ON THE EXCHANGE OF INTERSECTION AND SUPREMUM OF σ-FIELDS IN FILTERING THEORY*

BY RAMON VAN HANDEL

Princeton University

We construct a stationary Markov process with trivial tail σ-field and a nondegenerate observation process such that the corresponding nonlinear filtering process is not uniquely ergodic. This settles in the negative a conjecture of the author in the ergodic theory of nonlinear filters arising from an erroneous proof in the classic paper of H. Kunita (1971), wherein an exchange of intersection and supremum of σ-fields is taken for granted.

1. Introduction and main result. Let E and F be Polish spaces, and consider an $E \times F$-valued stochastic process $(X_k, Y_k)_{k\in\mathbb{Z}}$ with the following properties:

1. $(X_k, Y_k)_{k\in\mathbb{Z}}$ is a stationary Markov process.
2. There exist transition kernels P from E to E and Φ from E to F such that $\mathbf{P}[(X_n, Y_n) \in A | X_{n-1}, Y_{n-1}] = \int \mathbf{1}_A(x, y)\, P(X_{n-1}, dx)\, \Phi(x, dy)$.

Such a process is called a stationary *hidden Markov model*; its dependence structure is illustrated schematically in Figure 1. In applications, $(X_k)_{k\in\mathbb{Z}}$ represents a "hidden" process which is not directly observable, while the observable process $(Y_k)_{k\in\mathbb{Z}}$ represents "noisy observations" of the hidden process [4].

Of fundamental importance in the theory of hidden Markov models is the *nonlinear filter* $(\pi_k)_{k\geq 0}$, defined as the regular conditional probability

$$\pi_n = \mathbf{P}[X_n \in \cdot\, | Y_1, \dots, Y_n].$$

That is, π_n is the conditional distribution of the current state of the hidden process given the observations to date. It is a basic fact in this theory that the filtering process $(\pi_k)_{k\geq 0}$ is itself a Markov process taking values in the space $\mathcal{P}(E)$ of probability measures on E, whose transition kernel $\mathsf{\Pi}$ can be expressed in terms of the transition kernels P and Φ that determine the model (this and other basic facts on nonlinear filters are reviewed in the appendix).

Following Kunita [12], we will be interested in the structure of the space of $\mathsf{\Pi}$-invariant probability measures in $\mathcal{P}(\mathcal{P}(E))$. It is easily seen that for every $\mathsf{\Pi}$-invariant measure $\mathsf{m} \in \mathcal{P}(\mathcal{P}(E))$, the barycenter $\mu \in \mathcal{P}(E)$ of m must be invariant

*This work was partially supported by NSF grant DMS-1005575.

AMS 2000 subject classifications: 37A25, 37A50, 60F20, 60G35, 60J05, 60K37

Keywords and phrases: nonlinear filtering, unique ergodicity, tail σ-field, exchange of intersection and supremum, random walk in random scenery

$$
\begin{array}{ccccccccc}
\cdots & \xrightarrow{P} & X_{n-1} & \xrightarrow{P} & X_n & \xrightarrow{P} & X_{n+1} & \xrightarrow{P} & \cdots \\
 & & \Phi\big\downarrow & & \Phi\big\downarrow & & \Phi\big\downarrow & & \\
\cdots & & Y_{n-1} & & Y_n & & Y_{n+1} & & \cdots
\end{array}
$$

FIG 1. *Dependence structure of a hidden Markov model.*

for the transition kernel P of the hidden process. Conversely, for every P-invariant measure $\mu \in \mathcal{P}(E)$, there exists at least one Π-invariant measure $\mathsf{m} \in \mathcal{P}(\mathcal{P}(E))$ whose barycenter is μ. However, the latter need not be unique.

THEOREM 1.1 (Kunita). *Let $\mathbf{P}[X_0 \in \cdot\,] := \mu$ be the P-invariant measure defined by the stationary hidden Markov model $(X_k, Y_k)_{k\in\mathbb{Z}}$ as above. If*

$$\bigcap_{n\le 0} (\mathcal{F}^Y_{-\infty,0} \vee \mathcal{F}^X_{-\infty,n}) = \mathcal{F}^Y_{-\infty,0} \quad \mathbf{P}\text{-}a.s., \tag{1.1}$$

then there exists a unique Π-invariant measure with barycenter μ. The converse holds if in addition Φ possesses a transition density with respect to some σ-finite reference measure. [Here $\mathcal{F}^Y_{-\infty,0} := \sigma\{Y_k : k \le 0\}$, $\mathcal{F}^X_{-\infty,n} := \sigma\{X_k : k \le n\}$.]

REMARK 1.2. Though the main ideas of the proof are implicitly contained in [12], this simple and general statement does not appear in the literature without various additional simplifying assumptions. For completeness, and in order to make this paper self-contained, we therefore include the proof in the appendix.

Theorem 1.1 is not actually stated as such by Kunita [12]. Instead, Kunita assumes that the hidden process $(X_k)_{k\in\mathbb{Z}}$ is purely nondeterministic:

DEFINITION 1.3. A stochastic process $(X_k)_{k\in\mathbb{Z}}$ is called *purely nondeterministic* if its past tail σ-field $\bigcap_{n\le 0} \mathcal{F}^X_{-\infty,n}$ is $\mathbf{P}$-a.s. trivial.

Kunita's main theorem states[1] that if the hidden process $(X_k)_{k\in\mathbb{Z}}$ is purely nondeterministic, then there exists a unique Π-invariant measure with barycenter μ. Kunita's proof, however, does not establish this claim. Indeed, at the crucial point

[1] In fact, Kunita's paper is written in the context of a continuous time model with white noise observations. None of these specific features are used in the proofs, however.

in the proof ([12], top of p. 384), Kunita implicitly takes for granted that the following exchange of intersection and supremum is permitted:

$$\bigcap_{n\le 0}\left(\mathcal{F}^Y_{-\infty,0}\vee\mathcal{F}^X_{-\infty,n}\right) \stackrel{?}{=} \mathcal{F}^Y_{-\infty,0}\vee\bigcap_{n\le 0}\mathcal{F}^X_{-\infty,n}\quad \mathbf{P}\text{-a.s.} \tag{1.2}$$

If this exchange were justified, then Kunita's result would indeed follow immediately from Theorem 1.1. However, in general, such an exchange of intersection and supremum is *not* permitted, as will be shown in section 1.1 below.

The goal of this paper is to settle, in the negative, a natural conjecture on the validity of the identity (1.2). Before we can describe the conjecture, we must review what is known about the validity of (1.2) in the filtering setting.

REMARK 1.4. Beyond the relevance of (1.2) to filtering theory, the problem studied in this paper provides a case study on an enigmatic problem: when is the exchange of countable intersection and supremum of σ-fields permitted? Such problems arise in remarkably diverse areas of probability theory. The following references provide some further context on this general problem.

1. Several distinguished mathematicians have given erroneous proofs related to the exchange of intersection and supremum of σ-fields, including Kolmogorov (see [22], p. 837) and Wiener (see [15], pp. 91–93).
2. A simple counterexample to the validity of the exchange of intersection and supremum due to Barlow and Perkins can be found in [31], p. 48. This example is closely related to the example given in section 1.1 below. See also [5], pp. 29–30 and the references therein.
3. The exchange of intersection and supremum appears in diverse probabilistic settings: see [29], section 5 and the references therein for various examples and counterexamples. In particular, the innovations problem and several variants of Tsirelson's celebrated counterexample provide a rich setting in which one can study the exchange of intersection and supremum problem; see [32, 9, 13, 2] and the references therein. See also [27] for a different connection to filtering theory.
4. Von Weizsäcker [29] gives a general necessary and sufficient condition for validity of the exchange of intersection and supremum, which is however often difficult to apply in practice. It is shown in [7] that the exchange of intersection and supremum is always valid in a given probability space if and only if its probability measure is purely atomic.

1.1. *A simple counterexample.* The gap in Kunita's proof was discovered in [1], where a simple counterexample to (1.2) was given. The following variant of this example will be helpful in understanding our main result.

Let $(\xi_k)_{k\in\mathbb{Z}}$ be an i.i.d. sequence of (Bernoulli) random variables uniformly distributed in $\{0,1\}$. Let $E=\{0,1\}\times\{0,1\}$ and $F=\{0,1\}$, and define the stochastic process $(X_k,Y_k)_{k\in\mathbb{Z}}$ taking values in $E\times F$ as follows:

$$X_n=(\xi_{n-1},\xi_n),\qquad\qquad Y_n=|\xi_n-\xi_{n-1}|.$$

It is evident that $(X_k,Y_k)_{k\in\mathbb{Z}}$ is a stationary hidden Markov model. Note that:

- Clearly $\xi_0=(\xi_{n-1}+Y_n+\cdots+Y_0)\bmod 2$ for any $n\le 0$. Therefore,
$$\xi_0\quad\text{is}\quad\bigcap_{n\le 0}(\mathcal{F}^Y_{-\infty,0}\vee\mathcal{F}^X_{-\infty,n})\text{-measurable.}$$
- On the other hand, as $\mathbf{P}[\xi_0=0|\mathcal{F}^Y_{-\infty,0}]=1/2$ by direct computation,
$$\xi_0\quad\text{is not}\quad\mathcal{F}^Y_{-\infty,0}\text{-measurable }\mathbf{P}\text{-a.s.}$$
- $(X_k)_{k\in\mathbb{Z}}$ is purely nondeterministic by the Kolmogorov zero-one law.

