The Annals of Applied Probability
2010, Vol. 20, No. 6, 2318–2345
DOI: 10.1214/10-AAP688

A COMPLETE SOLUTION TO BLACKWELL'S UNIQUE ERGODICITY PROBLEM FOR HIDDEN MARKOV CHAINS

BY PAVEL CHIGANSKY[1] AND RAMON VAN HANDEL

Hebrew University and Princeton University

We develop necessary and sufficient conditions for uniqueness of the invariant measure of the filtering process associated to an ergodic hidden Markov model in a finite or countable state space. These results provide a complete solution to a problem posed by Blackwell (1957), and subsume earlier partial results due to Kaijser, Kochman and Reeds. The proofs of our main results are based on the stability theory of nonlinear filters.

1. Introduction. The interest in the stationary behavior of hidden Markov models dates back at least to a 1957 paper by Blackwell [2], who was motivated by the following problem from information theory. Suppose that $(X_n)_{n\geq 0}$ is a stationary Markov chain which takes values in a finite set. The entropy rate of such a chain admits a simple expression in terms of its transition probabilities and stationary distribution. The purpose of the paper by Blackwell was to obtain a similar expression for the entropy rate of the stochastic process $Y_n = h(X_n)$, where h is a noninvertible function. The latter expression does not involve directly the stationary distribution of the process $(X_n)_{n\geq 0}$, but rather a particular stationary distribution of the associated filtering process $(\pi_n)_{n\geq 0}$, which is a measure-valued Markov process defined as $\pi_n := \mathbf{P}(X_n \in \cdot | Y_1, \dots, Y_n)$.

The result of Blackwell raises a natural question: does the filtering process possess a unique stationary measure or, in other words, is the filtering process *uniquely ergodic*? Blackwell conjectured that the filter is uniquely ergodic, provided that the underlying Markov chain $(X_n)_{n\geq 0}$ is irreducible. However, as is pointed out by Kaijser [8], one of Blackwell's own counterexamples demonstrates that this conjecture is incorrect. The problem of finding a complete characterization of the unique ergodicity of the filtering process has hitherto remained open. The present paper provides one solution to this problem (in a more general setting).

1.1. *The contributions of Kaijser, Kochman and Reeds.* To our knowledge, the only direct contributions to the problem studied in this paper are contained in Blackwell's 1957 paper [2], in a 1975 paper by Kaijser [8] and in two recent papers by Kochman and Reeds [10] and by Kaijser [9], which we presently review.

Received October 2009.
[1]Supported by ISF Grant 314/09.
AMS 2000 subject classifications. Primary 93E11; secondary 37A50, 60J05, 60J10, 93E15.
Key words and phrases. Hidden Markov models, filtering, unique ergodicity, asymptotic stability.

In the 1975 paper [8], Kaijser observes that the filtering process can be expressed as the ratio of two quantities which are defined in terms of products of random matrices. Therefore, the unique ergodicity problem can be studied by means of the Furstenberg–Kesten theory of random matrix products. Such an analysis leads Kaijser to introduce a certain subrectangularity condition on the matrices that define the filter [Condition (K) in Section 6]. This rather strong condition is sufficient, but not necessary for unique ergodicity. It should be noted that Blackwell's original paper [2] already gives a sufficient condition for unique ergodicity, which is, however, even stronger than Kaijser's subrectangularity condition.

In their 2006 paper [10], Kochman and Reeds introduce a weaker sufficient condition for unique ergodicity of the filter, which requires that the closure of a certain cone of matrices contains an element of rank one [Condition (KR) in Section 2.3]. Kochman and Reeds demonstrate by means of an explicit computation that Kaijser's condition implies the rank one condition, but a counterexample shows that the latter condition is strictly weaker. Besides providing a generalization of Kaijser's result, Kochman and Reeds employ a different method of proof that is based on a general result in the ergodic theory of Markov chains in topological state spaces (which is applied to the filtering process).

Finally, in a recent paper [9], Kaijser presents an extension of the result of Kochman and Reeds to hidden Markov models where the underlying Markov chain $(X_n)_{n\geq 0}$ takes values in a countable state space. (It should be noted that Kochman and Reeds, as well as Kaijser, admit a more general observation structure than in Blackwell's original problem.) The extension is far from straightforward, as the ergodic theory employed by Kochman and Reeds is restricted to Markov chains in locally compact state spaces, while the space of probability measures on a countable set is certainly not locally compact. A large part of this lengthy paper is taken up with the development of a rather specialized ergodic theorem for Markov chains in Polish spaces, from which a condition similar in spirit to Kochman and Reeds' rank one condition [Condition (B1) in Section 6] can be derived.

1.2. *The approach of Kunita and filter stability.* Independently from Blackwell's unique ergodicity problem, a general study of the ergodic theory of nonlinear filtering processes was initiated in the seminal 1971 paper of Kunita [11]. Kunita studies a somewhat different problem, in continuous time and with white noise type observations, but which otherwise bears strong similarities to the problem studied by Blackwell. In contrast to the approaches developed by Kaijser, Kochman and Reeds, who study the *equations* that define the filter using general methods (products of random matrices and ergodic theory of Markov chains), Kunita studies the nonlinear filter directly through its characterization as a conditional expectation (an approach we called *intrinsic* in [5]). The techniques developed by Kunita are in fact extremely general and can be applied also to Blackwell's problem, though this approach has not previously been systematically exploited.

Kunita characterizes the invariant measures of the filtering process by means of the convex ordering. When the signal $(X_n)_{n\geq 0}$ is uniquely ergodic, all invariant measures of the filter are sandwiched between two distinguished invariant measures which are minimal and maximal with respect to the convex order, respectively (see Remark 3.2 below for a more precise statement). The filter is uniquely ergodic precisely when the minimal and maximal invariant measures coincide. The main result of Kunita's paper claims that this is always the case, when the signal is ergodic in a certain sense. Unfortunately, the proof of this result contains a serious gap [1]; indeed, the correctness of the proof is already contradicted by the counterexample given in Kaijser [8] (see [1, 4] for extensive discussion).

The gap in Kunita's main result is now largely resolved [14], but under an additional *nondegeneracy* assumption on the observation structure [Condition (N) of Section 6 in the present setting]. This assumption holds, for example, if $Y_n = h(X_n) + \varepsilon\xi_n$ where ξ_n is an independent Gaussian random variable and $\varepsilon > 0$ is an arbitrarily small noise strength, but breaks down in the noiseless case $\varepsilon = 0$. The nondegeneracy assumption evidently captures the phenomenon that observation noise has a stabilizing effect on the filter, as is the case in a large number of interesting applications. Unfortunately, it is the degenerate case that is chiefly of interest in Blackwell's problem, and unique ergodicity turns out to be more delicate in this setting as is demonstrated by various counterexamples [1, 8, 10].

In recent years, there has been considerable interest in the somewhat different problem of *filter stability* (see the survey [5]). Roughly speaking, the filtering process is called stable if π_n becomes independent of its initial condition π_0 as $n \to \infty$ in a certain pathwise sense (e.g., as in Theorem 3.1). However, it is now well established that when the signal $(X_n)_{n\geq 0}$ is ergodic, filter stability and unique ergodicity of the filter are essentially equivalent properties [4, 6, 12]. In the present setting, this has two important consequences. First, filter stability can be used as a tool to study unique ergodicity of the filter, a fact that is heavily exploited in this paper. Second, previous work on the filter stability problem provides a set of sufficient conditions for Blackwell's unique ergodicity problem which are distinct from those proposed by Kaijser, Kochman and Reeds.

1.3. *Contributions of this paper.* The present paper was inspired by the observation that the conditions of Kochman and Reeds [10] and Kaijser [9] are reminiscent of the filter stability property, albeit along a single sample path. It is therefore a natural step to make the connection with filter stability theory and Kunita's ergodic theory. Our results demonstrate that this approach is both natural and fruitful.

Our first main result, Theorem 2.6, establishes that a certain Condition (C) is necessary and sufficient for unique ergodicity of the filter in the case where X_n and Y_n both take values in an (at most) countable state space. It is easily shown, as we do in Section 6, that the sufficient conditions given in Kaijser's recent paper [9] imply Condition (C). It should be noted that the proof of Theorem 2.6 is surprisingly easy and natural—that is, provided the connection between filter stability and Kunita's ergodic theory (given in Theorem 3.1) is taken for granted.

Our second main result, Theorem 2.7, shows that the rank one Condition (KR) of Kochman and Reeds is necessary and sufficient for unique ergodicity of the filter in the case where X_n takes values in a finite state space. Sufficiency was already proved by Kochman and Reeds [10], though we give here an entirely different proof of this fact by showing that Condition (KR) implies Condition (C). The necessity of Condition (KR) is new, and answers in the affirmative the question posed on the last page of [10]. Thus the necessity and sufficiency of Condition (KR) provides a complete solution to the original problem posed by Blackwell [2].

Our main results subsume all of the sufficient conditions introduced in the papers of Kaijser, Kochman and Reeds. In addition, we discuss in Section 6 some sufficient conditions of a different nature which are inherited from results in the filter stability literature. Though these conditions are not necessary, they may be substantially easier to check in practice than Condition (C) or (KR). Moreover, such conditions remain of independent interest, as we were not able to verify by an explicit computation that they imply Condition (C) or (KR) (of course, this implication follows indirectly from the necessity of these conditions).

1.4. *Organization of the paper.* The remainder of this paper is organized as follows. In Section 2 we introduce the basic hidden Markov model, and we fix once and for all the notation and standing assumptions that are presumed to be in force throughout the paper. We also state our main results, Theorems 2.6 and 2.7. In Section 3, we introduce the connection between filter stability and unique ergodicity of the filter. The main result of this section, Theorem 3.1, adapts the necessary theory to the setting of this paper and forms the foundation for the proofs of our main results. Section 4 is devoted to the proof of Theorem 2.6, while Section 5 is devoted to the proof of Theorem 2.7. Section 6 develops various sufficient conditions for unique ergodicity within the setting of this paper. Finally, the Appendix is devoted to the proofs of various results that were omitted from the body of the paper.

2. Preliminaries and main results.

2.1. *The canonical setup and standing assumptions.* Throughout this paper, we operate in the following setup. We consider the stochastic process $(X_n, Y_n)_{n\in\mathbb{Z}}$, where X_n takes values in the state space E, and Y_n takes values in the state space F. We will always presume that the following assumptions are in force:

- E is either finite ($E = \{1, \dots, p\}$) or countable ($E = \mathbb{N}$).
- F is a Polish space [endowed with its Borel σ-field $\mathcal{B}(F)$].

We realize the stochastic process $(X_n, Y_n)_{n\in\mathbb{Z}}$ on the canonical path space $\Omega = \Omega^X \times \Omega^Y$ with $\Omega^X = E^{\mathbb{Z}}$ and $\Omega^Y = F^{\mathbb{Z}}$, such that $X_n(x, y) = x(n)$ and $Y_n(x, y) = y(n)$. Denote by $\mathcal{F}$ the Borel σ-field on Ω, and introduce 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 $m, n \in \mathbb{Z}$, $m \le n$. The σ-fields $\mathcal{F}^X_{-\infty,n}$, $\mathcal{F}^X_{m,\infty}$, etc., are defined in the usual fashion (e.g., $\mathcal{F}^X_{-\infty,n} = \bigvee_{m\le n} \mathcal{F}^X_{m,n}$). For future reference, we define

$$\mathcal{G}_{m,n} = \mathcal{F}^X_{-\infty,m} \vee \mathcal{F}^Y_{-\infty,n}, \qquad \mathcal{G}_{-\infty,n} = \bigcap_{m\le n} \mathcal{G}_{m,n}$$

(note that $\mathcal{F}^Y_{-\infty,n} \subset \mathcal{G}_{-\infty,n}$, a fact that will be used frequently in the following). Finally, the shift $\Theta : \Omega \to \Omega$ is defined as $\Theta(x, y)(m) = (x(m+1), y(m+1))$.

We now define a probability measure on $(\Omega, \mathcal{F})$ under which $(X_k, Y_k)_{k\in\mathbb{Z}}$ is a *hidden Markov model*. Our model is specified by the following ingredients:

(1) A σ-finite reference measure φ on F.
(2) A nonnegative matrix function $M : F \to \mathbb{R}_+^{E\times E}$ such that

$$\sup_{i\in E} \sum_{j\in E} M_{ij}(y) < \infty \qquad \text{for } \varphi\text{-a.e. } y \in F,$$

and such that the matrix

$$P = (P_{ij})_{i,j\in E}, \qquad P_{ij} := \int M_{ij}(y)\varphi(dy)$$

defines the transition matrix of an irreducible and positive recurrent (but not necessarily aperiodic) Markov chain in the state space E.

As P is irreducible and positive recurrent, there is a unique probability measure λ on E that is invariant $\lambda P = \lambda$ (as is usual, we identify measures and functions on a countable space with row and column vectors, respectively). A standard extension argument allows us to construct a probability measure $\mathbf{P}$ on $(\Omega, \mathcal{F})$ under which $(X_k, Y_k)_{k\in\mathbb{Z}}$ is a stationary Markov chain with transition probabilities

$$\mathbf{P}(X_k = j, Y_k \in A | X_{k-1} = i, Y_{k-1} = y) = \int_A M_{ij}(y')\varphi(dy')$$

for $i, j \in E$, $y \in F$, $A \in \mathcal{B}(F)$. It should be noted that under $\mathbf{P}$, the process $(X_k)_{k\in\mathbb{Z}}$ is a stationary Markov chain with transition matrix P, and $(Y_k)_{k\in\mathbb{Z}}$ are conditionally independent given $(X_k)_{k\in\mathbb{Z}}$. This is precisely the defining property of a hidden Markov model. The process $(X_k)_{k\in\mathbb{Z}}$ represents an unobserved or "hidden" signal process, while $(Y_k)_{k\in\mathbb{Z}}$ is the observation process. The canonical probability space $(\Omega, \mathcal{F}, \mathbf{P})$ thus constructed will remain fixed throughout the paper.

