Whiteley, N. P., & Lee, A. (2016). Perfect sampling for nonhomogeneous Markov chains and hidden Markov models. Annals of Applied Probability, 26(5), 3044-3077. DOI: 10.1214/15-AAP1169

The Annals of Applied Probability
2016, Vol. 26, No. 5, 3044–3077
DOI: 10.1214/15-AAP1169

PERFECT SAMPLING FOR NONHOMOGENEOUS MARKOV CHAINS AND HIDDEN MARKOV MODELS

BY NICK WHITELEY[1] AND ANTHONY LEE

University of Bristol and University of Warwick

We obtain a perfect sampling characterization of weak ergodicity for backward products of finite stochastic matrices, and equivalently, simultaneous tail triviality of the corresponding nonhomogeneous Markov chains. Applying these ideas to hidden Markov models, we show how to sample exactly from the finite-dimensional conditional distributions of the signal process given infinitely many observations, using an algorithm which requires only an almost surely finite number of observations to actually be accessed. A notion of "successful" coupling is introduced and its occurrence is characterized in terms of conditional ergodicity properties of the hidden Markov model and related to the stability of nonlinear filters.

1. Introduction. With the introduction of their famous Coupling From the Past (CFTP) algorithm, Propp and Wilson (1996) showed how to use a form of backward coupling to simulate exact samples from the invariant distribution of an ergodic Markov chain in a.s. finite time. Foss and Tweedie (1998), in part appealing to a construction of Murdoch and Green (1998), showed that existence of an a.s. finite backward coupling time characterizes uniform geometric ergodicity of the Markov chain in question. The present papers extends these ideas in the context of nonhomogeneous Markov chains, a setting which to date has received little attention, perhaps due to a lack of appropriate formulation or applications.

Our contribution is to present such a formulation and apply the insight which we develop about nonhomogeneous chains to hidden Markov models (HMMs), for which we obtain a perfect sampling characterization of *conditional ergodicity* phenomena, that is, ergodic properties of the signal process in the HMM under its conditional law given the observations, along the lines of those addressed by van Handel (2009). Even for HMMs with finite state space, conditional ergodicity and the connection to perfect sampling can be subtle, due to the delicate interplay between the observations and signal in the HMM, and the fact that the ergodic theory of nonhomogeneous Markov chains, which governs the behavior of the signal when conditioned on observations, is considerably more complicated than that of homogeneous chains.

Received October 2014; revised December 2015.

[1]Supported by EPSRC Grant EP/K023330/1.

MSC2010 subject classifications. Primary 60J10; secondary 60G35.

Key words and phrases. Coupling, conditional ergodicity, nonhomogeneous Markov chains, perfect simulation.

1.1. *Nonhomogeneous Markov chains and backward products*. One of the key notions underlying CFTP is that if an ergodic Markov chain were initialized infinitely far in the past and run forward in time, its state at the present would be distributed exactly according to the invariant distribution of the chain. In order to give an overview of our main results, we need to identify a suitable and somewhat elementary generalization of this notion.

Let $\mathbb{T} := \mathbb{Z}^- \cup \{0\}$ be the set of nonpositive integers and let $M = (M_n)_{n\in\mathbb{T}}$ be a sequence of Markov kernels on a finite set $E = \{1, \dots, s\}$, so for each $x \in E$, $M_n(x, \cdot)$ is probability distribution on E. One may construct a nonhomogeneous Markov chain $(X_n)_{n\in\mathbb{T}}$ with paths in $E^{\mathbb{T}}$ and transitions given by M in the sense that $X_n \mid X_{n-1} \sim M_n(X_{n-1}, \cdot)$, as soon as there exists a sequence $\pi = (\pi_n)_{n\in\mathbb{T}}$ of *absolute probabilities*: a family of probability distributions with the property that for all $n \in \mathbb{T}$ and $x \in E$,

$$\sum_{z\in E} \pi_{n-1}(z) M_n(z, x) = \pi_n(x).$$

Indeed one can then readily define a consistent family of finite dimensional distributions $(\mathbf{P}_\pi^{(n)})_{n\in\mathbb{T}}$,

$$\mathbf{P}_\pi^{(n)}(X_n = x_n, \dots, X_0 = x_0) := \pi_n(x_n) \prod_{k=n+1}^{0} M_k(x_{k-1}, x_k), \tag{1.1}$$

giving rise via the usual Kolmogorov extension argument to a probability measure $\mathbf{P}_\pi$ over paths in $E^{\mathbb{T}}$; one can think of $(X_n)_{n\in\mathbb{T}}$ running forward in time from the distant past toward zero.

Now for $k \in \mathbb{T}$ define recursively

$$M_{k,k} := \mathrm{Id}, \qquad M_{n-1,k}(x, x') := \sum_{z\in E} M_n(x, z) M_{n,k}(z, x'), \qquad n \le k. \tag{1.2}$$

With $k \in \mathbb{T}$ fixed, $(M_{n,k})_{n\le k}$ are called *backward products*, since they can be written in terms of matrix multiplications to the left: $M_{n-1,k} = M_n M_{n,k}$.

Questions of existence and uniqueness of π, and thus of $\mathbf{P}_\pi$, are answered with the following long-established facts, which hold for any sequence of Markov kernels $M = (M_n)_{n\in\mathbb{T}}$ on a finite state space E; see Seneta [(2006), Section 4.6] and references therein for an accessible introduction.

FACT 1. There always exists at least one sequence of absolute probabilities

FACT 2. There exists a unique sequence of absolute probabilities if and only if the backward products of M are *weakly ergodic*, meaning

$$\lim_{n\to-\infty} M_{n,k}(x, z) - M_{n,k}(x', z) = 0 \qquad \forall k \in \mathbb{T}, (z, x, x') \in E^3, \tag{1.3}$$

in which case

$$\lim_{n\to-\infty} M_{n,k}(x,z) - \pi_k(z) = 0 \qquad \forall k \in \mathbb{T}, (z,x) \in E^2, \tag{1.4}$$

where $\pi = (\pi_n)_{n\in\mathbb{T}}$ is the unique sequence of absolute probabilities for M.