Therefore, evidently the identity (1.2) does not hold in this example.

1.2. *A positive result and a conjecture.* In view of the counterexample above, one might expect that the gap in Kunita's proof cannot be resolved in general. However, it turns out that such counterexamples are extremely fragile. For example, let $(\gamma_k)_{k\in\mathbb{Z}}$ be an i.i.d. sequence of standard Gaussian random variables, and let us modify the observation model in the above example to

$$Y_n=|\xi_n-\xi_{n-1}|+\varepsilon\gamma_n.$$

Then it can be verified that for arbitrarily small $\varepsilon>0$, the identity (1.2) holds. It is only in the degenerate case $\varepsilon=0$ that (1.2) fails. This suggests that the presence of some amount of noise, however small, is sufficient in order to ensure the validity of (1.2). This intuition can be made precise in a surprisingly general setting, which is established by the following result due to the author [25]. Here the notion of nondegeneracy formalizes the presence of observation noise.

DEFINITION 1.5. The hidden Markov model $(X_k,Y_k)_{k\in\mathbb{Z}}$ is said to possess *nondegenerate observations* if there exist a σ-finite reference measure φ on F and a strictly positive measurable function $g:E\times F\to\,]0,\infty[$ such that

$$\Phi(x,A)=\int\mathbf{1}_A(y)\,g(x,y)\,\varphi(dy)\quad\text{for all }x\in E,\ A\in\mathcal{B}(F).$$

THEOREM 1.6 ([25]). *Given a stationary hidden Markov model $(X_k,Y_k)_{k\in\mathbb{Z}}$ as defined in this section, with P-invariant measure $\mathbf{P}[X_0\in\,\cdot\,]:=\mu$, assume that:*

1. *The hidden process $(X_k)_{k\in\mathbb{Z}}$ is absolutely regular:*

$$\mathbf{E}[\|\mathbf{P}[X_n \in \cdot\,|X_0] - \mu\|_{\mathrm{TV}}] \xrightarrow{n\to\infty} 0. \tag{1.3}$$

2. *The observations are nondegenerate.*

Then the identity (1.1) holds true.

This result resolves the validity of (1.1) in many cases of interest. Indeed, the mixing assumption (1.3) holds in a very broad class of applications, and a well-established theory provides a powerful set of tools to verify this assumption [19]. Nonetheless, the assumption (1.3) is strictly stronger than the assumption that the hidden process is purely nondeterministic; the latter is equivalent to

$$\mathbf{E}[|\mathbf{P}[X_n \in A|X_0] - \mu(A)|] \xrightarrow{n\to\infty} 0 \quad \text{for all } A \in \mathcal{B}(E)$$

(see [24], Proposition 3). If, as one might conjecture, nondegeneracy of the observations suffices to justify the exchange of intersection and supremum (1.2), then Theorem 1.6 should already hold when the hidden process is only purely nondeterministic, i.e., Kunita's claim would hold true whenever the observations are nondegenerate. This stronger result was conjectured in [25], pp. 1877–1878.

CONJECTURE 1.7. *If the hidden process is purely nondeterministic and the observations are nondegenerate, then (1.1) holds true.*

Conjecture 1.7 seems tantalizingly close to Theorem 1.6, particularly if we rephrase (1.3) in terms of tail σ-fields. Indeed, let $\mathbf{P}^x$ be a version of the regular conditional probability $\mathbf{P}^{X_0} = \mathbf{P}[\,\cdot\,|X_0]$. Then, from the results of [25], for example, one may read off the following equivalent formulation of (1.3):

> There exists a set $E_0 \in \mathcal{B}(E)$ such that $\mu(E_0) = 1$ and
> for all $A \in \bigcap_{n\le 0} \mathcal{F}^X_{-\infty,n}$ and $x, y \in E_0$ $\quad \mathbf{P}^x[A] = \mathbf{P}^y[A] \in \{0, 1\}$.

On the other hand, clearly $(X_k)_{k\in\mathbb{Z}}$ is purely nondeterministic if and only if

> For any $A \in \bigcap_{n\le 0} \mathcal{F}^X_{-\infty,n}$, there exists $E_0 \in \mathcal{B}(E)$ (depending possibly on A) such that $\mu(E_0) = 1$ and for all $x, y \in E_0$ $\quad \mathbf{P}^x[A] = \mathbf{P}^y[A] \in \{0, 1\}$.

Thus the difference between the assumptions is that in the latter, the set E_0 may depend on A, while in the former E_0 does not depend on A.

1.3. *Main result.* The main result of this paper is that Conjecture 1.7 is false. We establish this by exhibiting a counterexample.

THEOREM 1.8. *There exists a stationary hidden Markov model $(X_k, Y_k)_{k\in\mathbb{Z}}$ in a Polish state space $E \times F$ such that the hidden process is purely nondeterministic and the observations are nondegenerate, but nonetheless (1.1) fails to hold.*

Moreover, this model may be constructed such that the transition kernel P of the hidden process is Feller, and such that the observations are of standard additive noise type $Y_n = h(X_n)+\varepsilon\gamma_n$ where $h : E \to \mathbb{R}^3$ is a bounded continuous function, $\varepsilon > 0$ and $(\gamma_k)_{k\in\mathbb{Z}}$ are standard Gaussian random variables in $\mathbb{R}^3$.

The counterexample to Conjecture 1.7, whose existence is guaranteed by this result, must surely yield a nasty filtering problem! Yet, Theorem 1.8 indicates the model need not even be *too* nasty: the example can be chosen to satisfy standard regularity assumptions and using a perfectly ordinary observation model. It therefore seems doubtful that the general result of Theorem 1.6 can be substantially weakened; absolute regularity (1.3) is evidently essential.

Let us briefly explain the intuition behind the counterexample. We aim to mimic the noiseless counterexample in section 1.1. The idea is to construct a variant of that model which has very long memory: we can then hope to average out the additional observation noise (needed to make the observations nondegenerate), reverting essentially to the noiseless case. On the other hand, we cannot give the process such long memory that it ceases to be purely nondeterministic. The following construction strikes a balance between these competing goals. We reconsider the example of section 1.1 not as a time series, but as a random scenery. We then construct a stochastic process by running a random walk on the integers, and reporting at each time the value of the scenery at the current location of the walk. The resulting *random walk in random scenery* [8, 11] is purely nondeterministic, yet has a very long memory due to the recurrence of the random walk. The latter is exploited by a remarkable *scenery reconstruction* result of Matzinger and Rolles [16] which allows us to average out the observation noise. Theorem 1.8 follows essentially by combining the scenery reconstruction with the example of section 1.1, except that we must work in a slightly larger state space for technical reasons.

REMARK 1.9. Random walks in random scenery are closely related to the T, T^{-1}-process, which was conjectured by Weiss ([30], p. 682) and later proved by Kalikow [10] to be a natural example of a K-process that is not a B-process. In the language of ergodic theory, a purely nondeterministic process is a K-process [20] while a process that satisfies (1.1) is an $\mathcal{F}^Y_{-\infty,0}$-relative K-process [21]. Our example may thus be interpreted as a K-process that is not K relative to a nondegenerate observation process. The absolute regularity property (1.3) is equivalent to the weak Bernoulli property in ergodic theory (cf. [28]).

We end this section with a brief discussion of the practical implications of Theorem 1.8. The mixing assumption (1.3) required by Theorem 1.6 states that the law of the hidden process converges in the sense of total variation to the invariant measure μ for almost every initial condition. This occurs in a wide variety of applications [19], as long as the hidden state space E is finite dimensional. In infinite dimensions, however, most probability measures are mutually singular and total variation convergence is rare. When the hidden process is defined by the solution of a stochastic partial differential equation, for example, typically the best we can hope for is weak convergence to the invariant measure. In this case (1.3) fails, though the process is still purely nondeterministic. Our main result indicates that nice ergodic properties of the nonlinear filter cannot be taken for granted in the infinite dimensional setting. This is unfortunate, as infinite dimensional filtering problems appear naturally in important applications such as weather prediction and geophysical or oceanographic data assimilation (see, e.g., [14]), while ergodicity of the nonlinear filter is essential to reliable performance of filtering algorithms [26]. The current state of knowledge on the ergodic theory of infinite dimensional filtering problems appears to be essentially nonexistent.

The remainder of this paper is organized as follows. In section 2 we introduce the various stochastic processes needed to construct our counterexample. Sections 3 and 4 are devoted to the proof of Theorem 1.8. The appendix reviews the ergodic theory of nonlinear filters (including a proof of Theorem 1.1).