REMARK 2.1. A hidden Markov model is often assumed to satisfy the additional assumption that Y_k is a (noisy) function of X_k only. In this case, one can factor $M_{ij}(y) = P_{ij}R_j(y)$, where $R_j(y)$ is the density of $\mathbf{P}(Y_k \in \cdot | X_k = j)$ with respect to φ. In the present setting, the conditional law of Y_k can depend on both X_k and X_{k-1}. The generalization afforded by this model is minor, but allows us to include the partitioned transition matrices of [9, 10] as a special case.

REMARK 2.2. The boundedness condition $\sup_{i\in E}\sum_{j\in E} M_{ij}(y) < \infty$ a.e. is automatically satisfied in the following cases:

- When E is a finite set, the condition holds trivially.
- When E is countable and F is at most countable, the condition always holds. Indeed, note that in this case $\sum_{y\in F}\sum_{j\in E} M_{ij}(y)\varphi(\{y\}) = \sum_{j\in E} P_{ij} = 1$, so that $\sup_{i\in E}\sum_{j\in E} M_{ij}(y) \le \varphi(\{y\})^{-1} < \infty$ for φ-a.e. $y \in F$.

The significance of this assumption is that it ensures the Feller property of the filter.

For any Polish space S we denote by $\mathcal{B}(S)$ the Borel σ-field of S, by $\mathcal{P}(S)$ the space of probability measures on S, and by $\mathcal{C}_b(S)$ the space of bounded continuous functions on S. We will always endow $\mathcal{P}(S)$ with the topology of weak convergence of probability measures [recall that $\mathcal{P}(S)$ is then itself Polish], and we write $\mu_n \Rightarrow \mu$ if the sequence $(\mu_n) \subset \mathcal{P}(S)$ converges weakly to $\mu \in \mathcal{P}(S)$. The total variation distance between probability measures $\mu, \nu \in \mathcal{P}(S)$ is defined as

$$\|\mu - \nu\| = \sup_{\|f\|_\infty \le 1} \left| \int f\,d\mu - \int f\,d\nu \right|.$$

Finally, let us recall that as E is at most countable and P is irreducible, the invariant measure λ must charge every point of E. Therefore $\mu \ll \lambda$ for every $\mu \in \mathcal{P}(E)$, and we can define the probability measures $\mathbf{P}^\mu$ on $(\Omega, \mathcal{F})$ as

$$\frac{d\mathbf{P}^\mu}{d\mathbf{P}} = \frac{d\mu}{d\lambda}(X_0), \qquad \mu \in \mathcal{P}(E).$$

The restriction of $\mathbf{P}^\mu$ to $\mathcal{F}^X_{0,\infty} \vee \mathcal{F}^Y_{1,\infty}$ defines a hidden Markov model with the same transition probabilities as under $\mathbf{P}$, but with the initial distribution $X_0 \sim \mu$. If the initial distribution is a point mass on $x \in E$, we will write $\mathbf{P}^x$ instead of $\mathbf{P}^{\delta_x}$.

2.2. *Nonlinear filtering*. The purpose of nonlinear filtering is to compute the conditional distribution of the hidden signal given the available observations. In this paper we will encounter several variants of the nonlinear filter, defined as follows:

$$\pi_n = \mathbf{P}(X_n \in \cdot|\mathcal{F}^Y_{1,n}), \qquad \pi^\mu_n = \mathbf{P}^\mu(X_n \in \cdot|\mathcal{F}^Y_{1,n}), \qquad \pi^x_n = \mathbf{P}^x(X_n \in \cdot|\mathcal{F}^Y_{1,n})$$

for $n \in \mathbb{Z}_+$, $\mu \in \mathcal{P}(E)$, $x \in E$ (here $\pi_0 = \lambda$, $\pi^\mu_0 = \mu$ and $\pi^x_0 = \delta_x$) and

$$\pi^{\min}_n = \mathbf{P}(X_n \in \cdot|\mathcal{F}^Y_{-\infty,n}), \qquad \pi^{\max}_n = \mathbf{P}(X_n \in \cdot|\mathcal{G}_{-\infty,n})$$

for $n \in \mathbb{Z}$. Though the relevance of $\pi^{\min}_n$ and $\pi^{\max}_n$ may not be entirely evident at present, their role will be clarified in Section 3 below.

The following elementary results are essentially known; short proofs are provided in Appendix A.1 for the reader's convenience.

LEMMA 2.3 (Filtering recursion). *For any $m, n \in \mathbb{Z}$, $n > m$ we have* **P***-a.s.*

$$\pi_n^{\min} = \frac{\pi_m^{\min} M(Y_{m+1}) \cdots M(Y_n)}{\pi_m^{\min} M(Y_{m+1}) \cdots M(Y_n) 1}, \qquad \pi_n^{\max} = \frac{\pi_m^{\max} M(Y_{m+1}) \cdots M(Y_n)}{\pi_m^{\max} M(Y_{m+1}) \cdots M(Y_n) 1}.$$

Similarly, for any $n > m \geq 0$, we have $\mathbf{P}^\mu$-a.s.

$$\pi_n^\mu = \frac{\pi_m^\mu M(Y_{m+1}) \cdots M(Y_n)}{\pi_m^\mu M(Y_{m+1}) \cdots M(Y_n) 1}.$$

The recursion for π_n, π_n^x is obtained by choosing $\mu = \lambda$ or $\mu = \delta_x$, respectively.

It should be noted that π_n^μ is defined only up to a $\mathbf{P}^\mu$-null set. Indeed,

$$\mathbf{P}^\mu\big((Y_1, \ldots, Y_n) \in A\big) = \int_A \mu M(y_1) \cdots M(y_n) 1 \varphi(dy_1) \cdots \varphi(dy_n),$$

that is, $\mu M(y_1) \cdots M(y_n) 1$ is the density of the law of $(Y_1, \ldots, Y_n)$ under $\mathbf{P}^\mu$. Similarly, it is easily seen that $\pi_m^\mu M(y_{m+1}) \cdots M(y_n) 1$ is the density of the law of $(Y_{m+1}, \ldots, Y_n)$ under the conditional measure $\mathbf{P}^\mu(\cdot | Y_0, \ldots, Y_m)$. Therefore, the denominator in the filtering recursion can only vanish on a $\mathbf{P}^\mu$-null set. Similar considerations hold for $\pi_n^{\min}, \pi_n^{\max}$, which are defined up to a **P**-null set.

LEMMA 2.4 (Markov property). $(\pi_n^{\min})_{n\in\mathbb{Z}}$, $(\pi_n^{\max})_{n\in\mathbb{Z}}$ *are stationary $\mathcal{P}(E)$-valued Markov chains under* **P**, *whose transition kernel* Π *is defined by*

$$\int f(\nu) \Pi(\mu, d\nu) = \int f\left(\frac{\mu M(y)}{\mu M(y) 1}\right) \mu M(y) 1 \varphi(dy), \qquad f \in \mathcal{C}_b(\mathcal{P}(E)).$$

Similarly, $(\pi_n^\mu)_{n\in\mathbb{Z}_+}$ is a Markov chain under $\mathbf{P}^\mu$ with transition kernel Π.

REMARK 2.5. As $(\pi_n^{\min})_{n\in\mathbb{Z}}$, $(\pi_n^{\max})_{n\in\mathbb{Z}}$ are stationary Markov chains with transition kernel Π, the laws of $\pi_0^{\max}$ and $\pi_0^{\min}$ must be invariant for Π. Therefore, the filter always possesses at least one invariant measure.

2.3. *Main results*. This paper aims to resolve the following question: when does the filter possess a *unique* invariant measure, that is, when does the equation $\mathsf{M}\Pi = \mathsf{M}$ possess a unique solution $\mathsf{M} \in \mathcal{P}(\mathcal{P}(E))$?

We begin by establishing a general sufficient condition for unique ergodicity, which is also necessary when the observation state space F is at most countable.

CONDITION (C). For every $\varepsilon > 0$, there exist an integer $N \in \mathbb{N}$ and subsets $\mathcal{S} \subset \mathcal{P}(E)$ and $\mathcal{O} \subset F^N$ such that the following hold:

(1) $\mathbf{P}(\pi_0^{\min} \in \mathcal{S}$ and $\pi_0^{\max} \in \mathcal{S}) > 0$ and $\varphi^{\otimes N}(\mathcal{O}) > 0$.
(2) $\mu M(y_1) \cdots M(y_N) 1 > 0$ for all $\mu \in \mathcal{S}$ and $(y_1, \ldots, y_N) \in \mathcal{O}$.

(3) For all $\mu, \nu \in \mathcal{S}$ and $(y_1, \ldots, y_N) \in \mathcal{O}$

$$\left\| \frac{\mu M(y_1) \cdots M(y_N)}{\mu M(y_1) \cdots M(y_N) 1} - \frac{\nu M(y_1) \cdots M(y_N)}{\nu M(y_1) \cdots M(y_N) 1} \right\| < \varepsilon.$$

THEOREM 2.6. *Suppose that Condition (C) holds. Then the filter admits a unique invariant measure* M, *and we have* $n^{-1} \sum_{k=1}^{n} \mathsf{M}_0 \Pi^k \Rightarrow \mathsf{M}$ *as* $n \to \infty$ *for any* $\mathsf{M}_0 \in \mathcal{P}(\mathcal{P}(E))$. *If, in addition, the signal transition matrix* P *is aperiodic, then we have* $\mathsf{M}_0 \Pi^n \Rightarrow \mathsf{M}$ *as* $n \to \infty$ *for any* $\mathsf{M}_0 \in \mathcal{P}(\mathcal{P}(E))$.

Conversely, suppose that the observation state space F *is a finite or countable set, and that the filter is uniquely ergodic. Then Condition (C) holds.*

The proof of this result is given in Section 4.

Next, we consider the following condition, due to Kochman and Reeds [10], for the case where the signal state space E is a finite set.

CONDITION (KR). Let E be a finite set, and define the cone of matrices

$$\mathcal{K} = \{c M(y_1) \cdots M(y_n) : n \in \mathbb{N}, y_1, \ldots, y_n \in F, c \in \mathbb{R}_+\}.$$

Then the closure $\operatorname{cl} \mathcal{K}$ contains a matrix of rank 1.

Kochman and Reeds prove that this condition is sufficient for uniqueness of the invariant measure of the filter (in [10], both E and F are presumed to be finite). The following result shows that Condition (KR) is in fact *equivalent* to unique ergodicity of the filter, as well as to Condition (C) above, when the signal state space is a finite set. This provides a complete solution to a problem posed by Blackwell [2], and answers in the affirmative the question posed at the end of [10].

THEOREM 2.7. *Suppose* E *is a finite set and that one of the following hold*:

- F *is a finite or countable set, and* φ *is the counting measure; or*
- $F = \mathbb{R}^d$, φ *is the Lebesgue measure, and* $y \mapsto M(y)$ *is continuous.*

Then the following are equivalent:

(1) *The filter admits a unique invariant measure* M.
(2) *Condition (KR) holds.*
(3) *Condition (C) holds.*

When any of these conditions hold, we have $n^{-1} \sum_{k=1}^{n} \mathsf{M}_0 \Pi^k \Rightarrow \mathsf{M}$ *as* $n \to \infty$ *for any* $\mathsf{M}_0 \in \mathcal{P}(\mathcal{P}(E))$. *If, in addition, the signal transition matrix* P *is aperiodic, then we have* $\mathsf{M}_0 \Pi^n \Rightarrow \mathsf{M}$ *as* $n \to \infty$ *for any* $\mathsf{M}_0 \in \mathcal{P}(\mathcal{P}(E))$.

The proof will be given in Section 5.

Finally, various sufficient conditions for unique ergodicity of the filter were given by Kaijser [8, 9]. These conditions are easily shown to imply Condition (C), as is discussed in Section 6. We therefore reproduce Kaijser's results using a much simpler proof. Similarly, various conditions that have been introduced in the context of filter stability [5, 14, 15] are shown in Section 6 to imply unique ergodicity of the filter. None of the latter sufficient conditions is also necessary; however, when they apply, they are often easier to check than Condition (C) or (KR).

3. Ergodic theory and stability of nonlinear filters. The proofs of our main results are based on a general circle of ideas connecting the ergodic theory [11, 12] and asymptotic stability [5, 14] of nonlinear filters. Indeed, it is by now well established [4, 6] that unique ergodicity and stability of the filter are essentially equivalent properties. The purpose of this section is to introduce the relevant results in this direction that will be needed in what follows. Though the results in this section are adapted to the setting of this paper, their proofs largely follow along the lines of [6, 11, 12, 14]. We have therefore relegated the proofs to the Appendix.

The following characterization will be of central importance.

