ERGODICITY AND STABILITY OF THE CONDITIONAL DISTRIBUTIONS OF NONDEGENERATE MARKOV CHAINS[1]

By Xin Thomson Tong and Ramon van Handel

Princeton University

We consider a bivariate stationary Markov chain $(X_n, Y_n)_{n \geq 0}$ in a Polish state space, where only the process $(Y_n)_{n \geq 0}$ is presumed to be observable. The goal of this paper is to investigate the ergodic theory and stability properties of the measure-valued process $(\Pi_n)_{n \geq 0}$, where Π_n is the conditional distribution of X_n given $Y_0, \ldots, Y_n$. We show that the ergodic and stability properties of $(\Pi_n)_{n \geq 0}$ are inherited from the ergodicity of the unobserved process $(X_n)_{n \geq 0}$ provided that the Markov chain $(X_n, Y_n)_{n \geq 0}$ is nondegenerate, that is, its transition kernel is equivalent to the product of independent transition kernels. Our main results generalize, subsume and in some cases correct previous results on the ergodic theory of nonlinear filters.

1. Introduction. In this paper we will consider a bivariate Markov chain $(X_n, Y_n)_{n \geq 0}$ taking values in a Polish state space. Only the process $(Y_n)_{n \geq 0}$ is presumed to be directly observable to us, and we aim to estimate the state X_n of the unobserved process given the observed data $Y_0, \ldots, Y_n$ to date. This is the quintessential setup in problems with partial information, and models of this type can therefore be found in a wide range of applications [6].

We will be concerned, in particular, with the ergodic theory and stability properties of the measure-valued process $(\Pi_n)_{n \geq 0}$ defined by the conditional distributions $\Pi_n = \mathbf{P}(X_n \in \cdot \,|\, Y_0, \ldots, Y_n)$, which is called the nonlinear filter. It is not difficult to show that, in general, the processes $(\Pi_n, Y_n)_{n \geq 0}$ as well as $(\Pi_n, X_n, Y_n)_{n \geq 0}$ are themselves Markovian, and a typical question that we will aim to answer is whether ergodicity of the underlying model $(X_n, Y_n)_{n \geq 0}$ implies ergodicity of the extended Markov chain $(\Pi_n, X_n, Y_n)_{n \geq 0}$ in a suitable sense. Questions of this type date back at least to the work of Blackwell [2] and Kunita [14]. Beside the intrinsic probabilistic interest in the development of a *conditional* ergodic theory of Markov chains, ergodicity of the filter has substantial practical relevance to understanding the performance of nonlinear filtering and its numerical approximations over a long time horizon; cf. [5, 14, 21], and see [8, 20] for further references.

Much of the literature on the topic of this paper is concerned with the setting of a classical hidden Markov model whose dependence structure is illustrated in Figure 1(a); here the unobserved process $(X_n)_{n \geq 0}$ is assumed to be itself Markovian
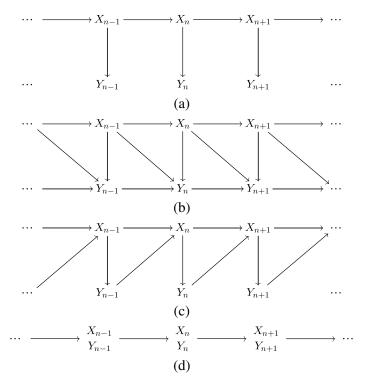
FIG. 1. *Dependence structure of* (a) *a classical hidden Markov model* [6]; (b) *a generalized hidden Markov model* [10, 11]; (c) *a hidden Markov model with correlated noise* [3]; (d) *a general Markov model.*

and the observations $(Y_n)_{n \geq 0}$ are conditionally independent.[2] In this special case $(\Pi_n)_{n \geq 0}$ is also Markovian, and two basic questions have been considered.

(1) Does $(\Pi_n)_{n \geq 0}$ possess a unique invariant measure, assuming $(X_n)_{n \geq 0}$ does?

For the second question, let $\tilde{\mathbf{P}}$ and $\mathbf{P}$ be the laws of the Markov chain $(X_n, Y_n)_{n \geq 0}$ with initial laws $\tilde{\mathbf{P}}(X_0 \in \cdot) \ll \mathbf{P}(X_0 \in \cdot)$, and let $\tilde{\Pi}_n = \tilde{\mathbf{P}}(X_n \in \cdot | Y_0, \ldots, Y_n)$.

(2) Is $(\Pi_n)_{n \geq 0}$ asymptotically stable in the sense that $|\tilde{\Pi}_n(f) - \Pi_n(f)| \xrightarrow{n \to \infty} 0$ in $\tilde{\mathbf{P}}$-probability for every bounded continuous function f?

These and related questions were studied in great generality by Kunita [14, 15], Stettner [19], and Ocone and Pardoux [17] (see [1, 4, 8, 20] for further references). Kunita and Stettner state that the answer to the first question is affirmative provided

---

[2]The continuous time version of this model, known as a Markov additive process, is also widely studied in the literature in various special cases (such as white noise or counting observations; see [25] for a unified view). We have restricted ourselves in this paper to discrete time models for simplicity. All our results are easily extended to the continuous time setting as in [20], Section 6.

that the stationary process $(X_n)_{n \in \mathbb{Z}}$ is purely nondeterministic, that is,

$$\bigcap_{n \geq 0} \mathcal{F}^X_{-n} \text{ is } \mathbf{P}\text{-trivial,}$$

where $\mathbf{P}$ is the stationary law of the two-sided process $(X_n, Y_n)_{n \in \mathbb{Z}}$ and $\mathcal{F}^X_n = \sigma\{X_k : -\infty < k \leq n\}$. Ocone and Pardoux state that the answer to the second question is affirmative under the same assumption. Unfortunately, the proofs of these results contain a serious error, as was pointed out by Baxendale, Chigansky and Liptser [1]. Indeed, the crucial step in the proofs is the identity

$$\bigcap_{n \geq 0} \mathcal{F}^Y_0 \vee \mathcal{F}^X_{-n} \stackrel{?}{=} \mathcal{F}^Y_0, \qquad \mathbf{P}\text{-a.s.,}$$

where $\mathcal{F}^Y_0 = \sigma\{Y_k : -\infty < k \leq 0\}$. It is tempting to exchange the order of the intersection $\cap$ and supremum $\vee$ of σ-fields, which would allow us to conclude this identity from the assumption that $(X_n)_{n \in \mathbb{Z}}$ is purely nondeterministic. But such an exchange cannot be taken for granted (see [7], page 30) and requires proof. In the filtering setting, various counterexamples given in [1, 22] show that the answers to the above questions may indeed be negative even when $(X_n)_{n \in \mathbb{Z}}$ is purely nondeterministic, in contradiction with the conclusions of [14, 15, 17, 19].

Before we proceed, let us briefly recall a simple counterexample from [1, 22] that will be helpful in understanding the problems addressed in this paper.

EXAMPLE 1.1. Let $(\xi_n)_{n \in \mathbb{Z}}$ be an i.i.d. sequence of random variables taking the values $\{0, 1\}$ with equal probability under $\mathbf{P}$, and define

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

Then $(X_n)_{n \in \mathbb{Z}}$ is an ergodic Markov chain in $\{00, 01, 10, 11\}$ that is purely nondeterministic by the Kolmogorov zero–one law, and $(X_n, Y_n)_{n \geq 0}$ is a hidden Markov model as in Figure 1(a). But clearly $\xi_n = (\xi_1 + Y_1 + \cdots + Y_{n-1}) \bmod 2$, so that

$$\tilde{\Pi}_n(f) = f(X_n), \qquad \Pi_n(f) = \frac{f(X_n) + f(11 - X_n)}{2}, \qquad \tilde{\mathbf{P}}\text{-a.s.}$$

where we defined $\tilde{\mathbf{P}}(\cdot) = \mathbf{P}(\cdot \mid X_0 = 00)$. Thus the filter is not asymptotically stable, and one may similarly establish that it admits distinct invariant measures.

One feature of the model of Example 1.1 is that it possesses degenerate observations in the sense that Y_n is a function of X_n without any additional noise. The phenomenon illustrated here turns out to disappear when some independent noise is added to the observations, for example, $Y_n = |\xi_{n+1} - \xi_n| + \eta_n$ where $(\eta_n)_{n \in \mathbb{Z}}$ is an i.i.d. sequence such that the law of η_0 has a nowhere vanishing density. In [20], one of the authors developed this idea to establish ergodicity and stability properties of the nonlinear filter under very general assumptions. To this end, let $(X_n, Y_n)_{n \in \mathbb{Z}}$ be a stationary hidden Markov model under $\mathbf{P}$, and assume that:

(1) $(X_n)_{n\in\mathbb{Z}}$ is absolutely regular: $\mathbf{E}(\|\mathbf{P}(X_n \in \cdot | X_0) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}}) \to 0$.

(2) The observations are nondegenerate: $\mathbf{P}(Y_n \in A | X_n) = \int_A g(X_n, y)\varphi(dy)$ for some strictly positive density $g(x, y) > 0$ and reference measure φ.

Then the above exchange of intersection and supremum of σ-fields is permitted, and the filter is stable [20] and uniquely ergodic [22]. Intuitively, nondegeneracy (which formalizes the notion of "noisy" observations) rules out the singular observation structure that causes the exchange of intersection and supremum to fail in Example 1.1. However, this intuition should not be taken too literally, as a more difficult example in [22] shows that the result may still fail if absolute regularity is replaced with the weaker purely nondeterministic assumption. Therefore, the assumptions in [14, 15, 17, 19] (which implicitly assume nondegeneracy, though this is not used in the proofs) are genuinely too weak to yield the desired results.

The results discussed above all assume the classical hidden Markov model setting illustrated in Figure 1(a). Such models are quite flexible and appear in a wide array of applications [6]. Nonetheless, there are many applications in which the need arises for more general classes of partially observed Markov models. For example, two common generalizations of the classical hidden Markov model are illustrated in Figure 1(b) and (c). The model of Figure 1(b) is a generalized hidden Markov model [10] or an autoregressive process with Markov regime [11]. This model is similar to a hidden Markov model in that the dynamics of $(X_n)_{n\geq 0}$ do not depend on the observations $(Y_n)_{n\geq 0}$; however, here the observations are not conditionally independent but may possess their own dynamics. Such models are common in financial mathematics, where $(Y_n)_{n\geq 0}$ might represent a sequence of investment returns while $(X_n)_{n\geq 0}$ models the state of the underlying economy. On the other hand, in the model of Figure 1(c) there is feedback from the observations to the dynamics of the unobserved process $(X_n)_{n\geq 0}$. Such models arise when the noise driving the unobserved process and the observation noise are correlated.

In these more general models, the process $(\Pi_n)_{n\geq 0}$ is no longer Markovian, but the pair $(\Pi_n, Y_n)_{n\geq 0}$ is still Markov. It is therefore natural, and of significant interest for applications, to investigate the ergodicity of $(\Pi_n, Y_n)_{n\geq 0}$ and the asymptotic stability of $(\Pi_n)_{n\geq 0}$ in a more general setting. It has been shown by Di Masi and Stettner [10] for the model of Figure 1(b), and by Budhiraja [3] for the model of Figure 1(c), that these problems can be reduced to establishing the validity of the exchange of intersection and supremum of σ-fields along the lines of the earlier approach for classical hidden Markov models in [14, 15, 17, 19]. The generalization of the positive results in [20] is far from straightforward, however.

To illustrate one of the complications that arises in generalized models, let us consider the setting of Budhiraja [3]. Budhiraja considers a model of the form

$$X_n = f(X_{n-1}, Y_{n-1}, \xi_n), \qquad Y_n = h(X_n) + \eta_n,$$

where $(\xi_n)_{n\geq 1}$ and $(\eta_n)_{n\geq 0}$ are independent i.i.d. sequences. It is assumed that f, h are continuous functions and that η_0 possesses a bounded and continuous density with respect to some reference measure φ. This is evidently a hidden Markov

model with correlated noise of the type illustrated in Figure 1(c). The main result in [3] states that if this model admits a unique stationary law $\mathbf{P}$ and if $(X_n)_{n \in \mathbb{Z}}$ is purely nondeterministic, then $(\Pi_n, Y_n)_{n \geq 0}$ possesses a unique invariant measure. Budhiraja's proof contains the same gap as in [14, 15, 19]; indeed, the result is clearly erroneous in light of Example 1.1. Nonetheless, it seems reasonable to guess that if we assume nondegeneracy of the observations (i.e., that the density of η_0 is strictly positive) and absolute regularity of the unobserved process, then the result will hold as in [20]. Even this, however, turns out to be false.

EXAMPLE 1.2. Define the $\{00, 01, 10, 11\}$-valued process $(X_n)_{n \in \mathbb{Z}}$ and real-valued process $(Y_n)_{n \in \mathbb{Z}}$ such that X_0 is uniformly distributed in $\{00, 01, 10, 11\}$,

$$(X_n^1, X_n^2) = (X_{n-1}^2, |X_{n-1}^2 - I_{[0,\infty[}(Y_{n-1})|), \qquad Y_n = \eta_n,$$

where $(\eta_n)_{n \in \mathbb{Z}}$ are i.i.d. $N(0, 1)$-distributed random variables. Then the process $(X_n, I_{[0,\infty[}(Y_{n-1}))_{n \in \mathbb{Z}}$ has the same law as the classical hidden Markov model of Example 1.1, so stability and unique ergodicity of the filter must fail.

Even though the observations are ostensibly nondegenerate in this example, the feedback from the observations affects the dynamics of the unobserved process in a singular fashion that recreates the problems of Example 1.1. We thus need at least a different notion of nondegeneracy in order to rule out such phenomena.

The goal of this paper is to develop a general ergodic and stability theory for nonlinear filters that subsumes all of the models discussed above. Indeed, we do not impose any structural assumptions other than that $(X_n, Y_n)_{n \geq 0}$ is a Markov chain that possesses a stationary law $\mathbf{P}$ [as is illustrated in Figure 1(d)]. The main assumptions of this paper generalize those of [20]; we assume that the model is:

(1) absolutely regular: $\mathbf{E}(\|\mathbf{P}((X_n, Y_n) \in \cdot | X_0, Y_0) - \mathbf{P}((X_n, Y_n) \in \cdot)\|_{\mathrm{TV}}) \to 0$;
(2) nondegenerate: there exist kernels P_0, Q and a density $g(x', y', x, y) > 0$ so that $\mathbf{P}((X_{n+1}, Y_{n+1}) \in A | X_n, Y_n) = \int_A g(X_n, Y_n, x, y) P_0(X_n, dx) Q(Y_n, dy)$.

The latter assumption states that the dynamics of the observed and unobserved processes can be made independent (on finite time intervals) by an equivalent change of measure. It is easily seen that the notion of nondegenerate observations for the classical hidden Markov model is a special case of this assumption; on the other hand, the present assumption also rules out the phenomenon observed in Example 1.2. This general nondegeneracy property appears to be precisely the right assumption required to generalize the results of [20], and seems very natural in view of Examples 1.1 and 1.2. The absolute regularity assumption on $(X_n, Y_n)_{n \in \mathbb{Z}}$ can in fact be weakened somewhat; see Sections 2.4 and 2.5 for a precise statement.

With the above assumptions in place, we will show that Kunita's exchange of intersection and supremum of σ-fields is permitted in our setting, and we can consequently develop general asymptotic stability and unique ergodicity results. The intuition behind the proofs is similar in spirit to the classical hidden Markov model

setting in [20, 22], and we refer to those papers for a discussion of the basic ideas. Nonetheless, to our surprise, key parts of the proofs in [20] break down completely in the generalized setting of this paper and almost all arguments in [20] require substantial modification, as we can no longer exploit many simplifying properties that hold trivially in classical hidden Markov models. The proofs in the present paper rely on the ergodic properties of nondegenerate Markov chains that are developed in Section 3 below. Though this paper is almost entirely self-contained, the reader may find it helpful to familiarize herself first with the simpler setting of [20].

This paper is organized as follows. Section 2 introduces the general model used throughout the paper and states our main results. We also give useful sufficient conditions for the models in Figure 1(a)–(d). Section 3 develops the ergodic properties of nondegenerate Markov chains that play a central role in our proofs. Sections 4–7 are devoted to the proofs of our main results. Appendices A and B collect auxiliary results and a notation list that is used throughout the paper.

2. Preliminaries and main results.

2.1. *The canonical setup.* Throughout this paper we consider the bivariate stochastic process $(X_n, Y_n)_{n\in\mathbb{Z}}$, where X_n takes values in the Polish space E and Y_n takes values in the Polish space F. We realize this process 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 of Ω, and define

$$\mathcal{F}_I^X = \sigma\{X_k : k \in I\}, \qquad \mathcal{F}_I^Y = \sigma\{Y_k : k \in I\}, \qquad \mathcal{F}_I = \mathcal{F}_I^X \vee \mathcal{F}_I^Y$$

for $I \subset \mathbb{Z}$. For simplicity of notation, we define the natural filtrations

$$\mathcal{F}_n^X = \mathcal{F}_{]-\infty,n]}^X, \qquad \mathcal{F}_n^Y = \mathcal{F}_{]-\infty,n]}^Y, \qquad \mathcal{F}_n = \mathcal{F}_{]-\infty,n]} \qquad (n \in \mathbb{Z})$$

and the σ-fields

$$\mathcal{F}^X = \mathcal{F}_{\mathbb{Z}}^X, \qquad \mathcal{F}^Y = \mathcal{F}_{\mathbb{Z}}^Y, \qquad \mathcal{F}_+^X = \mathcal{F}_{[0,\infty[}^X, \qquad \mathcal{F}_+^Y = \mathcal{F}_{[0,\infty[}^Y.$$

Finally, we denote by Y the $F^{\mathbb{Z}}$-valued random variable $(Y_k)_{k\in\mathbb{Z}}$, and the canonical shift $\Theta : \Omega \to \Omega$ is defined as $\Theta(x, y)(m) = (x(m + 1), y(m + 1))$.

For any Polish space Z, we denote by $\mathcal{B}(Z)$ its Borel σ-field and by $\mathcal{P}(Z)$ the space of all probability measures on Z endowed with the weak convergence topology [thus $\mathcal{P}(Z)$ is again Polish]. Let us recall that any probability kernel $\rho : Z \times \mathcal{B}(Z') \to [0, 1]$ may be equivalently viewed as a $\mathcal{P}(Z')$-valued random variable $z \mapsto \rho(z, \cdot)$ on $(Z, \mathcal{B}(Z))$. For notational convenience, we will implicitly identify probability kernels and random probability measures in the sequel.

2.2. *The model.*   The basic model of this paper is defined by a Markov transition kernel $P : E \times F \times \mathcal{B}(E \times F) \to [0, 1]$ and a P-invariant probability measure π on $(E \times F, \mathcal{B}(E \times F))$, which we presume to be fixed throughout the paper. We now define the probability measure $\mathbf{P}$ on $(\Omega, \mathcal{F})$ such that, under $\mathbf{P}$, the process $(X_n, Y_n)_{n \in \mathbb{Z}}$ is the stationary Markov chain with transition kernel P and stationary distribution π. We interpret Y_n to be the *observable component* of the model, while X_n is the *unobservable component*.

As $(X_n, Y_n)_{n \in \mathbb{Z}}$ is a stationary Markov chain under $\mathbf{P}$, the reverse time process $(X_{-n}, Y_{-n})_{n \in \mathbb{Z}}$ is again a stationary Markov chain. We fix throughout the paper a version $P' : E \times F \times \mathcal{B}(E \times F) \to [0, 1]$ of the regular conditional probability $\mathbf{P}((X_{-1}, Y_{-1}) \in \cdot \,|X_0, Y_0)$. Thus, by construction, the process $(X_{-n}, Y_{-n})_{n \in \mathbb{Z}}$ is a stationary Markov chain with transition kernel P' and invariant measure π.

In addition to the probability measure $\mathbf{P}$, we introduce the probability kernel $\mathbf{P}^{\cdot} : E \times F \times \mathcal{F} \to [0, 1]$ with the following properties: under $\mathbf{P}^{z,w}$,

(1) $(X_n, Y_n)_{n \geq 0}$ is Markov with transition kernel P and initial measure $\delta_z \otimes \delta_w$;
(2) $(X_{-n}, Y_{-n})_{n \geq 0}$ is Markov with transition kernel P' and initial measure $\delta_z \otimes \delta_w$;
(3) $(X_n, Y_n)_{n \geq 0}$ and $(X_{-n}, Y_{-n})_{n \geq 0}$ are independent.