2. Construction. In the following, we will work on the canonical probability space $(\Omega, \mathcal{F}, \mathbf{P})$ which supports the following independent random variables.

- $(\eta_k)_{k\in\mathbb{Z}}$, ξ_0 are i.i.d. random variables, uniformly distributed in $\{0, 1, 2\}$.
- $(\delta_k)_{k\in\mathbb{Z}}$ are i.i.d. random variables, uniformly distributed in $\{-1, 1\}$.
- $(\gamma_k)_{k\in\mathbb{Z}}$ are i.i.d. standard Gaussian random variables in $\mathbb{R}^3$.

Denote by $\{e(0), e(1), e(2)\} \subset \mathbb{R}^3$ the canonical basis in $\mathbb{R}^3$.

We now proceed to define various stochastic processes. Define recursively

$$\xi_n = \begin{cases} (\xi_{n-1} + \eta_n) \bmod 3 & \text{for } n > 0, \\ (\xi_{n+1} - \eta_{n+1}) \bmod 3 & \text{for } n < 0. \end{cases}$$

Note that $(\xi_k)_{k\in\mathbb{Z}}$ is an i.i.d. sequence uniformly distributed in $\{0, 1, 2\}$, and

$$\eta_n = (\xi_n - \xi_{n-1}) \bmod 3.$$

Next, we define the simple random walk $(N_k)_{k\in\mathbb{Z}}$ on $\mathbb{Z}$ as

$$N_n = \begin{cases} \sum_{k=1}^{n} \delta_k & \text{for } n \geq 0, \\ -\sum_{k=n+1}^{0} \delta_k & \text{for } n < 0. \end{cases}$$

We can now define the random walk in random scenery $(Z_k)_{k\in\mathbb{Z}}$ which takes values in the set $\{-1,1\}\times\{0,1,2\}\times\{0,1,2\} := I$ as follows:

$$Z_n = (Z_{n,0}, Z_{n,1}, Z_{n,2}) = (\delta_{n+1}, \xi_{N_n-1}, \xi_{N_n}).$$

It is not difficult to see that $(Z_n)_{n\in\mathbb{Z}}$ is a stationary process. We finally make the process Markovian by defining the $I^{\mathbb{Z}_+}$-valued process $(X_n)_{n\in\mathbb{Z}}$ as

$$X_n = (Z_k)_{k\geq n} \qquad \text{(that is, } X_{n,k} = Z_{n+k} \text{ for } k\in\mathbb{Z}_+\text{)},$$

and we define the $\mathbb{R}^3$-valued observation process $(Y_k)_{k\in\mathbb{Z}}$ as

$$Y_n = h(X_n) + \varepsilon\gamma_n = e(\eta_{N_n}) + \varepsilon\gamma_n,$$

where $\varepsilon > 0$ is a fixed constant and $h : I^{\mathbb{Z}_+} \to \mathbb{R}^3$ is defined as

$$h(x) = e((x_{0,2} - x_{0,1}) \bmod 3).$$

It is evident that the pair $(X_n, Y_n)_{n\in\mathbb{Z}}$ defines a stationary hidden Markov model taking values in the Polish space $I^{\mathbb{Z}_+}\times\mathbb{R}^3$ and with nondegenerate observations.

Let us define the σ-fields

$$\mathcal{F}^X_{m,n} = \sigma\{X_k : k\in[m,n]\}, \qquad \mathcal{F}^Y_{m,n} = \sigma\{Y_k : k\in[m,n]\},$$

for $m,n\in\mathbb{Z}$, $m\leq n$. The σ-fields $\mathcal{F}^X_{-\infty,n}$, $\mathcal{F}^X_{m,\infty}$, etc., are defined in the usual fashion (for example, $\mathcal{F}^X_{-\infty,n} = \bigvee_{m\leq n}\mathcal{F}^X_{m,n}$). Our main result is now as follows.

THEOREM 2.1. *For the hidden Markov model $(X_k, Y_k)_{k\in\mathbb{Z}}$ with nondegenerate observations, as defined in this section, the following hold:*

1. *The future tail σ-field*

$$\mathcal{T} := \bigcap_{n\geq 0}\mathcal{F}^X_{n,\infty} \quad \textit{is } \mathbf{P}\textit{-a.s. trivial.}$$

2. *We have the strict inclusion*

$$\bigcap_{n\geq 0}(\mathcal{F}^Y_{0,\infty}\vee\mathcal{F}^X_{n,\infty}) \supsetneq \mathcal{F}^Y_{0,\infty} \quad \mathbf{P}\textit{-a.s.},$$

provided that $\varepsilon > 0$ is chosen sufficiently small.

The proof of this result, given in section 3 below, is based on mixing and reconstruction results for random walks in random scenery [17, 16].

The model of Theorem 2.1 is time-reversed from the counterexample to be provided by Theorem 1.8. It is immediate from the Markov property, however, that the time reversal of a stationary hidden Markov model yields again a stationary hidden Markov model. Therefore, the following corollary is immediate:

COROLLARY 2.2. *For $\varepsilon > 0$ sufficiently small, the time-reversed model*

$$(\tilde{X}_k, \tilde{Y}_k)_{k\in\mathbb{Z}} := (X_{-k}, Y_{-k})_{k\in\mathbb{Z}}$$

is purely nondeterministic and has nondegenerate observations, but (1.1) fails.

This proves the first part of Theorem 1.8 and settles Conjecture 1.7. However, when constructed in this manner, the transition kernel of $(\tilde{X}_k)_{k\in\mathbb{Z}}$ cannot be chosen to satisfy the Feller property on $I^{\mathbb{Z}_+}$. Some further effort is therefore required to complete the proof of Theorem 1.8, which we postpone to section 4.

3. Proof of Theorem 2.1.

3.1. *First part.* Consider the stochastic process $\tilde{\xi}_n := (\xi_{n-1}, \xi_n)$. It is easily seen that this is a stationary, irreducible and aperiodic Markov chain taking values in the space $\{0,1,2\}\times\{0,1,2\}$, so that $(\tilde{\xi}_k)_{k\in\mathbb{Z}}$ is an ergodic process. The triviality of $\mathcal{T}$ now follows from the Theorem in [17], p. 267 (this follows in particular from equation (3) in [17] using [23], Theorem 7.9).

3.2. *Second part.* Consider the modified observation process $(Y'_k)_{k\in\mathbb{Z}}$ taking values in $\{0,1,2\}$, defined as follows:

$$Y'_n = \operatorname*{argmax}_{i=0,1,2} Y_{n,i}.$$

That is, Y'_n is the coordinate index of the largest component of the vector $Y_n \in \mathbb{R}^3$. By symmetry, it is easily seen that for some $\delta > 0$ depending on ε

$$\mathbf{P}[Y'_n = i | \eta_{N_n} = j] = \frac{\delta}{3} \quad \forall\, i \neq j, \qquad \mathbf{P}[Y'_n = i | \eta_{N_n} = i] = 1 - \frac{2\delta}{3} \quad \forall i,$$

where $\delta \downarrow 0$ as $\varepsilon \downarrow 0$. The conditional law of Y'_n can therefore be generated as follows: draw a Bernoulli random variable with parameter δ; if it is zero, set $Y'_n = \eta_{N_n}$, otherwise let Y'_n be a random draw from the uniform distribution on $\{0,1,2\}$. We can now apply the scenery reconstruction result from [16].

DEFINITION 3.1. Let $x, y \in \{0,1,2\}^{\mathbb{Z}}$. We write $x \approx y$ if there exist $a \in \{-1,1\}$ and $b \in \mathbb{Z}$ such that $x_n = y_{an+b}$ for all $n \in \mathbb{Z}$ (that is, $x \approx y$ iff the sequences x and y agree up to translation and/or reflection).

THEOREM 3.2 ([16]). *There is a measurable map $\iota : \{0,1,2\}^{\mathbb{Z}_+} \to \{0,1,2\}^{\mathbb{Z}}$ such that $\mathbf{P}[\iota((Y'_k)_{k\geq 0}) \approx (\eta_k)_{k\in\mathbb{Z}}] = 1$ provided $\varepsilon > 0$ is sufficiently small.*

From now on, let us fix $\varepsilon > 0$ sufficiently small and the map ι as in Theorem 3.2. By the definition of the equivalence relation $\approx$, there exist $\mathcal{F}^Y_{0,\infty} \vee \mathcal{F}^\eta_{-\infty,\infty}$-measurable random variables A and B, taking values in $\{-1,1\}$ and $\mathbb{Z}$, respectively, such that $\iota((Y'_k)_{k\geq 0})_n = \eta_{An+B}$ $\mathbf{P}$-a.s. for all $n \in \mathbb{Z}$.

REMARK 3.3. Let us note that, even though by construction $(\eta_{Ak+B})_{k\in\mathbb{Z}}$ is a.s. $\mathcal{F}^Y_{0,\infty}$-measurable, it is not possible for the random variables A and B to be $\mathcal{F}^Y_{0,\infty}$-measurable; see [11], Remark (ii). This will not be a problem for us.