THEOREM 3.1. *Consider the following conditions*:

(1) *The filter possesses a unique invariant measure* $\mathsf{M} \in \mathcal{P}(\mathcal{P}(E))$.
(2) $\pi_0^{\max} = \pi_0^{\min}$ $\mathbf{P}$-*a.s.*
(3) $\|\pi_n^\mu - \pi_n^\nu\| \to 0$ *as* $n \to \infty$ $\mathbf{P}^\mu$-*a.s. whenever* $\mu \ll \nu$.
(4) $n^{-1} \sum_{k=1}^n \mathsf{M}_0 \Pi^k \Rightarrow \mathsf{M}$ *as* $n \to \infty$ *for any* $\mathsf{M}_0 \in \mathcal{P}(\mathcal{P}(E))$.
(5) $\mathsf{M}_0 \Pi^n \Rightarrow \mathsf{M}$ *as* $n \to \infty$ *for any* $\mathsf{M}_0 \in \mathcal{P}(\mathcal{P}(E))$.

Conditions 1–4 *are equivalent*. *If*, *in addition*, *the signal transition matrix* P *is aperiodic*, *then conditions* 1–5 *are equivalent*.

The proof is given in Appendix A.2.

REMARK 3.2. Condition 1 is the desired unique ergodicity property of the filter. Condition 3 is the filter stability property. Conditions 4 and 5 characterize the convergence of the law of the filter to the invariant measure.

Condition 2 in Theorem 3.1 stems from an ingenious device introduced by Kunita in the seminal paper [11] and used in the proof of Theorem 3.1. By Lemma 2.4, $(\pi_n^{\min})_{n\in\mathbb{Z}}$ and $(\pi_n^{\max})_{n\in\mathbb{Z}}$ are stationary Markov processes. Therefore, the laws $\mathsf{M}^{\max}, \mathsf{M}^{\min} \in \mathcal{P}(\mathcal{P}(E))$ of the $\mathcal{P}(E)$-valued random variables $\pi_0^{\max}, \pi_0^{\min}$ are invariant for the filter transition kernel Π. Kunita shows that *any* invariant measure M for Π is sandwiched between $\mathsf{M}^{\max}$ and $\mathsf{M}^{\min}$ in the sense that

$$\int f(\mu)\mathsf{M}^{\min}(d\mu) \le \int f(\mu)\mathsf{M}(d\mu) \le \int f(\mu)\mathsf{M}^{\max}(d\mu)$$

for every convex function $f \in \mathcal{C}_b(\mathcal{P}(E))$. In other words, within the family of Π-invariant measures, $\mathsf{M}^{\min}$ is minimal and $\mathsf{M}^{\max}$ is maximal with respect to the convex ordering. The identity $\pi_0^{\max} = \pi_0^{\min}$ ensures that the maximal and minimal invariant measures are identical, so that there can be only one invariant measure.

EXAMPLE 3.3. Some intuition may be obtained from the following simple example [10], which is a typical case where the filter fails to be uniquely ergodic. Let $E = F = \{0, 1\}$ (endowed with the counting measure), and let

$$M(0) = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}, \qquad M(1) = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Note that $P = M(0) + M(1)$ is irreducible and aperiodic with invariant measure $\lambda_0 = \lambda_1 = 1/2$, and $Y_k = I_{\{X_{k-1}=X_k\}}$ for all $k \geq 1$.

As $(Y_k)_{k\geq 1}$ reveals exactly when the transitions of $(X_k)_{k\geq 0}$ occur, we evidently have $X_k \in \sigma\{X_m, Y_{m+1}, Y_{m+2}, \dots, Y_n\}$ for every $m \leq k \leq n$. It follows that $\mathcal{G}_{m,n} = \mathcal{F}^X_{-\infty,n}$ for every $m \leq n$, so that in particular $\mathcal{G}_{-\infty,n} = \mathcal{F}^X_{-\infty,n}$. Therefore

$$\pi_n^{\max} = \delta_{X_n}, \qquad \mathsf{M}^{\max} = \tfrac{1}{2}\{\delta_{\delta_0} + \delta_{\delta_1}\}.$$

On the other hand, it follows immediately from the filtering recursion that $\pi_n = \lambda$ for all n. It is therefore not difficult to show that $\pi_n^{\min} = \lambda$ also, so that

$$\pi_n^{\min} = \lambda, \qquad \mathsf{M}^{\min} = \delta_\lambda = \delta_{(\delta_0+\delta_1)/2}.$$

With a little more work, one can show that any invariant measure M is of the form

$$\mathsf{M} = \int_0^{1/2} \frac{\delta_{\epsilon\delta_0+(1-\epsilon)\delta_1} + \delta_{(1-\epsilon)\delta_0+\epsilon\delta_1}}{2} m(d\epsilon)$$

for some probability measure m on $[0, 1/2]$. It is easily seen that any such M does indeed lie between $\mathsf{M}^{\max}$ and $\mathsf{M}^{\min}$ in the convex ordering.

Besides the characterization of unique ergodicity in Theorem 3.1, we will require the following convergence property which holds regardless of unique ergodicity.

LEMMA 3.4. $\lim_{n\to\infty} \|\pi_n^{\max} - \pi_n^{\min}\|$ *exists* **P**-*a.s.*

The proof of this result is also given in Appendix A.2. Its relevance is due to the following observation. In order to prove $\pi_0^{\max} = \pi_0^{\min}$ (hence unique ergodicity by Theorem 3.1), it suffices to show that $\lim_{n\to\infty} \|\pi_n^{\max} - \pi_n^{\min}\| = 0$ **P**-a.s., as $(\pi_n^{\min})_{n\in\mathbb{Z}}$ and $(\pi_n^{\max})_{n\in\mathbb{Z}}$ are stationary processes. But by virtue of Lemma 3.4, it then suffices to show only that $\|\pi_n^{\max} - \pi_n^{\min}\|$ converges to zero along a sequence of stopping times. The main idea behind the proof of Theorem 2.6 is that Condition (C) allows us to construct explicitly such a sequence stopping times.

4. Proof of Theorem 2.6.

4.1. *Sufficiency of Condition* (C). We will need the following lemma.

LEMMA 4.1. *The sequence $(X_k, Y_k)_{k\in\mathbb{Z}}$ is ergodic under* $\mathbf{P}$.

PROOF. As the signal transition matrix P is presumed to be irreducible and positive recurrent, it is easily established that the pair $(X_k, Y_k)_{k\in\mathbb{Z}}$ is a Markov process that possesses a unique invariant measure. This measure is therefore trivially an extreme point of the set of invariant measures, hence ergodic. $\square$

We will also use the following simple result.

LEMMA 4.2. *Let the set $\mathcal{S} \subset \mathcal{P}(E)$ and $\mathcal{O} \subset F^N$ be as in Condition* (C). *Then we have* $\mathbf{P}(\pi_0^{\min} \in \mathcal{S}, \pi_0^{\max} \in \mathcal{S}, (Y_1, \dots, Y_N) \in \mathcal{O}) > 0$.

PROOF. Let us write for simplicity $Y = (Y_1, \dots, Y_N)$. As $\pi_0^{\min}$ and $\pi_0^{\max}$ are $\mathcal{G}_{-\infty,0}$-measurable by construction, we have

$$\begin{aligned}
&\mathbf{P}(\pi_0^{\min} \in \mathcal{S}, \pi_0^{\max} \in \mathcal{S}, Y \in \mathcal{O}) \\
&\quad = \mathbf{E}\big(I_{\mathcal{S}}(\pi_0^{\min}) I_{\mathcal{S}}(\pi_0^{\max}) \mathbf{P}(Y \in \mathcal{O} | \mathcal{G}_{-\infty,0})\big) \\
&\quad = \mathbf{E}\biggl(I_{\mathcal{S}}(\pi_0^{\min}) I_{\mathcal{S}}(\pi_0^{\max}) \int_{\mathcal{O}} \pi_0^{\max} M(y_1) \cdots M(y_N) 1 \varphi(dy_1) \cdots \varphi(dy_N)\biggr).
\end{aligned}$$

It is now easily seen that Condition (C) implies the result. $\square$

We now proceed with the proof of the sufficiency part of Theorem 2.6. Suppose that Condition (C) holds, and fix an arbitrary decreasing sequence $\varepsilon_k \searrow 0$. Then for every k we can find $N_k \in \mathbb{N}$, $\mathcal{S}_k \subset \mathcal{P}(E)$, and $\mathcal{O}_k \subset F^{N_k}$ such that the properties 1–3 of Condition (C) are satisfied. Define the events

$$A_{n,k} = \{\pi_n^{\min} \in \mathcal{S}_k, \pi_n^{\max} \in \mathcal{S}_k, (Y_{n+1}, \dots, Y_{n+N_k}) \in \mathcal{O}_k\}.$$

Then, by the stationarity of $(X_n, Y_n, \pi_n^{\min}, \pi_n^{\max})_{n\in\mathbb{Z}}$ (Lemma A.1), we have

$$\lim_{T\to\infty} \frac{1}{T} \sum_{n=1}^{T} I_{A_{n,k}} = \lim_{T\to\infty} \frac{1}{T} \sum_{n=1}^{T} \{I_{A_{0,k}} \circ \Theta^n\} = \mathbf{P}(A_{0,k}) > 0 \qquad \mathbf{P}\text{-a.s.},$$

where we have used Birkhoff's ergodic theorem together with Lemmas 4.1 and 4.2. Thus, for any k, the event $A_{n,k}$ occurs at a positive rate, so that certainly

$$\mathbf{P}\Biggl(\bigcap_{k=1}^{\infty} \limsup_{n\to\infty} A_{n,k}\Biggr) = 1.$$

Now define the stopping times $\tau_0 = 0$ and

$$\tau_k = \min\{n > \tau_{k-1} : \pi^{\min}_{n-N_k} \in \mathcal{S}_k, \pi^{\max}_{n-N_k} \in \mathcal{S}_k, (Y_{n-N_k+1}, \dots, Y_n) \in \mathcal{O}_k\}$$

for any $k \geq 1$. It follows directly that

$$\mathbf{P}(\tau_k < \infty \text{ for all } k) \geq \mathbf{P}\left(\bigcap_{k=1}^{\infty} \limsup_{n\to\infty} A_{n,k}\right) = 1.$$

Moreover, by Condition (C) and Lemma 2.3, we have

$$\begin{aligned}\|\pi^{\max}_{\tau_k} - \pi^{\min}_{\tau_k}\| = \Bigg\| &\frac{\pi^{\max}_{\tau_k-N_k} M(Y_{\tau_k-N_k+1}) \cdots M(Y_{\tau_k})}{\pi^{\max}_{\tau_k-N_k} M(Y_{\tau_k-N_k+1}) \cdots M(Y_{\tau_k})1} \\ &- \frac{\pi^{\min}_{\tau_k-N_k} M(Y_{\tau_k-N_k+1}) \cdots M(Y_{\tau_k})}{\pi^{\min}_{\tau_k-N_k} M(Y_{\tau_k-N_k+1}) \cdots M(Y_{\tau_k})1}\Bigg\| \leq \varepsilon_k\end{aligned}$$

for all $k \geq 1$ $\mathbf{P}$-a.s. Therefore, Lemma 3.4 shows that $\|\pi^{\max}_n - \pi^{\min}_n\| \to 0$ as $n \to \infty$ $\mathbf{P}$-a.s. But using the stationarity of $(\pi^{\min}_n, \pi^{\max}_n)_{n\in\mathbb{Z}}$ (Lemma A.1) and the dominated convergence theorem, we find that

$$\mathbf{E}(\|\pi^{\max}_0 - \pi^{\min}_0\|) = \mathbf{E}(\|\pi^{\max}_n - \pi^{\min}_n\|) \xrightarrow{n\to\infty} 0,$$

so that evidently $\pi^{\max}_0 = \pi^{\min}_0$ $\mathbf{P}$-a.s. The sufficiency part of Theorem 2.6 now follows immediately from Theorem 3.1.

4.2. *Necessity of Condition* (C). Throughout this subsection, we assume that the observation state space F is finite or countable, and that the filter possesses a unique invariant measure. We aim to show that Condition (C) must hold.

Denote by M the law of $\pi^{\min}_0$. Thus M is invariant, hence the unique invariant measure of the filter. Fix an arbitrary state $i \in E$, and note that

$$\mathbf{E}((\pi^{\min}_0)_i) = \lambda_i > 0 \quad \text{implies} \quad \mathbf{P}\big((\pi^{\min}_0)_i \geq \lambda_i/2\big) > 0.$$

Therefore, writing $\mathcal{R} = \{\mu \in \mathcal{P}(E) : \mu_i \geq \lambda_i/2\}$, we have $\mathsf{M}(\mathcal{R}) > 0$. We can thus define the probability measure $\mathsf{M}_{\mathcal{R}}(\cdot) := \mathsf{M}(\cdot \cap \mathcal{R})/\mathsf{M}(\mathcal{R})$.