1.2. *Perfect sampling and characterizations of weak ergodicity.* Our basic algorithmic goal, when weak ergodicity holds, is to obtain exact draws from each π_n. This can be achieved with a very modest generalization of Propp and Wilson's method. The only existing works on perfect simulation for nonhomogeneous chains which we know of are Glynn and Thorisson (2001) and Stenflo (2008), which respectively provide perfect sampling methods for Markov chains conditioned to avoid certain states, and products of transition matrices subject to a particular uniform regularity assumption, which we discuss in more detail later. Our first goal is to develop more general insight into how the feasibility of CFTP for nonhomogeneous chains is related to various ergodic properties of M and $\mathbf{P}_\pi$.

Inspired by Foss and Tweedie's (1998) characterization of uniform geometric ergodicity for a homogeneous chain in terms of the existence of a successful (meaning a.s. finite) backward coupling time, we assert that "success" in the nonhomogeneous case is for not just one, but all of a particular countably infinite family of coupling times to be a.s. finite. Our first main result, Theorem 1, shows that success so-defined of our coupling is equivalent to weak ergodicity, as in (1.3), which is weaker than the assumption of Stenflo (2008), and if successful our coupling delivers a sample from each member of the then unique sequence of absolute probabilities $(\pi_n)_{n\in\mathbb{T}}$ in a.s. finite time.

We extend this ergodic characterization in Theorem 2, by showing that unicity of a sequence of absolute probabilities, hence weak ergodicity, hence success of our coupling, is also equivalent to the *simultaneous tail triviality* condition

$$\begin{aligned} &\mathbf{P}_\pi(A) = \mathbf{P}_\pi(A)^2 = \mathbf{P}_{\tilde{\pi}}(A) \\ &\qquad \forall (\pi, \tilde{\pi}, A) \in \Pi_M \times \Pi_M \times \bigcap_{n\in\mathbb{T}} \sigma(X_k; k \le n), \end{aligned} \tag{1.5}$$

where Π_M is the set of all sequences of absolute probabilities for M.

1.3. *Hidden Markov models and conditional ergodicity.* Our motivation for considering nonhomogeneous chains and condition (1.5) stems from hidden Markov models, which are widely applied across econometrics, genomics, signal processing and many other disciplines as they provide flexible and interpretable means to model dependence between observed data in terms of an unobserved Markov chain. We take a slightly nonstandard perspective in that an HMM is for us a process $(X_n, Y_n)_{n\in\mathbb{T}}$ on a nonpositive time horizon, where the *signal* $X = (X_n)_{n\in\mathbb{T}}$ is a possibly homogeneous Markov chain with paths in $E^{\mathbb{T}}$, and the *observations* $Y = (Y_n)_{n\in\mathbb{T}}$ are conditionally independent given X, with each

Y_n valued in a Polish space F and having a conditional distribution given X which depends only on X_n.

With the law of $(X_n, Y_n)_{n\in\mathbb{T}}$ then written as $\mathbf{P}$, a standard task in applications of HMMs is to calculate conditional distributions of the form $\mathbf{P}(X_n \in \cdot \mid \sigma(Y_k; k \in I))$ where I is some finite subset of $\mathbb{T}$. These distributions are immediately useful for inference about the signal process and for making predictions about future observations given those recorded up to the present. For example, if one extends the HMM onto a positive time horizon by introducing (X_1, Y_1) such that $\mathbf{P}(X_1 = x, Y_1 \in \cdot \mid \mathcal{F}^X \vee \mathcal{F}^Y) = M(X_0, x)G(x, \cdot)$, where $\mathcal{F}^X = \sigma(X_k; k \in \mathbb{T})$, $\mathcal{F}^Y = \sigma(Y_k; k \in \mathbb{T})$, and M and G are probability kernels respectively from E to itself and from E to F, then for any $I \subset \mathbb{T}$,

$$\begin{aligned} &\mathbf{P}\big(Y_1 \in \cdot \mid \sigma(Y_k; k \in I)\big) \\ &\quad = \sum_{x,x'} \mathbf{P}\big(X_0 = x \mid \sigma(Y_k; k \in I)\big) M\big(x, x'\big) G\big(x', \cdot\big). \end{aligned} \tag{1.6}$$

When calculating such distributions in order to make predictions, it is desirable to impart as much information from the past as possible. For example, the mean-square optimal $\mathcal{F}^Y$-measurable predictor of Y_1 is, of course, the conditional expectation $\mathbf{E}[Y_1 \mid \mathcal{F}^Y]$. However, exact calculation of $\mathbf{P}(X_0 \in \cdot \mid \mathcal{F}^Y)$, or indeed $\mathbf{P}(X_n \in \cdot \mid \mathcal{F}^Y)$ for any $n \in \mathbb{T}$, requires an infinite number of observations to be recorded and in general cannot be accomplished in finite time.

Nevertheless, under certain conditions, our perfect sampling method makes it possible to obtain an exact draw from $\mathbf{P}(X_n \in \cdot \mid \mathcal{F}^Y)$ using an algorithm which runs for $|T_n|$ time-steps and uses only $(Y_k; k = T_n, T_n + 1, \dots, 0)$, where T_n is a $\mathbb{T}$-valued random time, the details of which we make precise later. If with $n = 0$, the resulting sample from $\mathbf{P}(X_0 \in \cdot \mid \mathcal{F}^Y)$ is denoted $X_0^\star$ and one also samples $Y_1^\star \sim \sum_x M(X_0^\star, x)G(x, \cdot)$, then by (1.6) with $I = \mathbb{T}$, $Y_1^\star$ is distributed exactly according to the "ideal" predictive conditional distribution $\mathbf{P}(Y_1 \in \cdot \mid \mathcal{F}^Y)$.

The connection to Sections 1.1–1.2 arises from the facts that

$$\begin{aligned} &\mathbf{P}\big(X_n \in \cdot \mid \mathcal{F}^Y \vee \sigma(X_k; k < n)\big) \\ &\quad = \mathbf{P}\big(X_n \in \cdot \mid \mathcal{F}^Y \vee \sigma(X_{n-1})\big) \\ &\quad = \mathbf{P}\big(X_n \in \cdot \mid \sigma(Y_k; k \geq n) \vee \sigma(X_{n-1})\big), \qquad \mathbf{P}\text{-a.s.,} \end{aligned}$$

that is, conditional on the observations, the signal process X is a nonhomogeneous Markov chain, and its conditional transition probabilities at time n depend on Y only through $(Y_k)_{k\geq n}$. Moreover, for any $y \in F^{\mathbb{T}}$ we can calculate and sample from each of a family of Markov kernels $M^y = (M_n^y)_{n\in\mathbb{T}}$, with $M_n^Y(X_{n-1}, \cdot)$ a version of $\mathbf{P}(X_n \in \cdot \mid \sigma(Y_k; k \geq n) \vee \sigma(X_{n-1}))$, for which conditional probabilities of the form $\mathbf{P}(X_n \in \cdot \mid \mathcal{F}^Y)$ define a sequence of absolute probabilities, and T_n as mentioned above is one of a collection of coupling times arising from CFTP applied to M^y.