The point of the above construction is the following claim: the random variable ξ_B is a.s. $\bigcap_n(\mathcal{F}^Y_{0,\infty} \vee \mathcal{F}^X_{n,\infty})$-measurable, but it is not a.s. $\mathcal{F}^Y_{0,\infty}$-measurable. This clearly suffices to prove the result. It thus remains to establish the claim.

LEMMA 3.4. *The random variable ξ_B is $\mathbf{P}$-a.s. $\bigcap_n(\mathcal{F}^Y_{0,\infty} \vee \mathcal{F}^X_{n,\infty})$-measurable.*

PROOF. Fix $n \in \mathbb{Z}$. Define the random variables $(\tau_k)_{k\in\mathbb{Z}}$ as

$$\tau_j = \inf\left\{k \geq 0 : \sum_{i=0}^{k-1} X_{n,i,0} = j\right\},$$

and define the random variables $(\xi'_k)_{k\in\mathbb{Z}}$ as

$$\xi'_j = X_{n,\tau_j,2}\,\mathbf{1}_{\tau_j<\infty}.$$

Then clearly $(\xi'_k)_{k\in\mathbb{Z}}$ is $\mathcal{F}^X_{n,\infty}$-measurable and $\mathbf{P}[(\xi'_k)_{k\in\mathbb{Z}} \approx (\xi_k)_{k\in\mathbb{Z}}] = 1$.

We now claim that we can "align" $(\xi'_k)_{k\in\mathbb{Z}}$ with $(\eta_{Ak+B})_{k\in\mathbb{Z}}$. Indeed, note that for any $b \in \mathbb{Z}$, we can estimate

$$\mathbf{P}\left[\eta_k = \eta_{k+b} \text{ for all } k \in \mathbb{Z}\right] \leq \mathbf{P}\left[\eta_0 = \eta_{kb} \text{ for all } k \geq 1\right] = 0,$$

$$\mathbf{P}\left[\eta_k = \eta_{-k+b} \text{ for all } k \in \mathbb{Z}\right] \leq \prod_{k=b}^{\infty} \mathbf{P}\left[\eta_k = \eta_{-k+b}\right] = 0,$$

where we have used that $(\eta_k)_{k\in\mathbb{Z}}$ are i.i.d. and nondeterministic. Therefore

$$\mathbf{P}\left[\text{there exist } a \in \{-1,1\},\ b \in \mathbb{Z} \text{ such that } \eta_k = \eta_{ak+b} \text{ for all } k \in \mathbb{Z}\right] = 0.$$

In particular, if we define $(\eta'_k)_{k\in\mathbb{Z}}$ as

$$\eta'_j = (\xi'_j - \xi'_{j-1}) \bmod 3,$$

it follows that there must exist $\mathbf{P}$-a.s. unique $\mathcal{F}^Y_{0,\infty} \vee \mathcal{F}^X_{n,\infty}$-measurable random variables A' and B', taking values in $\{-1,1\}$ and $\mathbb{Z}$, respectively, such that

$$\eta'_{A'j+B'} = \eta_{Aj+B} \quad \text{for all } j \in \mathbb{Z} \quad \mathbf{P}\text{-a.s.}$$

It follows by uniqueness that

$$\xi'_{A'j+B'} = \xi_{Aj+B} \quad \text{for all } j \in \mathbb{Z} \quad \mathbf{P}\text{-a.s.}$$

In particular, $\xi'_{B'} = \xi_B$ $\mathbf{P}$-a.s. But $\xi'_{B'}$ is $\mathcal{F}^Y_{0,\infty} \vee \mathcal{F}^X_{n,\infty}$-measurable by construction. Therefore, we have shown that ξ_B is $\mathbf{P}$-a.s. $\mathcal{F}^Y_{0,\infty} \vee \mathcal{F}^X_{n,\infty}$-measurable. As the choice of n was arbitrary, the proof is easily completed. $\square$

LEMMA 3.5. *The random variable ξ_B is not $\mathbf{P}$-a.s. $\mathcal{F}^Y_{0,\infty}$-measurable.*

PROOF. Note that $\mathbf{P}$-a.s.

$$\begin{aligned}
&\mathbf{P}[\xi_B = i,\ B = j | \mathcal{F}^\eta_{-\infty,\infty} \vee \mathcal{F}^\delta_{-\infty,\infty} \vee \mathcal{F}^\gamma_{-\infty,\infty}] \\
&\quad = \mathbf{1}_{B=j}\, \mathbf{P}[\xi_j = i | \mathcal{F}^\eta_{-\infty,\infty} \vee \mathcal{F}^\delta_{-\infty,\infty} \vee \mathcal{F}^\gamma_{-\infty,\infty}] \\
&\quad = \mathbf{1}_{B=j}\, \{\mathbf{P}[\xi_0 = i | \mathcal{F}^\eta_{-\infty,\infty} \vee \mathcal{F}^\delta_{-\infty,\infty} \vee \mathcal{F}^\gamma_{-\infty,\infty}] \circ \Theta^j\} \\
&\quad = \mathbf{1}_{B=j}\, \mathbf{P}[\xi_0 = i].
\end{aligned}$$

Here we have used that B is $\mathcal{F}^Y_{0,\infty} \vee \mathcal{F}^\eta_{-\infty,\infty}$-measurable for the first equality, stationarity of the law of $(\xi_k, \eta_k, \delta_k, \gamma_k)_{k\in\mathbb{Z}}$ for the second equality (Θ denotes the canonical shift), and independence of ξ_0 and $(\eta_k, \delta_k, \gamma_k)_{k\in\mathbb{Z}}$ for the third equality. Summing over j, and conditioning on $\mathcal{F}^Y_{0,\infty}$, we obtain

$$\mathbf{P}[\xi_B = i | \mathcal{F}^Y_{0,\infty}] = \mathbf{P}[\xi_0 = i] = 1/3 \quad \mathbf{P}\text{-a.s.}$$

Thus ξ_B is independent from $\mathcal{F}^Y_{0,\infty}$, hence not $\mathbf{P}$-a.s. $\mathcal{F}^Y_{0,\infty}$-measurable. $\square$

REMARK 3.6. The additive noise model $Y_n = h(X_n) + \varepsilon\gamma_n$ is inessential to the proof; we could have just as easily started from the $\{0, 1, 2\}$-valued observation model Y'_n as in [16]. The only reason we have chosen to construct our example with the additive noise model is to make the point that there is nothing special about the choice of observations: one does not have to "cook up" a complicated observation model to make the counterexample work. All the unpleasantness arises from the ergodic theory of random walks in random scenery.

4. Proof of Theorem 1.8. For any $x \in I^{\mathbb{Z}_+}$, define

$$\tau_j(x) = \inf\left\{k \geq 0 : \sum_{i=0}^{k-1} x_{i,0} = j\right\}.$$

Now define the space

$$E := \left\{x \in I^{\mathbb{Z}_+} : \tau_j(x) < \infty \text{ for all } j \in \mathbb{Z}\right\} \subset I^{\mathbb{Z}_+}.$$

We endow E with the topology of pointwise convergence (inherited from $I^{\mathbb{Z}_+}$).

LEMMA 4.1. *E is Polish.*

PROOF. For $x, x' \in E$, define the metric

$$d(x,x') := \sum_{k=0}^{\infty} 2^{-k}\,\mathbf{1}_{x_k \neq x'_k} + \sum_{j=-\infty}^{\infty} 2^{-|j|}\,\{|\tau_j(x) - \tau_j(x')| \wedge 1\}.$$

It suffices to prove that d metrizes the topology of pointwise convergence in E (which is certainly separable) and that (E, d) is a complete metric space.

We first prove that d metrizes the topology of pointwise convergence. Clearly $d(x_n, x) \to 0$ as $n \to \infty$ implies that $x_n \to x$ pointwise. Conversely, suppose that $x_n \to x$ as $n \to \infty$ pointwise. It suffices to show that $\tau_j(x_n) \to \tau_j(x)$ as $n \to \infty$ for all $j \in \mathbb{Z}$. But as $\tau_j(x) < \infty$ by assumption (as $x \in E$), it follows that $\tau_j(x_n) = \tau_j(x)$ whenever $x_{n,k} = x_k$ for all $k \leq \tau_j(x)$, which is the case for n sufficiently large by pointwise convergence. This establishes the claim.

It remains to show that (E, d) is complete. To this end, let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence for the metric d. Then it is clearly Cauchy for

$$\tilde{d}(x,x') := \sum_{k=0}^{\infty} 2^{-k}\,\mathbf{1}_{x_k \neq x'_k},$$

which defines a complete metric for the topology of pointwise convergence on $I^{\mathbb{Z}_+} \supset E$. Therefore, there exists $x \in I^{\mathbb{Z}_+}$ such that $x_n \to x$ as $n \to \infty$ pointwise. It suffices to show that $x \in E$. Indeed, when this is the case, it follows immediately that $d(x_n, x) \to 0$ as $n \to \infty$ (as we have shown that d metrizes pointwise convergence in E), thus proving completeness of (E, d).