Now note that, by Theorem 3.1, we have $\|\pi^i_n - \pi^\mu_n\| \to 0$ $\mathbf{P}^i$-a.s. for all $\mu \in \mathcal{R}$. In particular, the set of points $((y_k)_{k\in\mathbb{N}}, \mu) \in F^{\mathbb{N}} \times \mathcal{P}(E)$ such that

$$\left\|\frac{\delta_i M(y_1) \cdots M(y_n)}{\delta_i M(y_1) \cdots M(y_n)1} - \frac{\mu M(y_1) \cdots M(y_n)}{\mu M(y_1) \cdots M(y_n)1}\right\| \xrightarrow{n\to\infty} 0$$

has $\mathbf{P}^i \otimes \mathsf{M}_{\mathcal{R}}$-full measure. It follows that for $\mathbf{P}^i$-a.e. path $(y_k)_{k\in\mathbb{N}}$, the above convergence holds for $\mathsf{M}_{\mathcal{R}}$-a.e. μ. Therefore, as F is at most countable [so the law of $(Y_1, \dots, Y_n)$ is atomic for all $n < \infty$], we can certainly find a single sequence $(\tilde{y}_k)_{k\in\mathbb{N}}$ with $\mathbf{P}^i(Y_1 = \tilde{y}_1, \dots, Y_n = \tilde{y}_n) > 0$ for all $n < \infty$ such that

$$\left\|\frac{\delta_i M(\tilde{y}_1) \cdots M(\tilde{y}_n)}{\delta_i M(\tilde{y}_1) \cdots M(\tilde{y}_n)1} - \frac{\mu M(\tilde{y}_1) \cdots M(\tilde{y}_n)}{\mu M(\tilde{y}_1) \cdots M(\tilde{y}_n)1}\right\| \xrightarrow{n\to\infty} 0 \qquad \text{for } \mathsf{M}_{\mathcal{R}}\text{-a.e. } \mu.$$

By Egorov's theorem, there is a subset $\mathcal{S} \subset \mathcal{R}$ with $\mathcal{M}_{\mathcal{R}}(\mathcal{S}) > 0$ such that

$$\sup_{\mu \in \mathcal{S}} \left\| \frac{\delta_i M(\tilde{y}_1) \cdots M(\tilde{y}_n)}{\delta_i M(\tilde{y}_1) \cdots M(\tilde{y}_n) 1} - \frac{\mu M(\tilde{y}_1) \cdots M(\tilde{y}_n)}{\mu M(\tilde{y}_1) \cdots M(\tilde{y}_n) 1} \right\| \xrightarrow{n \to \infty} 0.$$

We are now in the position to show that Condition (C) holds true. Given $\varepsilon > 0$, we first choose the integer $N \in \mathbb{N}$ large enough so that

$$\sup_{\mu \in \mathcal{S}} \left\| \frac{\delta_i M(\tilde{y}_1) \cdots M(\tilde{y}_N)}{\delta_i M(\tilde{y}_1) \cdots M(\tilde{y}_N) 1} - \frac{\mu M(\tilde{y}_1) \cdots M(\tilde{y}_N)}{\mu M(\tilde{y}_1) \cdots M(\tilde{y}_N) 1} \right\| \le \frac{\varepsilon}{2}.$$

We let $\mathcal{S}$ be as above and define the singleton $\mathcal{O} = \{(\tilde{y}_1, \dots, \tilde{y}_N)\}$. By Theorem 3.1, we have $\pi_0^{\max} = \pi_0^{\min}$ and therefore

$$\mathbf{P}(\pi_0^{\max} \in \mathcal{S} \text{ and } \pi_0^{\min} \in \mathcal{S}) = \mathbf{P}(\pi_0^{\min} \in \mathcal{S}) = \mathsf{M}(\mathcal{S}) \ge \mathsf{M}(\mathcal{R}) \mathsf{M}_{\mathcal{R}}(\mathcal{S}) > 0.$$

Next, we note that as $\mathbf{P}^i(Y_1 = \tilde{y}_1, \dots, Y_N = \tilde{y}_N) > 0$,

$$\mathbf{P}^i\big((Y_1, \dots, Y_N) \in \mathcal{O}\big) = \delta_i M(\tilde{y}_1) \cdots M(\tilde{y}_N) 1 \varphi(\{\tilde{y}_1\}) \cdots \varphi(\{\tilde{y}_N\}) > 0,$$

so $\varphi^{\otimes N}(\mathcal{O}) > 0$ and $\mu M(\tilde{y}_1) \cdots M(\tilde{y}_N) 1 > 0$ for all $\mu \in \mathcal{S}$ (this holds by the definition of $\mathcal{R}$ and as $\mathcal{S} \subset \mathcal{R}$). Finally, by the triangle inequality,

$$\sup_{\mu, \nu \in \mathcal{S}} \left\| \frac{\mu M(\tilde{y}_1) \cdots M(\tilde{y}_N)}{\mu M(\tilde{y}_1) \cdots M(\tilde{y}_N) 1} - \frac{\nu M(\tilde{y}_1) \cdots M(\tilde{y}_N)}{\nu M(\tilde{y}_1) \cdots M(\tilde{y}_N) 1} \right\| \le \varepsilon.$$

Thus Condition (C) is satisfied, and the proof is complete.

5. Proof of Theorem 2.7. The implication $3 \Rightarrow 1$ is already established by Theorem 2.6. It therefore suffices to prove the implications $1 \Rightarrow 2$ and $2 \Rightarrow 3$.

5.1. *Proof of* $1 \Rightarrow 2$. We will need the following lemma.

LEMMA 5.1. *Let $(\Xi, \mathcal{X}, (\mathcal{X}_n)_{n \in \mathbb{N}}, \mathbf{\Phi})$ be a filtered probability space, and let $\mathbf{Q}, \mathbf{Q}'$ be mutually singular probability measures on Ξ. Suppose that $\mathbf{Q}, \mathbf{Q}'$ are locally absolutely continuous with respect to $\mathbf{\Phi}$, that is, $\mathbf{Q}|_{\mathcal{X}_n} \ll \mathbf{\Phi}|_{\mathcal{X}_n}$ and $\mathbf{Q}'|_{\mathcal{X}_n} \ll \mathbf{\Phi}|_{\mathcal{X}_n}$ with densities q_n and q_n', respectively. Then $q_n / q_n' \to 0$ as $n \to \infty$ $\mathbf{Q}'$-a.s.*

PROOF. Let $\tilde{\mathbf{\Phi}} = (\mathbf{\Phi} + \mathbf{Q} + \mathbf{Q}')/3$, and let r_n be the density of $\mathbf{\Phi}|_{\mathcal{X}_n}$ with respect to $\tilde{\mathbf{\Phi}}|_{\mathcal{X}_n}$. Then we have $q_n r_n \to d\mathbf{Q}/d\tilde{\mathbf{\Phi}}$ and $q_n' r_n \to d\mathbf{Q}'/d\tilde{\mathbf{\Phi}}$ $\tilde{\mathbf{\Phi}}$-a.s., hence $\mathbf{Q}'$-a.s. as $\mathbf{Q}' \ll \tilde{\mathbf{\Phi}}$. But $d\mathbf{Q}/d\tilde{\mathbf{\Phi}} = 0$ $\mathbf{Q}'$-a.s. and $d\mathbf{Q}'/d\tilde{\mathbf{\Phi}} > 0$ $\mathbf{Q}'$-a.s. by the mutual singularity of $\mathbf{Q}$ and $\mathbf{Q}'$. The claim follows directly. $\square$

We consider the finite state space $E = \{1, \dots, p\}$. Let us write

$$\Omega_1, \dots, \Omega_p \subseteq F^{\mathbb{N}}, \qquad \Omega_i = \operatorname{supp} \mathbf{P}^i|_{\mathcal{F}^Y_{1,\infty}}.$$

There exists a finite partition $\{A_1,\dots,A_K\}$ of $F^{\mathbb{N}}$ such that $\sigma\{A_1,\dots,A_K\} = \sigma\{\Omega_1,\dots,\Omega_p\}$. We may assume without loss of generality that

$$\mathbf{P}^i\big((Y_k)_{k\in\mathbb{N}} \in A_1\big) > 0 \qquad \text{for } i = 1,\dots,q,$$
$$\mathbf{P}^i\big((Y_k)_{k\in\mathbb{N}} \in A_1\big) = 0 \qquad \text{for } i = q+1,\dots,p$$

for some $q \in E$ (this can always be accomplished by relabeling the points of the state space). Define $\tilde{\mathbf{P}}(\cdot) = \mathbf{P}^1(\cdot|(Y_k)_{k\in\mathbb{N}} \in A_1)$. Then, by construction, $\tilde{\mathbf{P}}|_{\mathcal{F}^Y_{1,\infty}} \ll \mathbf{P}^i|_{\mathcal{F}^Y_{1,\infty}}$ for $i = 1,\dots,q$ and $\tilde{\mathbf{P}}|_{\mathcal{F}^Y_{1,\infty}} \perp \mathbf{P}^i|_{\mathcal{F}^Y_{1,\infty}}$ for $i = q+1,\dots,p$.

We assume that the filter is uniquely ergodic, so that $\|\pi_n^x - \pi_n\| \to 0$ $\mathbf{P}^x$-a.s. for every $x \in E$ by Theorem 3.1 (this follows as λ charges all points in E, so we certainly have $\delta_x \ll \lambda$ for all $x \in E$). Therefore, we find that

$$\lim_{n\to\infty} \|\pi_n^x - \pi_n\| = 0 \qquad \text{for } x = 1,\dots,q,\ \tilde{\mathbf{P}}\text{-a.s.}$$

Denote by q_n^i the density of $\mathbf{P}^i((Y_1,\dots,Y_n) \in \cdot)$ with respect to $\varphi^{\otimes n}$, and similarly denote by $\tilde{q}_n$ the density of $\tilde{\mathbf{P}}((Y_1,\dots,Y_n) \in \cdot)$ with respect to $\varphi^{\otimes n}$. Then

$$\tilde{q}_n, q_n^x > 0 \qquad \text{for all } n \in \mathbb{N}, x = 1,\dots,q \quad \text{and} \quad \lim_{n\to\infty} \frac{\tilde{q}_n}{q_n^1} < \infty \qquad \tilde{\mathbf{P}}\text{-a.s.}$$

(the latter follows as $\tilde{q}_n/q_n^1$ is a uniformly integrable martingale under $\mathbf{P}^1$), while

$$\lim_{n\to\infty} \frac{q_n^i}{\tilde{q}_n} = 0 \qquad \text{for } i = q+1,\dots,p,\ \tilde{\mathbf{P}}\text{-a.s.}$$

by Lemma 5.1. (The fact that φ may be any σ-finite measure does not preclude the application of Lemma 5.1, as φ can always be transformed into a probability measure by means of an equivalent change of measure.)

As all of the above statements hold $\tilde{\mathbf{P}}$-a.s., we can certainly find one sample path $(\tilde{y}_k)_{k\in\mathbb{N}}$ on which all these statements hold simultaneously. In particular, we have

$$\left\| \frac{\delta_x M(\tilde{y}_1)\cdots M(\tilde{y}_n)}{\delta_x M(\tilde{y}_1)\cdots M(\tilde{y}_n)1} - \frac{\lambda M(\tilde{y}_1)\cdots M(\tilde{y}_n)}{\lambda M(\tilde{y}_1)\cdots M(\tilde{y}_n)1} \right\| \xrightarrow{n\to\infty} 0 \qquad \text{for } x = 1,\dots,q,$$

as well as

$$\frac{\delta_i M(\tilde{y}_1)\cdots M(\tilde{y}_n)1}{\delta_1 M(\tilde{y}_1)\cdots M(\tilde{y}_n)1} = \frac{q_n^i(\tilde{y}_1,\dots,\tilde{y}_n)}{q_n^1(\tilde{y}_1,\dots,\tilde{y}_n)} \xrightarrow{n\to\infty} 0 \qquad \text{for } i = q+1,\dots,p.$$

Now define the matrix norm $\|M\| := \sup_{\|f\|_\infty\le 1} \sup_{\|\mu\|_1\le 1} \mu M f$. As the set of matrices of unit norm is compact (as we are in the finite-dimensional setting), there must be a subsequence $n_k \nearrow \infty$ and a matrix M_∞ such that

$$\frac{M(\tilde{y}_1)\cdots M(\tilde{y}_{n_k})}{\|M(\tilde{y}_1)\cdots M(\tilde{y}_{n_k})\|} \xrightarrow{k\to\infty} M_\infty, \qquad \|M_\infty\| = 1.$$

We claim that M_∞ is a rank 1 matrix. Indeed, for $i = q+1, \dots, p$ we have

$$\|\delta_i M_\infty\| = \lim_{k\to\infty} \frac{\|\delta_i M(\tilde y_1)\cdots M(\tilde y_{n_k})\|}{\|M(\tilde y_1)\cdots M(\tilde y_{n_k})\|} \le \lim_{k\to\infty} \frac{\delta_i M(\tilde y_1)\cdots M(\tilde y_{n_k})1}{\delta_1 M(\tilde y_1)\cdots M(\tilde y_{n_k})1} = 0.$$

On the other hand, consider a state $x \in \{1, \dots, q\}$ such that $\|\delta_x M_\infty\| > 0$. Then $\delta_x M_\infty 1 = \|\delta_x M_\infty\| > 0$, and thus also $\lambda M_\infty 1 > 0$. But then

$$\begin{aligned}&\left\| \frac{\delta_x M_\infty}{\delta_x M_\infty 1} - \frac{\lambda M_\infty}{\lambda M_\infty 1} \right\| \\ &\qquad = \lim_{k\to\infty} \left\| \frac{\delta_x M(\tilde y_1)\cdots M(\tilde y_{n_k})}{\delta_x M(\tilde y_1)\cdots M(\tilde y_{n_k})1} - \frac{\lambda M(\tilde y_1)\cdots M(\tilde y_{n_k})}{\lambda M(\tilde y_1)\cdots M(\tilde y_{n_k})1} \right\| = 0.\end{aligned}$$

Therefore, we have shown that for every $j = 1, \dots, p$, the jth row of M_∞ is either zero, or a multiple of the row vector λM_∞. Moreover, M_∞ is not identically zero as $\|M_\infty\| = 1$. Thus M_∞ is a rank 1 matrix, and Condition (KR) follows.