Building from the considerations of Section 1.2, our attention then turns to the question of how success of our HMM sampling scheme, meaning that every T_n is conditionally a.s.-finite given Y, is related to the ergodic properties of the HMM. Compared to the setup of Sections 1.1–1.2, we have to handle the additional complication here that each $\mathbf{P}(X_n \in \cdot \mid \mathcal{F}^Y)$, M_n^Y and success itself depend on Y. Our main result in this regard, Theorem 3, establishes that success for $\mathbf{P}$-almost all Y is equivalent to the following condition, which can be considered the HMM-counterpart of (1.5): there exists an event $H \in \mathcal{F}^X \otimes \mathcal{F}^Y$ with $\mathbf{P}(H) = 1$ such that for all $\omega = (x, y) \in H$,

$$\mathbf{P}^{\mathcal{F}^Y}(\omega, A) = \mathbf{P}^{\mathcal{F}^Y}(\omega, A)^2 \qquad \forall A \in \bigcap_{n \in \mathbb{T}} \sigma(X_k; k \leq n) \tag{1.7}$$

and

$$\begin{aligned} &\mathbf{P}_{\pi^{Y(\omega)}}(A) = \mathbf{P}^{\mathcal{F}^Y}(\omega, A) \\ &\qquad\qquad \forall \big(\pi^{Y(\omega)}, A\big) \in \Pi_{M^{Y(\omega)}} \times \bigcap_{n \in \mathbb{T}} \sigma(X_k; k \leq n). \end{aligned} \tag{1.8}$$

Here, with $\Omega := E^{\mathbb{T}} \times F^{\mathbb{T}}$, $\mathbf{P}^{\mathcal{F}^Y} : \Omega \times \mathcal{F}^X \to [0, 1]$ is a probability kernel such that for all $A \in \mathcal{F}^X$, $\mathbf{P}^{\mathcal{F}^Y}(\omega, A) = \mathbf{P}(A \mid \mathcal{F}^Y)(\omega)$, $\mathbf{P}$-a.s.; $\Pi_{M^{Y(\omega)}}$ is the set of all sequences of absolute probabilities for $M^{Y(\omega)}$; and $\mathbf{P}_{\pi^{Y(\omega)}}(\cdot)$ is the measure on $\mathcal{F}^X$ under which $(X_n)_{n \in \mathbb{T}}$ is a Markov chain with transitions $M^{Y(\omega)}$ and absolute probabilities $\pi^{Y(\omega)}$.

Condition (1.7) can be interpreted as meaning that the signal process is *conditionally ergodic* given the observations, and is a key condition in studies of stability with respect to initial conditions of nonlinear filters; see van Handel (2009) and references therein. Condition (1.8) can be understood as meaning that any probability measure which makes the signal process a Markov chain with transitions $M^{Y(\omega)}$ must have the same tail behavior as $\mathbf{P}(\cdot \mid \mathcal{F}^Y)(\omega)$.

The remainder of the paper is structured as follows. Some notation and other preliminaries are given in Section 2. Section 3 reviews existing literature on perfect sampling for nonhomogeneous chains, gives the details of the coupling method and describes its connection to weak ergodicity. Section 4 addresses tail triviality. Section 5 addresses the HMM setup. In Section 6, we discuss examples of HMMs for which our sampling method is not successful, either through failure of (1.7) or (1.8). We provide verifiable sufficient conditions for successful coupling, discuss sampling when only finitely many observations are available and numerically illustrate how the coupling can be influenced by the observation sequences. We discuss an approach to the simulation of multiple dependent samples using a single run of the perfect simulation and numerically investigate its computational efficiency.

2. Preliminaries. Throughout the paper, $E = \{1, \ldots, s\}$ is a finite set, which we endow with the discrete topology, and the corresponding Borel σ-algebra, that is, the power set of E, is denoted by $\mathcal{B}(E)$. For any two probability distributions μ, ν on E we write the total variation distance as

$$\begin{aligned}\|\mu - \nu\| &:= \sup_{A \subset E} \bigl|\mu(A) - \nu(A)\bigr| \\ &= \frac{1}{2} \sum_{x \in E} \bigl|\mu(x) - \nu(x)\bigr|,\end{aligned}$$

and for a Markov kernel K on E we write Dobrushin's coefficient

$$\begin{aligned}\beta(K) &:= \max_{(x,x') \in E^2} \bigl\|K(x, \cdot) - K(x', \cdot)\bigr\|, \\ &= 1 - \min_{(x,x') \in E^2} \sum_{z \in E} \min\bigl\{K(x, z), K(x', z)\bigr\}.\end{aligned} \tag{2.1}$$

Throughout Sections 2–4, we fix an arbitrary collection of Markov kernels $M = (M_n)_{n \in \mathbb{T}}$ on E and we add slightly to the definitions of (1.2) the convention that $M_{n,k} = \mathrm{Id}$ whenever $k \le n$, $n \in \mathbb{T}$.

We shall make extensive use of the following proposition, which expands on Fact 2, providing characterizations of weak ergodicity in terms of Dobrushin's coefficient.

PROPOSITION 1. *The following are equivalent*:

1. *for all* $k \in \mathbb{T}$ *and* $(x, x', z) \in E^3$, $\lim_{n \to -\infty} M_{n,k}(x, z) - M_{n,k}(x', z) = 0$,
2. $\mathrm{card}(\Pi_M) = 1$,
3. *for all* $k \in \mathbb{T}$, $\lim_{n \to -\infty} \beta(M_{n,k}) = 0$,
4. *there exists a strictly decreasing subsequence* $(n_i)_{i \in \mathbb{N}}$ *of* $\mathbb{T}$ *such that*

$$\sum_{i=0}^{\infty} 1 - \beta(M_{n_{i+1}, n_i}) = \infty,$$

and when any (*and then all*) *of conditions* 1–4 *hold*,

$$\lim_{n \to -\infty} M_{n,k}(x, z) - \pi_k(z) = 0 \qquad \forall k \in \mathbb{T}, (z, x) \in E^2, \tag{2.2}$$

where $\pi = (\pi_n)_{n \in \mathbb{T}}$ *is the unique sequence of absolute probabilities for* M.