To complete the proof, suppose that $x \notin E$. Then there exists $j \in \mathbb{Z}$ such that $\tau_j(x) = \infty$. In particular, if $x_{n,k} = x_k$ for all $k \leq N < \infty$, then $\tau_j(x_n) > N$. As this is the case for n sufficiently large by pointwise convergence, it follows that

$$\sup_{m\geq n} d(x_m, x_n) \geq 2^{-|j|} \sup_{m\geq n} |\tau_j(x_m) - \tau_j(x_n)| \wedge 1 = 2^{-|j|} \quad \text{for all } n \geq 1.$$

This contradicts the Cauchy property of $(x_n)_{n\in\mathbb{N}}$. $\square$

Denote by $\mathbf{P}[X_0 \in \cdot\,] := \mu$ the invariant measure of the $I^{\mathbb{Z}_+}$-valued Markov process $(X_k)_{k\in\mathbb{Z}}$ defined in section 2. It is clear that E is measurable as a subset of $I^{\mathbb{Z}_+}$ and that $\mu(E) = 1$. We are going to construct a Feller transition kernel $\tilde{P}$ from E to E with stationary measure μ (restricted to E), such that the corresponding stationary E-valued Markov process coincides a.s. with the stationary $I^{\mathbb{Z}_+}$-valued Markov process $(\tilde{X}_k)_{k\in\mathbb{Z}}$ defined in section 2.

LEMMA 4.2. *Define the transition kernel $\tilde{P} : E \times \mathcal{B}(E) \to [0,1]$ as follows:*

$$\tilde{P}(x, \{T_1(x)\}) = \tilde{P}(x, \{T_{-1}(x)\}) = \frac{1}{2},$$

where $T_a : E \to E$, $a \in \{-1, 1\}$ are defined as

$$T_a(x) = [(a, x_{\tau_{-a}(x),1}, x_{\tau_{-a}(x),2}), x].$$

Then the law under $\mathbf{P}$ of the process $(\tilde{X}_k)_{k\in\mathbb{Z}}$ defined in section 2 is that of a stationary Markov process taking values in E with transition kernel $\tilde{P}$ and invariant measure μ. Moreover, $\tilde{P}$ satisfies the Feller property.

PROOF. It follows along the lines of the proof of Lemma 4.1 that the functions T_1 and T_{-1} are continuous. Therefore, the Feller property of $\tilde{P}$ is immediate.

To complete the proof, it suffices (as clearly $\tilde{X}_n \in E$ $\mathbf{P}$-a.s. for all $n \in \mathbb{Z}$ and as $(\tilde{X}_k)_{k\in\mathbb{Z}}$ is a stationary Markov process) to show that

$$\mathbf{P}[\tilde{X}_1 \in A | \tilde{X}_0] = \tilde{P}(\tilde{X}_0, A) \quad \mathbf{P}\text{-a.s.} \quad \text{for all } A \in \mathcal{B}(E).$$

To this end, note that

$$\tilde{X}_1 = [(\delta_0, \xi_{-\delta_0-1}, \xi_{-\delta_0}), \tilde{X}_0] = [(\delta_0, \tilde{X}_{0,\tau_{-\delta_0}(\tilde{X}_0),1}, \tilde{X}_{0,\tau_{-\delta_0}(\tilde{X}_0),2}), \tilde{X}_0] \quad \mathbf{P}\text{-a.s.}$$

Moreover, as $\tilde{X}_0$ is $\mathcal{F}^{\xi}_{-\infty,\infty} \vee \mathcal{F}^{\delta}_{1,\infty}$-measurable, it follows from the construction in section 2 that δ_0 is independent of $\tilde{X}_0$. The result follows directly. □

PROOF OF THEOREM 1.8. Construct the canonical $E \times \mathbb{R}^3$-valued stationary hidden Markov model $(X'_k, Y'_k)_{k\in\mathbb{Z}}$ such that the hidden process $(X'_k)_{k\in\mathbb{Z}}$ has transition kernel $\tilde{P}$ and invariant measure $X'_0 \sim \mu$, and with the observation model $Y'_n = h(X'_n) + \varepsilon\gamma_n$ where $(\gamma_k)_{k\in\mathbb{Z}}$ is an i.i.d. sequence of standard Gaussian random variables in $\mathbb{R}^3$ independent of $(X'_k)_{k\in\mathbb{Z}}$. Clearly E and $\mathbb{R}^3$ are Polish by Lemma 4.1, the observations are nondegenerate, $h : E \to \mathbb{R}^3$ (defined in section 2) is bounded and continuous, and $\tilde{P}$ is Feller by Lemma 4.2. Moreover, the law of the model $(X'_k, Y'_k)_{k\in\mathbb{Z}}$ coincides with that of $(\tilde{X}_k, \tilde{Y}_k)_{k\in\mathbb{Z}}$ as defined in section 2. Therefore, by Corollary 2.2, $(X'_k)_{k\in\mathbb{Z}}$ is purely nondeterministic but (1.1) fails for this model when $\varepsilon > 0$ is chosen sufficiently small. □

APPENDIX A: ERGODIC THEORY OF NONLINEAR FILTERS

The goal of the appendix is to collect a few basic results on the ergodic theory of nonlinear filters. Similar results appear in various forms in the literature, see, for example, [3, 6] and the references therein. However, all known proofs require various simplifying assumptions, such as the Feller property or irreducibility of the hidden process, nondegenerate observations, etc. As a general result does not appear to be readily available in the literature, we provide here a largely self-contained treatment culminating in the proof of Theorem 1.1.

Let us note that analogous results can be obtained for continuous time, either by direct arguments (cf. [33]) or by reduction to discrete time (as in [25]).

A.1. Markov property of the filter. As in the introduction, we let E and F be Polish spaces, let $P : E \times \mathcal{B}(E) \to [0,1]$ and $\Phi : E \times \mathcal{B}(F) \to [0,1]$ be the transition kernels, and let $\mu : \mathcal{B}(E) \to [0,1]$ be the P-invariant measure defining the law of the stationary hidden Markov model $(X_k, Y_k)_{k\in\mathbb{Z}}$. We denote by $\mathcal{P}(G)$ the space of probability measures on the Polish space G, endowed with the topology of weak convergence of probability measures.

LEMMA A.1 ([18], Lemma 1). *For $\nu \in \mathcal{P}(E)$, define the probability measure*

$$P_\nu(A) = \int \mathbf{1}_A(x,y)\,\nu(dx')\,P(x',dx)\,\Phi(x,dy) \quad \textit{for all } A \in \mathcal{B}(E\times F).$$

Denote by $X : E\times F \to E$ and $Y : E\times F \to F$ the canonical projections. There exists a measurable map $\Pi : \mathcal{P}(E)\times F \to \mathcal{P}(E)$ such that $\Pi(\nu, Y)$ is a version of the regular conditional probability $P_\nu(X \in \cdot\,|Y)$ for every $\nu \in \mathcal{P}(E)$.

We now define the transition kernel $\sqcap : \mathcal{P}(E)\times\mathcal{B}(\mathcal{P}(E)) \to [0,1]$ as follows:

$$\sqcap(\nu, A) = \int \mathbf{1}_A(\Pi(\nu,y))\,\nu(dx')\,P(x',dx)\,\Phi(x,dy).$$

We claim that the nonlinear filter $(\pi_k)_{k\geq 0}$ is a $\mathcal{P}(E)$-valued Markov process with transition kernel $\sqcap$. To prove this we will need the following result on conditioning under a regular conditional probability due to von Weizsäcker.

LEMMA A.2 ([29]). *Let G, G' and H be Polish spaces, and denote by g, g' and h the canonical projections from $G\times G'\times H$ on G, G' and H, respectively. Let $\mathbf{Q}$ be a probability measure on $G\times G'\times H$, and let $q_{\cdot,\cdot} : G\times G'\times\mathcal{B}(H) \to [0,1]$ and $q_\cdot : G\times\mathcal{B}(G'\times H) \to [0,1]$ be versions of the regular conditional probabilities $\mathbf{Q}[h \in \cdot\,|g,g']$ and $\mathbf{Q}[(g',h) \in \cdot\,|g]$, respectively. Then for $\mathbf{Q}$-a.e. $x \in G$, the kernel $q_{x,g'}[\,\cdot\,]$ is a version of the regular conditional probability $q_x[h \in \cdot\,|g']$.*

PROPOSITION A.3. *For $n \geq 0$, let the nonlinear filter π_n be a version of the regular conditional probability $\mathbf{P}[X_n \in \cdot\,|Y_1,\ldots,Y_n]$. Then $(\pi_k)_{k\geq 0}$ is a $\mathcal{P}(E)$-valued Markov process with transition kernel $\sqcap$ and initial measure $\pi_0 \sim \delta_\mu$.*

PROOF. Fix $n \geq 1$. It is easily seen that for any $B \in \mathcal{B}(E \times F)$

$$\mathbf{P}[(X_n, Y_n) \in B | Y_1, \dots, Y_{n-1}] = \int \mathbf{1}_A(x, y)\, \pi_{n-1}(dx')\, P(x', dx)\, \Phi(x, dy).$$

Using Lemmas A.2 and A.1 and uniqueness of regular conditional probabilities, we find the recursive formula $\pi_n = \Pi(\pi_{n-1}, Y_n)$ $\mathbf{P}$-a.s. It follows easily that

$$\mathbf{P}[\pi_n \in A | Y_1, \dots, Y_{n-1}] = \mathsf{\Pi}(\pi_{n-1}, A) \quad \mathbf{P}\text{-a.s.} \quad \text{for all } A \in \mathcal{B}(\mathcal{P}(E)),$$

completing the proof. □

We now establish the two elementary facts stated in the introduction.