5.2. *Proof of* $2 \Rightarrow 3$. We assume that Condition (KR) holds. Therefore, there exists a nonnegative column vector u (which is not identically zero), and a probability measure ϱ, such that the rank 1 matrix $u\varrho$ is in the closure of the cone $\mathcal{C}$. In particular, for any $\delta > 0$, we can choose $N \in \mathbb{N}$, $y_1, \dots, y_N \in F$, $c > 0$ such that

$$\|cM(y_1)\cdots M(y_N) - u\varrho\| < \delta.$$

Let $\alpha > 0$ (to be chosen below), and define the set

$$\mathcal{S} = \{\mu \in \mathcal{P}(E) : \mu u > \alpha\}.$$

Then we can estimate

$$\sup_{\mu\in\mathcal{S}} \left\| \frac{\mu M(y_1)\cdots M(y_N)}{\mu M(y_1)\cdots M(y_N)1} - \varrho \right\| \le \frac{2\delta}{\alpha}.$$

In particular, by the triangle inequality,

$$\sup_{\mu,\nu\in\mathcal{S}} \left\| \frac{\mu M(y_1)\cdots M(y_N)}{\mu M(y_1)\cdots M(y_N)1} - \frac{\nu M(y_1)\cdots M(y_N)}{\nu M(y_1)\cdots M(y_N)1} \right\| \le \frac{4\delta}{\alpha}.$$

Moreover, note that

$$c\mu M(y_1)\cdots M(y_N)1 \ge \mu u - \|cM(y_1)\cdots M(y_N) - u\varrho\| > \alpha - \delta$$

for all $\mu \in \mathcal{S}$.

We aim to show that Condition (C) is satisfied. To this end, let $\varepsilon > 0$ be given (and $\varepsilon < 1$ without loss of generality). As $\lambda u > 0$, we may choose $\alpha = \lambda u/2$ and $\delta = \alpha\varepsilon/4$. Now choose $N \in \mathbb{N}$, $y_1, \dots, y_N \in F$, $c > 0$ as above. When F is at most countable, the above choices of N and $\mathcal{S}$, together with the singleton $\mathcal{O} =$

$\{(y_1, \dots, y_N)\}$, satisfy properties 1–3 of Condition (C). Indeed, properties 2 and 3 are immediate from the above computations. To prove property 1, note that

$$\begin{aligned}
&\mathbf{P}(\pi_0^{\min} \in \mathcal{S} \text{ and } \pi_0^{\max} \in \mathcal{S}) \\
&\qquad \geq \mathbf{P}(\pi_0^{\max} u \geq \pi_0^{\min} u > \lambda u/2) \\
&\qquad = \mathbf{E}\big(\mathbf{P}(\pi_0^{\max} u \geq \pi_0^{\min} u | \mathcal{F}^Y_{-\infty,0}) I_{]\lambda u/2,\infty[}(\pi_0^{\min} u)\big).
\end{aligned}$$

As trivially $\mathbf{P}(X \geq \mathbf{E}(X)) > 0$ for any random variable X, we have $\mathbf{P}$-a.s.

$$\mathbf{P}(\pi_0^{\max} u \geq \pi_0^{\min} u | \mathcal{F}^Y_{-\infty,0}) = \mathbf{P}\big(\pi_0^{\max} u \geq \mathbf{E}(\pi_0^{\max} u | \mathcal{F}^Y_{-\infty,0}) | \mathcal{F}^Y_{-\infty,0}\big) > 0,$$

while $\mathbf{P}(\pi_0^{\min} u > \lambda u/2) > 0$ by virtue of the fact that $\mathbf{E}(\pi_0^{\min} u) = \lambda u$. Therefore $\mathbf{P}(\pi_0^{\min} \in \mathcal{S}$ and $\pi_0^{\max} \in \mathcal{S}) > 0$, and the claim is established.

In the case where $F = \mathbb{R}^d$, we cannot choose $\mathcal{O}$ to be a singleton as this set has Lebesgue measure zero. However, note that by the assumed continuity, all the above computations extend to a sufficiently small neighborhood of the path $(y_1, \dots, y_N)$. Choosing $\mathcal{O}$ to be such a neighborhood, we have $\varphi^{\otimes N}(\mathcal{O}) > 0$ by construction, and the remainder of the proof proceeds as in the countable case.

6. Sufficient conditions. Our main results, Theorems 2.6 and 2.7, establish necessary and sufficient conditions for unique ergodicity of the filter. The purpose of this section is to discuss various sufficient conditions that have appeared in the literature, and their relations to our main results. First, we discuss the sufficient conditions introduced by Kaijser [8, 9] and show how these can be obtained directly from our Theorem 2.6. Then, we discuss various conditions that have been introduced in the context of the filter stability problem [5, 14, 15].

6.1. *Kaijser's sufficient conditions.* In Kaijser's 1975 paper [8], the following condition is shown to be sufficient for unique ergodicity of the filter.

CONDITION (K). Let E and F be finite sets, let φ be the counting measure on F and let the signal transition matrix P be aperiodic. There exist $y_1, \dots, y_n \in F$ such that the matrix $M = M(y_1) \cdots M(y_n)$ is nonzero and subrectangular, that is, $M_{ij} > 0$ and $M_{kl} > 0$ imply $M_{il} > 0$ and $M_{kj} > 0$.

Kaijser's proof of sufficiency is based on the Furstenberg–Kesten theory of products of random matrices. A much simpler proof was given by Kochman and Reeds in [10], Section 5, where Condition (K) is shown to imply Condition (KR) through an explicit computation. Kochman and Reeds prove the sufficiency of Condition (KR) by invoking a general result in the ergodic theory of Markov chains in topological state spaces. We would argue that the proof of sufficiency given here is even simpler, at least if one takes for granted the (essentially known) characterization of unique ergodicity of the filter provided by Theorem 3.1.

Kaijser showed already in [8] by means of a counterexample that the subrectangularity condition cannot be dropped, that is, that irreducibility and aperiodicity of the signal need not imply unique ergodicity of the filter. Kochman and Reeds provide two further counterexamples [10]. They demonstrate that the assumption of aperiodicity cannot be dropped in Condition (K), that is, that subrectangularity and irreducibility need not imply unique ergodicity of the filter. Moreover, they provide a counterexample where Condition (KR) is satisfied and the signal is irreducible and aperiodic, but Condition (K) is not satisfied. Theorem 2.7 in this paper completes these results by establishing the necessity of Condition (KR).

In a recent paper, Kaijser [9] introduces two sufficient conditions for unique ergodicity of the filter in the case where E and F are countable.

CONDITION (B1). Let E and F be countable, and let φ be the counting measure. There exists a nonnegative function $u : E \to \mathbb{R}_+$ with $\|u\|_\infty = 1$, a probability measure ϱ on E, a sequence of integers $(n_k)_{k\in\mathbb{N}}$ and a sequence of observation paths $(y_1^k, \dots, y_{n_k}^k)_{k\in\mathbb{N}}$ with $\|M(y_1^k)\cdots M(y_{n_k}^k)\| > 0$, such that

$$\left\| \frac{\delta_x M(y_1^k)\cdots M(y_{n_k}^k)}{\|M(y_1^k)\cdots M(y_{n_k}^k)\|} - u(x)\varrho \right\| \xrightarrow{k\to\infty} 0 \qquad \text{for all } x \in E.$$

(Here we have defined the norm $\|M\| := \sup_{\|f\|_\infty\le 1} \sup_{\|\mu\|_1\le 1} \mu M f$.)

CONDITION (B). Let E and F be countable, and let φ be the counting measure. For every $\beta > 0$, there exists an $x_0 \in E$ such that the following holds: given any tight set $\mathcal{T} \subset \mathcal{P}(E)$ such that, for any $\mathsf{M}_0 \in \mathcal{P}(\mathcal{P}(E))$ with $\int \nu \mathsf{M}_0(d\nu) = \lambda$,

$$\mathsf{M}_0(\mathcal{T} \cap \{\nu \in \mathcal{P}(E) : \nu_{x_0} > \lambda_{x_0}/2\}) \ge \lambda_{x_0}/3,$$

there exist $N \in \mathbb{N}$ and $y_1, \dots, y_N \in F$ such that $\delta_{x_0} M(y_1)\cdots M(y_N)1 > 0$ and

$$\left\| \frac{\mu M(y_1)\cdots M(y_N)}{\mu M(y_1)\cdots M(y_N)1} - \frac{\delta_{x_0} M(y_1)\cdots M(y_N)}{\delta_{x_0} M(y_1)\cdots M(y_N)1} \right\| < \beta$$

for all $\mu \in \mathcal{T} \cap \{\nu \in \mathcal{P}(E) : \nu_{x_0} > \lambda_{x_0}/2\}$.

Kaijser shows that either of these conditions implies unique ergodicity of the filter, provided the signal transition matrix P is aperiodic. Kaijser's proof is very long and requires the development of some dedicated ergodicity results for Markov chains in nonlocally compact spaces. We will presently show that Condition (B1) and Condition (B) imply our Condition (C), so that Kaijser's results follow easily from Theorem 2.6 (even in the case where P is not aperiodic).

LEMMA 6.1. *Condition (B1) implies Condition (C).*

PROOF. Suppose that Condition (B1) holds. We can estimate

$$\begin{aligned}
&\left\| \frac{\mu M(y_1^k)\cdots M(y_{n_k}^k)}{\|M(y_1^k)\cdots M(y_{n_k}^k)\|} - \mu u \varrho \right\| \\
&\quad \le \int \left\| \frac{\delta_x M(y_1^k)\cdots M(y_{n_k}^k)}{\|M(y_1^k)\cdots M(y_{n_k}^k)\|} - u(x)\varrho \right\| \mu(dx) \\
&\quad \le \sum_{x=1}^{J} \mu_x \left\| \frac{\delta_x M(y_1^k)\cdots M(y_{n_k}^k)}{\|M(y_1^k)\cdots M(y_{n_k}^k)\|} - u(x)\varrho \right\| + 2 \sum_{x=J+1}^{\infty} \mu_x.
\end{aligned}$$

Let $\mathcal{T} \subset \mathcal{P}(E)$ be a tight set. Then the first term converges to zero uniformly in $\mu \in \mathcal{T}$ by assumption, while the second term can be made arbitrarily small uniformly in $\mu \in \mathcal{T}$ by choosing J sufficiently large. Therefore,

$$\sup_{\mu \in \mathcal{T}} \left\| \frac{\mu M(y_1^k)\cdots M(y_{n_k}^k)}{\|M(y_1^k)\cdots M(y_{n_k}^k)\|} - \mu u \varrho \right\| < \delta$$

for any tight set $\mathcal{T} \subset \mathcal{P}(E)$, $\delta > 0$, and k sufficiently large. Let $\alpha > 0$ and define

$$\mathcal{S} = \mathcal{T} \cap \{\mu \in \mathcal{P}(E) : \mu u > \alpha\}.$$

Then we obtain

$$\sup_{\mu,\nu \in \mathcal{S}} \left\| \frac{\mu M(y_1^k)\cdots M(y_{n_k}^k)}{\mu M(y_1^k)\cdots M(y_{n_k}^k)1} - \frac{\nu M(y_1^k)\cdots M(y_{n_k}^k)}{\nu M(y_1^k)\cdots M(y_{n_k}^k)1} \right\| \le \frac{4\delta}{\alpha}.$$

We now show that Condition (C) is satisfied. Let $\varepsilon > 0$ be given, and choose $\alpha = \lambda u/2$ and $\delta = \alpha\varepsilon/4$. As in the proof of Theorem 2.7, we can show that

$$\mathbf{P}(\pi_0^{\min} u > \alpha \text{ and } \pi_0^{\max} u > \alpha) > 0.$$

Moreover, we can find an increasing sequence of tight sets $\mathcal{T}_n \subset \mathcal{P}(E)$ such that

$$\mathbf{P}(\pi_0^{\min} \in \mathcal{T}_n \text{ and } \pi_0^{\max} \in \mathcal{T}_n) \overset{n\to\infty}{\longrightarrow} 1,$$

as $\mathcal{P}(\mathcal{P}(E))$ is Polish. Therefore, we can choose $\mathcal{T}$ sufficiently large such that

$$\mathbf{P}(\pi_0^{\min} \in \mathcal{S} \text{ and } \pi_0^{\max} \in \mathcal{S}) > 0.$$

The remainder of the proof is identical to that of Theorem 2.7. □

LEMMA 6.2. *Condition (B) implies Condition (C).*

PROOF. Suppose that Condition (B) holds. We claim that Condition (C) holds with $\varepsilon = 2\beta$, $\mathcal{S} = \mathcal{T} \cap \{\nu \in \mathcal{P}(E) : \nu_{x_0} > \lambda_{x_0}/2\}$, and $\mathcal{O} = \{(y_1, \ldots, y_N)\}$, provided that $\mathcal{T} \subset \mathcal{P}(E)$ is chosen sufficiently large. Indeed, as the family $\mathcal{M} = \{\mathsf{M}_0 \in \mathcal{P}(\mathcal{P}(E)) : \int \nu \mathsf{M}_0(d\nu) = \lambda\}$ is tight (e.g., [7]) it is easily seen that

$$\mathsf{M}_0\big(\mathcal{T} \cap \{\nu \in \mathcal{P}(E) : \nu_{x_0} > \lambda_{x_0}/2\}\big) \ge \lambda_{x_0}/3 \qquad \text{for all } \mathsf{M}_0 \in \mathcal{M}$$

is satisfied for every sufficiently large tight set $\mathcal{T} \subset \mathcal{P}(E)$. Moreover,

$$\mathbf{P}(\pi_0^{\min} \in \mathcal{S} \text{ and } \pi_0^{\max} \in \mathcal{S}) > 0$$

when $\mathcal{T}$ is chosen sufficiently large, as is shown in the proof of Lemma 6.1. It remains to note that as $\delta_{x_0} M(y_1) \cdots M(y_N) 1 > 0$, we have $\mu M(y_1) \cdots M(y_N) 1 > 0$ for all $\mu \in \mathcal{S}$. The remainder of Condition (C) now follows immediately. $\square$

Though Condition (B1) is strongly reminiscent of Condition (KR), we did not succeed in extending the proof of the necessity of Condition (KR) to the countable case. Whether Conditions (B1), (B) or some variant of thereof are necessary and sufficient for unique ergodicity in the countable case remains an open problem.