For proof of (1) $\Leftrightarrow$ (2) and (2.2), see Seneta (2006), Theorem 4.20, (1) $\Leftrightarrow$ (3) is immediate from the definition of $\beta(\cdot)$ and for proof of (1)$\Leftrightarrow$(4) see Seneta (2006), Theorem 4.18.

To connect with the perhaps more familiar case of homogeneous chains, consider the case $E = \{0, 1\}$ and $M_n(x, 1 - x) = M(x, 1 - x) = 1$. For all $\alpha \in (0, 1)$, $\pi_{2n}(0) = \alpha = 1 - \pi_{2n}(1)$, $\pi_{2n-1}(0) = 1 - \alpha = 1 - \pi_{2n-1}(1)$, $n \in \mathbb{T}$ is a sequence of absolute probabilities, so that there are infinitely many sequences of absolute probabilities for M, even though there is a unique stationary distribution.

3. Perfect sampling for nonhomogeneous chains.

3.1. *Background.* We now review the existing literature on perfect sampling for nonhomogeneous Markov chains. Glynn and Thorisson [(2001), Section 5] formulated a perfect sampling algorithm for a finite state-space chain conditioned to remain in some set over a given time window and termination of their algorithm in a.s. finite time follows from assumptions they make about the conditioned process. Stenflo (2008) devised a perfect sampling procedure for nonhomogeneous backward products of stochastic matrices and showed that, when there exists a constant $c > 0$ such that

$$\inf_{n\in\mathbb{T}} \sum_{x'\in E} \min_{x\in E} M_n(x, x') \geq c, \tag{3.1}$$

the limit $\lim_{n\to-\infty} M_{n,0}(x, \cdot)$ exists, is independent of x, and defines a probability distribution on E, from which the algorithm of Stenflo (2008) produces a sample.

When (3.1) holds, straightforward calculations show that $c \leq 1$ and by (2.1), $\sup_{n\in\mathbb{T}} \beta(M_n) \leq 1 - c$, so part 4 of Proposition 1 holds and $\lim_{n\to-\infty} M_{n,0}(x, \cdot)$ is of course a member of the unique sequence of absolute probabilities. However, part 4 of Proposition 1 is clearly a weaker condition than (3.1).

3.2. *The coupling.* Consider $\Omega^\xi := (E^s)^{\mathbb{T}}$ with product σ-algebra $\mathcal{F}^\xi := (\mathcal{B}(E)^{\otimes s})^{\otimes \mathbb{T}}$. Define the coordinate process $\xi = (\xi_n^x; x \in E, n \in \mathbb{T})$, $\xi_n^x : \Omega^\xi \to E$, and let $\mathbf{Q}$ be the probability measure on $(\Omega^\xi, \mathcal{F}^\xi)$,

$$\mathbf{Q}(d\xi) := \bigotimes_{n\in\mathbb{T}} \bigotimes_{x\in E} M_n(x, \xi_n^x)\, d\xi_n^x,$$

where $d\xi_n^x$ is counting measure on E, so that

$$(\xi_n^x; x \in E, n \in \mathbb{T}) \text{ are independent under } \mathbf{Q}, \tag{3.2}$$

$$\mathbf{Q}(\xi_n^x = x') = M_n(x, x'), \qquad (n, x, x') \in \mathbb{T} \times E^2. \tag{3.3}$$

For each $n \in \mathbb{T}$, define the random map

$$\Phi_n : x \in E \longmapsto \Phi_n(x) := \xi_n^x \in E, \tag{3.4}$$

and the compositions

$$\Phi_{n,k} := \Phi_k \circ \Phi_{k-1} \circ \cdots \circ \Phi_{n+1}, \qquad n < k \in \mathbb{T}, \tag{3.5}$$

so, for example, $\Phi_{n,n+2}(x) = \Phi_{n+2}(\Phi_{n+1}(x)) = \Phi_{n+2}(\xi_{n+1}^x) = \xi_{n+2}^{\xi_{n+1}^x}$, etc., and it is easily checked that

$$\mathbf{Q}(\Phi_n(x) = x') = M_n(x, x') \quad \text{and} \quad \mathbf{Q}(\Phi_{n,k}(x) = x') = M_{n,k}(x, x'). \tag{3.6}$$

Now define the $\{-\infty\} \cup \mathbb{T}$-valued coalescence times

$$T_k := \sup\{n < k : \text{ image of } \Phi_{n,k} \text{ is a singleton}\}, \qquad k \in \mathbb{T}, \tag{3.7}$$

with $T_k := -\infty$ when the set $\{n < k : \text{ image of } \Phi_{n,k} \text{ is a singleton}\}$ is empty.

REMARK 1. Note that in the time-homogeneous case, $M_n = M_0$ for all $n \in \mathbb{T}$, the coalescence times T_n are identically distributed and the random maps Φ_n are i.i.d.

The main result of this section is the following theorem.

THEOREM 1. *Any* (*and then all*) *of Proposition* 1 *conditions* 1–4 *hold if and only if the coupling is successful*, *meaning that for all* $n \in \mathbb{T}$,

$$\mathbf{Q}(T_n > -\infty) = 1 \qquad \forall n \in \mathbb{T}. \tag{3.8}$$

Furthermore, *if* (3.8) *holds*, *then* $\mathbf{Q}(\Phi_{T_n,n}(x) \in \cdot) = \pi_n(\cdot)$ *for all* $x \in E$ *and* $n \in \mathbb{T}$, *where* $(\pi_n)_{n\in\mathbb{T}}$ *is the unique sequence of absolute probabilities for* M.

REMARK 2. As in the case of CFTP for time-homogeneous chains, if instead of (3.2) one allows dependence between $(\xi_n^x; x \in E)$, then it is possible to construct M and $\mathbf{Q}$ such that the backward products of M are weakly ergodic, but $\mathbf{Q}(T_n = -\infty) = 1$ for all n; see, for example Häggström (2002), Chapter 10. On the other hand, under (3.2), the situation is more clear-cut in the sense that the "if and only if" part of Theorem 1 holds. However, it should be noted that couplings involving dependence between $(\xi_n^x; x \in E)$ may lead to more computationally efficient algorithms in some situations, especially when the number of states s is large.