Clearly $\mathbf{P}^{z,w}$ is a version of the regular conditional probability $\mathbf{P}(\cdot \,|X_0, Y_0)$. Finally, for any probability measure ν on $(E \times F, \mathcal{B}(E \times F))$, we define

$$\mathbf{P}^{\nu}(A) = \int I_A(x, y) \mathbf{P}^{z,w}(dx, dy) \nu(dz, dw) \qquad \text{for all } A \in \mathcal{F}.$$

Note, in particular, that $\mathbf{P}^{\pi}$ coincides with $\mathbf{P}$ by construction.

2.3. *The nonlinear filter.*   As X_n is not directly observable, we are interested in the conditional distribution of X_n given the history of observations to date $Y_0, \ldots, Y_n$. To this end, we define for every probability measure μ on $E \times F$ and $n \geq 0$ the *nonlinear filter* $\Pi_n^{\mu} : \Omega^Y \times \mathcal{B}(E) \to [0, 1]$ to be a version of the regular conditional probability $\mathbf{P}^{\mu}(X_n \in \cdot \,|\mathcal{F}_{[0,n]}^Y)$. The nonlinear filter is the central object of interest throughout this paper.

We now state some basic properties of the nonlinear filter. The first property establishes that the filter can be computed recursively.

LEMMA 2.1.   *There is a measurable map $U : \mathcal{P}(E) \times F \times F \to \mathcal{P}(E)$ such that $\Pi_n^{\mu} = U(\Pi_{n-1}^{\mu}, Y_{n-1}, Y_n)$ $\mathbf{P}^{\mu}$-a.s. for every $n \geq 1$ and $\mu \in \mathcal{P}(E \times F)$.*

REMARK 2.2.   In the proof of our main results, it will be convenient to assume that the identity $\Pi_n^{\mu} = U(\Pi_{n-1}^{\mu}, Y_{n-1}, Y_n)$ holds everywhere on Ω^Y and not just $\mathbf{P}^{\mu}$-a.s. This corresponds to the choice of a particular version of the nonlinear filter. However, as none of our results will depend on the choice of version of the filter, there is clearly no loss of generality in fixing such a convenient version for the purposes of our proofs, as we will do in Section 5.

We now consider $(\Pi_n^\mu)_{n\geq 0}$ as a $\mathcal{P}(E)$-valued stochastic process. The second property establishes that this measure-valued process inherits certain Markovian properties from the underlying model $(X_n, Y_n)_{n\geq 0}$.

LEMMA 2.3. *There exist Markov transition kernels* Γ *on* $\mathcal{P}(E) \times F$ *and* Λ *on* $\mathcal{P}(E) \times E \times F$ *such that the following hold*: *for every* $\mu \in \mathcal{P}(E \times F)$,

(1) $(\Pi_n^\mu, Y_n)_{n\geq 0}$ *is a Markov chain under* $\mathbf{P}^\mu$ *with transition kernel* Γ; *and*
(2) $(\Pi_n^\mu, X_n, Y_n)_{n\geq 0}$ *is a Markov chain under* $\mathbf{P}^\mu$ *with transition kernel* Λ.

For any $\mathsf{m} \in \mathcal{P}(\mathcal{P}(E) \times F)$, define the *barycenter* $b\mathsf{m} \in \mathcal{P}(E \times F)$ as

$$b\mathsf{m}(A \times B) = \int \nu(A) I_B(w)\mathsf{m}(d\nu, dw).$$

We finally state some properties of Γ- and Λ-invariant measures.

LEMMA 2.4. *For any* Γ-*invariant probability measure* $\mathsf{m} \in \mathcal{P}(\mathcal{P}(E) \times F)$, *the barycenter* $b\mathsf{m}$ *is a* P-*invariant probability measure. Conversely, there exists at least one* Γ-*invariant probability measure with barycenter* π.

Similarly, for any Λ-*invariant probability measure* $\mathsf{M} \in \mathcal{P}(\mathcal{P}(E) \times E \times F)$, *the marginal* $\mathsf{M}(\mathcal{P}(E) \times \cdot)$ *is a* P-*invariant probability measure. Conversely, there exists at least one* Λ-*invariant probability measure with marginal* π.

In general, there may be multiple Γ-invariant measures with barycenter π, etc. Our main results will establish uniqueness under suitable assumptions.

REMARK 2.5. For the purposes of this paper it suffices to establish the above results for the case where Assumption 2.8 below is assumed to hold. In this setting, these results will be proved in Sections 6.1 and 7.1. In fact, the results in this subsection hold very generally as stated without any further assumptions, but the proofs in the general setting are somewhat more abstract. Such generality will not be needed in this paper, and we therefore leave the generalization of the proofs (along the lines of [22], Appendix A.1) to the interested reader.

2.4. *Main results.* We begin by introducing the fundamental model assumptions that are required by our main results. Let us emphasize that we will at no point in the paper automatically assume that any of these assumptions is in force; all assumptions will be imposed explicitly where they are needed. Some useful sufficient conditions will be given in Section 2.5 below.

ASSUMPTION 2.6 (Marginal ergodicity). The following holds:

$$\int \|\mathbf{P}^{z,w}(X_n \in \cdot) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}}\pi(dz, dw) \xrightarrow{n\to\infty} 0.$$

ASSUMPTION 2.7 (Reversed marginal ergodicity). The following holds:

$$\int \|\mathbf{P}^{z,w}(X_{-n} \in \cdot) - \mathbf{P}(X_{-n} \in \cdot)\|_{\mathrm{TV}} \pi(dz, dw) \xrightarrow{n \to \infty} 0.$$

ASSUMPTION 2.8 (Nondegeneracy). There exist transition probability kernels $P_0 : E \times \mathcal{B}(E) \to [0, 1]$ and $Q : F \times \mathcal{B}(F) \to [0, 1]$ such that

$$P(z, w, dz', dw') = g(z, w, z', w') P_0(z, dz') Q(w, dw')$$

for some strictly positive measurable function $g : E \times F \times E \times F \to]0, \infty[$.

We now proceed to state the main results of this paper. Our results address in turn each of the problems discussed in the Introduction: the exchange of intersection and supremum of σ-fields, asymptotic stability of the nonlinear filter and unique ergodicity of the processes $(\Pi_n^\mu, Y_n)_{n \geq 0}$ and $(\Pi_n^\mu, X_n, Y_n)_{n \geq 0}$.

Our first result establishes the validity of Kunita's exchange of intersection and supremum, and its time-reversed cousin, in the generalized setting of this paper.

THEOREM 2.9. *Suppose that Assumptions 2.6–2.8 are in force. Then*

$$\bigcap_{n \geq 0} \mathcal{F}_+^Y \vee \mathcal{F}_{[n,\infty[}^X = \mathcal{F}_+^Y \quad \text{and} \quad \bigcap_{n \geq 0} \mathcal{F}_0^Y \vee \mathcal{F}_{-n}^X = \mathcal{F}_0^Y, \qquad \mathbf{P}\text{-}a.s.$$

Our second result concerns filter stability which can be established in our setting (as in [20]) in a very strong sense: pathwise and in the total variation topology.

THEOREM 2.10. *Suppose that Assumptions 2.6–2.8 are in force. Let μ be a probability measure on $E \times F$ such that $\mu(E \times \cdot) \ll \pi(E \times \cdot)$ and*

$$\mathbf{E}^\mu \big(\|\mathbf{P}^\mu(X_n \in \cdot \,|\, Y_0) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}} \big) \xrightarrow{n \to \infty} 0.$$

Then $\|\Pi_n^\mu - \Pi_n^\pi\|_{\mathrm{TV}} \xrightarrow{n \to \infty} 0$ $\mathbf{P}^\mu$-a.s. [and $\mathbf{P}$-a.s. if $\mu(E \times \cdot) \sim \pi(E \times \cdot)$].

REMARK 2.11. The assumptions of Theorem 2.10 may be more intuitive when phrased in terms of the filtering recursion in Lemma 2.1. Let $\rho : F \times \mathcal{B}(E) \to [0, 1]$ be a probability kernel, and define the random measures $(\Pi_n)_{n \geq 0}$ by the recursion

$$\Pi_0 = \rho(Y_0, \cdot), \qquad \Pi_n = U(\Pi_{n-1}, Y_{n-1}, Y_n).$$

Suppose that the dynamics of $(X_n)_{n \geq 0}$ are such that the random initial law ρ is in the domain of attraction of the stationary distribution π in the sense that

$$\big\|\mathbf{P}^{\rho(w,\cdot) \otimes \delta_w}(X_n \in \cdot) - \mathbf{P}(X_n \in \cdot)\big\|_{\mathrm{TV}} \xrightarrow{n \to \infty} 0 \qquad \text{in } \pi(E \times dw)\text{-probability}.$$

Then $\|\Pi_n - \Pi_n^\pi\|_{\mathrm{TV}} \xrightarrow{n\to\infty} 0$ **P**-a.s. Indeed, this follows immediately from Theorem 2.10 by setting $\mu(dz, dw) = \rho(w, dz)\pi(E \times dw)$. Therefore, we may interpret Theorem 2.10 as follows: the filtering recursion of Lemma 2.1 is asymptotically stable inside the domain of attraction of the stationary distribution.

The result of Theorem 2.10 is easily extended to show $\|\Pi_n^\mu - \Pi_n^\nu\|_{\mathrm{TV}} \xrightarrow{n\to\infty} 0$ **P**$^\gamma$-a.s. whenever all three initial measures μ, ν, γ are in the domain of attraction of the stationary distribution in the above sense, using Corollary 3.6 below.

Our third result concerns uniqueness of the Γ-invariant measure.

THEOREM 2.12. *Suppose that Assumptions 2.6–2.8 are in force. Then there exists a unique Γ-invariant probability measure with barycenter π. In particular, if P has a unique invariant probability measure, then so does Γ.*

Our fourth result concerns uniqueness of the Λ-invariant measure. The situation here is a little more complicated; Assumptions 2.6–2.8 only ensure uniqueness within a restricted class of measures (cf. [15]), while a somewhat stronger variant of Assumption 2.6 yields uniqueness in the class of all probability measures.

THEOREM 2.13. *Suppose that Assumptions 2.6–2.8 hold. Then there exists a unique Λ-invariant probability measure with marginal π on $E \times F$ in the class*

$$\left\{ \mathsf{M} \in \mathcal{P}(\mathcal{P}(E) \times E \times F) : \textit{for every } A \in \mathcal{B}(\mathcal{P}(E)), B \in \mathcal{B}(E), C \in \mathcal{B}(F), \right.$$

$$\left. \mathsf{M}(A \times B \times C) = \int \nu(B) I_{A \times C}(\nu, w) \mathsf{M}(d\nu, dz, dw) \right\}.$$

If, in addition, we have

$$\int \|\mathbf{P}^{z,w}(X_n \in \cdot) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}} \mu(dz, dw) \xrightarrow{n\to\infty} 0$$

for every probability measure μ on $E \times F$ such that $\mu(E \times \cdot) = \pi(E \times \cdot)$, then there exists a unique Λ-invariant probability measure with marginal π among all probability measures in $\mathcal{P}(\mathcal{P}(E) \times E \times F)$. If we assume even further that P has a unique invariant probability measure, then so does Λ.

The following sections are devoted to the proofs of these results: Theorems 2.9, 2.10, 2.12 and 2.13 are proved in Sections 4, 5, 6 and 7, respectively.

2.5. *Sufficient conditions.* Our main results rely on the fundamental Assumptions 2.6–2.8. In most applications, the form of the transition kernel P is explicitly (or semi-explicitly) given. Existence and uniqueness of an invariant measure π and the ergodicity Assumption 2.6 can often be verified in terms of P only (cf. [16]),

while the nondegeneracy Assumption 2.8 can be read off directly from the explicit form of P. On the other hand, explicit expressions for the invariant measure π or the reversed transition kernel P' are often not available, so that Assumption 2.7 may be difficult to verify directly. The goal of this section is to provide sufficient conditions for our main results that are easily verified in practice.

2.5.1. *General sufficient conditions.*   Our main sufficient condition is absolute regularity (cf. [23]), of the process $(X_n, Y_n)_{n \in \mathbb{Z}}$, which was the assumption stated in the Introduction. This is slightly stronger than Assumptions 2.6 and 2.7, but has the benefit that it is automatically time-reversible and therefore easily verifiable.

LEMMA 2.14.   *Suppose that $(X_n, Y_n)_{n \in \mathbb{Z}}$ is absolutely regular,*

$$\int \left\| \mathbf{P}^{z,w}\big((X_n, Y_n) \in \cdot\big) - \pi \right\|_{\mathrm{TV}} \pi(dz, dw) \xrightarrow{n \to \infty} 0.$$

Then both Assumptions 2.6 and 2.7 hold true.

PROOF.   Absolute regularity trivially yields Assumption 2.6. On the other hand, the absolute regularity property of a stationary Markov chain is invariant under time reversal by [20], Proposition 4.4, so that Assumption 2.7 follows. $\square$

Similarly, the convergence assumption in Theorem 2.10 also admits a slightly stronger but potentially more easily verified counterpart.

LEMMA 2.15.   *Suppose that Assumption 2.6 holds. Let μ be a probability measure on $E \times F$ such that $\|\mathbf{P}^\mu((X_n, Y_n) \in \cdot) - \pi\|_{\mathrm{TV}} \to 0$ as $n \to \infty$. Then*

$$\mathbf{E}^\mu\big(\|\mathbf{P}^\mu(X_n \in \cdot \,|\, Y_0) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}}\big) \xrightarrow{n \to \infty} 0.$$

PROOF.   Define the quantity

$$\Delta_k(x, y) = \|\mathbf{P}^{x,y}(X_k \in \cdot) - \mathbf{P}(X_k \in \cdot)\|_{\mathrm{TV}}.$$

By the stationarity of $\mathbf{P}$, the Markov property and $\|\Delta_k - 1\|_\infty \leq 1$, we can estimate

$$
\begin{aligned}
\mathbf{E}^\mu\big(&\|\mathbf{P}^\mu(X_{n+k} \in \cdot \,|\, Y_0) - \mathbf{P}(X_{n+k} \in \cdot)\|_{\mathrm{TV}}\big) \\
&\leq \mathbf{E}^\mu\big(\|\mathbf{P}^{X_n, Y_n}(X_k \in \cdot) - \mathbf{P}(X_k \in \cdot)\|_{\mathrm{TV}}\big) \\
&= \mathbf{E}\big(\Delta_k(X_n, Y_n)\big) + \big\{\mathbf{E}^\mu\big(\Delta_k(X_n, Y_n) - 1\big) - \mathbf{E}\big(\Delta_k(X_n, Y_n) - 1\big)\big\} \\
&\leq \mathbf{E}\big(\|\mathbf{P}^{X_0, Y_0}(X_k \in \cdot) - \mathbf{P}(X_k \in \cdot)\|_{\mathrm{TV}}\big) + \|\mathbf{P}^\mu\big((X_n, Y_n) \in \cdot\big) - \pi\|_{\mathrm{TV}}.
\end{aligned}
$$

This expression converges to zero as $k, n \to \infty$ by our assumptions. $\square$

2.5.2. *Generalized hidden Markov models.* We now consider the special case where the underlying model $(X_n, Y_n)_{n \in \mathbb{Z}}$ is a generalized hidden Markov model, whose dependence structure is illustrated in Figure 1(b). Under Assumption 2.8, this dependence structure is enforced by the additional requirement that

$$\int g(z, w, z', w') Q(w, dw') = 1 \qquad \text{for all } w \in F, z, z' \in E.$$

This implies that $(X_n)_{n \in \mathbb{Z}}$ is itself Markovian under $\mathbf{P}$ with transition kernel P_0, and the probability measure $\pi_0 = \pi(\cdot \times F)$ must then be P_0-invariant. In this setting, it suffices to consider the ergodic properties of the unobserved process, provided that the reference transition kernel $Q(w, dw')$ does not depend on w.

LEMMA 2.16. *Suppose that Assumption 2.8 holds with $Q(w, dw') = \varphi(dw')$ for some probability measure φ on F, and that $(X_n, Y_n)_{n \in \mathbb{Z}}$ is a generalized hidden Markov model in the above sense. If $(X_n)_{n \in \mathbb{Z}}$ is absolutely regular*

$$\int \|P_0^n(z, \cdot) - \pi_0\|_{\mathrm{TV}} \pi_0(dz) \xrightarrow{n \to \infty} 0,$$

then both Assumptions 2.6 and 2.7 hold true.

PROOF. We reduce to the case of Lemma 2.14. A stationary Markov chain is absolutely regular if and only if for almost every pair of initial conditions, there is a finite time n at which the laws of the chain are not mutually singular (e.g., this is a special case of Theorem 4.1 below). Therefore, our assumption implies that for $\pi_0 \otimes \pi_0$-a.e. (z, z'), there is an $n \geq 0$ such that $P_0^n(z, \cdot)$ and $P_0^n(z', \cdot)$ are not mutually singular. But as $Q(w, dw') = \varphi(dw')$ and by Assumption 2.8, we have $P^n(z, w, \cdot) \sim P_0^n(z, \cdot) \otimes \varphi$ and $P^n(z', w', \cdot) \sim P_0^n(z', \cdot) \otimes \varphi$ for every z, w, z', w'. It follows that for $\pi \otimes \pi$-a.e. $((z, w), (z', w'))$ there is an $n \geq 0$ such that $P^n(z, w, \cdot)$ and $P^n(z', w', \cdot)$ are not mutually singular. We have therefore shown that the absolutely regularity assumption of Lemma 2.14 holds. □

REMARK 2.17. By the generalized hidden Markov structure $\mathbf{P}^{z,w}(X_n \in \cdot) = P_0^n(z, \cdot)$ is independent of w, so that Assumption 2.6 follows immediately from the absolute regularity of $(X_n)_{n \in \mathbb{Z}}$. Unfortunately, the generalized hidden Markov property is not invariant under time reversal, so this argument does not guarantee that Assumption 2.7 holds. The additional assumption that $Q(w, dw') = \varphi(dw')$ allows us to circumvent this problem by reducing to the case of Lemma 2.14.

We also have a counterpart of Lemma 2.15.

LEMMA 2.18. *Suppose the assumptions of Lemma 2.16 hold. Let μ be a probability measure on $E \times F$ so that $\|\mu(\cdot \times F) P_0^n - \pi_0\|_{\mathrm{TV}} \to 0$ as $n \to \infty$. Then*

$$\mathbf{E}^\mu \big(\|\mathbf{P}^\mu(X_n \in \cdot \,|\, Y_0) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}} \big) \xrightarrow{n \to \infty} 0.$$

PROOF.   We reduce to the case of Lemma 2.15. As $Q(w, dw') = \varphi(dw')$, we obtain $\pi \sim \pi_0 \otimes \varphi$ and $\mu P^n \sim \mu(\cdot \times F) P_0^n \otimes \varphi$ for all $n > 0$ by Assumption 2.8. Choose $S_n \in \mathcal{B}(E)$ such that $\mu(\cdot \times F) P_0^n(\cdot \cap S_n) \ll \pi_0$ and $\pi_0(S_n^c) = 0$ for all n [so S_n defines the Lebesgue decomposition of $\mu(\cdot \times F) P_0^n$ with respect to π_0]. Then clearly $\mu P^n(\cdot \cap S_n \times F) \ll \pi$ and $\pi(S_n^c \times F) = 0$. Therefore

$$\|\mu P^{k+n} - \pi\|_{\mathrm{TV}} \leq \mu P^n(S_n \times F)\|\nu_n P^k - \pi\|_{\mathrm{TV}} + 2\mu P^n(S_n^c \times F)$$

$$\leq \|\nu_n P^k - \pi\|_{\mathrm{TV}} + \|\mu(\cdot \times F) P_0^n - \pi_0\|_{\mathrm{TV}},$$

where we have defined $\nu_n = \mu P^n(\cdot \cap S_n \times F)/\mu P^n(S_n \times F)$. But as $(X_n, Y_n)_{n \in \mathbb{Z}}$ is absolutely regular (cf. Lemma 2.16) and $\nu_n \ll \pi$, the first term converges to zero as $k \to \infty$. Letting $n \to \infty$ and applying Lemma 2.15 yields the result. $\square$

2.5.3. *Hidden Markov models with correlated noise.*   We now turn to the special case where the underlying model $(X_n, Y_n)_{n \in \mathbb{Z}}$ is a hidden Markov model with correlated noise, whose dependence structure is illustrated in Figure 1(c). Under Assumption 2.8, this dependence structure is enforced by the following requirement: there is a probability measure φ on F such that $Q(w, dw') = \varphi(dw')$, and there are measurable functions $g_X : E \times F \times E \to \mathbb{R}_+$ and $g_Y : E \times F \to \mathbb{R}_+$ such that

$$g(z, w, z', w') = g_X(z, w, z') g_Y(z', w'), \qquad \int g_Y(z, w)\varphi(dw) = 1.$$

Unlike in the case of a generalized hidden Markov model, in the present model the probabilities $\mathbf{P}^{z,w}(X_n \in \cdot)$ do depend on w. Nonetheless, in the present case the unobserved process $(X_n)_{n \in \mathbb{Z}}$ is still Markov under the stationary measure $\mathbf{P}$ with respect to its own filtration, with transition kernel $\tilde{P}_0$ given for $A \in \mathcal{B}(E)$ by

$$\tilde{P}_0(z, A) = \int P(z, w, A \times F) g_Y(z, w)\varphi(dw).$$

To see this, note that $\pi(dz, dw) = g_Y(z, w)\pi(dz \times F)\varphi(dw)$ by our assumption on P and $\pi P = \pi$, so we can compute $\mathbf{P}(X_{n+1} \in A|\mathcal{F}_n^X) = \tilde{P}_0(X_n, A)$.