6.2. *Nondegeneracy and observability.* Conditions of a rather different kind than are considered by Kaijser, Kochman and Reeds relate to the filter stability problem (see the survey [5]). By Theorem 3.1, however, filter stability and unique ergodicity are essentially equivalent, so that also these conditions can be brought to bear on the problem considered in this paper. In this section, we consider the following conditions that are borrowed from [14, 15]: nondegeneracy [Condition (N)], uniform observability [Condition (UO)] and observability [Condition (O)].

CONDITION (N). If $i, j \in E$ and $P_{ij} > 0$, then $M_{ij}(y) > 0$ for all $y \in F$.

CONDITION (UO). For every $\varepsilon > 0$, there is a $\delta > 0$ such that

$$\|\mathbf{P}^\mu|_{\mathcal{F}^Y_{1,\infty}} - \mathbf{P}^\nu|_{\mathcal{F}^Y_{1,\infty}}\| < \delta \quad \text{implies} \quad \|\mu - \nu\| < \varepsilon \qquad [\text{for any } \mu, \nu \in \mathcal{P}(E)].$$

CONDITION (O). If $\mu, \nu \in \mathcal{P}(E)$ and $\mathbf{P}^\mu|_{\mathcal{F}^Y_{1,\infty}} = \mathbf{P}^\nu|_{\mathcal{F}^Y_{1,\infty}}$, then $\mu = \nu$.

THEOREM 6.3. *Suppose that one of the following holds*:

- *Condition* (N) *holds, and the signal transition matrix P is aperiodic*; *or*
- *Condition* (UO) *holds*; *or*
- *Condition* (O) *holds, and E is a finite set.*

Then the filter admits a unique invariant measure M, *and* $n^{-1} \sum_{k=1}^n \mathsf{M}_0 \Pi^k \Rightarrow \mathsf{M}$ *as* $n \to \infty$ *for any* $\mathsf{M}_0 \in \mathcal{P}(\mathcal{P}(E))$. *If, in addition, the signal transition matrix P is aperiodic, then we have* $\mathsf{M}_0 \Pi^n \Rightarrow \mathsf{M}$ *as* $n \to \infty$ *for any* $\mathsf{M}_0 \in \mathcal{P}(\mathcal{P}(E))$.

SKETCH OF PROOF. First, suppose that Condition (N) holds and that P is aperiodic. Consider the stochastic process $(\mathbf{X}_n, \mathbf{Y}_n)_{n \in \mathbb{Z}}$ defined as

$$\mathbf{X}_n = (X_n, X_{n+1}) \in \mathbf{E}, \qquad \mathbf{Y}_n = Y_{n+1} \in F,$$

where $\mathbf{E} = \{x \in E^2 : \mathbf{P}(\mathbf{X}_0 = x) > 0\}$. Then $(\mathbf{X}_n, \mathbf{Y}_n)_{n\in\mathbb{Z}}$ is a stationary Markov chain, $(\mathbf{X}_n)_{n\in\mathbb{Z}}$ is an irreducible and aperiodic Markov chain, and $(\mathbf{Y}_n)_{n\in\mathbb{Z}}$ are conditionally independent given $(\mathbf{X}_n)_{n\in\mathbb{Z}}$. Moreover,

$$\mathbf{P}\big(\mathbf{Y}_n \in A|(\mathbf{X}_k)_{k\in\mathbb{Z}}\big) = \int_A M_{X_n X_{n+1}}(y)\varphi(dy) := \int_A \Upsilon(\mathbf{X}_n, y)\varphi(dy),$$

where $\Upsilon(x, y) > 0$ for all $x \in \mathbf{E}$ and $y \in F$ by Condition (N). Therefore,

$$\|\mathbf{P}^{\mu}(\mathbf{X}_n \in \cdot|\mathbf{Y}_0, \dots, \mathbf{Y}_n) - \mathbf{P}^{\nu}(\mathbf{X}_n \in \cdot|\mathbf{Y}_0, \dots, \mathbf{Y}_n)\| \xrightarrow{n\to\infty} 0 \qquad \mathbf{P}^{\mu}\text{-a.s.}$$

for all $\mu, \nu \in \mathcal{P}(E)$ by [14], Corollary 5.5. It follows immediately that $\|\pi_n^{\mu} - \pi_n^{\nu}\| \to 0$ as $n \to \infty$ $\mathbf{P}^{\mu}$-a.s. The proof is completed by invoking Theorem 3.1.

Next, suppose Condition (UO) holds. By a result of Blackwell and Dubins [3],

$$\big\|\mathbf{P}^{\mu}\big((Y_k)_{k>n} \in \cdot|\mathcal{F}_{1,n}^{Y}\big) - \mathbf{P}^{\nu}\big((Y_k)_{k>n} \in \cdot|\mathcal{F}_{1,n}^{Y}\big)\big\| \xrightarrow{n\to\infty} 0 \qquad \mathbf{P}^{\mu}\text{-a.s.}$$

whenever $\mu \ll \nu$. But one can show (e.g., [15]) that

$$\mathbf{P}^{\rho}\big((Y_k)_{k>n} \in \cdot|\mathcal{F}_{1,n}^{Y}\big) = \mathbf{P}^{\pi_n^{\rho}}\big((Y_k)_{k>0} \in \cdot\big) = \mathbf{P}^{\pi_n^{\rho}}|_{\mathcal{F}_{1,\infty}^{Y}} \qquad \text{for all } \rho \in \mathcal{P}(E).$$

Using Condition (UO), it therefore follows that $\|\pi_n^{\mu} - \pi_n^{\nu}\| \to 0$ as $n \to \infty$ $\mathbf{P}^{\mu}$-a.s. whenever $\mu \ll \nu$. The proof is completed by invoking Theorem 3.1.

Finally, suppose that E is finite, and that Condition (O) holds. Then it is not difficult to establish, along the lines of [15], Proposition 3.5, that Condition (UO) is satisfied. The result therefore follows as above. □

REMARK 6.4. When the signal transition kernel P is periodic, Condition (N) by itself does not ensure unique ergodicity of the filter (this can be seen, e.g., by considering the example of a periodic signal in $E = \{1, 2\}$ with the trivial observation state space $F = \{1\}$). However, if Condition (N) holds and E is a finite set, a *detectability* condition [which is weaker than Condition (O)] is necessary and sufficient for stability of the filter, and hence for unique ergodicity. The necessary arguments can be adapted from [13], Section 6.2, with some care. As this is somewhat outside the scope of this paper, we omit the details.

It should be noted that none of the conditions of Theorem 6.3 are necessary. Indeed, Condition (N) is not satisfied by the examples given by Kochman and Reeds [10]. That Condition (UO) [hence Condition (O)] is not necessary can be seen from the trivial counterexample, where P is aperiodic and $F = \{1\}$. In this case the observations are completely noninformative, so that the point mass at $\lambda \in \mathcal{P}(E)$ is the unique invariant measure for the filter, but Condition (UO) is not satisfied.

Nonetheless, the sufficient conditions of Theorem 6.3 can be useful in practice, as they may be substantially easier to check than Condition (C) or (KR). For example, in the case where E is a finite set, verifying Condition (O) is simply a matter of linear algebra (see [5] for an example), while verifying Condition (KR) involves taking limits. Moreover, despite that Conditions (C) and (KR) are both necessary and sufficient in many cases, we did not succeed in our attempt to prove Theorem 6.3 by directly verifying that Condition (C) or (KR) hold. Therefore, such sufficient but not necessary conditions remain of independent interest.

APPENDIX: SUPPLEMENTARY PROOFS

A.1. Proof of Lemmas 2.3 and 2.4. We will need the following.

LEMMA A.1. *$\pi_n^{\min} = \pi_m^{\min} \circ \Theta^{n-m}$ and $\pi_n^{\max} = \pi_m^{\max} \circ \Theta^{n-m}$ for $m, n \in \mathbb{Z}$.*

PROOF. By stationarity of $\mathbf{P}$, it is easily seen that

$$\mathbf{E}(f(X_n)|\mathcal{F}^Y_{n-\ell,n}) = \mathbf{E}(f(X_m)|\mathcal{F}^Y_{m-\ell,m}) \circ \Theta^{n-m},$$

$$\mathbf{E}\big(f(X_n)|\mathcal{F}^Y_{n-\ell,n} \vee \mathcal{F}^X_{n-\ell,n-k}\big) = \mathbf{E}\big(f(X_m)|\mathcal{F}^Y_{m-\ell,m} \vee \mathcal{F}^X_{m-\ell,m-k}\big) \circ \Theta^{n-m}.$$

The result follows by letting $\ell \to \infty$, then $k \to \infty$. □

We begin by proving Lemma 2.3 for the case π_n^μ. It clearly suffices to prove

$$\pi_n^\mu = \frac{\mu M(Y_1)\cdots M(Y_n)}{\mu M(Y_1)\cdots M(Y_n)1}, \qquad \mathbf{P}^\mu\text{-a.s. for all } n \geq 1.$$

Let $f \in \mathcal{C}_b(E)$ and $A \in \mathcal{B}(F^n)$. Then

$$\begin{aligned}
&\mathbf{E}^\mu\left(\frac{\mu M(Y_1)\cdots M(Y_n) f}{\mu M(Y_1)\cdots M(Y_n)1} I_A(Y_1,\ldots,Y_n)\right)\\
&\quad = \int_A \frac{\mu M(y_1)\cdots M(y_n) f}{\mu M(y_1)\cdots M(y_n)1} \mu M(y_1)\cdots M(y_n)1 \varphi(dy_1)\cdots\varphi(dy_n)\\
&\quad = \int_A \mu M(y_1)\cdots M(y_n) f \varphi(dy_1)\cdots\varphi(dy_n)\\
&\quad = \mathbf{E}^\mu(I_A(Y_1,\ldots,Y_n) f(X_n)).
\end{aligned}$$

As this holds for any $f \in \mathcal{C}_b(E)$ and $A \in \mathcal{B}(F^n)$, the above expression for π_n^μ follows from the definition of the conditional expectation.

To prove Lemma 2.3 for $\pi_n^{\min}$, let $k, n \geq 1$. Note that

$$\begin{aligned}
\mathbf{E}(f(X_n)|\mathcal{F}^Y_{-k+1,n}) &= \pi_{n+k} f \circ \Theta^{-k}\\
&= \frac{\lambda M(Y_{-k+1})\cdots M(Y_n) f}{\lambda M(Y_{-k+1})\cdots M(Y_n)1}\\
&= \frac{(\pi_k \circ \Theta^{-k}) M(Y_1)\cdots M(Y_n) f}{(\pi_k \circ \Theta^{-k}) M(Y_1)\cdots M(Y_n)1}.
\end{aligned}$$

But $\mathbf{E}(f(X_n)|\mathcal{F}^Y_{-k+1,n}) \to \pi^{\min}_n f$ and $\pi_k f \circ \Theta^{-k} = \mathbf{E}(f(X_0)|\mathcal{F}^Y_{-k+1,0}) \to \pi^{\min}_0 f$ as $k \to \infty$ $\mathbf{P}$-a.s. by the martingale convergence theorem. Therefore

$$\pi^{\min}_n = \frac{\pi^{\min}_0 M(Y_1)\cdots M(Y_n)}{\pi^{\min}_0 M(Y_1)\cdots M(Y_n)1}, \qquad \mathbf{P}\text{-a.s. for all } n \geq 1,$$

and the result follows for arbitrary $m, n \in \mathbb{Z}$, $n \geq m$ by Lemma A.1.

To prove Lemma 2.3 for $\pi^{\max}_n$, let $n \geq 1$ and $k \geq \ell \geq 0$. Note that

$$\begin{aligned}\mathbf{E}\big(f(X_n)|\mathcal{F}^Y_{-k,n} \vee \mathcal{F}^X_{-k,-\ell}\big) &= \mathbf{E}\big(f(X_{n+\ell})|\mathcal{F}^Y_{-k+\ell,n+\ell} \vee \mathcal{F}^X_{-k+\ell,0}\big) \circ \Theta^{-\ell} \\ &= \mathbf{E}\big(f(X_{n+\ell})|\mathcal{F}^Y_{1,n+\ell} \vee \sigma\{X_0\}\big) \circ \Theta^{-\ell},\end{aligned}$$

where we have used the Markov property. Moreover, it is easily seen that

$$\mathbf{E}\big(f(X_{n+\ell})|\mathcal{F}^Y_{1,n+\ell} \vee \sigma\{X_0\}\big) = \pi^{X_0}_{n+\ell} f = \frac{\delta_{X_0} M(Y_1)\cdots M(Y_{n+\ell}) f}{\delta_{X_0} M(Y_1)\cdots M(Y_{n+\ell})1}.$$

Therefore, we can write

$$\begin{aligned}\mathbf{E}\big(f(X_n)|\mathcal{F}^Y_{-k,n} \vee \mathcal{F}^X_{-k,-\ell}\big) &= \frac{\delta_{X_{-\ell}} M(Y_{-\ell+1})\cdots M(Y_n) f}{\delta_{X_{-\ell}} M(Y_{-\ell+1})\cdots M(Y_n)1} \\ &= \frac{(\pi^{X_0}_\ell \circ \Theta^{-\ell}) M(Y_1)\cdots M(Y_n) f}{(\pi^{X_0}_\ell \circ \Theta^{-\ell}) M(Y_1)\cdots M(Y_n)1}.\end{aligned}$$

Letting $k \to \infty$, then $\ell \to \infty$ and applying the martingale convergence theorem, we obtain the desired recursion for $\pi^{\max}_n$.