The proof of Theorem 1 is composed of Propositions 2 and 3, which follow Lemma 1.

LEMMA 1. *If for some* $x^\star \in E$, $\min_{x\in E} M_{n,k}(x, x^*) \geq \varepsilon > 0$, *then* $\mathbf{Q}(T_k \geq n) \geq \varepsilon^s$.

PROOF. We have, with $A_{n,k}(j) := \{\Phi_{n,k}(1) = x^*, \dots, \Phi_{n,k}(j) = x^*\}$,

$$\mathbf{Q}(T_k \geq n) \geq \mathbf{Q}(\Phi_{n,k}(1) = x^*, \dots, \Phi_{n,k}(s) = x^*) = \mathbf{Q}(A_{n,k}(s)).$$

We shall prove by an inductive argument that $\mathbf{Q}(A_{n,k}(s)) \geq \varepsilon^s$, the inductive hypothesis being that, with $j \in \{2, \dots, s\}$,

$$\mathbf{Q}(A_{n,k}(j-1)) \geq \varepsilon^{j-1}, \tag{3.9}$$

which is validated in the case $j = 2$ by the assumption of the lemma. Since $\mathbf{Q}(A_{n,k}(j) \mid A_{n,k}(j-1)) = \mathbf{Q}(\Phi_{n,k}(j) = x^* \mid A_{n,k}(j-1))$, to show that (3.9) holds with $j-1$ replaced by j, it is enough to establish

$$\mathbf{Q}(\Phi_{n,k}(j) \neq x^* \mid A_{n,k}(j-1)) \leq 1 - \varepsilon. \tag{3.10}$$

To this end, we need more some notation. Define $Q_{n,k} : E \times (2^E)^{k-n} \to [0,1]$ as

$$
\begin{aligned}
&Q_{n,k}(x, S_{n+1}, \dots, S_k) \\
&\quad := \mathbf{Q}\big(\Phi_{n+1}(x) \notin S_{n+1}, \Phi_{n,n+2}(x) \notin S_{n+2}, \dots, \Phi_{n,k}(x) \notin S_k\big) \\
&\quad = \sum_{(x_{n+1},\dots,x_k)\in S_{n+1}^{\complement}\times\cdots\times S_k^{\complement}} M_{n+1}(x, x_{n+1}) \prod_{i=n+2}^{k} M_i(x_{i-1}, x_i),
\end{aligned}
$$

where the final equality is easily deduced from (3.3). Thus, $Q_{n,k}(x, S_{n+1}, \dots, S_k)$ is the probability that a Markov chain evolving according to M from time n to k starting at x avoids, for each $i \in \{n+1, \dots, k\}$, the set S_i at time i. It follows from the nonnegativity of each M_i that for any sequence of subsets $S_{n+1}, \dots, S_{k-1}$ of E and any points $x, x' \in E$,

$$
\begin{aligned}
Q_{n,k}\big(x, S_{n+1}, \dots, S_{k-1}, \{x'\}\big) &\le Q_{n,k}\big(x, \varnothing, \dots, \varnothing, \{x'\}\big) \\
&= \mathbf{Q}\big(\Phi_{n,k}(x) \neq x'\big).
\end{aligned}
\tag{3.11}
$$

Now with $j \in \{2, \dots, s\}$ as in (3.10), introduce the notation

$$
\mathbf{x}_{n,k,j} := (x_{n+1,1}, \dots, x_{n+1,j-1}, \dots, x_{k-1,1}, \dots, x_{k-1,j-1}),
$$

which is a point in $E^{(k-n-1)(j-1)} =: E_{n,k,j}$, and

$$
B(\mathbf{x}_{n,k,j}) := \big\{\Phi_{n+1}(1) = x_{n+1,1}, \dots, \Phi_{n,k-1}(j-1) = x_{k-1,j-1}\big\}.
$$

Using (3.2), (3.3) and the fact that for any $x, x' \in E$ and $m \in \{n+1, \dots, k\}$, $\{\Phi_{n,m}(x) = \Phi_{n,m}(x')\} \subset \{\Phi_{n,k}(x) = \Phi_{n,k}(x')\}$, it follows by some elementary but tedious manipulations that

$$
\begin{aligned}
&\mathbf{Q}\big(\Phi_{n,k}(j) \neq x^* \mid B(\mathbf{x}_{n,k,j}) \cap A_{n,k}(j-1)\big) \\
&\quad = Q_{n,k}\left(j, \bigcup_{i=1}^{j-1}\{x_{n+1,i}\}, \dots, \bigcup_{i=1}^{j-1}\{x_{k-1,i}\}, \{x^*\}\right),
\end{aligned}
$$

so,

$$
\begin{aligned}
&\mathbf{Q}\big(\Phi_{n,k}(j) \neq x^* \mid A_{n,k}(j-1)\big) \\
&\quad = \sum_{\mathbf{x}_{n,k,j}\in E_{n,k,j}} Q_{n,k}\left(j, \bigcup_{i=1}^{j-1}\{x_{n+1,i}\}, \dots, \bigcup_{i=1}^{j-1}\{x_{k-1,i}\}, \{x^*\}\right) \\
&\qquad \times \mathbf{Q}\big(B(\mathbf{x}_{n,k,j}) \mid A_{n,k}(j-1)\big) \\
&\quad \le Q_{n,k}\big(j, \varnothing, \dots, \varnothing, \{x^*\}\big) = \mathbf{Q}\big(\Phi_{n,k}(j) \neq x^*\big) \le 1 - \varepsilon,
\end{aligned}
\tag{3.12}
$$

where the penultimate inequality and final equality are from (3.11), and the final inequality follows from the hypothesis of the lemma. Thus the inductive hypothesis (3.9) holds with $j-1$ replaced with j, and the proof of the lemma is complete. □

PROPOSITION 2. *If any of Proposition 1's conditions* 1–4 *hold*, *then for all* $n \in \mathbb{T}$, $\mathbf{Q}(T_n > -\infty) = 1$.