REMARK 2.19.   Unlike in the case of a generalized hidden Markov model, where $Q(w, dw') = \varphi(dw')$ is an additional assumption, in the present setting the assumption $Q(w, dw') = \varphi(dw')$ entails no loss of generality. Indeed, the hidden Markov structure with correlated noise can be generally formulated by the requirement that $P(z, w, dz', dw') = P_X(z, w, dz')P_Y(z', dw')$ for some probability kernels P_X and P_Y. It is easily seen that any such model that also satisfies Assumption 2.8 must have the above form for a suitable choice of φ.

The idea is now that in the present setting, it suffices to consider the ergodic properties of the unobserved process (i.e., the transition kernel $\tilde{P}_0$).

LEMMA 2.20. *Suppose that Assumption* 2.8 *holds and that* $(X_n, Y_n)_{n \in \mathbb{Z}}$ *is a hidden Markov model with correlated noise in the above sense. If also*

$$\int \|\tilde{P}_0^n(z, \cdot) - \pi_0\|_{\mathrm{TV}} \pi_0(dz) \xrightarrow{n \to \infty} 0,$$

where $\pi_0 = \pi(\cdot \times F)$, *then both Assumptions* 2.6 *and* 2.7 *hold true.*

PROOF. Note that for all $(z, w) \in E \times F$, $B \in \mathcal{B}(E \times F)$ and $n \geq 1$ we have

$$P^n(z, w, B) = \int \left\{ \int_B \tilde{P}_0^{n-1}(z', d\tilde{z}) g_Y(\tilde{z}, \tilde{w}) \varphi(d\tilde{w}) \right\} g_X(z, w, z') P_0(z, dz').$$

Therefore, we have for $n \geq 1$

$$\int \|P^n(z, w, \cdot) - \pi\|_{\mathrm{TV}} \pi(dz, dw) \leq \int \|\tilde{P}_0^{n-1}(z, \cdot) - \pi_0\|_{\mathrm{TV}}.$$

The result now follows directly from Lemma 2.14. □

In the present setting (as in the case of a classical hidden Markov model), the most natural initial measures μ are those that are compatible with the observation model in the sense that $\mu(dz, dw) = g_Y(z, w)\mu_0(dz)\varphi(dw)$ for some probability measure μ_0 on E. This yields the following counterpart of Lemma 2.18, whose proof (by reduction to Lemma 2.15) is trivial and is therefore omitted.

LEMMA 2.21. *Suppose the assumptions of Lemma* 2.20 *hold. Let* μ_0 *be a probability measure on* E *such that* $\|\mu_0 \tilde{P}_0^n - \pi_0\|_{\mathrm{TV}} \to 0$ *as* $n \to \infty$. *Then*

$$\mathbf{E}^\mu \big(\|\mathbf{P}^\mu(X_n \in \cdot \,|Y_0) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}} \big) \xrightarrow{n \to \infty} 0,$$

where we have defined $\mu(dz, dw) = g_Y(z, w)\mu_0(dz)\varphi(dw)$.

REMARK 2.22. Let us note that in all of the special cases discussed above the process $(X_n, Y_n)_{n \in \mathbb{Z}}$ is absolutely regular so that Assumptions 2.6 and 2.7 hold by virtue of Lemma 2.14. Absolute regularity of $(X_n, Y_n)_{n \in \mathbb{Z}}$ is not necessary, however, for Assumptions 2.6 and 2.7 to hold. For example, in the trivial case that Assumption 2.8 holds with $g \equiv 1$, it is easily seen that Assumptions 2.6 and 2.7 hold if and only if the unobserved process $(X_n)_{n \in \mathbb{Z}}$ is absolutely regular, while the pair process $(X_n, Y_n)_{n \in \mathbb{Z}}$ need not even be ergodic [e.g., when $Q(w, dw') = \delta_w(dw')$]. Thus Assumptions 2.6 and 2.7 are strictly weaker than the absolute regularity of the pair process $(X_n, Y_n)_{n \in \mathbb{Z}}$. Nonetheless, the latter assumption is very mild and will likely hold in most applications of practical interest.

3. Nondegenerate Markov chains. The nondegeneracy Assumption 2.8 will play an essential role in our theory. Before we can turn to the proofs of our main results, we must therefore begin by establishing some general consequences of the nondegeneracy assumption that will be needed throughout the paper.

3.1. *Product structure of the invariant measure.* Assumption 2.8 states that the transition kernel P of the Markov chain $(X_n, Y_n)_{n \in \mathbb{Z}}$ is equivalent to a product of transition kernels of two independent Markov chains. Our first question is, therefore, whether this forces the invariant measure π to possess a similar product structure; that is, if a stationary Markov chain is nondegenerate, then is its invariant measure necessarily equivalent to the product of its marginals? In general, of course, the answer is negative (e.g., consider the case where P is the identity and π is any probability measure that is not equivalent to a product measure). However, we will presently show that if, in addition to nondegeneracy, we assume that the marginal process $(X_n)_{n \in \mathbb{Z}}$ is ergodic in a suitable sense, then π is forced to possess the desired product structure.

We need two lemmas. The first states that the nondegeneracy of the transition kernel P implies that the iterates P^n are also nondegenerate; in fact, we show that $\mathbf{P}((X_n, Y_n) \in \cdot \,|X_0, Y_0) \sim \mathbf{P}(X_n \in \cdot \,|X_0) \otimes \mathbf{P}(Y_n \in \cdot \,|Y_0)$.

LEMMA 3.1. *Suppose that Assumption 2.8 is in force. Choose fixed versions $\pi^Y(w, dz)$, $\pi^X(z, dw)$ of the regular conditional probabilities $\mathbf{P}(X_0 \in \cdot \,|Y_0)$, $\mathbf{P}(Y_0 \in \cdot \,|X_0)$, respectively, and define the probability kernels*

$$P_n^X(z, A) = \int \mathbf{1}_A(z') P^n(z, w, dz', dw') \pi^X(z, dw),$$

$$P_n^Y(w, B) = \int \mathbf{1}_B(w') P^n(z, w, dz', dw') \pi^Y(w, dz).$$

Then we have for all $n \in \mathbb{N}$

$$P^n(z, w, dz', dw') = G_n(z, w, z', w') P_n^X(z, dz') P_n^Y(w, dw'),$$

where $G_n : E \times F \times E \times F \to \,]0, \infty[$ are strictly positive measurable functions.

PROOF. From the Assumption 2.8, it follows directly that

$$P^n(z, w, dz', dw') = g_n(z, w, z', w') P_0^n(z, dz') Q^n(w, dw')$$

for some strictly positive measurable function $g_n : E \times F \times E \times F \to \,]0, \infty[$. But then the result follows directly from the definition of P_n^X, P_n^Y with

$$
\begin{aligned}
G_n&(z, w, z', w') \\
&= g_n(z, w, z', w') \\
&\quad \times \left(\int g_n(z, \tilde{w}, z', \tilde{w}') Q^n(\tilde{w}, d\tilde{w}') \pi^X(z, d\tilde{w}) \right. \\
&\qquad \left. \times \int g_n(\tilde{z}, w, \tilde{z}', w') P_0^n(\tilde{z}, d\tilde{z}') \pi^Y(w, d\tilde{z}) \right)^{-1}.
\end{aligned}
$$

The proof is complete. $\quad\square$

The second lemma states that if the unobserved process $(X_n)_{n \in \mathbb{Z}}$ is ergodic in a suitable sense, and if the nondegeneracy assumption holds, then every P-invariant function is independent of its unobserved component.

LEMMA 3.2.   *Suppose that Assumption* 2.8 *is in force, and that*

$$\int \|P_n^X(z, \cdot) - \pi(\cdot \times F)\|_{\mathrm{TV}} \pi(dz \times F) \xrightarrow{n \to \infty} 0.$$

Then for any bounded measurable function $f : E \times F \to \mathbb{R}$ *that is* P-*invariant* (*i.e.,* $f = Pf$), *there exists a bounded measurable function* $g : F \to \mathbb{R}$ *such that* $f(z, w) = g(w)$ *for* π-*a.e.* $(z, w) \in E \times F$.

PROOF.   As f is P-invariant, the process $(f(X_n, Y_n))_{n \geq 0}$ is a martingale under $\mathbf{P}$. By stationarity and the martingale convergence theorem,

$$\mathbf{E}(|f(X_n, Y_n) - f(X_0, Y_0)|) = \mathbf{E}(|f(X_{n+k}, Y_{n+k}) - f(X_k, Y_k)|) \xrightarrow{k \to \infty} 0.$$

In particular, we have

$$\int \mathbf{P}^{z,w}(f(X_0, Y_0) = f(X_n, Y_n) \text{ for all } n \geq 0) \pi(dz, dw) = 1.$$

Therefore, we may choose a set $H_1 \in \mathcal{B}(E \times F)$ with $\pi(H_1) = 1$ such that

$$\mathbf{P}^{z,w}(f(z, w) = f(X_n, Y_n) \text{ for all } n \geq 0) = 1 \qquad \text{for all } (z, w) \in H_1.$$

Next, let $\rho : F \times \mathcal{B}(E) \to [0, 1]$ be a version of the regular conditional probability $\mathbf{P}(X_0 \in \cdot | Y_0)$. Then by our assumption and the triangle inequality,

$$\int \|P_n^X(z, \cdot) - P_n^X(z', \cdot)\|_{\mathrm{TV}} \rho(w, dz) \rho(w, dz') \pi(E \times dw)$$

$$\leq 2 \int \|P_n^X(z, \cdot) - \pi(\cdot \times F)\|_{\mathrm{TV}} \pi(dz \times F) \xrightarrow{n \to \infty} 0.$$

Therefore, using Fatou's lemma, we can choose a set $H_2 \in \mathcal{B}(E \times E \times F)$ of $(\rho \otimes \rho)\pi(E \times \cdot)$-full measure such that

$$\liminf_{n \to \infty} \|P_n^X(z, \cdot) - P_n^X(z', \cdot)\|_{\mathrm{TV}} = 0 \qquad \text{for all } (z, z', w) \in H_2.$$

Now define the set $H \in \mathcal{B}(E \times E \times F)$ as follows:

$$H = \{(z, z', w) \in E \times E \times F : (z, w), (z', w) \in H_1\} \cap H_2.$$

Then it is easily seen that the set H has $(\rho \otimes \rho)\pi(E \times \cdot)$-full measure.

We now claim that $f(z, w) = f(z', w)$ for every $(z, z', w) \in H$. To see this, let us fix some point $(z, z', w) \in H$, and choose $n \geq 0$ such that

$$\|P_n^X(z, \cdot) - P_n^X(z', \cdot)\|_{\mathrm{TV}} < 1.$$

Thus $P_n^X(z, \cdot)$ and $P_n^X(z', \cdot)$ are not mutually singular. By Lemma 3.1

$$P^n(z, w, \cdot) \sim P_n^X(z, \cdot) \otimes P_n^Y(w, \cdot), \qquad P^n(z', w, \cdot) \sim P_n^X(z', \cdot) \otimes P_n^Y(w, \cdot).$$

Therefore, $P^n(z, w, \cdot)$ and $P^n(z', w, \cdot)$ are not mutually singular. But note that, by the definition of H, $P^n(z, w, \cdot)$ is supported on the set

$$\Xi_1 = \{(\tilde{z}, \tilde{w}) \in E \times F : f(z, w) = f(\tilde{z}, \tilde{w})\},$$

while $P^n(z', w, \cdot)$ is supported on the set

$$\Xi_2 = \{(\tilde{z}, \tilde{w}) \in E \times F : f(z', w) = f(\tilde{z}, \tilde{w})\}.$$

Thus the fact that $P^n(z, w, \cdot)$ and $P^n(z', w, \cdot)$ are not mutually singular implies that $\Xi_1 \cap \Xi_2 \neq \varnothing$, which establishes the claim.

To complete the proof, define $g(w) = \int f(z, w) \rho(w, dz)$. Then

$$\int |f(z, w) - g(w)| \pi(dz, dw)$$

$$\leq \int |f(z, w) - f(z', w)| \rho(w, dz) \rho(w, dz') \pi(E \times dw) = 0.$$

Thus $f(z, w) = g(w)$ for π-a.e. $(z, w) \in E \times F$ as desired. $\square$

We can now prove the main result of this subsection: if the nondegeneracy assumption holds, and if, in addition, the unobserved component $(X_n)_{n \in \mathbb{Z}}$ is ergodic, then the invariant measure π is necessarily equivalent to the product of its marginals. Note that the ergodicity assumption in this result automatically holds when Assumption 2.6 is in force.

PROPOSITION 3.3.   *Suppose that Assumption* 2.8 *is in force, and that*

$$\int \|P_n^X(z, \cdot) - \pi(\cdot \times F)\|_{\mathrm{TV}} \pi(dz \times F) \xrightarrow{n \to \infty} 0.$$

Then there exists a strictly positive measurable function $h : E \times F \to]0, \infty[$ *such that* $\pi(dz, dw) = h(z, w) \pi(dz \times F) \pi(E \times dw)$.

PROOF.   We begin by noting that

$$\pi(A \times F) = \int \pi(dz \times F) P_n^X(z, A), \qquad \pi(E \times B) = \int \pi(E \times dw) P_n^Y(w, B)$$

by the invariance of π. Now let $C \in \mathcal{B}(E \times F)$ be a set such that $\pi(C) = 0$. As $\pi P^n = \pi$, it follows from Lemma 3.1 that

$$\int \mathbf{1}_C(z', w') P_n^X(z, dz') P_n^Y(w, dw') \pi(dz, dw) = 0$$

for all $n \in \mathbb{N}$. But note that

$$\int \mathbf{1}_C(z, w) \pi(dz \times F) \pi(E \times dw)$$

$$= \int \mathbf{1}_C(z', w') \pi(dz' \times F) P_n^Y(w, dw') \pi(dz, dw)$$

$$\leq \int \|P_n^X(z, \cdot) - \pi(\cdot \times F)\|_{\mathrm{TV}} \pi(dz \times F)$$

$$+ \int \mathbf{1}_C(z', w') P_n^X(z, dz') P_n^Y(w, dw') \pi(dz, dw).$$

Letting $n \to \infty$ and using the ergodicity assumption gives

$$\int \mathbf{1}_C(z, w) \pi(dz \times F) \pi(E \times dw) = 0.$$

As this holds for any set C such that $\pi(C) = 0$, we have evidently shown that $\pi(dz \times F) \pi(E \times dw) \ll \pi(dz, dw)$. Conversely, choose a set C such that

$$\int \mathbf{1}_C(z, w) \pi(dz \times F) \pi(E \times dw) = 0.$$

Then, by Lemma 3.1, we have

$$\int P^n(z, w, C) \pi(dz \times F) \pi(E \times dw) = 0$$

for all $n \in \mathbb{N}$. By the Birkhoff ergodic theorem,

$$\frac{1}{N} \sum_{n=1}^{N} P^n(z, w, C) \xrightarrow{N \to \infty} f(z, w) \qquad \text{for } \pi\text{-a.e. } (z, w) \in E \times F,$$

where f is a P-invariant function with $\pi(f) = \pi(C)$. Moreover, by Lemma 3.2 we have $f(z, w) = g(w)$ for π-a.e. $(z, w) \in E \times F$ for some function g. But as we have already shown that $\pi(dz \times F) \pi(E \times dw) \ll \pi(dz, dw)$, these statements hold $\pi(dz \times F) \pi(E \times dw)$-a.e. also. Therefore,

$$0 = \frac{1}{N} \sum_{n=1}^{N} \int P^n(z, w, C) \pi(dz \times F) \pi(E \times dw)$$

$$\xrightarrow{N \to \infty} \int g(w) \pi(E \times dw) = \int f(z, w) \pi(dz, dw) = \pi(C).$$

As this holds for any C such that $\int \mathbf{1}_C(z, w) \pi(dz \times F) \pi(E \times dw) = 0$, we evidently have $\pi(dz, dw) \ll \pi(dz \times F) \pi(E \times dw)$, and the proof is complete. $\qquad \square$

3.2. *Reversed nondegeneracy.* One important consequence of Proposition 3.3 is that, if the unobserved process $(X_n)_{n \in \mathbb{Z}}$ is ergodic and the transition kernel P is nondegenerate, then the nondegeneracy Assumption 2.8 holds also in reverse time (i.e., the backward transition kernel P' must be nondegenerate also). In particular, this implies that the Assumptions 2.6–2.8 are invariant under time reversal.

LEMMA 3.4. *Suppose that Assumption 2.8 is in force, and that*

$$\int \| P_n^X(z, \cdot) - \pi(\cdot \times F) \|_{\mathrm{TV}} \pi(dz \times F) \xrightarrow{n \to \infty} 0.$$

Then P' is also nondegenerate; that is, there exist transition probability kernels $P_0': E \times \mathcal{B}(E) \to [0, 1]$ and $Q': F \times \mathcal{B}(F) \to [0, 1]$ such that

$$P'(z, w, dz', dw') = g'(z, w, z', w') P_0'(z, dz') Q'(w, dw')$$

for some strictly positive measurable function $g': E \times F \times E \times F \to \,]0, \infty[$.

PROOF. Note that by Proposition 3.3 and Assumption 2.8

$$\mathbf{E}((X_0, Y_0, X_1, Y_1) \in B) = \int_B h(x_0, y_0) g(x_0, y_0, x_1, y_1) \rho(dx_0, dx_1) \kappa(dy_0, dy_1),$$

where $\rho(dx, dx') = \pi(dx \times F) P_0(x, dx')$, $\kappa(dy, dy') = \pi(E \times dy) Q(y, dy')$, and where g, h are strictly positive measurable functions. Let us now fix any versions $r(x_1, dx_0)$ and $k(y_1, dy_0)$ of the regular conditional probabilities $\rho(X_0 \in \cdot | X_1)$ and $\kappa(Y_0 \in \cdot | Y_1)$, respectively. Then by the Bayes formula,

$$\mathbf{P}((X_0, Y_0) \in A | X_1, Y_1) = \frac{\int_A h(z, w) g(z, w, X_1, Y_1) r(X_1, dz) k(Y_1, dw)}{\int h(z, w) g(z, w, X_1, Y_1) r(X_1, dz) k(Y_1, dw)}.$$

As P' is a version of $\mathbf{P}((X_0, Y_0) \in \cdot | X_1, Y_1)$, the result follows. $\square$

3.3. *Equivalence of the observations.* We now turn to a different consequence of the nondegeneracy assumption. It is easily seen that when Assumption 2.8 holds, the laws of $(Y_0, \ldots, Y_n)$ under $\mathbf{P}^{z,w}$ and $\mathbf{P}^{z',w}$ are equivalent for any $z, z' \in E$, $w \in F$, $n < \infty$. That is, the laws of the observed process under different initializations of the unobserved process are equivalent on any finite time horizon. To prove our main results, however, we will require such an equivalence to hold on the *infinite* time horizon. The following result is therefore of central importance.