We now turn to the proof of Lemma 2.4. The stationarity of $(\pi^{\max}_n)_{n\in\mathbb{Z}}$ and $(\pi^{\min}_n)_{n\in\mathbb{Z}}$ follows directly from Lemma A.1 and the stationarity of $(X_n, Y_n)_{n\in\mathbb{Z}}$. It only remains to prove the Markov property. For $f \in \mathcal{C}_b(\mathcal{P}(E))$, we can compute

$$\begin{aligned}\mathbf{E}(f(\pi^{\max}_{n+1})|\mathcal{G}_{-\infty,n}) &= \mathbf{E}\left(f\left(\frac{\pi^{\max}_n M(Y_{n+1})}{\pi^{\max}_n M(Y_{n+1})1}\right)\middle|\mathcal{G}_{-\infty,n}\right) \\ &= \mathbf{E}\left(f\left(\frac{\mu M(Y_{n+1})}{\mu M(Y_{n+1})1}\right)\middle|\mathcal{G}_{-\infty,n}\right)\Bigg|_{\mu=\pi^{\max}_n},\end{aligned}$$

where we have used Lemma 2.3 and the fact that $\pi^{\max}_n$ is $\mathcal{G}_{-\infty,n}$-measurable. But for any bounded measurable function $g : F \to \mathbb{R}$, we have

$$\mathbf{E}(g(Y_{n+1})|\mathcal{G}_{-\infty,n}) = \int g(y)\pi^{\max}_n M(y)1\varphi(dy).$$

The Markov property and the expression for the transition kernel Π follows immediately. The Markov property of $\pi^{\min}_n$ and π^μ_n follows along similar lines.

A.2. Proof of Theorem 3.1 and Lemma 3.4. The proof of Theorem 3.1 follows closely along the lines of [6, 11, 12]. We will sketch the necessary arguments, concentrating on the special features of the countable setting.

We begin by establishing the Feller property.

LEMMA A.2. *Let $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(E)$ and $\mu \in \mathcal{P}(E)$ be such that $\mu_n \Rightarrow \mu$. Then $\int f(\nu)\Pi(\mu_n, d\nu) \to \int f(\nu)\Pi(\mu, d\nu)$ for every $f \in \mathcal{C}_b(\mathcal{P}(E))$.*

PROOF. Let $N \subset F$ be a φ-null set such that $\sup_i \sum_j M_{ij}(y) < \infty$ for all $y \notin N$. Then $\mu_n M(y)1 \to \mu M(y)1$ for all $y \notin N$, and $\mu_n M(y)/\mu_n M(y)1 \Rightarrow \mu M(y)/\mu M(y)1$ whenever $y \notin N$ and $\mu M(y)1 > 0$. It follows that

$$f\left(\frac{\mu_n M(y)}{\mu_n M(y)1}\right)\mu_n M(y)1 \xrightarrow{n\to\infty} f\left(\frac{\mu M(y)}{\mu M(y)1}\right)\mu M(y)1 \qquad \text{for all } y \notin N.$$

But the family $\{f(\mu_n M(y)/\mu_n M(y)1)\mu_n M(y)1 : n \in \mathbb{N}\}$ is uniformly integrable (under φ), as $|f(\mu_n M(y)/\mu_n M(y)1)\mu_n M(y)1| \le \|f\|_\infty \mu_n M(y)1$ and by Scheffé's lemma $\int |\mu_n M(y)1 - \mu M(y)1|\varphi(dy) \to 0$. The result therefore follows from the expression for Π in Lemma 2.4. $\square$

We will need some basic elements from Choquet theory.

DEFINITION A.3. Let S be Polish. For $\mathsf{M}, \mathsf{M}' \in \mathcal{P}(\mathcal{P}(S))$ we write $\mathsf{M} \prec \mathsf{M}'$ if

$$\int f(\nu)\mathsf{M}(d\nu) \le \int f(\nu)\mathsf{M}'(d\nu) \qquad \text{for every convex } f \in \mathcal{C}_b(\mathcal{P}(S)).$$

For any $\mathsf{M} \in \mathcal{P}(\mathcal{P}(S))$, the *barycenter* $b(\mathsf{M}) \in \mathcal{P}(S)$ is defined as

$$b(\mathsf{M})u = \int \nu u \mathsf{M}(d\nu) \qquad \text{for all } u \in \mathcal{C}_b(S).$$

For any $\mu \in \mathcal{P}(S)$, define $\mathsf{m}_\mu, \tilde{\mathsf{m}}_\mu \in \mathcal{P}(\mathcal{P}(S))$ as

$$\int f(\nu)\mathsf{m}_\mu(d\nu) = f(\mu), \qquad \int f(\nu)\tilde{\mathsf{m}}_\mu(d\nu) = \int f(\delta_x)\mu(dx)$$

for every $f \in \mathcal{C}_b(\mathcal{P}(S))$.

LEMMA A.4. *Let S be a Polish space. The following hold*:

(1) *Given $\mathsf{M} \in \mathcal{P}(\mathcal{P}(S))$, we have $f(b(\mathsf{M})) \le \int f(\nu)\mathsf{M}(d\nu)$ for every convex function $f \in \mathcal{C}_b(\mathcal{P}(S))$ (Jensen's inequality).*
(2) *For any $\mathsf{M} \in \mathcal{P}(\mathcal{P}(S))$, we have $\mathsf{m}_{b(\mathsf{M})} \prec \mathsf{M} \prec \tilde{\mathsf{m}}_{b(\mathsf{M})}$.*
(3) *If $\mathsf{M}, \mathsf{M}' \in \mathcal{P}(\mathcal{P}(S))$, $\mathsf{M} \prec \mathsf{M}'$ and $\mathsf{M}' \prec \mathsf{M}$, we have $\mathsf{M} = \mathsf{M}'$.*

In particular, $\prec$ defines a partial order on $\mathcal{P}(\mathcal{P}(S))$.

PROOF. Jensen's inequality is proved as in [11], Lemma 3.1. The second property follows easily from Jensen's inequality. The third property follows from the fact that the family of convex functions in $\mathcal{C}_b(\mathcal{P}(S))$ is a measure determining class (see, e.g., Proposition A1 in [12]). $\square$

We now need some basic convexity properties of the filter.

LEMMA A.5. *The following hold for any* $\mathsf{M} \in \mathcal{P}(\mathcal{P}(E))$:

(1) *If* $f \in \mathcal{C}_b(\mathcal{P}(E))$ *is convex, then* $\Pi f \in \mathcal{C}_b(\mathcal{P}(E))$ *is also convex.*
(2) $b(\mathsf{M}\Pi) = b(\mathsf{M})P$.
(3) *If* $\mathsf{M}\Pi = \mathsf{M}$, *then* $b(\mathsf{M}) = \lambda$.
(4) $\mathsf{m}_{b(\mathsf{M})P^n}\Pi^m \prec \mathsf{M}\Pi^{m+n} \prec \tilde{\mathsf{m}}_{b(\mathsf{M})P^n}\Pi^m$ *for any* $m, n \geq 0$.

PROOF. The first claim follows as in [11], Lemma 3.2. The second claim follows directly from Lemma 2.4. The third claim follows from the second claim and the fact that λ is the unique invariant measure for P. The fourth claim follows from the first and second claims, together with the second claim of Lemma A.4. $\square$

The following lemma connects $\pi_0^{\min}, \pi_0^{\max}$ to the filter transition kernel Π.

LEMMA A.6. *Denote by* $\mathsf{M}^{\max}, \mathsf{M}^{\min} \in \mathcal{P}(\mathcal{P}(E))$ *the laws of* $\pi_0^{\max}$ *and* $\pi_0^{\min}$, *respectively. Then* $\mathsf{m}_\lambda \Pi^n \Rightarrow \mathsf{M}^{\min}$ *and* $\tilde{\mathsf{m}}_\lambda \Pi^n \Rightarrow \mathsf{M}^{\max}$ *as* $n \to \infty$.

PROOF. Let $f \in \mathcal{C}_b(\mathcal{P}(E))$. Then

$$\mathsf{m}_\lambda \Pi^n f = \mathbf{E}[f(\pi_n)] = \mathbf{E}\big[f\big(\mathbf{P}(X_0 \in \cdot | \mathcal{F}^Y_{-n+1,0})\big)\big] \xrightarrow{n\to\infty} \mathbf{E}[f(\pi_0^{\min})],$$

where we have used stationarity and martingale convergence. Similarly,

$$\begin{aligned}\tilde{\mathsf{m}}_\lambda \Pi^n f &= \mathbf{E}[f(\pi_n^{X_0})] = \mathbf{E}\big[f\big(\mathbf{P}(X_0 \in \cdot | \mathcal{F}^Y_{-n+1,0} \vee \sigma\{X_{-n}\})\big)\big] \\ &= \mathbf{E}\big[f\big(\mathbf{P}(X_0 \in \cdot | \mathcal{F}^Y_{-\infty,0} \vee \mathcal{F}^X_{-\infty,-n})\big)\big] \xrightarrow{n\to\infty} \mathbf{E}[f(\pi_0^{\max})],\end{aligned}$$

where we have additionally used the Markov property of $(X_n, Y_n)_{n\in\mathbb{Z}}$. $\square$

Finally, we will need the following convergence property:

LEMMA A.7. $\lim_{n\to\infty} \|\pi_n^\mu - \pi_n^\nu\|$ *exists* $\mathbf{P}^\mu$*-a.s. whenever* $\mu \ll \nu$.

PROOF. It is not difficult to show along the lines of [14], Corollary 5.7, that

$$\begin{aligned}&\|\pi_n^\mu - \pi_n^\nu\| \\ &\quad= \frac{\mathbf{E}^\nu(|\mathbf{E}^\nu((d\mu/d\nu)(X_0)|\mathcal{F}^X_{n,\infty} \vee \mathcal{F}^Y_{1,\infty}) - \mathbf{E}^\nu((d\mu/d\nu)(X_0)|\mathcal{F}^Y_{1,n})||\mathcal{F}^Y_{1,n})}{\mathbf{E}^\nu((d\mu/d\nu)(X_0)|\mathcal{F}^Y_{1,n})}\end{aligned}$$

$\mathbf{P}^\mu$-a.s. whenever $\mu \ll \nu$. The denominator converges $\mathbf{P}^\nu$-a.s., hence $\mathbf{P}^\mu$-a.s. (as $\mu \ll \nu$), to a random variable which is strictly positive $\mathbf{P}^\mu$-a.s.

To prove convergence of the numerator, let $\varepsilon > 0$ and define

$$\begin{aligned} M_n &= \mathbf{E}^\nu\biggl(\frac{d\mu}{d\nu}(X_0) I_{d\mu/d\nu(X_0)<\varepsilon} | \mathcal{F}^Y_{1,n}\biggr),\\ M'_n &= \mathbf{E}^\nu\biggl(\frac{d\mu}{d\nu}(X_0) I_{d\mu/d\nu(X_0)\geq\varepsilon} | \mathcal{F}^Y_{1,n}\biggr),\\ L_n &= \mathbf{E}^\nu\biggl(\frac{d\mu}{d\nu}(X_0) I_{d\mu/d\nu(X_0)<\varepsilon} | \mathcal{F}^X_{n,\infty} \vee \mathcal{F}^Y_{1,\infty}\biggr),\\ L'_n &= \mathbf{E}^\nu\biggl(\frac{d\mu}{d\nu}(X_0) I_{d\mu/d\nu(X_0)\geq\varepsilon} | \mathcal{F}^X_{n,\infty} \vee \mathcal{F}^Y_{1,\infty}\biggr). \end{aligned}$$

Clearly M_n and M'_n are uniformly integrable martingales, while L_n and L'_n are reverse martingales. Moreover, the numerator can be written as $\mathbf{E}^\nu(Z_n|\mathcal{F}^Y_{1,n})$ where $Z_n = |L_n + L'_n - M_n - M'_n|$. We proceed to estimate as follows:

$$|\mathbf{E}^\nu(Z_n - Z_\infty|\mathcal{F}^Y_{1,n})| \leq \mathbf{E}^\nu(|L_n - L_\infty||\mathcal{F}^Y_{1,n}) + \mathbf{E}^\nu(|M_n - M_\infty||\mathcal{F}^Y_{1,n}) + 4M'_n.$$

The first two terms converge to zero $\mathbf{P}^\nu$-a.s. as $n \to \infty$ by Hunt's lemma ([3], Theorem 2), while $\lim_{n\to\infty} M'_n$ vanishes if we let $\varepsilon \to \infty$. Therefore $\mathbf{E}^\nu(Z_n - Z_\infty|\mathcal{F}^Y_{1,n}) \to 0$ as $n \to \infty$ $\mathbf{P}^\nu$-a.s., and the proof is easily completed. $\square$

We now proceed to the proof of Theorem 3.1 and Lemma 3.4.