PROOF. Under the hypothesis of the proposition, there is only one member of Π_M, denote it by $\pi = (\pi_n)_{n\in\mathbb{T}}$. Since the $(\pi_n)_{n\in\mathbb{T}}$ are probability distributions, for each n there must exist some $x_n^\star \in E$ such that $\pi_n(x_n^\star) \geq s^{-1}$ (recall $E = \{1, \dots, s\}$). Now fix $\varepsilon \in (0, s^{-1})$. By (2.2), for each $k \in \mathbb{T}$ there exists $n < k$ such that

$$M_{n,k}(x, x_k^\star) \geq \pi_k(x_k^\star) - (s^{-1} - \varepsilon) \geq \varepsilon > 0 \qquad \forall x \in E.$$

We may then define $(k_i)_{i\in\mathbb{N}}$ a strictly decreasing subsequence of $\mathbb{T}$, with $k_0 := 0$ and

$$k_{i+1} := \sup\{n < k_i : M_{n,k_i}(x, x_{k_i}^\star) \geq \varepsilon \ \forall x \in E\},$$

so that by construction

$$\inf_{i\in\mathbb{N}} \min_{x\in E} M_{k_{i+1},k_i}(x, x_{k_i}^\star) \geq \varepsilon > 0.$$

Lemma 1 then gives

$$\sup_{i\in\mathbb{N}} \mathbf{Q}(T_{k_i} \leq k_{i+1}) \leq 1 - \varepsilon^s. \tag{3.13}$$

We now wish to apply this bound to control the tails of the coalescence times T_n. To this end, first note that for any $n, k, k' \in \mathbb{T}$, $k' < k < n$,

$$\{T_n \geq k\} \cup \{T_k \geq k'\} \subseteq \{T_n \geq k'\}$$

and since the events $\{T_n \geq k\}$ and $\{T_k \geq k'\}$ are independent, we have

$$\mathbf{Q}(T_n < k') \leq \mathbf{Q}(T_n < k)\mathbf{Q}(T_k < k'). \tag{3.14}$$

Now fix $n \in \mathbb{T}$. Since $(k_i)_{i\in\mathbb{N}}$ is strictly decreasing, there exists some $i(n)$ such that $k_{i(n)} < n$. Then by repeated application of (3.14), we find that for any $\ell > i(n)$,

$$\mathbf{Q}(T_n < k_\ell) \leq \mathbf{Q}(T_n < k_{i(n)}) \prod_{j=i(n)}^{\ell-1} \mathbf{Q}(T_{k_j} < k_{j+1}). \tag{3.15}$$

Now (3.13) provides an upper bound for the j-indexed terms in (3.15), and then taking $\ell \to \infty$ we find $\mathbf{Q}(T_n > -\infty) = 1$, which completes the proof of the proposition. $\square$

REMARK 3. If one has available quantitative convergence information in addition to (2.2), then the inequalities (3.14), (3.15) and Lemma 1 could be used to bound the moments of the T_n.

PROPOSITION 3. *If for all* $n \in \mathbb{T}$, $\mathbf{Q}(T_n > -\infty) = 1$, *then both of the following hold*:

1. *There exists a strictly decreasing subsequence* $(n_i)_{i\in\mathbb{N}}$ *of* $\mathbb{T}$ *such that* $\sum_{i=0}^{\infty} 1-\beta(M_{n_{i+1},n_i})=\infty$, *that is, condition* 4 *of Proposition* 1 *holds.*

2. *For all* $n\in\mathbb{T}$, $\mathbf{Q}(X_{T_n}^x(n)\in\cdot)=\pi_n(\cdot)$ *for all* $x\in E$, *where* $(\pi_n)_{n\in\mathbb{T}}$ *is the unique sequence of absolute probabilities for* M.

PROOF. Fix some $\delta>0$. Under the hypothesis, we have that for each n there exists $k\in\mathbb{T}$ such that $\mathbf{Q}(T_n<k)<\delta$. We may therefore define $(k_i)_{i\in\mathbb{N}}$ a strictly decreasing subsequence of $\mathbb{T}$ with $k_0:=0$,

$$k_{i+1}:=\sup\{n<k_i:\mathbf{Q}(T_{k_i}<n)<\delta\},$$

so that by construction

$$\sup_{i\in\mathbb{N}}\mathbf{Q}(T_{k_i}<k_{i+1})<\delta. \tag{3.16}$$

Now for any $x,x'\in E$, $\{T_{k_i}\geq k_{i+1}\}\subseteq\{\Phi_{k_{i+1},k_i}(x)=\Phi_{k_{i+1},k_i}(x')\}$ so

$$\mathbf{Q}(\Phi_{k_{i+1},k_i}(x)\neq\Phi_{k_{i+1},k_i}(x'))\leq\mathbf{Q}(T_{k_i}<k_{i+1}).$$

Combining this observation with (3.16) and (3.6), we obtain

$$\begin{aligned}\beta(M_{k_{i+1},k_i})&=\max_{x,x'}\|M_{k_{i+1},k_i}(x,\cdot)-M_{k_{i+1},k_i}(x',\cdot)\|\\&=\max_{x,x'}\sup_{A\subset E}|\mathbf{Q}(\Phi_{k_{i+1},k_i}(x)\in A)-\mathbf{Q}(\Phi_{k_{i+1},k_i}(x')\in A)|\\&\leq\max_{x,x'}\mathbf{Q}(\Phi_{k_{i+1},k_i}(x)\neq\Phi_{k_{i+1},k_i}(x'))<\delta\qquad\forall i\in\mathbb{N},\end{aligned}\tag{3.17}$$

where we have used the fact that for any two E-valued random variables X,X' defined on a common probability space, $\sup_{A\subset E}|\mathbf{P}(X\in A)-\mathbf{P}(X'\in A)|\leq\mathbf{P}(X\neq X')$ [Lindvall (2002), page 12]. From (3.17), we immediately have $\sum_i 1-\beta(M_{k_{i+1},k_i})=\infty$, which completes the proof of part (1).