PROPOSITION 3.5. *Suppose that Assumption 2.8 holds. Let ξ, ξ' be probability measures on $(E, \mathcal{B}(E))$, let η be a probability measure on $(F, \mathcal{B}(F))$ and let $v: E \times F \to \,]0, \infty[$ and $v': E \times F \to \,]0, \infty[$ be strictly positive measurable functions. Define the probability measures on $(E \times F, \mathcal{B}(E \times F))$*

$$v(dx, dy) = v(x, y) \xi(dx) \eta(dy), \qquad v'(dx, dy) = v'(x, y) \xi'(dx) \eta(dy).$$

If $\liminf_{n \to \infty} \| \mathbf{P}^v(X_n \in \cdot) - \mathbf{P}^{v'}(X_n \in \cdot) \|_{\mathrm{TV}} = 0$, then $\mathbf{P}^v|_{\mathcal{F}_+^Y} \sim \mathbf{P}^{v'}|_{\mathcal{F}_+^Y}$.

PROOF.   Choose any $A \in \mathcal{F}_+^Y$ such that $\mathbf{P}^{\nu'}(A) = 0$. It suffices to prove that $\mathbf{P}^\nu(A) = 0$. Indeed, this shows that $\mathbf{P}^\nu|_{\mathcal{F}_+^Y} \ll \mathbf{P}^{\nu'}|_{\mathcal{F}_+^Y}$, while the reverse statement follows as the assumptions are symmetric in ν and ν'.

Fix for the time being $n \in \mathbb{N}$. Note that by construction

$$I_A(x, y) = I_A(y(0), \ldots, y(n), (y(k))_{k>n}).$$

Let us define the measurable function

$$a(y_0, \ldots, y_n, x_n) = \mathbf{E}^{x_n, y_n}(I_A(y_0, \ldots, y_n, (Y_k)_{k \geq 1})).$$

Then, by the Markov property,

$$a(Y_0, \ldots, Y_n, X_n) = \mathbf{P}^\rho\big(A|\mathcal{F}_{[0,n]}^X \vee \mathcal{F}_{[0,n]}^Y\big), \qquad \mathbf{P}^\rho\text{-a.s.}$$

for any initial probability measure ρ. In particular,

$$a(Y_0, \ldots, Y_n, X_n) = 0, \qquad \mathbf{P}^{\nu'}\text{-a.s.}$$

Let $\mathbf{Q}^\eta$ be the law of the Markov chain $(Y_k)_{k \geq 0}$ with initial measure η and transition kernel Q, and $\mathbf{P}_0^{\xi'}$ be the law of the Markov chain $(X_k)_{k \geq 0}$ with initial measure ξ' and transition kernel P_0. By our assumptions,

$$\mathbf{P}^{\nu'}(A) = (\mathbf{P}_0^{\xi'} \otimes \mathbf{Q}^\eta)\left[I_A \nu'(X_0, Y_0) \prod_{i=0}^{n-1} g(X_i, Y_i, X_{i+1}, Y_{i+1})\right]$$

for every $A \in \mathcal{F}_{[0,n]}^X \vee \mathcal{F}_{[0,n]}^Y$. In particular, the law of $(Y_0, \ldots, Y_n, X_n)$ under $\mathbf{P}^{\nu'}$ and the law of $\mathbf{Q}^\eta|_{\mathcal{F}_{[0,n]}^Y} \otimes \xi' P_0^n$ are equivalent. Therefore,

$$a(Y_0, \ldots, Y_n, X_n) = 0, \qquad (\mathbf{Q}^\eta|_{\mathcal{F}_{[0,n]}^Y} \otimes \xi' P_0^n)\text{-a.s.}$$

Choose $S_n \in \mathcal{B}(E)$ such that $(\xi P_0^n)(\cdot \cap S_n) \ll \xi' P_0^n$ and $(\xi' P_0^n)(S_n^c) = 0$ (so S_n defines the Lebesgue decomposition of ξP_0^n with respect to $\xi' P_0^n$). Then

$$I_{S_n}(X_n) a(Y_0, \ldots, Y_n, X_n) = 0, \qquad (\mathbf{Q}^\eta|_{\mathcal{F}_{[0,n]}^Y} \otimes \xi P_0^n)\text{-a.s.}$$

Therefore,

$$a(Y_0, \ldots, Y_n, X_n) \leq I_{S_n^c}(X_n), \qquad (\mathbf{Q}^\eta|_{\mathcal{F}_{[0,n]}^Y} \otimes \xi P_0^n)\text{-a.s.}$$

But, as above, we find that the law of $(Y_0, \ldots, Y_n, X_n)$ under $\mathbf{P}^\nu$ is equivalent to $\mathbf{Q}^\eta|_{\mathcal{F}_{[0,n]}^Y} \otimes \xi P_0^n$. Therefore, we obtain immediately

$$a(Y_0, \ldots, Y_n, X_n) \leq I_{S_n^c}(X_n), \qquad \mathbf{P}^\nu\text{-a.s.}$$

Taking the expectation, we find that $\mathbf{P}^\nu(A) \leq \mathbf{P}^\nu(X_n \in S_n^c)$.

At this point, we note that $n \in \mathbb{N}$ in the above construction was arbitrary. Moreover, we have already shown that for any $n \in \mathbb{N}$, the law of X_n under $\mathbf{P}^{\nu'}$ is equivalent to $\xi' P_0^n$. Therefore $\mathbf{P}^{\nu'}(X_n \in S_n^c) = 0$, and we find that

$$\mathbf{P}^{\nu}(A) \le \liminf_{n \to \infty} \mathbf{P}^{\nu}(X_n \in S_n^c) \le \liminf_{n \to \infty} \|\mathbf{P}^{\nu}(X_n \in \cdot) - \mathbf{P}^{\nu'}(X_n \in \cdot)\|_{\mathrm{TV}} = 0.$$

Thus the proof is complete. $\square$

A useful corollary is the following result.

COROLLARY 3.6.   *Suppose that Assumptions 2.6 and 2.8 hold, and let μ be a probability measure on $E \times F$ such that $\mu(E \times \cdot) \ll \pi(E \times \cdot)$ and*

$$\mathbf{E}^{\mu}\big(\|\mathbf{P}^{\mu}(X_n \in \cdot \,|Y_0) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}}\big) \xrightarrow{n \to \infty} 0.$$

Then $\mathbf{P}^{\mu}|_{\mathcal{F}_+^Y} \ll \mathbf{P}|_{\mathcal{F}_+^Y}$. If $\mu(E \times \cdot) \sim \pi(E \times \cdot)$, then $\mathbf{P}^{\mu}|_{\mathcal{F}_+^Y} \sim \mathbf{P}|_{\mathcal{F}_+^Y}$.

PROOF.   We begin by noting that

$$\mathbf{E}\big(\|\mathbf{P}(X_n \in \cdot \,|Y_0) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}}\big) \le \mathbf{E}\big(\|\mathbf{P}^{X_0, Y_0}(X_n \in \cdot) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}}\big).$$

Therefore, by Assumption 2.6,

$$\|\mathbf{P}(X_n \in \cdot \,|Y_0) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}} \xrightarrow{n \to \infty} 0 \qquad \text{in } \mathbf{P}\text{-probability.}$$

As $\mathbf{P}^{\mu}(Y_0 \in \cdot) \ll \mathbf{P}(Y_0 \in \cdot)$, this convergence is also in $\mathbf{P}^{\mu}$-probability. Therefore, using dominated convergence and the triangle inequality,

$$\mathbf{E}^{\mu}\big(\|\mathbf{P}^{\mu}(X_n \in \cdot \,|Y_0) - \mathbf{P}(X_n \in \cdot \,|Y_0)\|_{\mathrm{TV}}\big) \xrightarrow{n \to \infty} 0.$$

By Fatou's lemma, we obtain

$$\liminf_{n \to \infty}\|\mathbf{P}^{\mu}(X_n \in \cdot \,|Y_0) - \mathbf{P}(X_n \in \cdot \,|Y_0)\|_{\mathrm{TV}} = 0, \qquad \mathbf{P}^{\mu}\text{-a.s.}$$

Let $\nu : F \times \mathcal{B}(E) \to [0, 1]$, $\nu' : F \times \mathcal{B}(E) \to [0, 1]$ be versions of the regular conditional probabilities $\mathbf{P}^{\mu}(X_0 \in \cdot \,|Y_0)$, $\mathbf{P}(X_0 \in \cdot \,|Y_0)$, respectively. Then

$$\liminf_{n \to \infty}\|\mathbf{P}^{\nu(w,\cdot) \otimes \delta_w}(X_n \in \cdot) - \mathbf{P}^{\nu'(w,\cdot) \otimes \delta_w}(X_n \in \cdot)\|_{\mathrm{TV}} = 0, \qquad \mu(E \times \cdot)\text{-a.e. } w.$$

By Proposition 3.5, it follows that

$$\mathbf{P}^{\nu(w,\cdot) \otimes \delta_w}|_{\mathcal{F}_+^Y} \sim \mathbf{P}^{\nu'(w,\cdot) \otimes \delta_w}|_{\mathcal{F}_+^Y}, \qquad \mu(E \times \cdot)\text{-a.e. } w.$$

By the Lebesgue decomposition for kernels ([9], Section V.58), there is a measurable version of the Radon–Nikodym derivative. It follows that

$$\mathbf{P}^{\mu}|_{\mathcal{F}_+^Y} = \mathbf{P}^{\nu \otimes \mu(E \times \cdot)}|_{\mathcal{F}_+^Y} \sim \mathbf{P}^{\nu' \otimes \mu(E \times \cdot)}|_{\mathcal{F}_+^Y} \ll \mathbf{P}^{\nu' \otimes \pi(E \times \cdot)}|_{\mathcal{F}_+^Y} = \mathbf{P}|_{\mathcal{F}_+^Y},$$

where we have used that $\mu(E \times \cdot) \ll \pi(E \times \cdot)$. If $\mu(E \times \cdot) \sim \pi(E \times \cdot)$, then clearly $\ll$ can be replaced by $\sim$ in the previous equation. $\square$

4. Proof of Theorem 2.9. The goal of this section is to prove Theorem 2.9. To this end, we begin by recalling the basic result from [20] on the ergodicity of Markov chains in random environments. This result will be used to establish that the unobservable process $(X_n)_{n \geq 0}$ has trivial tail σ-field under the conditional measure $\mathbf{P}(\cdot|\mathcal{F}^Y)$. Finally, we show that $\mathbf{P}(X_0 \in \cdot|\mathcal{F}^Y) \sim \mathbf{P}(X_0 \in \cdot|\mathcal{F}^Y_+)$, which allows us to complete the proof by applying a result of von Weizsäcker [24].

4.1. *Markov chains in random environments.* We begin by recalling the relevant notions from [20], Section 2. A Markov chain in a random environment is defined by the following three ingredients:

(1) A probability kernel $P^X : E \times \Omega^Y \times \mathcal{B}(E) \to [0, 1]$.
(2) A probability kernel $\varpi : \Omega^Y \times \mathcal{B}(E) \to [0, 1]$ such that

$$\int P^X(z, y, A)\varpi(y, dz) = \varpi(\Theta y, A) \qquad \text{for all } y \in \Omega^Y, A \in \mathcal{B}(E).$$

(3) A stationary probability measure $\mathbf{P}^Y$ on $(\Omega^Y, \mathcal{F}^Y)$.

The process X_n is called a Markov chain in a random environment when

$$P^X(X_n, Y \circ \Theta^n, A) = \mathbf{P}(X_{n+1} \in A|\mathcal{F}^X_n \vee \mathcal{F}^Y), \qquad \mathbf{P}\text{-a.s.},$$

$$\varpi(Y \circ \Theta^n, A) = \mathbf{P}(X_n \in A|\mathcal{F}^Y), \qquad \mathbf{P}\text{-a.s.}$$

for every $A \in \mathcal{B}(E)$ and $n \in \mathbb{Z}$, and $\mathbf{P}^Y = \mathbf{P}|_{\mathcal{F}^Y}$. One should think of a Markov chain in a random environment X_n as a process that is Markov *conditionally* on the environment Y. The conditional chain is time-inhomogeneous but must satisfy certain stationarity properties: the environment is stationary and the (time-dependent) conditional transition probabilities $P^X(\cdot, Y \circ \Theta^n, \cdot)$ and quasi-invariant measure $\varpi(Y \circ \Theta^n, \cdot)$ are themselves stationary processes with respect to the environment. The stationarity properties ensure that Markov chains in random environments behave "almost" like time-homogeneous Markov chains; cf. Theorem 4.1 below.

Let us introduce a probability kernel $\mathbf{P}_{\cdot} : E \times \Omega^Y \times \mathcal{F}^X_+ \to [0, 1]$ so that

$$\mathbf{P}_{z, y}(A) = \int I_A(x) P^X(x(n-1), \Theta^{n-1}y, dx(n)) \times \cdots$$

$$\times P^X(x(1), \Theta y, dx(2)) P^X(x(0), y, dx(1))\delta_z(dx(0))$$

for $A \in \mathcal{F}^X_{[0,n]}$. It is easily seen that $\mathbf{P}_{z, y}$ is a version of the regular conditional probability $\mathbf{P}((X_k)_{k \geq 0} \in \cdot|\mathcal{F}^X_0 \vee \mathcal{F}^Y)$. We can now state the following ergodic theorem for Markov chains in random environments ([20], Theorem 2.3).

THEOREM 4.1. *The following are equivalent.*

(1) $\|\mathbf{P}_{z, y}(X_n \in \cdot) - \mathbf{P}_{z', y}(X_n \in \cdot)\|_{\mathrm{TV}} \xrightarrow{n \to \infty} 0$ *for* $(\varpi \otimes \varpi)\mathbf{P}^Y$*-a.e.* (z, z', y).

(2) *The tail σ-field $\mathcal{T}^X = \bigcap_{n \geq 0} \mathcal{F}^X_{[n,\infty[}$ is a.s. trivial in the following sense*:

$$\mathbf{P}_{z,y}(A) = \mathbf{P}_{z,y}(A)^2 = \mathbf{P}_{z',y}(A) \qquad \text{for all } A \in \mathcal{T}^X \text{ and } (z, z', y) \in H,$$

where H is a fixed set (independent of A) of $(\varpi \otimes \varpi)\mathbf{P}^Y$-full measure.

(3) *For $(\varpi \otimes \varpi)\mathbf{P}^Y$-a.e. (z, z', y), there is an $n \in \mathbb{N}$ such that the measures $\mathbf{P}_{z,y}(X_n \in \cdot)$ and $\mathbf{P}_{z',y}(X_n \in \cdot)$ are not mutually singular.*

4.2. *Weak ergodicity of the conditional process.*   Our first order of business is to establish that, under the model defined in this paper, X_n is indeed a Markov chain in a random environment in the sense of Section 4.1, where the observations Y play the role of the environment; that is, we must show that the unobserved process X_n is still a Markov chain *conditionally* on the observations Y satisfying the requisite stationarity properties. This is the statement of the following lemma, whose proof is omitted as it is identical to that in [20]. As everything that follows is based on this elementary fact, however, let us briefly sketch why the result is true for the convenience of the reader. It is easily seen that

$$\mathbf{P}(X_{n+1} \in \cdot \,|\mathcal{F}^X_n \vee \mathcal{F}^Y) \circ \Theta^{-n} = \mathbf{P}(X_1 \in \cdot \,|\mathcal{F}^X_0 \vee \mathcal{F}^Y) = \mathbf{P}(X_1 \in \cdot \,|\sigma\{X_0\} \vee \mathcal{F}^Y).$$

The first equality follows from stationarity of $\mathbf{P}$, and the second equality follows as $\mathcal{F}_{[1,\infty[}$ is conditionally independent of $\mathcal{F}_{-1}$ given $\sigma\{X_0, Y_0\}$ by the Markov property of $(X_n, Y_n)_{n \in \mathbb{Z}}$. We can therefore choose P^X to be a regular version of $\mathbf{P}(X_1 \in \cdot \,|\sigma\{X_0\} \vee \mathcal{F}^Y)$. Similarly, we can choose ϖ to be a regular version of $\mathbf{P}(X_0 \in \cdot \,|\mathcal{F}^Y)$, and $\mathbf{P}^Y$ to be the law of Y. It is now an elementary exercise to check that these kernels do indeed characterize the process X_n as a Markov chain in a random environment in the sense of Section 4.1.

LEMMA 4.2.   *There exist probability kernels $P^X : E \times \Omega^Y \times \mathcal{B}(E) \to [0, 1]$ and $\varpi : \Omega^Y \times \mathcal{B}(E) \to [0, 1]$, and a probability measure $\mathbf{P}^Y$ on $(\Omega^Y, \mathcal{F}^Y)$, such that the conditions of Section 4.1 are satisfied.*

PROOF.   The proof is identical to that of [20], Lemma 3.3.   □

The main goal of this subsection is to prove the following theorem.

THEOREM 4.3.   *Suppose that both Assumptions 2.6 and 2.8 are in force. Then any (hence, all) of the conditions of Theorem 4.1 hold true.*

The strategy of the proof of Theorem 4.3 is to show that condition (3) of Theorem 4.1 follows from Assumptions 2.6 and 2.8. To this end, we begin by proving that Theorem 4.3 would follow if we can establish equivalence of the conditional and unconditional transition kernels P^X and P.

LEMMA 4.4. *Suppose that Assumptions* 2.6 *and* 2.8 *are in force, and that there exists a strictly positive measurable function* $h : E \times \Omega^Y \times E \to \,]0, \infty[$ *such that*

$$P^X(z, y, A) = \int_{E \times F} I_A(\tilde{z}) h(z, y, \tilde{z}) P(z, y(0), d\tilde{z}, d\tilde{w}) \qquad \textit{for all } A \in \mathcal{B}(E)$$

for $\varpi \mathbf{P}^Y$*-a.e.* (z, y). *Then condition* 3 *of Theorem* 4.1 *holds.*

PROOF. By Assumption 2.6 and the triangle inequality

$$\int \|\mathbf{P}^{z, y(0)}(X_n \in \cdot) - \mathbf{P}^{z', y(0)}(X_n \in \cdot)\|_{\mathrm{TV}} \varpi(y, dz) \varpi(y, dz') \mathbf{P}^Y(dy) \xrightarrow{n \to \infty} 0.$$

By Fatou's lemma, there is a set H_1 of $(\varpi \otimes \varpi) \mathbf{P}^Y$-full measure such that

$$\liminf_{n \to \infty} \|\mathbf{P}^{z, y(0)}(X_n \in \cdot) - \mathbf{P}^{z', y(0)}(X_n \in \cdot)\|_{\mathrm{TV}} = 0 \qquad \text{for all } (z, z', y) \in H_1.$$

In particular, there is for every $(z, z', y) \in H_1$ an $n \in \mathbb{N}$ such that $\mathbf{P}^{z, y(0)}(X_n \in \cdot)$ and $\mathbf{P}^{z', y(0)}(X_n \in \cdot)$ are not mutually singular.

Now let H_2 be a set of $\varpi \mathbf{P}^Y$-full measure such that the absolute continuity condition in the statement of the lemma holds true for all $(z, y) \in H_2$. By Lemma A.1, there is a subset $H_3 \subset H_2$ of $\varpi \mathbf{P}^Y$-full measure such that for every $(z, y) \in H_3$ we have $\mathbf{P}_{z,y}((X_n, \Theta^n y) \in H_3 \text{ for all } n \geq 0) = 1$. It follows directly that for every $(z, y) \in H_3$, $n \in \mathbb{N}$ and $A \in \mathcal{B}(E)$, we have

$$\mathbf{P}_{z,y}(X_n \in A) = \int I_A(x_n) f(x_0, \ldots, x_n, y) \prod_{i=0}^{n-1} P_0(x_i, dx_{i+1}) \delta_z(dx_0),$$

where we have defined the strictly positive measurable function

$$f(x_0, \ldots, x_n, y) = \prod_{i=0}^{n-1} h(x_i, \Theta^i y, x_{i+1}) \int g(x_i, y(i), x_{i+1}, \tilde{w}) Q(y(i), d\tilde{w}).$$

On the other hand, we have for every z, y

$$\mathbf{P}^{z, y(0)}(X_n \in A) = \int I_A(x_n) f'(x_0, \ldots, x_n, y(0)) \prod_{i=0}^{n-1} P_0(x_i, dx_{i+1}) \delta_z(dx_0),$$

where we have defined the strictly positive measurable function

$$f'(x_0, \ldots, x_n, y_0) = \int \prod_{i=0}^{n-1} g(x_i, y_i, x_{i+1}, y_{i+1}) Q(y_i, dy_{i+1}).$$

Therefore $\mathbf{P}_{z, y}(X_n \in \cdot) \sim \mathbf{P}^{z, y(0)}(X_n \in \cdot)$ for all $(z, y) \in H_3$ and $n \in \mathbb{N}$.