A.2.1. *Proof of Theorem* 3.1 (1 $\Leftrightarrow$ 2). First suppose $\mathbf{P}(\pi_0^{\max} = \pi_0^{\min}) < 1$. Then $\mathbf{E}(\{(\pi_0^{\max})_i - (\pi_0^{\min})_i\}^2) > 0$ for some $i \in E$. Now note that

$$\begin{aligned} \mathbf{E}\bigl(\{(\pi_0^{\max})_i - (\pi_0^{\min})_i\}^2\bigr) &= \mathbf{E}(\{(\pi_0^{\max})_i\}^2) - \mathbf{E}(\{(\pi_0^{\min})_i\}^2)\\ &= \int \{\nu_i\}^2 \mathsf{M}^{\max}(d\nu) - \int \{\nu_i\}^2 \mathsf{M}^{\min}(d\nu), \end{aligned}$$

so that $\mathbf{P}(\pi_0^{\max} = \pi_0^{\min}) < 1$ implies $\mathsf{M}^{\max} \neq \mathsf{M}^{\min}$. But $\mathsf{M}^{\max}$ and $\mathsf{M}^{\min}$ are invariant measures for Π by Lemma 2.4, so we have shown that the filter admits two distinct invariant measures. Conversely, if the invariant measure is unique, then $\mathbf{P}(\pi_0^{\max} = \pi_0^{\min}) = 1$. Thus we have proved the implication $1 \Rightarrow 2$.

Now suppose that $\pi_0^{\max} = \pi_0^{\min}$, so that in particular $\mathsf{M}^{\max} = \mathsf{M}^{\min}$. Let M be any invariant measure for Π. We claim that $\mathsf{M}^{\min} \prec \mathsf{M} \prec \mathsf{M}^{\max}$, so that necessarily $\mathsf{M} = \mathsf{M}^{\max} = \mathsf{M}^{\min}$ by Lemma A.4. To prove the claim, note that $\mathsf{m}_\lambda \Pi^n \prec \mathsf{M}\Pi^n = \mathsf{M} \prec \tilde{\mathsf{m}}_\lambda \Pi^n$ for any $n \geq 0$ by Lemmas A.4 and A.5. The claim therefore follows directly from Lemma A.6. Thus we have proved the implication $2 \Rightarrow 1$.

A.2.2. *Proof of Theorem* 3.1 ($2 \Leftrightarrow 3$). Proceeding along the same lines as in the proof of Lemma A.6 (and taking into account the fact that weak convergence and total variation convergence of probability measures coincide when the state space is countable), one can show that

$$\mathbf{E}(\|\pi_n^{X_0} - \pi_n\|) \overset{n\to\infty}{\longrightarrow} \mathbf{E}(\|\pi_0^{\max} - \pi_0^{\min}\|).$$

Suppose first that property 3 holds. Then

$$\mathbf{E}(\|\pi_n^{X_0} - \pi_n\|) = \sum_{i \in E} \lambda_i \mathbf{E}^i(\|\pi_n^i - \pi_n\|) \overset{n\to\infty}{\longrightarrow} 0.$$

Therefore $\pi_0^{\max} = \pi_0^{\min}$, and we have proved the implication $3 \Rightarrow 2$.

Conversely, suppose that $\pi_0^{\max} = \pi_0^{\min}$. Let $\mu, \nu \in \mathcal{P}(E)$ such that $\mu \ll \nu$. Note that we can write $\pi_n^\mu = \mathbf{E}^\mu(\pi_n^{X_0} | \mathcal{F}_{1,n}^Y)$. Therefore, we have

$$\mathbf{E}^\mu(\|\pi_n^\mu - \pi_n\|) = \mathbf{E}^\mu\big(\|\mathbf{E}^\mu(\pi_n^{X_0} - \pi_n | \mathcal{F}_{1,n}^Y)\|\big) \le \mathbf{E}^\mu(\|\pi_n^{X_0} - \pi_n\|).$$

But $\mathbf{E}(\|\pi_n^{X_0} - \pi_n\|) \to 0$, so $\|\pi_n^{X_0} - \pi_n\| \to 0$ in probability. As $\mu \ll \lambda$, we find that $\|\pi_n^{X_0} - \pi_n\| \to 0$ in $\mathbf{P}^\mu$-probability also, and by dominated convergence

$$\mathbf{E}^\mu(\|\pi_n^\mu - \pi_n\|) \le \mathbf{E}^\mu(\|\pi_n^{X_0} - \pi_n\|) \overset{n\to\infty}{\longrightarrow} 0.$$

By Lemma A.7, it follows that $\|\pi_n^\mu - \pi_n\| \to 0$ $\mathbf{P}^\mu$-a.s. Similarly, we find that $\|\pi_n^\nu - \pi_n\| \to 0$ $\mathbf{P}^\nu$-a.s., hence $\mathbf{P}^\mu$-a.s. as $\mu \ll \nu$. Therefore $\|\pi_n^\mu - \pi_n^\nu\| \to 0$ $\mathbf{P}^\mu$-a.s., and we have evidently proved the implication $2 \Rightarrow 3$.

A.2.3. *Proof of Theorem* 3.1 ($1 \Leftrightarrow 4$). The implication $4 \Rightarrow 1$ follows immediately by choosing M_0 and M to be distinct invariant measures of the filter and applying property 4, which leads to a contradiction.

To prove the converse implication, choose M_0 arbitrarily and define the measures $\mathsf{M}_n = n^{-1} \sum_{k=1}^n \mathsf{M}_0 \Pi^n$. Note that $b(\mathsf{M}_n) = n^{-1} \sum_{k=1}^n b(\mathsf{M}_0) P^n \Rightarrow \lambda$ as the signal is irreducible and positive recurrent. It follows from [7] that the sequence $(\mathsf{M}_n)_{n\in\mathbb{N}}$ is tight. It therefore suffices to prove that every convergent subsequence has the same limit. But it is easily seen that any convergent subsequence converges to an invariant measure of Π, so that the result follows from the uniqueness of the invariant measure. Thus we have proved the implication $1 \Rightarrow 4$.

A.2.4. *Proof of Theorem* 3.1 ($1 \Leftrightarrow 5$). The implication $5 \Rightarrow 1$ follows immediately by choosing M_0 and M to be distinct invariant measures of the filter and applying property 5, which leads to a contradiction.

We prove the converse implication under the assumption that the signal transition matrix P is aperiodic. Choose M_0 arbitrarily, and note that $b(\mathsf{M}_0\Pi^n) = b(\mathsf{M}_0) P^n \Rightarrow \lambda$. It follows from [7] that the sequence $(\mathsf{M}_0\Pi^n)_{n\in\mathbb{N}}$ is tight. It therefore suffices to prove that any convergent subsequence converges to the unique invariant measure of the filter M. Let $n(k) \nearrow \infty$ be a subsequence such that $\mathsf{M}_0\Pi^{n(k)} \Rightarrow \mathsf{M}_\infty$, and let $f \in \mathcal{C}_b(\mathcal{P}(E))$ be convex. By Lemma A.5, we have

$$\mathsf{m}_{b(\mathsf{M}_0)P^{n(k)-m}}\Pi^m f \le \mathsf{M}_0\Pi^{n(k)} f \le \tilde{\mathsf{m}}_{b(\mathsf{M}_0)P^{n(k)-m}}\Pi^m f \qquad \text{for all } m \le n(k).$$

In particular, letting $k \to \infty$ and using the Feller property, we have

$$\mathsf{m}_\lambda \Pi^m f \le \mathsf{M}_\infty f \le \tilde{\mathsf{m}}_\lambda \Pi^m f \qquad \text{for all } m \ge 0.$$

But letting $m \to \infty$ and using Lemma A.6, we find that $\mathsf{M}^{\min} \prec \mathsf{M}_\infty \prec \mathsf{M}^{\max}$. As the invariant measure M is presumed to be unique, we have $\mathsf{M}^{\min} = \mathsf{M}^{\max} = \mathsf{M}$ by the implication $1 \Rightarrow 2$. Therefore, we find that $\mathsf{M}_\infty = \mathsf{M}$ by Lemma A.4. This completes the proof of the implication $1 \Rightarrow 5$.

A.2.5. *Proof of Lemma* 3.4. First, we note that $\mathbf{E}((\pi_0^{\max})_i|\mathcal{F}^Y_{-\infty,0}) = (\pi_0^{\min})_i$. Therefore, $\mathbf{P}((\pi_0^{\min})_i = 0 \text{ and } (\pi_0^{\max})_i > 0) = 0$ for every $i \in E$. In particular, this implies that we have $\pi_0^{\max} \ll \pi_0^{\min}$ with unit probability under $\mathbf{P}$. Now note that $\pi_k^{\max} = \pi_k^\mu|_{\mu=\pi_0^{\max}}$ and $\pi_k^{\min} = \pi_k^\mu|_{\mu=\pi_0^{\min}}$ by Lemma 2.3. Therefore,

$$\begin{aligned}
&\mathbf{P}\Big(\lim_{k\to\infty} \|\pi_k^{\max} - \pi_k^{\min}\| \text{ exists}\big|\mathcal{G}_{-\infty,0}\Big) \\
&\quad = \mathbf{P}\Big(\lim_{k\to\infty} \|\pi_k^\mu - \pi_k^\nu\| \text{ exists}\big|\mathcal{G}_{-\infty,0}\Big)\Big|_{\mu=\pi_0^{\max},\nu=\pi_0^{\min}} \\
&\quad = \mathbf{P}^\mu\Big(\lim_{k\to\infty} \|\pi_k^\mu - \pi_k^\nu\| \text{ exists}\Big)\Big|_{\mu=\pi_0^{\max},\nu=\pi_0^{\min}} = 1 \qquad \mathbf{P}\text{-a.s.},
\end{aligned}$$

where we have used the fact that $\pi_0^{\min}$ and $\pi_0^{\max}$ are $\mathcal{G}_{-\infty,0}$-measurable in the first step, the Markov property in the second step, and Lemma A.7 in the third step. The result now follows by taking the expectation with respect to $\mathbf{P}$.

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** 643–669 (electronic). MR2086177

[2] BLACKWELL, D. (1957). The entropy of functions of finite-state Markov chains. In *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes Held at Liblice Near Prague from November 28 to 30, 1956* 13–20. Publishing House of the Czechoslovak Academy of Sciences, Prague. MR0100297

[3] BLACKWELL, D. and DUBINS, L. (1962). Merging of opinions with increasing information. *Ann. Math. Statist.* **33** 882–886. MR0149577

[4] BUDHIRAJA, A. (2003). Asymptotic stability, ergodicity and other asymptotic properties of the nonlinear filter. *Ann. Inst. H. Poincaré Probab. Statist.* **39** 919–941. MR2010391

[5] CHIGANSKY, P., LIPTSER, R. and VAN HANDEL, R. (2010). Intrinsic methods in filter stability. In *The Oxford University Handbook of Nonlinear Filtering* (D. Crisan and B. Rozovsky, eds.). Oxford Univ. Press. To appear.

[6] DI MASI, G. B. and STETTNER, Ł. (2005). Ergodicity of hidden Markov models. *Math. Control Signals Systems* **17** 269–296. MR2177319

[7] JAKUBOWSKI, A. (1988). Tightness criteria for random measures with application to the principle of conditioning in Hilbert spaces. *Probab. Math. Statist.* **9** 95–114. MR945679

[8] KAIJSER, T. (1975). A limit theorem for partially observed Markov chains. *Ann. Probab.* **3** 677–696. MR0383536

[9] KAIJSER, T. (2009). On Markov chains induced by partitioned transition probability matrices. Preprint. Available at http://arxiv.org/abs/0907.4502.

[10] KOCHMAN, F. and REEDS, J. (2006). A simple proof of Kaijser's unique ergodicity result for hidden Markov α-chains. *Ann. Appl. Probab.* **16** 1805–1815. MR2288705

[11] KUNITA, H. (1971). Asymptotic behavior of the nonlinear filtering errors of Markov processes. *J. Multivariate Anal.* **1** 365–393. MR0301812

[12] STETTNER, Ł. (1989). On invariant measures of filtering processes. In *Stochastic Differential Systems* (*Bad Honnef*, *1988*). *Lecture Notes in Control and Inform. Sci.* **126** 279–292. Springer, Berlin. MR1236074

[13] VAN HANDEL, R. (2009). Observability and nonlinear filtering. *Probab. Theory Related Fields* **145** 35–74. MR2520121

[14] VAN HANDEL, R. (2009). The stability of conditional Markov processes and Markov chains in random environments. *Ann. Probab.* **37** 1876–1925. MR2561436

[15] VAN HANDEL, R. (2009). Uniform observability of hidden Markov models and filter stability for unstable signals. *Ann. Appl. Probab.* **19** 1172–1199. MR2537203

DEPARTMENT OF STATISTICS
HEBREW UNIVERSITY
MOUNT SCOPUS
JERUSALEM 91905
ISRAEL
E-MAIL: pchiga@mscc.huji.ac.il

DEPARTMENT OF OPERATIONS RESEARCH
AND FINANCIAL ENGINEERING
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544
USA
E-MAIL: rvan@princeton.edu