LEMMA A.4. *Let* $\mathsf{m} \in \mathcal{P}(\mathcal{P}(E))$ *be any* $\mathsf{\Pi}$*-invariant probability measure. Then the barycenter of* m *is a* P*-invariant probability measure.*

PROOF. Let $m \in \mathcal{P}(E)$ be the barycenter of m. By definition,

$$m(A) = \int \nu(A)\, \mathsf{m}(d\nu) = \int \nu(A)\, \mathsf{\Pi}(\nu', d\nu)\, \mathsf{m}(d\nu') \quad \text{for } A \in \mathcal{B}(E).$$

But note that $\int \nu(A)\, \mathsf{\Pi}(\nu', d\nu) = \mathbf{E}_{P_{\nu'}}[P_{\nu'}(X \in A|Y)] = \int P(x, A)\, \nu'(dx)$ by the definition of $\mathsf{\Pi}$. It follows directly that $mP = m$, that is, m is P-invariant. □

LEMMA A.5. *There is at least one* $\mathsf{\Pi}$*-invariant measure with barycenter* μ.

PROOF. For $n \in \mathbb{Z}$, let $\tilde{\pi}_n$ be a version of the regular conditional probability $\mathbf{P}[X_n \in \cdot\, | \mathcal{F}^Y_{-\infty,n}]$. Proceeding exactly as in the proof of Proposition A.3, we find that $(\tilde{\pi}_k)_{k \in \mathbb{Z}}$ is a $\mathcal{P}(E)$-valued Markov process with transition kernel $\mathsf{\Pi}$. But as the underlying hidden Markov model $(X_k, Y_k)_{k \in \mathbb{Z}}$ is stationary, clearly $(\tilde{\pi}_k)_{k \in \mathbb{Z}}$ is also stationary. Therefore, the law of $\tilde{\pi}_0$ is a $\mathsf{\Pi}$-invariant measure, and its barycenter is μ by the tower property of the conditional expectation. □

A.2. Proof of Theorem 1.1: sufficiency. The proof is essentially contained in Kunita [12], though we are careful here not to exploit any unnecessary assumptions. The idea is to introduce a suitable randomization, which is most conveniently done in a canonical probability model. To this end, define the Polish space $\Omega_0 = \mathcal{P}(E) \times E \times (E \times F)^{\mathbb{N}}$ with the canonical projections $m_0 : \Omega_0 \to \mathcal{P}(E)$ and (with a slight abuse of notation) $X_0 : \Omega_0 \to E$, $(X_k, Y_k)_{k \geq 1} : \Omega_0 \to (E \times F)^{\mathbb{N}}$. Given $\mathsf{m} \in \mathcal{P}(\mathcal{P}(E))$, we define a probability measure $\mathbf{P}_{\mathsf{m}}$ on Ω_0 with the finite dimensional distributions

$$\begin{aligned} &\mathbf{P}_{\mathsf{m}}((m_0, X_0, \dots, X_n, Y_1, \dots, Y_n) \in A) = \\ &\qquad \int \mathbf{1}_A(\nu, x_0, \dots, x_n, y_1, \dots, y_n)\, \nu(dx_0)\, P(x_0, dx_1)\, \Phi(x_1, dy_1) \cdots \\ &\qquad\qquad P(x_{n-1}, dx_n)\, \Phi(x_n, dy_n)\, \mathsf{m}(d\nu). \end{aligned}$$

We now define for $n \geq 0$ three distinguished nonlinear filters:

$$\begin{aligned} \pi_n^{\min} &:= \mathbf{P}_{\mathsf{m}}[X_n \in \cdot\,|Y_1, \ldots, Y_n], \\ \pi_n^{\mathsf{m}} &:= \mathbf{P}_{\mathsf{m}}[X_n \in \cdot\,|m_0, Y_1, \ldots, Y_n], \\ \pi_n^{\max} &:= \mathbf{P}_{\mathsf{m}}[X_n \in \cdot\,|m_0, X_0, Y_1, \ldots, Y_n]. \end{aligned}$$

We now have the following easy result. Here $\delta_\mu, \varepsilon_\mu \in \mathcal{P}(\mathcal{P}(E))$ are defined by $\delta_\mu(A) = \mathbf{1}_{\mu \in A}$ (as usual) and $\varepsilon_\mu(A) = \int \mathbf{1}_{\delta_x \in A}\, \mu(dx)$.

LEMMA A.6. *Let $\mathsf{m} \in \mathcal{P}(\mathcal{P}(E))$ be any probability measure with barycenter μ. Then $(\pi_n^{\min})_{n \geq 0}$, $(\pi_n^{\mathsf{m}})_{n \geq 0}$, $(\pi_n^{\max})_{n \geq 0}$ are $\mathcal{P}(E)$-valued Markov processes under $\mathbf{P}_{\mathsf{m}}$ with transition kernel $\sqcap$ and initial measures δ_μ, m, ε_μ, respectively.*

PROOF. The proof is identical to that of Proposition A.3. □

The following result completes the proof of sufficiency.

PROPOSITION A.7. *Let $p \in \mathbb{N}$, let $f_i : E \to \mathbb{R}$, $i = 1, \ldots, p$ be bounded measurable functions, and let $\kappa : \mathbb{R}^p \to \mathbb{R}$ be convex. Define the bounded measurable function $F : \mathcal{P}(E) \to \mathbb{R}$ as $F(\nu) = \kappa\left(\int f_1(x)\, \nu(dx), \ldots, \int f_p(x)\, \nu(dx)\right)$. Finally, let $\mathsf{m} \in \mathcal{P}(\mathcal{P}(E))$ be any $\sqcap$-invariant measure with barycenter μ. Then*

$$\begin{aligned} \mathbf{E}\left[\kappa\left(\mathbf{E}[f_1(X_0)|\mathcal{F}^Y_{-\infty,0}], \ldots, \mathbf{E}[f_p(X_0)|\mathcal{F}^Y_{-\infty,0}]\right)\right] &\leq \int F(\nu)\, \mathsf{m}(d\nu) \\ &\leq \mathbf{E}\left[\kappa\left(\mathbf{E}[f_1(X_0)|\mathcal{G}_{-\infty,0}], \ldots, \mathbf{E}[f_p(X_0)|\mathcal{G}_{-\infty,0}]\right)\right], \end{aligned}$$

where $\mathcal{G}_{-\infty,0} := \bigcap_n (\mathcal{F}^Y_{-\infty,0} \vee \mathcal{F}^X_{-\infty,n})$. In particular, if (1.1) holds, m coincides with the distinguished $\sqcap$-invariant measure defined in the proof of Lemma A.5.

PROOF. Note that as κ is convex, it is continuous, hence F is bounded and measurable. It is an immediate consequence of Jensen's inequality that

$$\mathbf{E}_{\mathsf{m}}[F(\pi_n^{\min})] \leq \mathbf{E}_{\mathsf{m}}[F(\pi_n^{\mathsf{m}})] = \int F(\nu)\, \mathsf{m}(d\nu) \leq \mathbf{E}_{\mathsf{m}}[F(\pi_n^{\max})]$$

for every $n \geq 0$, where we have used Lemma A.6 and the $\sqcap$-invariance of m to obtain the middle equality. Using Lemma A.6 and the stationarity of $(X_k, Y_k)_{k \in \mathbb{Z}}$ under $\mathbf{P}$, it is also easily seen that the laws of $\pi_n^{\min}(f)$, $\pi_n^{\max}(f)$ under $\mathbf{P}_{\mathsf{m}}$ coincide with the laws of $\mathbf{E}[f(X_0)|Y_{-n+1}, \ldots, Y_0]$, $\mathbf{E}[f(X_0)|X_{-n}, Y_{-n+1}, \ldots, Y_0]$ under $\mathbf{P}$, respectively. We therefore have for every $n \geq 0$

$$\begin{aligned} \mathbf{E}\left[\kappa\left(\mathbf{E}[f_1(X_0)|\mathcal{F}^Y_{-n+1,0}], \ldots, \mathbf{E}[f_p(X_0)|\mathcal{F}^Y_{-n+1,0}]\right)\right] &\leq \int F(\nu)\, \mathsf{m}(d\nu) \\ &\leq \mathbf{E}\left[\kappa\left(\mathbf{E}[f_1(X_0)|\mathcal{G}_{-n,0}], \ldots, \mathbf{E}[f_p(X_0)|\mathcal{G}_{-n,0}]\right)\right], \end{aligned}$$

where $\mathcal{G}_{-n,0} := \mathcal{F}^Y_{-\infty,0} \vee \mathcal{F}^X_{-\infty,-n}$ and we have used the fact that

$$\mathbf{E}[f(X_0)|X_{-n}, Y_{-n+1}, \ldots, Y_0] = \mathbf{E}[f(X_0)|\mathcal{G}_{-n,0}] \quad \mathbf{P}\text{-a.s.}$$

as $\mathcal{F}^X_{-n+1,0} \vee \mathcal{F}^Y_{-n+1,0}$ is conditionally independent of $\mathcal{F}^X_{-\infty,-n-1} \vee \mathcal{F}^Y_{-\infty,-n}$ given X_{-n}. But as κ is continuous, the equation display in the statement of the result follows by letting $n \to \infty$ using the martingale convergence theorem.