To complete the proof, define the following set:

$$H_4 = \{(z, z', y) : (z, z', y) \in H_1, (z, y), (z', y) \in H_3\}.$$

Then H_4 has $(\varpi \otimes \varpi)\mathbf{P}^Y$-full measure, and for every $(z, z', y) \in H_4$, there is an $n \in \mathbb{N}$ such that $\mathbf{P}_{z,y}(X_n \in \cdot)$ and $\mathbf{P}_{z',y}(X_n \in \cdot)$ are not mutually singular. This establishes condition (3) of Theorem 4.1. $\square$

We now proceed to prove the following lemma, which verifies the assumption of Lemma 4.4. This completes the proof of Theorem 4.3.

LEMMA 4.5. *Suppose that Assumptions 2.6 and 2.8 hold. Then there exists a strictly positive measurable function* $h : E \times \Omega^Y \times E \to \,]0, \infty[$ *such that*

$$P^X(z, y, A) = \int_{E \times F} I_A(\tilde{z}) h(z, y, \tilde{z}) P(z, y(0), d\tilde{z}, d\tilde{w}) \qquad \text{for all } A \in \mathcal{B}(E),$$

for $\varpi \mathbf{P}^Y$*-a.e.* (z, y).

PROOF. By definition, P^X is a version of the regular conditional probability $\mathbf{P}(X_1 \in \cdot | \mathcal{F}_0^X \vee \mathcal{F}^Y)$. But by the Markov property of $(X_n, Y_n)_{n \in \mathbb{Z}}$, the σ-fields $\mathcal{F}_{[1,\infty[}$ and $\mathcal{F}_{-1}$ are conditionally independent given $\sigma(X_0, Y_0)$. Therefore, P^X is, in fact, a version of the regular conditional probability $\mathbf{P}(X_1 \in \cdot | \sigma(X_0, Y_0) \vee \mathcal{F}_{[1,\infty[}^Y)$. Moreover, clearly the kernel $\tilde{P}$ defined as

$$\tilde{P}(z, w, A) = \int I_A(\tilde{z}) P(z, w, d\tilde{z}, d\tilde{w}) \qquad \text{for all } A \in \mathcal{B}(E), (z, w) \in E \times F$$

is a version of the regular conditional probability $\mathbf{P}(X_1 \in \cdot | \sigma(X_0, Y_0))$. Finally, we fix throughout the proof arbitrary versions $R : E \times F \times \mathcal{F}_+^Y \to [0, 1]$ and $R^X : E \times F \times E \times \mathcal{F}_+^Y \to [0, 1]$ of the regular conditional probabilities $P((Y_k)_{k \geq 1} \in \cdot | \sigma(X_0, Y_0))$ and $P((Y_k)_{k \geq 1} \in \cdot | \sigma(X_0, Y_0, X_1))$, respectively. To complete the proof, it suffices to show that $R^X(z, w, z', \cdot) \sim R(z, w, \cdot)$ for $(z, w, z') \in H$ with $P((X_0, Y_0, X_1) \in H) = 1$. Indeed, if this is the case, then by the Lebesgue decomposition for kernels ([9], Section V.58), there is a strictly positive measurable function $h : E \times \Omega^Y \times E \to \,]0, \infty[$ such that

$$R^X(z, y(0), \tilde{z}, A) = \int I_A((y(i))_{i \geq 1}) h(z, y, \tilde{z}) R(z, y(0), d(y(i))_{i \geq 1})$$

for all $A \in \mathcal{F}_{[1,\infty[}^Y$ and $(z, y(0), z') \in H'$ with $P((X_0, Y_0, X_1) \in H') = 1$. It remains to apply Lemma A.2 to the law of the triple $((X_0, Y_0), X_1, (Y_k)_{k \geq 1})$.

It therefore remains to show that $R^X(z, w, z', \cdot) \sim R(z, w, \cdot)$. To this end, let us introduce convenient versions of the regular conditional probabilities R and R^X. Note that we can write for $A \in \mathcal{F}_+^Y$

$$R(X_0, Y_0, A \circ \Theta) = \mathbf{E}(\mathbf{P}^{X_1, Y_1}(A) | \sigma(X_0, Y_0)) = \mathbf{P}^{\nu_{X_0}, Y_0}(A)$$

by the Markov property of $(X_n, Y_n)_{n \geq 0}$, where we have defined

$$\nu_{z,w}(d\tilde{z}, d\tilde{w}) = P(z, w, d\tilde{z}, d\tilde{w}) = g(z, w, \tilde{z}, \tilde{w}) P_0(z, d\tilde{z}) Q(w, d\tilde{w}).$$

On the other hand, using the Bayes formula, we can compute for $A \in \mathcal{F}_+^Y$

$$R^X(X_0, Y_0, X_1, A \circ \Theta) = \mathbf{P}(\mathbf{P}^{X_1, Y_1}(A) | \sigma(X_0, Y_0, X_1)) = \mathbf{P}^{\nu_{X_0, Y_0, X_1}}(A),$$

where we have defined

$$\nu_{z, w, z'}(d\tilde{z}, d\tilde{w}) = \frac{g(z, w, z', \tilde{w})}{\int g(z, w, z', w')Q(w, dw')} \delta_{z'}(d\tilde{z})Q(w, d\tilde{w}).$$

It therefore suffices to show that $\mathbf{P}^{\nu_{z,w,z'}}|_{\mathcal{F}_+^Y} \sim \mathbf{P}^{\nu_{z,w}}|_{\mathcal{F}_+^Y}$ for $(z, w, z') \in H$ with $\mathbf{P}((X_0, Y_0, X_1) \in H) = 1$. By Proposition 3.5, it suffices to show that

$$\liminf_{n \to \infty} \| \mathbf{P}^{\nu_{z,w,z'}}(X_n \in \cdot) - \mathbf{P}^{\nu_{z,w}}(X_n \in \cdot) \|_{TV} = 0$$

for $(z, w, z') \in H$ with $\mathbf{P}((X_0, Y_0, X_1) \in H) = 1$. But

$$\mathbf{E}\big(\|\mathbf{P}^{\nu_{X_0, Y_0, X_1}}(X_n \in \cdot) - \mathbf{P}^{\nu_{X_0, Y_0}}(X_n \in \cdot)\|_{TV}\big)$$

$$= \mathbf{E}\big(\big\|\mathbf{P}(X_{n+1} \in \cdot \,|\, X_0, Y_0, X_1) - \mathbf{P}(X_{n+1} \in \cdot \,|\, X_0, Y_0)\big\|_{TV}\big)$$

$$\leq \mathbf{E}\big(\big\|\mathbf{P}(X_{n+1} \in \cdot \,|\, X_0, Y_0, X_1, Y_1) - \mathbf{P}(X_{n+1} \in \cdot)\big\|_{TV}\big)$$

$$\quad + \mathbf{E}\big(\big\|\mathbf{P}(X_{n+1} \in \cdot \,|\, X_0, Y_0) - \mathbf{P}(X_{n+1} \in \cdot)\big\|_{TV}\big)$$

$$= \mathbf{E}\big(\big\|\mathbf{P}(X_n \in \cdot \,|\, X_0, Y_0) - \mathbf{P}(X_n \in \cdot)\big\|_{TV}\big)$$

$$\quad + \mathbf{E}\big(\big\|\mathbf{P}(X_{n+1} \in \cdot \,|\, X_0, Y_0) - \mathbf{P}(X_{n+1} \in \cdot)\big\|_{TV}\big),$$

where we have used the triangle inequality and the stationarity of $\mathbf{P}$. Thus the result follows from Assumption 2.6 and Fatou's lemma. $\quad\square$

4.3. *Exchange of intersection and supremum of σ-fields.* Fix a version $\varpi^+ : \Omega^Y \times \mathcal{B}(E) \to [0, 1]$ of the regular conditional probability $\mathbf{P}(X_0 \in \cdot \,|\, \mathcal{F}_+^Y)$. We begin by establishing the validity of the exchange of intersection and supremum in Theorem 2.9 assuming that ϖ^+ has a positive density with respect to ϖ.

PROPOSITION 4.6. *Suppose Assumptions 2.6 and 2.8 hold, and that there exists a strictly positive measurable function $k : \Omega^Y \times E \to \,]0, \infty[$ such that*

$$\varpi(y, A) = \int I_A(z)k(y, z)\varpi^+(y, dz) \qquad \text{for all } A \in \mathcal{B}(E)$$

for $\mathbf{P}^Y$-a.e. $y \in \Omega^Y$. Then $\bigcap_{n \geq 0} \mathcal{F}_+^Y \vee \mathcal{F}_{[n, \infty[}^X = \mathcal{F}_+^Y$ $\mathbf{P}$-a.s.

PROOF. By Theorem 4.3, there is a set H of $(\varpi \otimes \varpi)\mathbf{P}^Y$-full measure with

$$\mathbf{P}_{z, y}(A) = \mathbf{P}_{z, y}(A)^2 = \mathbf{P}_{z', y}(A) \qquad \text{for all } A \in \mathcal{T}^X \text{ and } (z, z', y) \in H.$$

As H has $(\varpi \otimes \varpi)\mathbf{P}^Y$-full measure, there clearly exists a set $H^Y \in \mathcal{B}(\Omega^Y)$ of $\mathbf{P}^Y$-full measure such that $\int I_H(z, z', y)\varpi(y, dz)\varpi(y, dz') = 1$ for all $y \in H^Y$. Let us now define $\mathbf{P}_y(A) = \int \mathbf{P}_{z, y}(A)\varpi(y, dz)$. Then

$$\mathbf{P}_y(A) - \mathbf{P}_y(A)^2 = \int I_H(z, z', y)\mathbf{P}_{z, y}(A)\big(1 - \mathbf{P}_{z', y}(A)\big)\varpi(y, dz)\varpi(y, dz') = 0$$

for every $y \in H^Y$ and $A \in \mathcal{T}^X$. Thus $\mathcal{T}^X$ is $\mathbf{P}_y$-trivial for all $y \in H^Y$. Therefore, defining $\mathbf{P}_y^+(A) = \int \mathbf{P}_{z,y}(A) \varpi^+(y, dz)$, our assumption that $\varpi^+(y, \cdot) \sim \varpi(y, \cdot)$ $\mathbf{P}^Y$-a.e. $y \in \Omega^Y$ implies that $\mathcal{T}^X$ is $\mathbf{P}_y^+$-trivial for $\mathbf{P}^Y$-a.e. $y \in \Omega^Y$.

Now recall that, by definition, $\mathbf{P}_{z,y}$ is a version of the regular conditional probability $\mathbf{P}((X_k)_{k \geq 0} \in \cdot \, | \sigma(X_0) \vee \mathcal{F}^Y)$. But as $\mathcal{F}_{-1}^Y$ is conditionally independent of $\mathcal{F}_+$ given $\sigma(X_0, Y_0)$ by the Markov property, it follows that $\mathbf{P}_{z,y}$ is even a version of $\mathbf{P}((X_k)_{k \geq 0} \in \cdot \, | \sigma(X_0) \vee \mathcal{F}_+^Y)$. It therefore follows that $\mathbf{P}_y^+$ is a version of the regular conditional probability $\mathbf{P}((X_k)_{k \geq 0} \in \cdot \, | \mathcal{F}_+^Y)$. We have therefore shown that $\mathcal{T}^X$ is $\mathbf{P}(\cdot \, | \mathcal{F}_+^Y)$-trivial $\mathbf{P}$-a.s., which implies

$$\bigcap_{n \geq 0} \mathcal{F}_+^Y \vee \mathcal{F}_{[n, \infty[}^X = \mathcal{F}_+^Y, \qquad \mathbf{P}\text{-a.s.}$$

by Lemma A.4 in Appendix A. $\quad\square$

To prove Theorem 2.9, we must therefore establish that ϖ^+ has a positive density with respect to ϖ. It is here that the time-reversed Assumption 2.7 enters the picture; indeed, the alert reader will not have failed to notice that we have only used Assumptions 2.6 and 2.8 up to this point.

LEMMA 4.7.   *Suppose that Assumptions 2.6–2.8 are in force. Then there exists a strictly positive measurable function $k : \Omega^Y \times E \to \,]0, \infty[$ such that*

$$\varpi(y, A) = \int I_A(z) k(y, z) \varpi^+(y, dz) \qquad \text{for all } A \in \mathcal{B}(E)$$

for $\mathbf{P}^Y$-a.e. $y \in \Omega^Y$.

PROOF.   By the Markov property of $(X_n, Y_n)_{n \in \mathbb{Z}}$, we find that $\mathbf{P}^{z, y_0}((Y_k)_{k < 0} \in \cdot)$ and $\mathbf{P}^{\varpi^+(y, \cdot) \otimes \delta_{y_0}}((Y_k)_{k < 0} \in \cdot)$ are versions of the regular conditional probabilities $\mathbf{P}((Y_k)_{k < 0} \in \cdot \, | \sigma(X_0) \vee \mathcal{F}_+^Y)$ and $\mathbf{P}((Y_k)_{k < 0} \in \cdot \, | \mathcal{F}_+^Y)$, respectively. Applying Lemma A.2 to $((Y_k)_{k \geq 0}, X_0, (Y_k)_{k < 0})$, it suffices to show that

$$\mathbf{P}^{z, y_0}((Y_k)_{k < 0} \in \cdot) \sim \mathbf{P}^{\varpi^+(y, \cdot) \otimes \delta_{y_0}}((Y_k)_{k < 0} \in \cdot)$$

for $\varpi \mathbf{P}^Y$-a.e. (z, y). By Lemma 3.4, we may apply Proposition 3.5 to the reverse-time model. Therefore, it suffices to prove that

$$\liminf_{n \to \infty} \left\| \mathbf{P}^{z, y_0}(X_{-n} \in \cdot) - \mathbf{P}^{\varpi^+(y, \cdot) \otimes \delta_{y_0}}(X_{-n} \in \cdot) \right\|_{\mathrm{TV}} = 0$$

for $\varpi \mathbf{P}^Y$-a.e. (z, y). To this end, let us note that

$$\mathbf{E}\big(\left\| \mathbf{P}^{X_0, Y_0}(X_{-n} \in \cdot) - \mathbf{P}^{\varpi^+(Y, \cdot) \otimes \delta_{Y_0}}(X_{-n} \in \cdot) \right\|_{\mathrm{TV}} \big)$$

$$\leq \mathbf{E}\big(\left\| \mathbf{P}^{X_0, Y_0}(X_{-n} \in \cdot) - \mathbf{P}(X_{-n} \in \cdot) \right\|_{\mathrm{TV}} \big)$$

$$+ \mathbf{E}\big(\left\| \mathbf{P}(X_{-n} \in \cdot \, | \mathcal{F}_+^Y) - \mathbf{P}(X_{-n} \in \cdot) \right\|_{\mathrm{TV}} \big)$$

$$\leq 2 \mathbf{E}\big(\left\| \mathbf{P}^{X_0, Y_0}(X_{-n} \in \cdot) - \mathbf{P}(X_{-n} \in \cdot) \right\|_{\mathrm{TV}} \big).$$

Thus the result follows by Assumption 2.7 and Fatou's lemma. $\quad\square$

We now complete the proof of Theorem 2.9.

PROOF. The first part of Theorem 2.9 follows immediately from Proposition 4.6 and Lemma 4.7. Now note that by Lemma 3.4, Assumptions 2.6–2.8 still hold if we replace the model $(X_n, Y_n)_{n \in \mathbb{Z}}$ by the time-reversed model $(X_{-n}, Y_{-n})_{n \in \mathbb{Z}}$. Therefore, the second part of Theorem 2.9 follows immediately from the first part by time reversal. $\quad\square$

5. Proof of Theorem 2.10. The goal of this section is to prove Theorem 2.10. We begin by recalling some basic properties of the filter. Then, we prove Theorem 2.10 first for a special case, then in the general case by a recursive argument.

5.1. *Preliminaries.* Recall that Π_n^μ is defined as a version of the regular conditional probability $\mathbf{P}^\mu(X_n \in \cdot \, | \mathcal{F}_{[0,n]}^Y)$. Of course, we are free to choose an arbitrary version of the filter, as the statement of Theorem 2.10 does not depend on the choice of version (this follows from Corollary 3.6). Nonetheless, we will find it convenient in our proofs to work with specific versions of these regular conditional probabilities, which we define presently.

For notational simplicity, we introduce the following device: for every probability measure ρ on $E \times F$, we fix a probability kernel $\rho. : F \times \mathcal{B}(E) \to [0, 1]$ such that $\rho_{Y_0}(A) = \mathbf{P}^\rho(X_0 \in A | Y_0)$ for all $A \in \mathcal{B}(E)$ [i.e., $\rho.$ is a version of the regular conditional probability $\mathbf{P}^\rho(X_0 \in \cdot \, | Y_0)$].

LEMMA 5.1. *Suppose that assumption* 2.8 *holds. For every probability measure μ on $E \times F$, we define a sequence of probability kernels $\Pi_n^\mu : \Omega^Y \times \mathcal{B}(E) \to [0, 1]$ ($n \geq 0$) through the following recursion:*

$$\Pi_n^\mu(y, A) = \frac{\int I_A(z) g(z', y(n-1), z, y(n)) P_0(z', dz) \Pi_{n-1}^\mu(y, dz')}{\int g(z', y(n-1), z, y(n)) P_0(z', dz) \Pi_{n-1}^\mu(y, dz')},$$

$$\Pi_0^\mu(y, A) = \mu_{y(0)}(A).$$

Then Π_n^μ is a version of the regular conditional probability $\mathbf{P}^\mu(X_n \in \cdot \, | \mathcal{F}_{[0,n]}^Y)$ for every $n \geq 0$. Moreover, $\Pi_n^\mu(y, \cdot) \sim \mathbf{P}^{\mu_{y(0)} \otimes \delta_{y(0)}}(X_n \in \cdot)$ for all y, n.

PROOF. By construction, we have

$\mathbf{P}^\mu(X_0 \in dx_0, \ldots, X_n \in dx_n, Y_0 \in dy_0, \ldots, Y_n \in dy_n)$

$$= \mu(E \times dy_0)\mu_{y_0}(dx_0) \prod_{i=0}^{n-1} g(x_i, y_i, x_{i+1}, y_{i+1}) P_0(x_i, dx_{i+1}) Q(y_i, dy_{i+1}).$$

Therefore, the Bayes formula gives for any $A \in \mathcal{B}(E)$

$$\mathbf{P}^{\mu}\big(X_n \in A | \mathcal{F}^Y_{[0,n]}\big)$$

$$= \frac{\int I_A(x_n) \mu_{Y_0}(dx_0) \prod_{i=0}^{n-1} g(x_i, Y_i, x_{i+1}, Y_{i+1}) P_0(x_i, dx_{i+1})}{\int \mu_{Y_0}(dx_0) \prod_{i=0}^{n-1} g(x_i, Y_i, x_{i+1}, Y_{i+1}) P_0(x_i, dx_{i+1})}.$$

This clearly coincides with the recursive definition of Π_n^{μ}. Moreover, it follows directly that $\Pi_n^{\mu}(y, \cdot) \sim \mu_{y(0)} P_0^n$ for all y, n. But note that

$$\mathbf{P}^{\mu_w \otimes \delta_w}(X_n \in A) = \int I_A(x_n) \mu_w(dx_0) f(w, x_0, \ldots, x_n) \prod_{i=0}^{n-1} P_0(x_i, dx_{i+1}),$$

where we have defined

$$f(y_0, x_0, \ldots, x_n) = \int \prod_{i=0}^{n-1} g(x_i, y_i, x_{i+1}, y_{i+1}) Q(y_i, dy_{i+1}).$$

Therefore, $\Pi_n^{\mu}(y, \cdot) \sim \mu_{y(0)} P_0^n \sim \mathbf{P}^{\mu_{y(0)} \otimes \delta_{y(0)}}(X_n \in \cdot)$ for every y, n. $\square$

Throughout the remainder of this section, the nonlinear filter Π_n^{μ} will always be assumed to be chosen according to the particular version defined in Lemma 5.1. This entails no loss of generality in our final results.