For part (2), fix any $n\in\mathbb{T}$, and note that on the event $\{T_n>-\infty\}$, $\Phi_{T_n,n}(x)$ is well-defined as a random variable. When $\mathbf{Q}(T_n>-\infty)=1$, we have by construction of the algorithm that $\lim_{k\to-\infty}\Phi_{n+k,n}(x)=\Phi_{T_n,n}(x)$, $\mathbf{Q}$-a.s. Using (2.2), we also have for any $z\in E$, $\mathbf{Q}(\Phi_{n+k,n}(x)=z)=M_{n+k,n}(x,z)\to\pi_n(z)$ as $k\to-\infty$, hence $\mathbf{Q}(\Phi_{T_n,n}(x)=z)=\pi_n(z)$. The proof is complete. $\square$

4. Tail triviality and unicity of absolute probabilities. Let $\Omega^X=E^{\mathbb{T}}$, let $\mathcal{F}^X=\mathcal{B}(E)^{\otimes\mathbb{T}}$ be the product σ-algebra. Let $X=(X_n)_{n\in\mathbb{T}}$ be the coordinate process on Ω^X and for $I\subset\mathbb{T}$, define $\mathcal{F}_I^X=\sigma(X_n;n\in I)$. As in Section 1, for any $\pi\in\Pi_M$ we let $\mathbf{P}_\pi$ be the probability measure on $(\Omega^X,\mathcal{F}^X)$ constructed from the finite dimensional distributions $(\mathbf{P}_\pi^{(n)})_{n\in\mathbb{T}}$ given by

$$\mathbf{P}_\pi^{(n)}(X_n=x_n,\dots,X_0=x_0):=\pi_n(x_n)\prod_{k=n+1}^{0}M_k(x_{k-1},x_k). \tag{4.1}$$

Expectation w.r.t. $\mathbf{P}_\pi$ is denoted by $\mathbf{E}_\pi$. The main result of this section is the following theorem, which via Proposition 1 gives an alternative characterization of the success of the coupling in the sense of Theorem 1.

THEOREM 2. *The following are equivalent*:

1. $\operatorname{card}(\Pi_M) = 1$.
2. $\mathbf{P}_\pi(A) = \mathbf{P}_\pi(A)^2 = \mathbf{P}_{\tilde\pi}(A)$, $\forall(\pi, \tilde\pi, A) \in \Pi_M \times \Pi_M \times \bigcap_{n\in\mathbb{T}} \mathcal{F}^X_{]-\infty,n]}$.

When 1 holds, then obviously $\mathbf{P}_\pi(A) = \mathbf{P}_{\tilde\pi}(A)$. The proof of $1 \Rightarrow 2$ is completed by Proposition 4. The implication $2 \Rightarrow 1$ is the subject of Proposition 5.

PROPOSITION 4. *If there exists* $\pi \in \Pi_M$ *and* $A \in \bigcap_{n\in\mathbb{T}} \mathcal{F}^X_{]-\infty,n]}$ *such that* $\mathbf{P}_\pi(A) \in]0, 1[$, *then* $\operatorname{card}(\Pi_M) > 1$.

PROOF. Let π and A be as in the statement of the proposition and with $Z(\omega) := \mathbb{I}_A(\omega)/\mathbf{P}_\pi(A)$, define a new probability measure $\tilde{\mathbf{P}}$ on $(\Omega^X, \mathcal{F}^X)$ by $\tilde{\mathbf{P}}(d\omega) := Z(\omega)\mathbf{P}_\pi(d\omega)$, that is, $\tilde{\mathbf{P}}(\cdot) = \mathbf{P}_\pi(\cdot \mid A)$. Define also the sequence of marginal distributions $\tilde\pi = (\tilde\pi_n)_{n\in\mathbb{T}}$, $\tilde\pi_n(\cdot) := \tilde{\mathbf{P}}(X_n \in \cdot)$. We are going to show that $\tilde\pi \in \Pi_M$ and $\tilde\pi \neq \pi$, thus proving $\operatorname{card}(\Pi_M) > 1$ as desired.

The Markov property of X under $\mathbf{P}_\pi$ and the fact that Z is measurable w.r.t. to $\bigcap_n \mathcal{F}^X_{]-\infty,n]}$ combine to give $\mathbf{E}_\pi[Z \mid \mathcal{F}^X_{[n,0]}] = \mathbf{E}_\pi[Z \mid \sigma(X_n)]$, $\mathbf{P}$-a.s., so, for each $n \in \mathbb{T}$, there exists a measurable function h_n on E, uniquely defined and nonnegative π_n-almost everywhere, such that $h_n(X_n) = \mathbf{E}_\pi[Z \mid \mathcal{F}^X_{[n,0]}]$, $\mathbf{P}$-a.s. We then have, for any $n \in \mathbb{T}$ and $(x_n, \dots, x_0) \in E^{|n|+1}$,

$$
\begin{aligned}
&\tilde{\mathbf{P}}(\{\omega : X_n(\omega) = x_n, \dots, X_0(\omega) = x_0\}) \\
&\quad = \int_{\{\omega: X_n(\omega)=x_n,\dots,X_0(\omega)=x_0\}} Z(\omega)\, d\mathbf{P}_\pi \\
&\quad = \int_{\{\omega: X_n(\omega)=x_n,\dots,X_0(\omega)=x_0\}} \mathbf{E}_\pi[Z \mid \mathcal{F}^X_{[n,0]}](\omega)\, d\mathbf{P}_\pi \\
&\quad = \int_{\{\omega: X_n(\omega)=x_n,\dots,X_0(\omega)=x_0\}} h_n(X_n(\omega))\, d\mathbb{P}_\pi \\
&\quad = h_n(x_n)\pi_n(x_n) \prod_{k=n+1}^{0} M_k(x_{k-1}, x_k).
\end{aligned} \tag{4.2}
$$

From (4.2), we immediately deduce three facts. First, for each $n \in \mathbb{T}$ and $x \in E$, $\tilde\pi_n(x) = h_n(x)\pi_n(x)$. Second, $\tilde\pi \in \Pi_M$. Third, the finite dimensional marginals of $\tilde{\mathbf{P}}$ coincide with those of $\mathbf{P}_{\tilde\pi}$, so by a monotone class argument, $\tilde{\mathbf{P}} = \mathbf{P}_{\tilde\pi}$.

