{-1}$. This in particular implies that $\overline{X}_0$ and $\overline{Y}_0 - h(\overline{X}_0)$ are independent with laws μ and μ_g respectively. Furthermore from (4.8) it follows that

$$\overline{P}\big(\overline{X}_0 \in \cdot \mid \overline{\nu}_0, \overline{Y}_0\big) = \overline{\nu}_0(\cdot). \tag{4.9}$$

For $n \in \mathbb{N}$ define $\mathcal{P}(E)$ valued random variables

$$\pi_n^{(1)}(\cdot) \doteq \overline{P}\big[\overline{X}_n \in \cdot \mid \overline{Y}_0, \dots, \overline{Y}_n\big],$$
$$\pi_n^{(2)}(\cdot) \doteq \overline{P}\big[\overline{X}_n \in \cdot \mid \overline{Y}_0, \dots, \overline{Y}_n, \overline{X}_0\big]$$

and

$$\pi_n^{*}(\cdot) \doteq \overline{P}\big[\overline{X}_n \in \cdot \mid \overline{Y}_0, \dots, \overline{Y}_n, \overline{\nu}_0\big].$$

Note that, by construction,

$$\overline{P}o\big(\{\overline{Y}_n, \overline{X}_n\}_{n\geq 0}\big)^{-1} = \tilde{P}o\big(\{\tilde{Y}_n, \tilde{X}_n\}_{n\geq 0}\big)^{-1},$$

and so, for $i = 1, 2$,

$$\overline{P}o\big(\overline{Y}_n, \overline{\pi}_n^{(i)}\big)^{-1} = M_n^{(i)}.$$

Thus, from Proposition 4.3,

$$\overline{\mathbb{E}}\Big(G\big(\overline{Y}_n, \overline{\pi}_n^{(1)}\big)\Big) = \int_{\mathbb{R}^d \times \mathcal{P}(E)} (T_n G)(z, \pi)\mu_1(dz, d\pi) \tag{4.10}$$

and

$$\overline{\mathbb{E}}\Big(G\big(\overline{Y}_n, \overline{\pi}_n^{(2)}\big)\Big) = \int_{E \times \mathbb{R}^d} (T_n G)(z, \delta_x)\mu_{XY}(dx, dz). \tag{4.11}$$

Also note that, for $\phi \in C_b(E)$,

$$\begin{aligned}\overline{\pi}_n^{*}(\phi) &= \overline{\mathbb{E}}\Big(\phi(\overline{X}_n) \mid \mathcal{F}_{0,n}^{\overline{Y}} \vee \sigma(\overline{\nu}_0)\Big)\\ &= \Lambda_n(\overline{\nu}_0)(\overline{Y}_{0+\cdot})(\phi).\end{aligned}$$

Next observe that

$$\begin{aligned}\overline{\mathbb{E}}\big[G(\overline{Y}_n, \overline{\pi}_n^{*}) \mid \overline{Y}_0 = z, \overline{\nu}_0 = \nu\big] &= \overline{\mathbb{E}}\big[G\big(\overline{Y}_n, \Lambda_n(\overline{\nu}_0)(\overline{Y}_{0+\cdot})\big) \mid \overline{Y}_0 = z, \overline{\nu}_0 = \nu\big]\\ &= \mathbb{E}\big[G(Y_n, \pi_n) \mid Y_0 = z, \pi_0 = \nu\big]\\ &= (T_n G)(z, \nu),\end{aligned}$$

where the second equality above is a consequence of (4.9). Therefore

$$\overline{\mathbb{E}}\big(G(\overline{Y}_n, \overline{\pi}_n^{*})\big) = \int_{\mathbb{R}^d \times \mathcal{P}(E)} (T_n G)(z, \nu)\Phi(dz, d\nu). \tag{4.12}$$

Finally in view of (4.10), (4.11) and (4.12) it suffices to show that

$$\overline{\mathbb{E}}[G(\overline{Y}_n,\overline{\pi}_n^{(1)})] \le \overline{\mathbb{E}}[G(\overline{Y}_n,\overline{\pi}_n^{*})] \le \overline{\mathbb{E}}[G(\overline{Y}_n,\overline{\pi}_n^{(2)})]. \tag{4.13}$$