REMARK 5.2.   From the recursive formula for Π_n^{μ}, we can read off that

$$\Pi_{n+m}^{\mu}(y, A) = \Pi_m^{\Pi_n^{\mu}(y, \cdot) \otimes \delta_{y(n)}}(\Theta^n y, A) \qquad \text{for all } n, m \geq 0, \; y \in \Omega^Y, \; A \in \mathcal{B}(E).$$

This recursive property will play an important role in our proof. One of the advantages of our specific choice of version of the filter is that this property holds pathwise, so that we need not worry about the joint measurability of $\Pi_n^{\mu}(y, \cdot)$ with respect to (y, μ). Of course, our choice of version is not essential and technicalities of this kind could certainly be resolved more generally if one were so inclined.

5.2. *The absolutely continuous case.*   We begin by obtaining an explicit formula for the limit of $\|\Pi_n^{\mu} - \Pi_n^{\nu}\|_{\mathrm{TV}}$ for absolutely continuous measures $\mu \ll \nu$. This result will be applied recursively in the proof of Theorem 2.10.

PROPOSITION 5.3.   *For any probability measures μ, ν on $E \times F$ with $\mu \ll \nu$*

$$\mathbf{E}^{\mu}\Big[\limsup_{n \to \infty} \|\Pi_n^{\mu} - \Pi_n^{\nu}\|_{\mathrm{TV}}\Big]$$

$$= \mathbf{E}^{\nu}\bigg[\bigg|\mathbf{E}^{\nu}\Big(\frac{d\mu}{d\nu}(X_0, Y_0)\Big| \bigcap_{n \geq 0} \mathcal{F}^Y_+ \vee \mathcal{F}^X_{[n,\infty[}\Big) - \mathbf{E}^{\nu}\Big(\frac{d\mu}{d\nu}(X_0, Y_0)\Big| \mathcal{F}^Y_+\Big)\bigg|\bigg].$$

PROOF.   As $d\mathbf{P}^\mu/d\mathbf{P}^\nu = (d\mu/d\nu)(X_0, Y_0)$ by the Markov property, we have

$$\mathbf{P}^\mu\big(X_n \in A | \mathcal{F}^Y_{[0,n]}\big) = \frac{\mathbf{E}^\nu(I_A(X_n)\mathbf{E}^\nu((d\mu/d\nu)(X_0, Y_0)|\sigma(X_n) \vee \mathcal{F}^Y_{[0,n]})|\mathcal{F}^Y_{[0,n]})}{\mathbf{E}^\nu((d\mu/d\nu)(X_0, Y_0)|\mathcal{F}^Y_{[0,n]})}$$

$\mathbf{P}^\mu$-a.s. by the Bayes formula. Therefore, we evidently have

$$\frac{d\Pi^\mu_n}{d\Pi^\nu_n}(X_n) = \frac{\mathbf{E}^\nu((d\mu/d\nu)(X_0, Y_0)|\sigma(X_n) \vee \mathcal{F}^Y_{[0,n]})}{\mathbf{E}^\nu((d\mu/d\nu)(X_0, Y_0)|\mathcal{F}^Y_{[0,n]})}, \qquad \mathbf{P}^\mu\text{-a.s.}$$

In particular, we can write $\mathbf{P}^\mu$-a.s.

$$\|\Pi^\mu_n - \Pi^\nu_n\|_{\mathrm{TV}} = \int \left|\frac{d\Pi^\mu_n}{d\Pi^\nu_n}(x) - 1\right| \Pi^\nu_n(dx) = \frac{\mathbf{E}^\nu(M_n|\mathcal{F}^Y_{[0,n]})}{\mathbf{E}^\nu((d\mu/d\nu)(X_0, Y_0)|\mathcal{F}^Y_{[0,n]})},$$

where we have defined

$$M_n = \left|\mathbf{E}^\nu\left(\frac{d\mu}{d\nu}(X_0, Y_0)\Big|\sigma(X_n) \vee \mathcal{F}^Y_{[0,n]}\right) - \mathbf{E}^\nu\left(\frac{d\mu}{d\nu}(X_0, Y_0)\Big|\mathcal{F}^Y_{[0,n]}\right)\right|.$$

Thus it is easily seen that

$$\mathbf{E}^\mu\Big[\limsup_{n\to\infty}\|\Pi^\mu_n - \Pi^\nu_n\|_{\mathrm{TV}}\Big] = \mathbf{E}^\nu\Big[\limsup_{n\to\infty}\mathbf{E}^\nu(M_n|\mathcal{F}^Y_{[0,n]})\Big].$$

Now note that, by the Markov property, $\mathcal{F}_{[n+1,\infty[}$ and $\sigma(X_0) \vee \mathcal{F}^Y_{[0,n-1]}$ are conditionally independent given $\sigma(X_n, Y_n)$. Therefore,

$$M_n = \left|\mathbf{E}^\nu\left(\frac{d\mu}{d\nu}(X_0, Y_0)\Big|\mathcal{F}^Y_+ \vee \mathcal{F}^X_{[n,\infty[}\right) - \mathbf{E}^\nu\left(\frac{d\mu}{d\nu}(X_0, Y_0)\Big|\mathcal{F}^Y_{[0,n]}\right)\right|.$$

If $d\mu/d\nu$ were uniformly bounded, the result would follow directly from the martingale convergence theorem and Hunt's lemma ([9], Theorem V.45).

In the case that $d\mu/d\nu$ is unbounded, define the truncated process

$$M^k_n = \left|\mathbf{E}^\nu\left(\frac{d\mu}{d\nu}(X_0, Y_0) \wedge k\Big|\mathcal{F}^Y_+ \vee \mathcal{F}^X_{[n,\infty[}\right) - \mathbf{E}^\nu\left(\frac{d\mu}{d\nu}(X_0, Y_0) \wedge k\Big|\mathcal{F}^Y_{[0,n]}\right)\right|.$$

By Hunt's lemma and dominated convergence,

$$\lim_{k\to\infty}\lim_{n\to\infty}\mathbf{E}^\nu(M^k_n|\mathcal{F}^Y_{[0,n]}) = \mathbf{E}^\nu(M_\infty|\mathcal{F}^Y_+), \qquad \mathbf{P}^\nu\text{-a.s.,}$$

where $M_\infty = \lim_{n\to\infty} M_n$. Therefore, we obtain $\mathbf{P}^\nu$-a.s.

$$\limsup_{n\to\infty}\mathbf{E}^\nu(M_n|\mathcal{F}^Y_{[0,n]}) = \mathbf{E}^\nu(M_\infty|\mathcal{F}^Y_+) + \limsup_{k\to\infty}\limsup_{n\to\infty}\mathbf{E}^\nu(M_n - M^k_n|\mathcal{F}^Y_{[0,n]}).$$

It remains to note that the second term vanishes $\mathbf{P}^\nu$-a.s.,

$$\left|\limsup_{k\to\infty}\limsup_{n\to\infty}\mathbf{E}^\nu(M_n - M^k_n|\mathcal{F}^Y_{[0,n]})\right|$$

$$\leq 2\limsup_{k\to\infty}\limsup_{n\to\infty}\mathbf{E}^\nu\left(\frac{d\mu}{d\nu}(X_0, Y_0) - \frac{d\mu}{d\nu}(X_0, Y_0) \wedge k\Big|\mathcal{F}^Y_{[0,n]}\right) = 0.$$

The proof is complete. $\square$

5.3. *The general case.* In the special case where $\mu \ll \pi$, Theorem 2.10 follows directly from Proposition 5.3 and Theorem 2.9. An additional step is needed, however, to prove Theorem 2.9 in the general case.

LEMMA 5.4. *Let* μ, ρ *be probability measures on* E, *and choose* $S \in \mathcal{B}(E)$ *such that* $\mu(S) > 0$. *Define the probability measure* $\nu = \mu(\cdot \cap S)/\mu(S)$. *Then*

$$\|\Pi_n^{\mu \otimes \delta_w} - \Pi_n^{\rho \otimes \delta_w}\|_{\mathrm{TV}} \leq 2\mathbf{P}^{\mu \otimes \delta_w}(X_0 \notin S|\mathcal{F}_{[0,n]}^Y)$$
$$+ \mathbf{P}^{\mu \otimes \delta_w}(X_0 \in S|\mathcal{F}_{[0,n]}^Y)\|\Pi_n^{\nu \otimes \delta_w} - \Pi_n^{\rho \otimes \delta_w}\|_{\mathrm{TV}}$$

$\mathbf{P}^{\mu \otimes \delta_w}$*-a.s. for any* $w \in F$.

PROOF. If $\mu(S) = 1$, the proof is trivial. Otherwise, by the Bayes formula,

$$\Pi_n^{\mu \otimes \delta_w} = \mathbf{P}^{\mu \otimes \delta_w}(X_0 \in S|\mathcal{F}_{[0,n]}^Y)\Pi_n^{\nu \otimes \delta_w} + \mathbf{P}^{\mu \otimes \delta_w}(X_0 \notin S|\mathcal{F}_{[0,n]}^Y)\Pi_n^{\nu^\perp \otimes \delta_w}$$

$\mathbf{P}^{\mu \otimes \delta_w}$-a.s., where $\nu^\perp = \mu(\cdot \cap S^c)/\mu(S^c)$. But obviously

$$\Pi_n^{\rho \otimes \delta_w} = \mathbf{P}^{\mu \otimes \delta_w}(X_0 \in S|\mathcal{F}_{[0,n]}^Y)\Pi_n^{\rho \otimes \delta_w} + \mathbf{P}^{\mu \otimes \delta_w}(X_0 \notin S|\mathcal{F}_{[0,n]}^Y)\Pi_n^{\rho \otimes \delta_w}$$

$\mathbf{P}^{\mu \otimes \delta_w}$-a.s., so the result follows directly. $\square$

REMARK 5.5. Even though we have fixed a version of the filter $\Pi_n^{\rho \otimes \delta_w}$, our results should ultimately not depend on the choice of version. In this light, Lemma 5.4 may appear somewhat suspicious as the regular conditional probability $\mathbf{P}^{\rho \otimes \delta_w}(X_n \in \cdot |\mathcal{F}_{[0,n]}^Y)$ is not $\mathbf{P}^{\mu \otimes \delta_w}$-a.s. uniquely defined. However, there is no problem here, as the proof shows that the inequality in Lemma 5.4 holds for *any* choice of version, even though different versions may be inequivalent. On the other hand, we will ultimately apply this result only when $\nu \ll \rho$, in which case the expression is in fact independent of the choice of version.

The idea is now to apply the recursive property of the filter:

$$\|\Pi_{m+n}^\mu - \Pi_{m+n}^\pi\|_{\mathrm{TV}} = \|\Pi_n^{\Pi_m^\mu \otimes \delta_{Y_m}}(Y \circ \Theta^m, \cdot) - \Pi_n^{\Pi_m^\pi \otimes \delta_{Y_m}}(Y \circ \Theta^m, \cdot)\|_{\mathrm{TV}}$$

for any $m \geq 0$. As $\Pi_m^\mu \sim \mathbf{P}^\mu(X_m \in \cdot |Y_0)$ and $\Pi_m^\pi \sim \mathbf{P}(X_m \in \cdot |Y_0)$ by Lemma 5.1, the assumption of Theorem 2.10 guarantees that the singular part of Π_m^μ with respect to Π_m^π vanishes as $m \to \infty$. We can therefore use Lemma 5.4 to replace Π_m^μ by its absolutely continuous part, so that we have reduced the limit as $n \to \infty$ to the special case of Proposition 5.3. In order to apply Proposition 5.3, however, we will require one additional result.

LEMMA 5.6. *Suppose that Assumptions 2.6–2.8 hold. Then for any* $m \geq 0$

$$\bigcap_{n \geq 0} \mathcal{F}_+^Y \vee \mathcal{F}_{[n,\infty[}^X = \mathcal{F}_+^Y, \qquad \mathbf{P}^{\Pi_m^\pi(y,\cdot) \otimes \delta_{y(m)}}\text{-a.s.} \qquad \text{for } \mathbf{P}^Y\text{-a.e. } y.$$

PROOF.   As in the proof of Proposition 4.6, it suffices to establish that $\mathcal{T}^X$ is $\mathbf{P}^{\Pi_m^\pi(y,\cdot)\otimes\delta_{y(m)}}(\cdot\,|\mathcal{F}_+^Y)$-trivial $\mathbf{P}^{\Pi_m^\pi(y,\cdot)\otimes\delta_{y(m)}}$-a.s. for $\mathbf{P}^Y$-a.e. y. Note that

$$\mathbf{P}^{\Pi_m^\pi(Y,\cdot)\otimes\delta_{Y_m}}(A) = \mathbf{E}(\mathbf{P}^{X_m,Y_m}(A)|\mathcal{F}_{[0,m]}^Y) = \mathbf{P}(A\circ\Theta^m|\mathcal{F}_{[0,m]}^Y)$$

for all $A\in\mathcal{F}_+$ by the Markov property. Therefore,

$$\mathbf{P}^{\Pi_m^\pi(Y\circ\Theta^{-m},\cdot)\otimes\delta_{Y_0}}(A) = \mathbf{P}(A\circ\Theta^m|\mathcal{F}_{[0,m]}^Y)\circ\Theta^{-m} = \mathbf{P}(A|\mathcal{F}_{[-m,0]}^Y),$$

where we used that $\mathbf{P}$ is stationary. It follows that $\mathbf{P}^{\Pi_m^\pi(Y\circ\Theta^{-m},\cdot)\otimes\delta_{Y_0}}|_{\mathcal{F}_+}$ is a version of $\mathbf{P}((X_n,Y_n)_{n\geq0}\in\cdot\,|\mathcal{F}_{[-m,0]}^Y)$. By Lemma A.3

$$\mathbf{P}((X_n)_{n\geq0}\in\cdot\,|\mathcal{F}_{[-m,\infty[}^Y) = \mathbf{P}^{\Pi_m^\pi(Y\circ\Theta^{-m},\cdot)\otimes\delta_{Y_0}}((X_n)_{n\geq0}\in\cdot\,|\mathcal{F}_+^Y), \qquad \mathbf{P}\text{-a.s.}$$

Thus it suffices to show that $\mathcal{T}^X$ is $\mathbf{P}(\cdot\,|\mathcal{F}_{[-m,\infty[}^Y)$-trivial $\mathbf{P}$-a.s., which is equivalent (by virtue of Lemma A.4 in Appendix A) to

$$\bigcap_{n\geq0}\mathcal{F}_{[-m,\infty[}^Y\vee\mathcal{F}_{[n,\infty[}^X = \mathcal{F}_{[-m,\infty[}^Y, \qquad \mathbf{P}\text{-a.s.}$$

But this follows directly from Theorem 2.9 and the stationarity of $\mathbf{P}$.   □

We can now complete the proof of Theorem 2.10.

PROOF.   By the recursive property of the filter,

$$\limsup_{k\to\infty}\|\Pi_k^\mu - \Pi_k^\pi\|_{\mathrm{TV}}$$

$$= \limsup_{k\to\infty}\|\Pi_k^{\Pi_n^\mu(Y,\cdot)\otimes\delta_{Y_n}}(Y\circ\Theta^n,\cdot) - \Pi_k^{\Pi_n^\pi(Y,\cdot)\otimes\delta_{Y_n}}(Y\circ\Theta^n,\cdot)\|_{\mathrm{TV}}$$

for all $n\geq0$. Therefore, we obtain $\mathbf{P}^\mu$-a.s.

$$\mathbf{E}^\mu\Big(\limsup_{k\to\infty}\|\Pi_k^\mu - \Pi_k^\pi\|_{\mathrm{TV}}|\mathcal{F}_{[0,n]}^Y\Big)$$

$$= \mathbf{E}^{\Pi_n^\mu(y,\cdot)\otimes\delta_{y(n)}}\Big(\limsup_{k\to\infty}\|\Pi_k^{\Pi_n^\mu(y,\cdot)\otimes\delta_{y(n)}} - \Pi_k^{\Pi_n^\pi(y,\cdot)\otimes\delta_{y(n)}}\|_{\mathrm{TV}}\Big)\Big|_{y=Y},$$

where we have used that $\Pi_n^\mu(Y,\cdot)$ and $\Pi_n^\pi(Y,\cdot)$ are $\mathcal{F}_{[0,n]}^Y$-measurable.
    To proceed, let us first recall that

$$\Pi_n^\mu(y,\cdot)\sim\mathbf{P}^{\mu_{y(0)}\otimes\delta_{y(0)}}(X_n\in\cdot) \quad\text{and}\quad \Pi_n^\pi(y,\cdot)\sim\mathbf{P}^{\pi_{y(0)}\otimes\delta_{y(0)}}(X_n\in\cdot)$$

for all y,n by Lemma 5.1. Choose a set $S_n\in\mathcal{B}(E\times F)$ such that

$$\mathbf{P}^{\mu_w\otimes\delta_w}(X_n\in\cdot\cap S_n(w)) \ll \mathbf{P}^{\pi_w\otimes\delta_w}(X_n\in\cdot)$$

and

$$\mathbf{P}^{\pi_w\otimes\delta_w}(X_n\in S_n(w)) = 1$$

for all $w \in F$, where $I_{S_n(w)}(z) = I_{S_n}(z, w)$ (the existence of such a set follows from the Lebesgue decomposition for kernels; [9], Section V.58). Define

$$\Sigma_n(y, \cdot) = \Pi_n^\mu\big(y, \cdot \cap S_n(y(0))\big) / \Pi_n^\mu\big(y, S_n(y(0))\big).$$

Then clearly $\Sigma_n(y, \cdot) \ll \Pi_n^\pi(y, \cdot)$ for all y, and by Lemma 5.4

$$\mathbf{E}^{\Pi_n^\mu(y, \cdot) \otimes \delta_{y(n)}} \Big(\limsup_{k \to \infty} \big\| \Pi_k^{\Pi_n^\mu(y, \cdot) \otimes \delta_{y(n)}} - \Pi_k^{\Pi_n^\pi(y, \cdot) \otimes \delta_{y(n)}} \big\|_{\mathrm{TV}} \Big)$$

$$\leq 2\mathbf{P}^{\Pi_n^\mu(y, \cdot) \otimes \delta_{y(n)}} \big(X_0 \notin S_n(y(0)) \big)$$

$$+ \mathbf{E}^{\Sigma_n(y, \cdot) \otimes \delta_{y(n)}} \Big(\limsup_{k \to \infty} \big\| \Pi_k^{\Sigma_n(y, \cdot) \otimes \delta_{y(n)}} - \Pi_k^{\Pi_n^\pi(y, \cdot) \otimes \delta_{y(n)}} \big\|_{\mathrm{TV}} \Big).$$

The last term vanishes for $\mathbf{P}^Y$-a.e. y by Proposition 5.3 and Lemma 5.6, hence, for $\mathbf{P}^\mu(Y \in \cdot)$-a.e. y by Corollary 3.6. We have therefore shown that

$$\mathbf{E}^\mu \Big(\limsup_{k \to \infty} \big\| \Pi_k^\mu - \Pi_k^\pi \big\|_{\mathrm{TV}} | \mathcal{F}_{[0,n]}^Y \Big) \leq 2\mathbf{P}^\mu \big(X_n \notin S_n(Y_0) | \mathcal{F}_{[0,n]}^Y \big), \qquad \mathbf{P}^\mu\text{-a.s.}$$

for every $n \geq 0$. In particular, we have

$$\mathbf{E}^\mu \Big(\limsup_{k \to \infty} \big\| \Pi_k^\mu - \Pi_k^\pi \big\|_{\mathrm{TV}} \Big) \leq 2\mathbf{P}^\mu \big(X_n \notin S_n(Y_0) \big) \qquad \text{for all } n \geq 0.$$

But as $\mathbf{P}(X_n \notin S_n(Y_0) | Y_0) = \mathbf{P}^{\pi_{Y_0} \otimes \delta_{Y_0}}(X_n \notin S_n(Y_0)) = 0$, we obtain

$$\mathbf{P}^\mu \big(X_n \notin S_n(Y_0) \big) = \mathbf{E}^\mu \big(\mathbf{P}^\mu \big(X_n \notin S_n(Y_0) | Y_0 \big) - \mathbf{P} \big(X_n \notin S_n(Y_0) | Y_0 \big) \big)$$

$$\leq \mathbf{E}^\mu \big(\big\| \mathbf{P}^\mu (X_n \in \cdot | Y_0) - \mathbf{P}(X_n \in \cdot | Y_0) \big\|_{\mathrm{TV}} \big) \xrightarrow{n \to \infty} 0,$$

where convergence follows as in the proof of Corollary 3.6. Therefore,

$$\limsup_{k \to \infty} \big\| \Pi_k^\mu - \Pi_k^\pi \big\|_{\mathrm{TV}} = 0, \qquad \mathbf{P}^\mu\text{-a.s.},$$

which completes the main part of the proof. To obtain $\mathbf{P}$-a.s. convergence (rather than $\mathbf{P}^\mu$-a.s. convergence) in the case where $\mu(E \times \cdot) \sim \pi(E \times \cdot)$, it suffices to note that in this case $\mathbf{P}^\mu|_{\mathcal{F}_+^Y} \sim \mathbf{P}|_{\mathcal{F}_+^Y}$ by Corollary 3.6. $\square$

6. Proof of Theorem 2.12.

The goal of this section is to prove Theorem 2.12. We begin by developing some details of the basic properties of $(\Pi_n^\mu, Y_n)_{n \geq 0}$ in Section 2.3 under Assumption 2.8. We then complete the proof of Theorem 2.12.