Now suppose that (1.1) holds, and denote by m_0 be the distinguished Π-invariant measure obtained in the proof of Lemma A.5. Then we have evidently shown that $\int F(\nu)\, \mathsf{m}(d\nu) = \int F(\nu)\, \mathsf{m}_0(d\nu)$ for all functions F of the form $F(\nu) = \kappa\left(\int f_1(x)\, \nu(dx), \ldots, \int f_p(x)\, \nu(dx)\right)$ for any p, bounded measurable $f_1, \ldots, f_p$ and convex κ. We claim that this class of functions is measure-determining, so we can conclude that $\mathsf{m} = \mathsf{m}_0$. To establish the claim, first note that by the Stone-Weierstrass theorem, any continuous function on $\mathbb{R}^p$ can be approximated uniformly on any compact set by the difference of convex functions. As $f_1, \ldots, f_p$ are bounded (hence take values in a compact subset of $\mathbb{R}^p$), it therefore suffices to assume that κ is continuous rather than convex. Next, note that the indicator function $\mathbf{1}_A$ of any open subset A of $\mathbb{R}^p$ can be obtained as the increasing limit of nonnegative continuous functions. It therefore suffices to assume that κ is the indicator of an open subset of $\mathbb{R}^p$. But any probability measure on a Polish space is regular, so it suffices to assume that κ is the indicator function of a Borel subset of $\mathbb{R}^p$. The proof is completed by an application of the Dynkin system lemma. □

A.3. Proof of Theorem 1.1: necessity. We will in fact prove necessity under a weaker assumption than stated in the theorem: the key assumption is

$$\text{(A.1)} \quad \bigcap_{n\le 0} (\mathcal{F}^Y_{-\infty,k} \vee \mathcal{F}^X_{-\infty,n}) = \mathcal{F}^Y_{1,k} \vee \bigcap_{n\le 0} (\mathcal{F}^Y_{-\infty,0} \vee \mathcal{F}^X_{-\infty,n}) \quad \mathbf{P}\text{-a.s.} \quad \forall\, k \in \mathbb{N}.$$

The assumption in the theorem that Φ possesses a transition density only enters the proof inasmuch as it guarantees the validity (A.1). Let us note that the assumption of the theorem is itself weaker than nondegeneracy of the observations, as the transition density is not required to be strictly positive here.

LEMMA A.8. *Suppose there exists a σ-finite reference measure φ on F and a transition density $g : E \times F \to [0, \infty[$ such that $\Phi(x, A) = \int \mathbf{1}_A(y)\, g(x, y)\, \varphi(dy)$ for all $x \in E$, $A \in \mathcal{B}(F)$. Then the identity (A.1) holds true.*

PROOF. It is easily seen that the assumption guarantees the existence of a probability measure $\mathbf{Q}$ such that $\mathbf{P} \ll \mathbf{Q}$ and $\mathcal{F}^Y_{1,k}$ is independent of $\mathcal{F}^X_{-\infty,0} \vee \mathcal{F}^Y_{-\infty,0}$ under $\mathbf{Q}$. Thus the identity in (A.1) holds $\mathbf{Q}$-a.s., and therefore $\mathbf{P}$-a.s. □

The proof is based on the following result.

LEMMA A.9. *Suppose there exists a unique Π-invariant measure with barycenter μ and that assumption (A.1) holds. Then we have for every $A \in \mathcal{B}(E)$*

$$\mathbf{P}[X_0 \in A | \textstyle\bigcap_n (\mathcal{F}^Y_{-\infty,0} \vee \mathcal{F}^X_{-\infty,n})] = \mathbf{P}[X_0 \in A | \mathcal{F}^Y_{-\infty,0}] \quad \mathbf{P}\text{-}a.s.$$

PROOF. Define the regular conditional probabilities $\pi^0_k = \mathbf{P}[X_k \in \cdot | \mathcal{F}^Y_{-\infty,k}]$ and $\pi^1_k = \mathbf{P}[X_k \in \cdot | \bigcap_n (\mathcal{F}^Y_{-\infty,k} \vee \mathcal{F}^X_{-\infty,n})]$, and denote by $\mathsf{m}_0, \mathsf{m}_1 \in \mathcal{P}(\mathcal{P}(E))$ the laws of π^0_0 and π^1_0, respectively. Then m_0 is the Π-invariant measure defined in the proof of Lemma A.5. We claim that m_1 is also Π-invariant. Indeed, this follows as a variant of Lemma A.2 (pp. 95–96 in [29]) and the assumption (A.1) imply that $\pi^1_k = \Pi(\pi^1_{k-1}, Y_k)$ $\mathbf{P}$-a.s., so that $(\pi^1_k)_{k \in \mathbb{Z}}$ is Markov with transition kernel Π, while $(\pi^1_k)_{k \in \mathbb{Z}}$ is easily seen to be a stationary process.

Clearly m_0 and m_1 both have barycenter μ, so by assumption $\mathsf{m}_0 = \mathsf{m}_1$. Thus

$$\mathbf{E}[(\pi^1_k(A) - \pi^0_k(A))^2] = \mathbf{E}[(\pi^1_k(A))^2] - \mathbf{E}[(\pi^0_k(A))^2] = \mathsf{m}_1(F_A) - \mathsf{m}_0(F_A) = 0$$

for every $A \in \mathcal{B}(E)$, where we defined $F_A : \nu \mapsto (\nu(A))^2$. It follows that $\pi^1_0(A) = \pi^0_0(A)$ $\mathbf{P}$-a.s. for every $A \in \mathcal{B}(E)$, which completes the proof. $\square$

To complete the proof, we require the following easy variant of Lemma A.1.

LEMMA A.10. *For $\nu \in \mathcal{P}(E)$ and $k \in \mathbb{N}$, define the probability measure*

$$P^k_\nu(A) = \int \mathbf{1}_A(x_0, y_1, \ldots, y_k)\, \nu(dx_0)\, P(x_0, dx_1)\, \Phi(x_1, dy_1) \cdots \\ P(x_{k-1}, dx_k)\, \Phi(x_k, dy_k) \quad \textit{for } A \in \mathcal{B}(E \times F^k).$$

Denote by $X : E \times F^k \to E$ and $Y^k : E \times F^k \to F^k$ the canonical projections. There exists a measurable map $\Sigma^k : \mathcal{P}(E) \times F^k \to \mathcal{P}(E)$ such that $\Sigma^k(\nu, Y^k)$ is a version of the regular conditional probability $P^k_\nu(X \in \cdot | Y^k)$ for every $\nu \in \mathcal{P}(E)$.

We now complete the proof.