Note that

$$\overline{\mathbb{E}}[\overline{\pi}_n^{(2)} \mid Y_0 \cdots Y_n, \overline{\nu}_0] = \overline{\pi}_n^{*}$$

and therefore by the assumed convexity properties of the function G and Lemma 4.6 we have that

$$\begin{aligned}\overline{\mathbb{E}}[G(\overline{Y}_n,\overline{\pi}_n^{(2)})] &= \overline{\mathbb{E}}[\overline{\mathbb{E}}[G(\overline{Y}_n,\overline{\pi}_n^{(2)}) \mid \overline{Y}_0,\dots,\overline{Y}_n,\overline{\nu}_0]]\\ &\ge \overline{\mathbb{E}}[G(\overline{Y}_n,\overline{\mathbb{E}}[\overline{\pi}_n^{(2)} \mid \overline{Y}_0,\dots,\overline{Y}_n,\overline{\nu}_0])]\\ &= \overline{\mathbb{E}}[G(\overline{Y}_n,\overline{\pi}_n^{*})].\end{aligned}$$

This proves the second inequality in (4.13). The first inequality is proved in an identical manner on using the observation that $\overline{\mathbb{E}}[\overline{\pi}_n^{*} \mid \overline{Y}_0 \cdots \overline{Y}_n] = \overline{\pi}_n^{(1)}$. □

We now come to our main result.

THEOREM 4.9. *Suppose that Assumptions* 4.1 *and* 4.7 *hold. Then there exists a unique* (T_m) *invariant measure.*

PROOF. From Theorem 4.4 we have that there is at least one (T_m) invariant measure. Now let Φ be an arbitrary (T_m) invariant measure and let $G \in \mathcal{G}$. Then from Proposition 4.8 we have that (4.7) holds for Φ. Taking the limit in (4.7) as $m \to \infty$ we have from Theorem 4.4 that

$$M^{(1)}(G) \le \int_{\mathbb{R}^d \times \mathcal{P}(E)} G(z,\nu)\Phi(dz,d\nu) \le M^{(2)}(G).$$

From Assumption 4.7 it now follows that $M^{(1)}$ equals $M^{(2)}$ and so

$$M^{(1)}(G) = \int_{\mathbb{R}^d \times \mathcal{P}(E)} G(z,\nu)\Phi(dz,d\nu).$$

Since $G \in \mathcal{G}$ is arbitrary and $\mathcal{G}$ is a measure determining class we have that $\Phi = M^{(1)}$. This proves the theorem. □

Acknowledgment. The author thanks a careful referee for pointing out errors and suggesting several improvements.

REFERENCES

[1] ATAR, R. and ZEITOUNI, O. (1997). Exponential stability for nonlinear filtering. *Ann. Inst. H. Poincaré Probab. Statist.* **33** 697–725.

[2] BHATT, A., BUDHIRAJA, A. and KARANDIKAR, R. (2000). Markov property and ergodicity of the nonlinear filter. *SIAM J. Control Optim.* **39** 928–949.

[3] CLARK, J. M. C., OCONE, D. L. and COUMARBATCH, C. (1999). Relative entropy and error bounds for filtering of Markov processes. *Math. Control Signals Systems* **12** 346–360.
[4] KALLIANPUR, G. (1980). *Stochastic Filtering Theory*. Springer, New York.
[5] KUNITA, H. (1971). Asymptotic behavior of the nonlinear filtering errors of Markov processes. *J. Multivariate Anal.* **1** 365–393.
[6] KUNITA, H. (1991). Ergodic properties of nonlinear filtering processes. In *Spatial Stochastic Processes* (K. C. Alexander and J. C. Watkins, eds.). Birkhäuser, Boston.
[7] LE GLAND, F. and MEVEL, L. (2000). Exponential forgetting and geometric ergodicity in hidden Markov models. *Math. Control Signals Systems* **13** 63–93.
[8] OCONE, D. and PARDOUX, E. (1996). Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM J. Control Optim.* **34** 226–243.
[9] STETTNER, L. (1989). On invariant measures of filtering processes. *Stochastic Differential Systems*. *Lecture Notes in Control and Inform Sci.* **126** 279–292. Springer, New York.
[10] STETTNER, L. (1991). Invariant measures of pair: State, approximate filtering process. *Colloq. Math.* **62** 347–352.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599-3260
E-MAIL: budhiraj@email.unc.edu