6.1. *Markov property of the pair* $(\Pi_n^\mu, Y_n)_{n \geq 0}$. Throughout this section, we assume that Assumption 2.8 is in force. We begin by defining a measurable map $U : \mathcal{P}(E) \times F \times F \to \mathcal{P}(E)$ as follows:

$$U(\nu, y_0, y_1)(A) = \frac{\int I_A(z) g(z', y_0, z, y_1) P_0(z', dz) \nu(dz')}{\int g(z', y_0, z, y_1) P_0(z', dz) \nu(dz')}.$$

It follows immediately from Lemma 5.1 that $\Pi_n^\mu = U(\Pi_{n-1}^\mu, Y_{n-1}, Y_n)$ $\mathbf{P}^\mu$-a.s. for every $n \geq 1$ and $\mu \in \mathcal{P}(E \times F)$.

Now define the transition kernel $\Gamma : \mathcal{P}(E) \times F \times \mathcal{B}(\mathcal{P}(E) \times F) \to [0, 1]$ as

$$\Gamma(\nu, y_0, A) = \int I_A(U(\nu, y_0, y_1), y_1) P(z, y_0, dz', dy_1)\nu(dz).$$

Then we have the following lemma.

LEMMA 6.1. *Suppose that Assumption 2.8 holds. Then the $(\mathcal{P}(E) \times F)$-valued process $(\Pi_n^\mu, Y_n)_{n \geq 0}$ is Markov under $\mathbf{P}^\mu$ with transition kernel Γ.*

PROOF.   It suffices to note that (Π_n^μ, Y_n) is $\mathcal{F}_{[0,n]}^Y$-measurable and

$$\mathbf{P}^\mu\big((\Pi_{n+1}^\mu, Y_{n+1}) \in A | \mathcal{F}_{[0,n]}^Y\big)$$

$$= \mathbf{P}^\mu\big((U(\Pi_n^\mu, Y_n, Y_{n+1}), Y_{n+1}) \in A | \mathcal{F}_{[0,n]}^Y\big)$$

$$= \int I_A(U(\Pi_n^\mu, Y_n, w), w) P(z, Y_n, dz', dw) \Pi_n^\mu(dz) = \Gamma(\Pi_n^\mu, Y_n, A)$$

for every $A \in \mathcal{B}(\mathcal{P}(E) \times F)$. $\square$

We can now establish some basic properties of Γ-invariant measures.

LEMMA 6.2.   *Suppose that Assumption 2.8 holds. Then for any Γ-invariant probability measure m, the barycenter $b\mathsf{m}$ is a P-invariant measure. Conversely, there is at least one Γ-invariant measure with barycenter π.*

PROOF.   First, let $\mathsf{m} \in \mathcal{P}(\mathcal{P}(E) \times F)$ be a Γ-invariant measure. Then

$$b\mathsf{m}(A \times B) = \int \nu(A) I_B(w) \Gamma(\nu', w', d\nu, dw) \mathsf{m}(d\nu', dw')$$

$$= \int U(\nu', w', w)(A) I_B(w) P(z, w', d\tilde{z}, dw)\nu'(dz)\mathsf{m}(d\nu', dw')$$

$$= \int \frac{\int I_A(\tilde{z}) g(z, w', \tilde{z}, w) P_0(z, d\tilde{z})\nu'(dz)}{\int g(z, w', \tilde{z}, w) P_0(z, d\tilde{z})\nu'(dz)}$$

$$\times g(z, w', \tilde{z}, w) P_0(z, d\tilde{z})\nu'(dz) I_B(w) Q(w', dw)\mathsf{m}(d\nu', dw')$$

$$= \int P(z, w', A \times B)\nu'(dz)\mathsf{m}(d\nu', dw')$$

$$= \int P(z, w', A \times B) b\mathsf{m}(dz, dw').$$

Thus the barycenter $b\mathsf{m}$ is P-invariant.

To prove the converse, let Π_n be a version of the regular conditional probability $\mathbf{P}(X_n \in \cdot | \mathcal{F}_n^Y)$, and let $\Pi_{k,n}$ be a version of the regular conditional probability

$\mathbf{P}(X_n \in \cdot | \mathcal{F}_{[k,n]}^Y)$. Applying the Bayes formula as in the proof of Lemma 5.1, we find that $U(\Pi_{k,n}, Y_n, Y_{n+1}) = \Pi_{k,n+1}$ $\mathbf{P}$-a.s. for every $k \le n$. By the martingale convergence theorem, it follows directly that

$$U(\Pi_n, Y_n, Y_{n+1})(A) = \lim_{k \to -\infty} U(\Pi_{k,n}, Y_n, Y_{n+1})(A) = \Pi_{n+1}(A)$$

$\mathbf{P}$-a.s. for every $A \in \mathcal{B}(E)$. As $\mathcal{B}(E)$ is countably generated, a standard monotone class argument shows that $U(\Pi_n, Y_n, Y_{n+1}) = \Pi_{n+1}$ $\mathbf{P}$-a.s. Therefore, the proof of Lemma 6.1 shows that $(\Pi_n, Y_n)_{n \in \mathbb{Z}}$ is Markov under $\mathbf{P}$ with transition kernel Γ. But as $\mathbf{P}$ is stationary, the process $(\Pi_n, Y_n)_{n \in \mathbb{Z}}$ is stationary also. Therefore, the law of (Π_0, Y_0) is a Γ-invariant measure whose barycenter is obviously π. $\square$

6.2. *Uniqueness of the Γ-invariant measure.*   Given $\mathsf{m} \in \mathcal{P}(\mathcal{P}(E) \times F)$, define the probability measure $\mathbf{P}_\mathsf{m}$ on the space $\mathcal{P}(E) \times E^\mathbb{N} \times F^\mathbb{N}$ as

$$\mathbf{P}_\mathsf{m}\big((m_0, X_0, \ldots, X_n, Y_0, \ldots, Y_n) \in A\big)$$
$$= \int I_A(\nu, x_0, \ldots, x_n, y_0, \ldots, y_n) \nu(dx_0) P(x_0, y_0, dx_1, dy_1)$$
$$\times \cdots \times P(x_{n-1}, y_{n-1}, dx_n, dy_n) \mathsf{m}(d\nu, dy_0).$$

We now choose regular versions of the following conditional probabilites:

$$\Pi_n^{\min} = \mathbf{P}_\mathsf{m}\big(X_n \in \cdot | \mathcal{F}_{[0,n]}^Y\big),$$
$$\Pi_n^\mathsf{m} = \mathbf{P}_\mathsf{m}\big(X_n \in \cdot | \sigma(m_0) \vee \mathcal{F}_{[0,n]}^Y\big),$$
$$\Pi_n^{\max} = \mathbf{P}_\mathsf{m}\big(X_n \in \cdot | \sigma(m_0, X_0) \vee \mathcal{F}_{[0,n]}^Y\big).$$

The following result is straightforward.

LEMMA 6.3.   *The laws of $(\Pi_n^{\min}, Y_n)$ and $(\Pi_n^{\max}, Y_n)$ under $\mathbf{P}_\mathsf{m}$ coincide with the laws of $(\mathbf{P}^{b\mathsf{m}}(X_n \in \cdot | \mathcal{F}_{[0,n]}^Y), Y_n)$ and $(\mathbf{P}^{b\mathsf{m}}(X_n \in \cdot | \sigma(X_0) \vee \mathcal{F}_{[0,n]}^Y), Y_n)$ under $\mathbf{P}^{b\mathsf{m}}$, respectively. Moreover, the process $(\Pi_n^\mathsf{m}, Y_n)_{n \ge 0}$ is Markov under $\mathbf{P}_\mathsf{m}$ with transition kernel Γ and initial measure m.*

PROOF.   By definition of the barycenter, the law of $(X_n, Y_n)_{n \ge 0}$ under $\mathbf{P}_\mathsf{m}$ coincides with the law of $(X_n, Y_n)_{n \ge 0}$ under $\mathbf{P}^{b\mathsf{m}}$. Moreover, it is easily seen that $\Pi_n^{\max} = \mathbf{P}_\mathsf{m}(X_n \in \cdot | \sigma(X_0) \vee \mathcal{F}_{[0,n]}^Y)$ by the Markov property, so $\Pi_n^{\max}$ and $\Pi_n^{\min}$ depend on $(X_n, Y_n)_{n \ge 0}$ only. This establishes the first part of the result. The second part follows as in the proof of Lemma 6.1. $\square$

We can now complete the proof of Theorem 2.12.

PROOF.   Throughout the proof, let m be a fixed Γ-invariant probability measure with barycenter π. We will show that, by virtue of Theorem 2.9, this invariant

measure must necessarily coincide with the invariant measure obtained in the proof of Lemma 6.2.

Let $p \in \mathbb{N}$, choose arbitrary bounded measurable functions $f : E \to \mathbb{R}^p$ and $g : F \to \mathbb{R}$, and let $\kappa : \mathbb{R}^{p+1} \to \mathbb{R}$ be a convex function. Then κ is necessarily continuous, so the function $F : \mathcal{P}(E) \times F \to \mathbb{R}$ defined by

$$F(\nu, w) = \kappa \left(g(w), \int f(x) \nu(dx) \right)$$

is bounded and measurable. By Jensen's inequality,

$$\mathbf{E}_{\mathsf{m}}(F(\Pi_n^{\min}, Y_n)) \leq \mathbf{E}_{\mathsf{m}}(F(\Pi_n^{\mathsf{m}}, Y_n)) \leq \mathbf{E}_{\mathsf{m}}(F(\Pi_n^{\max}, Y_n))$$

for all $n \geq 0$. Therefore, by Lemma 6.3 and the Γ-invariance of m, we obtain

$$\mathbf{E}\big(\kappa\big(g(Y_n), \mathbf{E}\big(f(X_n)|\mathcal{F}_{[0,n]}^Y\big)\big)\big) \leq \int F(\nu, w) \mathsf{m}(d\nu, dw)$$
$$\leq \mathbf{E}\big(\kappa\big(g(Y_n), \mathbf{E}\big(f(X_n)|\sigma(X_0) \vee \mathcal{F}_{[0,n]}^Y\big)\big)\big).$$

But using stationarity of $\mathbf{P}$ and the Markov property of $(X_n, Y_n)_{n \in \mathbb{Z}}$,

$$\mathbf{E}\big(\kappa\big(g(Y_n), \mathbf{E}\big(f(X_n)|\mathcal{F}_{[0,n]}^Y\big)\big)\big) = \mathbf{E}\big(\kappa\big(g(Y_0), \mathbf{E}\big(f(X_0)|\mathcal{F}_{[-n,0]}^Y\big)\big)\big),$$
$$\mathbf{E}\big(\kappa\big(g(Y_n), \mathbf{E}\big(f(X_n)|\sigma(X_0) \vee \mathcal{F}_{[0,n]}^Y\big)\big)\big) = \mathbf{E}\big(\kappa\big(g(Y_0), \mathbf{E}\big(f(X_0)|\mathcal{F}_0^Y \vee \mathcal{F}_{-n}^X\big)\big)\big)$$

for all $n \geq 0$. Thus martingale convergence and Theorem 2.9 yield

$$\int F(\nu, w) \mathsf{m}(d\nu, dw) = \mathbf{E}(\kappa(g(Y_0), \mathbf{E}(f(X_0)|\mathcal{F}_0^Y))) = \int F(\nu, w) \mathsf{m}^0(d\nu, dw),$$

where m^0 denotes the distinguished Γ-invariant measure obtained in the proof of Lemma 6.2. But a standard approximation argument shows that class of functions of the form $F(\nu, w) = \kappa(g(w), \int f(x)\nu(dx))$ is measure-determining (see, e.g., the proof of Proposition A.7 [22]), so we can conclude that $\mathsf{m} = \mathsf{m}^0$. Thus we have shown that any Γ-invariant probability measure with barycenter π must coincide with m^0, which establishes uniqueness.

To complete the proof, it remains to consider the case when P has unique invariant probability measure (i.e., π is the only P-invariant probability measure). As the barycenter of any Γ-invariant probability measure must be P-invariant, this implies that any Γ-invariant measure must have barycenter π. Therefore, in this case, Γ has a unique invariant probability measure. $\square$

7. Proof of Theorem 2.13.   The goal of this section is to prove Theorem 2.13. We begin by developing some details of the basic properties of $(\Pi_n^\mu, X_n, Y_n)_{n \geq 0}$ in Section 2.3 under Assumption 2.8. We then complete the proof of Theorem 2.13.

7.1. *Markov property of the triple* $(\Pi_n^\mu, X_n, Y_n)_{n\geq 0}$. In this section we use the notation of Section 6.1, and we again assume that Assumption 2.8 is in force. Define the transition kernel $\Lambda : \mathcal{P}(E) \times E \times F \times \mathcal{B}(\mathcal{P}(E) \times E \times F) \to [0, 1]$ as

$$\Lambda(\nu, x_0, y_0, A) = \int I_A(U(\nu, y_0, y_1), x_1, y_1) P(x_0, y_0, dx_1, dy_1).$$

Then we have the following lemma.

LEMMA 7.1. *Suppose that Assumption 2.8 holds. Then* $(\Pi_n^\mu, X_n, Y_n)_{n\geq 0}$ *is a* $(\mathcal{P}(E) \times E \times F)$-*valued Markov chain under* $\mathbf{P}^\mu$ *with transition kernel* Λ.

PROOF. It suffices to note that (Π_n^μ, X_n, Y_n) is $\mathcal{F}_{[0,n]}$-measurable and

$$\mathbf{P}^\mu\big((\Pi_{n+1}^\mu, X_{n+1}, Y_{n+1}) \in A | \mathcal{F}_{[0,n]}\big)$$
$$= \int I_A(U(\Pi_n^\mu, Y_n, w), z, w) P(X_n, Y_n, dz, dw)$$
$$= \Lambda(\Pi_n^\mu, X_n, Y_n, A)$$

for every $A \in \mathcal{B}(\mathcal{P}(E) \times E \times F)$. □

For any probability measure $\mathsf{M} \in \mathcal{P}(\mathcal{P}(E) \times E \times F)$, we define probability measures $m\mathsf{M} \in \mathcal{P}(E \times F)$ and $\gamma\mathsf{M} \in \mathcal{P}(\mathcal{P}(E) \times F)$ as follows:

$$m\mathsf{M}(A \times B) = \mathsf{M}\big(\mathcal{P}(E) \times A \times B\big), \qquad \gamma\mathsf{M}(C \times B) = \mathsf{M}(C \times E \times B).$$

Moreover, we define the class

$$\mathfrak{M} = \Big\{\mathsf{M} \in \mathcal{P}\big(\mathcal{P}(E) \times E \times F\big) : \forall A \in \mathcal{B}(\mathcal{P}(E)), B \in \mathcal{B}(E), C \in \mathcal{B}(F),$$

$$\mathsf{M}(A \times B \times C) = \int \nu(B) I_{A \times C}(\nu, w) \mathsf{M}(d\nu, dz, dw)\Big\}.$$

We can now establish some basic properties of Λ-invariant measures.

LEMMA 7.2. *Suppose that Assumption 2.8 holds. Then for any* Λ-*invariant probability measure* M, *the marginal* $m\mathsf{M}$ *is a* P-*invariant measure. If, in addition,* $\mathsf{M} \in \mathfrak{M}$, *then* $\gamma\mathsf{M}$ *is a* Γ-*invariant measure with barycenter* $m\mathsf{M}$. *Conversely, there is at least one* Λ-*invariant* $\mathsf{M} \in \mathfrak{M}$ *with marginal* π.

PROOF. Let $\mathsf{M} \in \mathcal{P}(\mathcal{P}(E) \times E \times F)$ be a Λ-invariant probability measure. It is trivial that $m\mathsf{M}$ is P-invariant. Now suppose that also $\mathsf{M} \in \mathfrak{M}$. Then

$$\gamma\mathsf{M}(A) = \int I_A(\nu, w) \mathsf{M}(d\nu, dz, dw)$$
$$= \int I_A(\nu', w') \Lambda(\nu, z, w, d\nu', dz', dw') \mathsf{M}(d\nu, dz, dw)$$

$$= \int I_A(U(v, w, w'), w') P(z, w, dz', dw') \mathsf{M}(dv, dz, dw)$$

$$= \int I_A(U(v, w, w'), w') P(z, w, dz', dw') v(dz) \gamma \mathsf{M}(dv, dw)$$

$$= \int I_A(v', w') \Gamma(v, w, dv', dw') \gamma \mathsf{M}(dv, dw),$$

where we have used that $\mathsf{M} \in \mathfrak{M}$ in the penultimate equality. Thus $\gamma \mathsf{M}$ is a Γ-invariant measure. Moreover, it follows from the definition of $\mathfrak{M}$ that

$$\int v(B) I_C(w) \gamma \mathsf{M}(dv, dw) = \mathsf{M}\big(\mathcal{P}(E) \times B \times C\big) = m \mathsf{M}(B \times C),$$

so $m\mathsf{M}$ is the barycenter of $\gamma \mathsf{M}$. Finally, let Π_0 be a version of the regular conditional probability $\mathbf{P}(X_0 \in \cdot \,|\mathcal{F}_0^Y)$. Then as in the proof of Lemma 6.2, the law of (Π_0, X_0, Y_0) is a Λ-invariant measure in $\mathfrak{M}$ with marginal π. $\square$

7.2. *Uniqueness of the Λ-invariant measure.* The first part of the proof of Theorem 2.13 follows easily from Theorem 2.12 and Lemma 7.2.

LEMMA 7.3. *Suppose that Assumptions 2.6–2.8 hold. Then there is a unique Λ-invariant probability measure with marginal π in the class $\mathfrak{M}$.*

PROOF. Lemma 7.2 guarantees the existence of a Λ-invariant measure in $\mathfrak{M}$ with marginal π. To prove uniqueness, note that every probability measure $\mathsf{M} \in \mathfrak{M}$ is uniquely determined by $\gamma \mathsf{M}$ as

$$\mathsf{M}(A \times B \times C) = \int v(B) I_{A \times C}(v, w) \gamma \mathsf{M}(dv, dw).$$

Therefore, by Lemma 7.2, if there were to exist two distinct Λ-invariant measures in $\mathfrak{M}$ with marginal π, then there must exist two distinct Γ-invariant measures with barycenter π, in contradiction with Theorem 2.12. $\square$

The second part of the proof of Theorem 2.13 relies on Theorem 2.10 instead of Theorem 2.12. To prepare for the proof, we begin by showing that the strengthened variant of Assumption 2.6 in Theorem 2.13 is equivalent to the requirement that the assumption of Theorem 2.10 holds universally.