PROPOSITION A.11. *Suppose there exists a unique Π-invariant measure with barycenter μ and that assumption (A.1) holds. Then (1.1) holds true.*

PROOF. As $\bigcup_{k \le 0} L^1(\mathcal{F}^X_{k,0} \vee \mathcal{F}^Y_{k,0}, \mathbf{P})$ is dense in $L^1(\mathcal{F}^X_{-\infty,0} \vee \mathcal{F}^Y_{-\infty,0}, \mathbf{P})$, it suffices to show that for every $k \le 0$ and $Z \in L^1(\mathcal{F}^X_{k,0} \vee \mathcal{F}^Y_{k,0}, \mathbf{P})$

$$\mathbf{E}[Z | \textstyle\bigcap_n (\mathcal{F}^Y_{-\infty,0} \vee \mathcal{F}^X_{-\infty,n})] = \mathbf{E}[Z | \mathcal{F}^Y_{-\infty,0}] \quad \mathbf{P}\text{-a.s.}$$

However, for $Z \in L^1(\mathcal{F}^X_{k,0} \vee \mathcal{F}^Y_{k,0}, \mathbf{P})$, we have by the Markov property

$$\mathbf{E}[Z | \textstyle\bigcap_n (\mathcal{F}^Y_{-\infty,0} \vee \mathcal{F}^X_{-\infty,n})] = \mathbf{E}[\, \mathbf{E}[Z | \sigma\{X_k\} \vee \mathcal{F}^Y_{k,0}] \, | \textstyle\bigcap_n (\mathcal{F}^Y_{-\infty,0} \vee \mathcal{F}^X_{-\infty,n})].$$

It therefore suffices to consider $Z \in L^1(\sigma\{X_k\} \vee \mathcal{F}^Y_{k,0}, \mathbf{P})$. But note that the class of random variables $\{Z^X Z^Y : Z^X \in L^\infty(\sigma\{X_k\}, \mathbf{P}),\ Z^Y \in L^\infty(\mathcal{F}^Y_{k,0}, \mathbf{P})\}$ is total in $L^1(\sigma\{X_k\} \vee \mathcal{F}^Y_{k,0}, \mathbf{P})$. Therefore, it suffices to show that

$$\mathbf{P}\big[X_k \in A \big| \textstyle\bigcap_n (\mathcal{F}^Y_{-\infty,0} \vee \mathcal{F}^X_{-\infty,n})\big] = \mathbf{P}\big[X_k \in A \big| \mathcal{F}^Y_{-\infty,0}\big] \quad \mathbf{P}\text{-a.s.}$$

for all $k \leq 0$ and $A \in \mathcal{B}(E)$. For $k = 0$, this follows directly from Lemma A.9.

For $k < 0$, we proceed as follows. Define π^0_k and π^1_k as in the proof of Lemma A.9. It is easily established using Lemma A.2 that

$$\mathbf{P}\big[X_k \in \cdot \big| \mathcal{F}^Y_{-\infty,0}\big] = \Sigma^k(\pi^0_k, Y_{k+1}, \ldots, Y_0) \quad \mathbf{P}\text{-a.s.}$$

Similarly, a variant of Lemma A.2 (pp. 95–96 in [29]) and (A.1) imply

$$\mathbf{P}\big[X_k \in \cdot \big| \textstyle\bigcap_n (\mathcal{F}^Y_{-\infty,0} \vee \mathcal{F}^X_{-\infty,n})\big] = \Sigma^k(\pi^1_k, Y_{k+1}, \ldots, Y_0) \quad \mathbf{P}\text{-a.s.}$$

But by Lemma A.9, applying the Dynkin system lemma with a countable generating system, and using that $(X_k, Y_k)_{k\in\mathbb{Z}}$ is stationary under $\mathbf{P}$, it follows directly that $\pi^0_k = \pi^1_k$ $\mathbf{P}$-a.s. This completes the proof. □

REFERENCES

[1] BAXENDALE, P., CHIGANSKY, P., AND LIPTSER, R. (2004). Asymptotic stability of the Wonham filter: ergodic and nonergodic signals. *SIAM J. Control Optim.* **43**, 2, 643–669.

[2] BROSSARD, J. AND LEURIDAN, C. (2007). Chaînes de Markov constructives indexées par **Z**. *Ann. Probab.* **35**, 2, 715–731.

[3] BUDHIRAJA, A. (2003). Asymptotic stability, ergodicity and other asymptotic properties of the nonlinear filter. *Ann. Inst. H. Poincaré Probab. Statist.* **39**, 6, 919–941.

[4] CAPPÉ, O., MOULINES, E., AND RYDÉN, T. (2005). *Inference in hidden Markov models*. Springer Series in Statistics. Springer, New York.

[5] CHAUMONT, L. AND YOR, M. (2003). *Exercises in probability*. Cambridge University Press, Cambridge.

[6] CHIGANSKY, P. AND VAN HANDEL, R. (2010). A complete solution to Blackwell's unique ergodicity problem for hidden Markov chains. *Ann. Appl. Probab.* **20**, 6, 2318–2345.

[7] CRIMALDI, I., LETTA, G., AND PRATELLI, L. (2007). Sur l'interversion de l'ordre entre deux opérations sur les tribus. *C. R. Math. Acad. Sci. Paris* **345**, 6, 341–344.

[8] DEN HOLLANDER, F. AND STEIF, J. E. (2006). Random walk in random scenery: a survey of some recent results. In *Dynamics & stochastics*. IMS Lecture Notes Monogr. Ser., Vol. **48**. Inst. Math. Statist., Beachwood, OH, 53–65.

[9] ÉMERY, M. AND SCHACHERMAYER, W. (2001). On Vershik's standardness criterion and Tsirelson's notion of cosiness. In *Séminaire de Probabilités, XXXV*. Lecture Notes in Math., Vol. **1755**. Springer, Berlin, 265–305.

[10] KALIKOW, S. A. (1982). T, T^{-1} transformation is not loosely Bernoulli. *Ann. of Math. (2)* **115**, 2, 393–409.

[11] KESTEN, H. (1998). Distinguishing and reconstructing sceneries from observations along random walk paths. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*. DIMACS Ser. Discrete Math. Theoret. Comput. Sci., Vol. **41**. Amer. Math. Soc., Providence, RI, 75–83.

[12] KUNITA, H. (1971). Asymptotic behavior of the nonlinear filtering errors of Markov processes. *J. Multivariate Anal. 1*, 365–393.

[13] LAURENT, S. (2010). Further comments on the representation problem for stationary processes. *Statist. Probab. Lett.* **80**, 7-8, 592–596.

[14] MAJDA, A. J., HARLIM, J., AND GERSHGORIN, B. (2010). Mathematical strategies for filtering turbulent dynamical systems. *Discrete Contin. Dyn. Syst.* **27**, 2, 441–486.

[15] MASANI, P. (1966). Wiener's contributions to generalized harmonic analysis, prediction theory and filter theory. *Bull. Amer. Math. Soc. 72*, 73–125.

[16] MATZINGER, H. AND ROLLES, S. W. W. (2003). Reconstructing a random scenery observed with random errors along a random walk path. *Probab. Theory Related Fields* **125**, 4, 539–577.

[17] MEILIJSON, I. (1974). Mixing properties of a class of skew-products. *Israel J. Math. 19*, 266–270.

[18] MEYER, P. A. (1976). La théorie de la prédiction de F. Knight. In *Séminaire de Probabilités, X (Première partie, Univ. Strasbourg, Strasbourg, année universitaire 1974/1975)*. Springer, Berlin, 86–103. Lecture Notes in Math., Vol. 511.

[19] MEYN, S. AND TWEEDIE, R. L. (2009). *Markov chains and stochastic stability*, Second ed. Cambridge University Press, Cambridge.

[20] ORNSTEIN, D. S. (1973). An application of ergodic theory to probability theory. *Ann. Probability* **1**, 1, 43–65.

[21] RAHE, M. (1978). Relatively finitely determined implies relatively very weak Bernoulli. *Canad. J. Math.* **30**, 3, 531–548.

[22] SINAI, Y. G. (1989). Kolmogorov's work on ergodic theory. *Ann. Probab.* **17**, 3, 833–839.

[23] SMORODINSKY, M. (1971). *Ergodic theory, entropy*. Lecture Notes in Mathematics, Vol. 214. Springer-Verlag, Berlin.

[24] TOTOKI, H. (1970). On a class of special flows. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 15*, 157–167.

[25] VAN HANDEL, R. (2009a). The stability of conditional Markov processes and Markov chains in random environments. *Ann. Probab.* **37**, 5, 1876–1925.

[26] VAN HANDEL, R. (2009b). Uniform time average consistency of Monte Carlo particle filters. *Stochastic Process. Appl.* **119**, 11, 3835–3861.

[27] VAN HANDEL, R. (2009/10). When do nonlinear filters achieve maximal accuracy? *SIAM J. Control Optim.* **48**, 5, 3151–3168.

[28] VOLKONSKII, V. A. AND ROZANOV, Y. A. (1959). Some limit theorems for random functions. I. *Theor. Probability Appl. 4*, 178–197.

[29] VON WEIZSÄCKER, H. (1983). Exchanging the order of taking suprema and countable intersections of σ-algebras. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **19**, 1, 91–100.

[30] WEISS, B. (1972). The isomorphism problem in ergodic theory. *Bull. Amer. Math. Soc. 78*, 668–684.

[31] WILLIAMS, D. (1991). *Probability with martingales*. Cambridge University Press.

[32] YANO, K. AND YOR, M. (2011). Around Tsirelson's equation, or: the evolution process may not explain everything. preprint, arxiv:0906.3442.

[33] YOR, M. (1977). Sur les théories du filtrage et de la prédiction. In *Séminaire de Probabilités, XI (Univ. Strasbourg, Strasbourg, 1975/1976)*. Springer, Berlin, 257–297. Lecture Notes in Math., Vol. 581.

SHERRERD HALL 227,
PRINCETON UNIVERSITY,
PRINCETON, NJ 08544, USA.
E-MAIL: rvan@princeton.edu