LEMMA 7.4. *The following are equivalent*:

(1) *For every probability measure μ on $E \times F$ such that $\mu(E \times \cdot) = \pi(E \times \cdot)$*

$$\int \|\mathbf{P}^{z, w}(X_n \in \cdot) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}} \mu(dz, dw) \xrightarrow{n \to \infty} 0.$$

(2) *For every probability measure μ on $E \times F$ such that $\mu(E \times \cdot) \ll \pi(E \times \cdot)$*

$$\mathbf{E}^\mu\big(\|\mathbf{P}^\mu(X_n \in \cdot \,|Y_0) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}}\big) \xrightarrow{n \to \infty} 0.$$

PROOF. $(1) \Rightarrow (2)$. Let μ be any probability measure on $E \times F$ with $\mu(E \times \cdot) \ll \pi(E \times \cdot)$, let $\mu_w(dz)$ be a version of the regular conditional probability $\mathbf{P}^\mu(X_0 \in \cdot | Y_0)$ and define $\mu'(dz, dw) = \mu_w(dz)\pi(E \times dw)$. Then $\mu'(E \times \cdot) = \pi(E \times \cdot)$, so the first statement of the lemma implies that we have

$$\|\mathbf{P}^{X_0, Y_0}(X_n \in \cdot) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}} \xrightarrow{n \to \infty} 0 \qquad \text{in } \mathbf{P}^{\mu'}\text{-probability.}$$

But $\mu \ll \mu'$ by construction, so the convergence also holds in $\mathbf{P}^\mu$-probability. Therefore, we obtain by dominated convergence

$$\mathbf{E}^\mu\big(\|\mathbf{P}^\mu(X_n \in \cdot | Y_0) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}}\big)$$
$$\leq \mathbf{E}^\mu\big(\|\mathbf{P}^{X_0, Y_0}(X_n \in \cdot) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}}\big) \xrightarrow{n \to \infty} 0.$$

Thus the second statement of the lemma follows.

$(2) \Rightarrow (1)$. Let μ be any probability measure on $E \times F$ such that $\mu(E \times \cdot) = \pi(E \times \cdot)$ and let $\mu_w(dz)$ be a version of the regular conditional probability $\mathbf{P}^\mu(X_0 \in \cdot | Y_0)$. By [13], Lemma 3.22, there is a measurable function $\iota \colon F \times [0, 1] \to E$ such that $\int f(z)\mu_w(dz) = \int_0^1 f(\iota(w, x)) \, dx$ for all w. Applying the second statement of the lemma to $\mu^x(dz, dw) = \delta_{\iota(w,x)}(dz)\mu(E \times dw) = \delta_{\iota(w,x)}(dz)\pi(E \times dw)$ gives

$$\int \|\mathbf{P}^{\iota(w,x), w}(X_n \in \cdot) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}}\mu(E \times dw) \xrightarrow{n \to \infty} 0 \qquad \text{for all } x \in [0, 1].$$

Thus the first statement of the lemma follows by integrating with respect to $\int_0^1 \cdot \, dx$ and applying the dominated convergence theorem. $\quad\square$

Let us note that only the first half of this result is needed in what follows. However, the equivalence of the two assumptions shows that we have not unnecessarily strengthened the assumptions of Theorem 2.13.

For the proof of Theorem 2.13, we require another lemma.

LEMMA 7.5. *Suppose that Assumptions 2.6–2.8 are in force and that*

$$\mathbf{E}^\mu\big(\|\mathbf{P}^\mu(X_n \in \cdot | Y_0) - \mathbf{P}(X_n \in \cdot)\|_{\mathrm{TV}}\big) \xrightarrow{n \to \infty} 0$$

for every probability measure μ on $E \times F$ with $\mu(E \times \cdot) \ll \pi(E \times \cdot)$. Then

$$\int \mathbf{E}^{z, w}\big(\|\Pi_n^{m(z,w) \otimes \delta_w} - \Pi_n^\pi\|_{\mathrm{TV}}\big)\pi(dz, dw) \xrightarrow{n \to \infty} 0$$

for any measurable function $m \colon E \times F \to \mathcal{P}(E)$.

PROOF. By Proposition 3.3 and the Bayes formula, there is a strictly positive measurable function $h \colon E \times F \to \mathbb{R}_+$ such that the probability kernel

$$\pi^X(z, A) = \frac{\int I_A(w)h(z, w)\pi(E \times dw)}{\int h(z, w)\pi(E \times dw)} \qquad \text{for all } z \in E, A \in \mathcal{B}(F)$$

is a version of the regular conditional probability $\mathbf{P}(Y_0 \in \cdot | X_0)$. In particular, $\pi^X(z, \cdot) \sim \pi(E \times \cdot)$ for all $z \in E$, so by our assumptions and Corollary 3.6 we obtain $\mathbf{P}^{\delta_z \otimes \pi^X(z, \cdot)}|_{\mathcal{F}_+^Y} \sim \mathbf{P}|_{\mathcal{F}_+^Y}$ for all $z \in E$.

Fix a measurable function $m : E \times F \to \mathcal{P}(E)$. For every $z \in E$, define $\mu^z(dz', dw) = m(z, w)(dz')\pi(E \times dw)$. Then by Theorem 2.10, we have

$$\|\Pi_n^{\mu^z} - \Pi_n^{\pi}\|_{\mathrm{TV}} \xrightarrow{n \to \infty} 0, \qquad \mathbf{P}\text{-a.s.}$$

for all $z \in E$. Thus by $\mathbf{P}^{\delta_z \otimes \pi^X(z, \cdot)}|_{\mathcal{F}_+^Y} \sim \mathbf{P}|_{\mathcal{F}_+^Y}$ and dominated convergence,

$$\int \mathbf{E}^{z, w}(\|\Pi_n^{\mu^z} - \Pi_n^{\pi}\|_{\mathrm{TV}})\pi^X(z, dw) \xrightarrow{n \to \infty} 0$$

for all $z \in E$. But by Lemma 5.1 we have $\Pi_n^{\mu^z} = \Pi_n^{m(z, w) \otimes \delta_w}$ $\mathbf{P}^{z, w}$-a.s. for all $n \geq 0$. Integrating with respect to $\pi(dz \times F)$ and applying the dominated convergence theorem completes the proof. $\square$

We now proceed to the proof of Theorem 2.13. Let Z be any Polish space endowed with the complete metric d_Z. Recall that the space $\mathcal{P}(Z)$ is Polish when endowed with the metric (cf. [12], Theorem 11.3.3 and Corollary 11.5.5)

$$d_{\mathcal{P}(Z)}(\nu, \nu') = \sup\left\{ \left| \int f(z)\nu(dz) - \int f(z)\nu'(dz) \right| : \sup_{x \in Z} |f(x)| \leq 1, \right.$$
$$\left. \sup_{x, y \in Z} \frac{|f(x) - f(y)|}{d_Z(x, y)} \leq 1 \right\}.$$

In particular, the complete metric

$$D((\nu, z, w), (\nu', z', w')) = d_{\mathcal{P}(E)}(\nu, \nu') + d_E(z, z') + d_F(w, w')$$

metrizes the topology of $\mathcal{P}(E) \times E \times F$.

PROOF OF THEOREM 2.13. The first part of the theorem was established in Lemma 7.3. For the remainder of the proof, let us assume that one of the equivalent assumptions in Lemma 7.4 is in force. We will show that any two Λ-invariant probability measures with marginal π must coincide.

To this end, let M and M' be two Λ-invariant probability measures with marginal π. By [13], Lemma 3.22, there exist measurable functions $m : E \times F \times [0, 1] \to \mathcal{P}(E)$ and $m' : E \times F \times [0, 1] \to \mathcal{P}(E)$ such that

$$\int f(\nu, z, w)\mathsf{M}(d\nu, dz, dw) = \int_0^1 \int f(m(z, w, x), z, w)\pi(dz, dw)\,dx,$$

$$\int f(\nu, z, w)\mathsf{M}'(d\nu, dz, dw) = \int_0^1 \int f(m'(z, w, x), z, w)\pi(dz, dw)\,dx$$

for every bounded measurable function $f : \mathcal{P}(E) \times E \times F \to \mathbb{R}$. Moreover, note that by the definition of Λ and Lemma 5.1

$$\int f(v', z', w') \Lambda^n(v, z, w, dv', dz', dw') = \mathbf{E}^{z,w}(f(\Pi_n^{v \otimes \delta_w}, X_n, Y_n)).$$

Let us now fix a bounded function f such that

$$|f(v, z, w) - f(v', z', w')| \le D((v, z, w), (v', z', w'))$$

for all $v, v' \in \mathcal{P}(E)$, $z, z' \in E$, $w, w' \in F$. We can now estimate

$$\left| \int f(v, z, w) \mathsf{M}(dv, dz, dw) - \int f(v, z, w) \mathsf{M}'(dv, dz, dw) \right|$$

$$\le \int_0^1 \int \mathbf{E}^{z,w} \left(d_{\mathcal{P}(E)} \left(\Pi_n^{m(z,w,x) \otimes \delta_w}, \Pi_n^{m'(z,w,x) \otimes \delta_w} \right) \right) \pi(dz, dw) \, dx$$

$$\le \int_0^1 \int \mathbf{E}^{z,w} \left(\left\| \Pi_n^{m(z,w,x) \otimes \delta_w} - \Pi_n^{m'(z,w,x) \otimes \delta_w} \right\|_{\mathrm{TV}} \right) \pi(dz, dw) \, dx$$

for every $n \ge 0$, where we used that $\mathsf{M}\Lambda^n = \mathsf{M}$ and $\mathsf{M}'\Lambda^n = \mathsf{M}'$. By the triangle inequality, Lemma 7.5 and the dominated convergence theorem, the right-hand side of this inequality converges to zero as $n \to \infty$. Therefore, we have shown that

$$\left| \int f(v, z, w) \mathsf{M}(dv, dz, dw) - \int f(v, z, w) \mathsf{M}'(dv, dz, dw) \right| = 0$$

for all bounded functions f that are 1-Lipschitz for the metric D. In other words, $d_{\mathcal{P}(\mathcal{P}(E) \times E \times F)}(\mathsf{M}, \mathsf{M}') = 0$, so $\mathsf{M} = \mathsf{M}'$. Thus we have shown that all Λ-invariant probability measures with marginal π must coincide, establishing uniqueness.

To complete the proof, it remains to consider the case when P has unique invariant probability measure (i.e., π is the only P-invariant probability measure). As the marginal of any Λ-invariant probability measure must be P-invariant, this implies that any Λ-invariant measure must have marginal π. Therefore, in this case, Λ has a unique invariant probability measure. $\square$

REMARK 7.6.   It is instructive to note that Assumptions 2.6–2.8 are not sufficient to ensure uniqueness of the Λ-invariant probability measure even in the case that P has a unique invariant probability measure. Let us briefly sketch a counterexample. Let $E = \mathbb{R} \times \{0, 1\}$ and $F = \mathbb{R}$, and consider the filtering model

$$X_n^1 = 2 X_{n-1}^1 X_{n-1}^2 + \xi_n, \qquad X_n^2 = X_{n-1}^2, \qquad Y_n = X_n^1 + \eta_n,$$

where $(\xi_n)_{n \ge 0}$, $(\eta_n)_{n \ge 0}$ are i.i.d. $N(0, 1)$-distributed random variables. It is clear that the corresponding transition kernel P has a unique invariant probability measure π [with $\pi(\cdot \times F) = N(0, 1) \otimes \delta_0$] and that Assumptions 2.6–2.8 hold.

Now let $\mu = \delta_0 \otimes \delta_1 \otimes N(0, 1)$. Then $\Pi_n^\mu = N(m_n, \sigma_n^2) \otimes \delta_1$, where m_n and σ_n^2 can be computed recursively using the Kalman filtering equations corresponding

to the model $X_n = 2X_{n-1} + \xi_n$, $Y_n = X_n + \eta_n$. It is easily verified by inspection of the Kalman filtering equations that the law of (Π_n^μ, X_n, Y_n) converges weakly as $n \to \infty$ under the stationary measure **P**. The limiting law is therefore a Λ-invariant probability measure that is supported on $\mathcal{P}(\mathbb{R} \times \{1\}) \times E \times F$. On the other hand, the Λ-invariant measure defined in the proof of Lemma 7.2 is clearly supported on $\mathcal{P}(\mathbb{R} \times \{0\}) \times E \times F$. Therefore, Λ has distinct invariant measures.

This example illustrates that the stronger assumption of Theorem 2.13 is indeed required to establish uniqueness of the Λ-invariant measure in the class of all probability measures. Of course, the first part of Theorem 2.13 is not contradicted as the additional Λ-invariant measure obtained in this example is not in $\mathfrak{M}$.

APPENDIX A: AUXILIARY RESULTS

The goal of this Appendix is to collect for easy reference a few auxiliary results that are used throughout the paper.

The following result on the existence of invariant sets for stationary Markov chains is given in [20], Lemma 2.6. The construction of the set H follows closely along the lines of [18], pages 1636 and 1637, so the proof is omitted.

LEMMA A.1. *Let* $\mathbf{P}^z$ *be the law of a Markov process* $(Z_k)_{k \geq 0}$ *given* $Z_0 = z$, *and let* ν *be a stationary probability for this Markov process. Then for any set* $\tilde{H}$ *of* ν*-full measure, there is a subset* $H \subset \tilde{H}$ *of* ν*-full measure such that*

$$\mathbf{P}^z(Z_n \in H \text{ for all } n \geq 0) = 1 \qquad \text{for all } z \in H.$$

The following elementary can be found in [20], Lemma 3.6.

LEMMA A.2. *Let* G_1, G_2 *and* K *be Polish spaces and set* $\Omega = G_1 \times G_2 \times K$. *We consider a probability measure* **P** *on* $(\Omega, \mathcal{B}(\Omega))$. *Denote by* $\gamma_1 : \Omega \to G_1$, $\gamma_2 : \Omega \to G_2$, *and* $\kappa : \Omega \to K$ *the coordinate projections, and let* $\mathcal{G}_1$, $\mathcal{G}_2$, *and* $\mathcal{K}$ *be the* σ*-fields generated by* γ_1, γ_2 *and* κ, *respectively. Choose fixed versions of the following regular conditional probabilities*:

$$\Xi_1^K(g_1, \cdot) = \mathbf{P}(\kappa \in \cdot |\mathcal{G}_1)(g_1), \qquad \Xi_{12}^K(g_1, g_2, \cdot) = \mathbf{P}(\kappa \in \cdot |\mathcal{G}_1 \vee \mathcal{G}_2)(g_1, g_2),$$

$$\Xi_1^2(g_1, \cdot) = \mathbf{P}(\gamma_2 \in \cdot |\mathcal{G}_1)(g_1), \qquad \Xi_{1K}^2(g_1, k, \cdot) = \mathbf{P}(\gamma_2 \in \cdot |\mathcal{G}_1 \vee \mathcal{K})(g_1, k),$$

where $g_1 \in G_1$, $g_2 \in G_2$, $k \in K$. *Suppose that there exists a nonnegative measurable function* $h : G_1 \times G_2 \times K \to [0, \infty[$ *and a set* $H \subset G_1 \times G_2$ *such that* $\mathbf{E}(I_H(\gamma_1, \gamma_2)) = 1$ *and for every* $(g_1, g_2) \in H$

$$\Xi_{12}^K(g_1, g_2, A) = \int I_A(k) h(g_1, g_2, k) \Xi_1^K(g_1, dk) \qquad \text{for all } A \in \mathcal{K}.$$

Then there is $H' \subset G_1 \times K$ *with* $\mathbf{E}(I_{H'}(\gamma_1, \kappa)) = 1$ *so that for all* $(g_1, k) \in H'$

$$\Xi_{1K}^2(g_1, k, B) = \int I_B(g_2) h(g_1, g_2, k) \Xi_1^2(g_1, dg_2) \qquad \text{for all } B \in \mathcal{G}_2.$$

We now recall two results of von Weizsäcker that are of central importance in our proofs. The first result is a special case of the result in [24], pages 95 and 96.

LEMMA A.3.   *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 $q_x[h \in \cdot \,|g']$.*

Though the second result is not given precisely in this form in [24], its proof follows easily from [24] modulo minor modifications (see also [20], Section 4.1).

LEMMA A.4.   *Let G and H be Polish spaces, let $(X_n)_{n \geq 0}$ be a sequence of random variables with values in G and let Y be a random variable with values in H on some underlying probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Define the σ-field $\mathcal{H} = \sigma\{Y\}$ and the decreasing filtration $\mathcal{G}_n = \sigma\{X_k : k \geq n\}$. Then*

$$\bigcap_{n \geq 0} \mathcal{H} \vee \mathcal{G}_n = \mathcal{H}, \qquad \mathbf{P}\text{-a.s.}$$

if and only if

$$\bigcap_{n \geq 0} \mathcal{G}_n \text{ is } \mathbf{P}^{\mathcal{H}}\text{-trivial}, \qquad \mathbf{P}\text{-a.s.},$$

where $\mathbf{P}^{\mathcal{H}}$ is a version of the regular conditional probability $\mathbf{P}((X_n)_{n \geq 0} \in \cdot \,|\mathcal{H})$.

APPENDIX B: NOTATION LIST

The following list of frequently used notation, together with the page numbers where they are defined, is included for easy reference.

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] BUDHIRAJA, A. (2002). On invariant measures of discrete time filters in the correlated signal-noise case. *Ann. Appl. Probab*. **12** 1096–1113. MR1925453

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

[5] BUDHIRAJA, A. and KUSHNER, H. J. (2001). Monte Carlo algorithms and asymptotic problems in nonlinear filtering. In *Stochastics in Finite and Infinite Dimensions* 59–87. Birkhäuser, Boston, MA. MR1797081

[6] CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer, New York. MR2159833

[7] CHAUMONT, L. and YOR, M. (2003). *Exercises in Probability. Cambridge Series in Statistical and Probabilistic Mathematics* **13**. Cambridge Univ. Press, Cambridge. MR2016344

[8] CRISAN, D. and ROZOVSKY, B. (2011). *The Oxford Handbook of Nonlinear Filtering*. Oxford Univ. Press, Oxford.

[9] DELLACHERIE, C. and MEYER, P.-A. (1982). *Probabilities and Potential. B. Theory of Martingales. North-Holland Mathematics Studies* **72**. North-Holland, Amsterdam. MR0745449

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

[11] DOUC, R., MOULINES, É. and RYDÉN, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Statist*. **32** 2254–2304. MR2102510

[12] DUDLEY, R. M. (2002). *Real Analysis and Probability. Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358

[13] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer, New York. MR1876169

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

[15] KUNITA, H. (1991). Ergodic properties of nonlinear filtering processes. In *Spatial Stochastic Processes. Progress in Probability* **19** 233–256. Birkhäuser, Boston, MA. MR1144099

[16] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2509253

[17] OCONE, D. and PARDOUX, E. (1996). Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM J. Control Optim*. **34** 226–243. MR1372912

[18] ORSTEIN, D. and SUCHESTON, L. (1970). An operator theorem on L_1 convergence to zero with applications to Markov kernels. *Ann. Math. Statist*. **41** 1631–1639. MR0272057

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

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

[21] VAN HANDEL, R. (2009). Uniform time average consistency of Monte Carlo particle filters. *Stochastic Process. Appl*. **119** 3835–3861. MR2552307

[22] VAN HANDEL, R. (2012). On the exchange of intersection and supremum of σ-fields in filtering theory. *Israel J. Math.* To appear.

[23] VOLKONSKIĬ, V. A. and ROZANOV, Y. A. (1959). Some limit theorems for random functions. I. *Theory Probab. Appl.* **4** 178–197. MR0121856

[24] 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** 91–100. MR0699981

[25] YOR, M. (1977). Sur les théories du filtrage et de la prédiction. In *Séminaire de Probabilités*, *XI* (*Univ. Strasbourg, Strasbourg*, 1975/1976). *Lecture Notes in Math.* **581** 257–297. Springer, Berlin. MR0471060

SHERRERD HALL
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544
USA
E-MAIL: xintong@princeton.edu
            rvan@princeton.edu