It remains to prove that $\tilde\pi \neq \pi$. Yet again by a monotone class argument, note that for any $\mu, \nu \in \Pi_M$, if for all $n \in \mathbb{T}$, $\mathbf{P}^{(n)}_\mu = \mathbf{P}^{(n)}_\nu$, then $\mathbf{P}_\mu = \mathbf{P}_\nu$. We have already seen that $\tilde{\mathbf{P}} = \mathbf{P}_{\tilde{\pi}}$, and by construction, $\tilde{\mathbf{P}} \neq \mathbf{P}_{\pi}$, so by applying the contrapositive of the implication in the previous sentence, there must exist some $n \in \mathbb{T}$ such that $\mathbf{P}_{\pi}^{(n)} \neq \mathbf{P}_{\tilde{\pi}}^{(n)}$, which is only possible if there exists some x such that $\pi_n(x) \neq \tilde{\pi}_n(x)$. This completes the proof. $\square$

PROPOSITION 5. *If for all* $(\pi, \tilde{\pi}, A) \in \Pi_M \times \Pi_M \times \bigcap_{n\in\mathbb{T}} \mathcal{F}^X_{]-\infty,n]}$, $\mathbf{P}_\pi(A) = \mathbf{P}_\pi(A)^2 = \mathbf{P}_{\tilde{\pi}}(A)$, *then* $\mathrm{card}(\Pi_M) = 1$.

PROOF. Fix arbitrarily $x \in E$, $k \in \mathbb{T}$ and let π be any member of Π_M. For any $n \leq k$, we have $\mathbf{P}_\pi(X_k = x \mid \mathcal{F}^X_{]-\infty,n]}) = M_{n,k}(X_n, x)$, $\mathbf{P}_\pi$-a.s. and since $\mathcal{F}^X_{]-\infty,n]} \searrow \bigcap_{n\in\mathbb{T}} \mathcal{F}^X_{]-\infty,n]}$, a classical martingale convergence theorem [Doob (1953), page 331, Theorem 4.3] dictates that $\lim_{n\to-\infty} \mathbf{P}_\pi(X_k = x \mid \mathcal{F}^X_{]-\infty,n]}) = \mathbf{P}_\pi(X_k = x \mid \bigcap_{n\in\mathbb{T}} \mathcal{F}^X_{]-\infty,n]})$, $\mathbf{P}_\pi$-a.s. Under the hypothesis of the proposition, $\mathbf{P}_\pi(A) \in \{0,1\}$ for all $A \in \bigcap_{n\in\mathbb{T}} \mathcal{F}^X_{]-\infty,n]}$, and by construction $\mathbf{P}_\pi(X_k = x) = \pi_k(x)$, so we obtain

$$\lim_{n\to-\infty} M_{n,k}(X_n, x) = \pi_k(x), \qquad \mathbf{P}_\pi\text{-a.s.} \tag{4.3}$$

Now choose any $\tilde{\pi} \in \Pi_M$. Repeating the above argument, we obtain

$$\lim_{n\to-\infty} M_{n,k}(X_n, x) = \tilde{\pi}_k(x), \qquad \mathbf{P}_{\tilde{\pi}}\text{-a.s.}, \tag{4.4}$$

and since $A_x := \{\lim_{n\to-\infty} M_{n,k}(X_n, x) = \pi_k(x)\} \in \bigcap_{n\in\mathbb{T}} \mathcal{F}^X_{]-\infty,n]}$, the hypothesis of the proposition dictates $\mathbf{P}_\pi(A_x) = \mathbf{P}_{\tilde{\pi}}(A_x)$, so from (4.3) and (4.4) we find $\pi_k(x) = \tilde{\pi}_k(x)$. Since x and k were arbitrary, we have thus established $\pi = \tilde{\pi}$, and since π and $\tilde{\pi}$ were arbitrary members of Π_M we have proved that $\mathrm{card}(\Pi_M) = 1$. $\square$

5. Perfect sampling for hidden Markov models.

5.1. *The model.* Throughout Section 5, we take $\Omega^X := E^{\mathbb{T}}$ equipped with the product σ-algebra $\mathcal{B}(E)^{\otimes\mathbb{T}}$, we introduce F a nonempty, Polish state-space with Borel σ-algebra denoted by $\mathcal{B}(F)$, and we consider $\Omega^Y := F^{\mathbb{T}}$ equipped with the product σ-algebra $\mathcal{B}(F)^{\otimes\mathbb{T}}$. Define $\Omega := \Omega^X \times \Omega^Y$ and the coordinate projections: ζ, η by

$$\zeta : (x, y) \in \Omega \mapsto x \in \Omega^X, \qquad \eta : (x, y) \in \Omega \mapsto y \in \Omega^Y,$$

and $(\tilde{X}_n)_{n\in\mathbb{T}}$, $(\tilde{Y}_n)_{n\in\mathbb{T}}$ by

$$\tilde{X}_n : x = (\dots, x_{-1}, x_0) \in \Omega^X \mapsto x_n \in E,$$
$$\tilde{Y}_n : y = (\dots, y_{-1}, y_0) \in \Omega^Y \mapsto y_n \in F.$$

Then let

$$X_n := \tilde{X}_n \circ \zeta, \qquad Y_n := \tilde{Y}_n \circ \eta,$$

so clearly $X_n : \omega = (x, y) \in \Omega \mapsto x_n \in E$ and $Y_n : \omega = (x, y) \in \Omega \mapsto y_n \in F$. We shall write X and Y for respectively the $E^{\mathbb{T}}$ and $F^{\mathbb{T}}$-valued random variables $(X_n)_{n\in\mathbb{T}}$ and $(Y_n)_{n\in\mathbb{T}}$.

Let $\mathcal{F}$ be the Borel σ-algebra on Ω and for $I \subset \mathbb{T}$, define

$$\mathcal{F}_I^{\tilde{X}} := \sigma(\tilde{X}_n; n \in I), \qquad \mathcal{F}_I^{X} := \sigma(X_n; n \in I),$$
$$\mathcal{F}_I^{\tilde{Y}} := \sigma(\tilde{Y}_n; n \in I), \qquad \mathcal{F}_I^{Y} := \sigma(Y_n; n \in I),$$
$$\mathcal{F}_I := \mathcal{F}_I^X \vee \mathcal{F}_I^Y,$$

and $\mathcal{F}^{\tilde{X}} := \mathcal{F}^{\tilde{X}}_{\mathbb{T}}$, $\mathcal{F}^{X} := \mathcal{F}^{X}_{\mathbb{T}}$, $\mathcal{F}^{\tilde{Y}} := \mathcal{F}^{\tilde{Y}}_{\mathbb{T}}$, $\mathcal{F}^{Y} := \mathcal{F}^{Y}_{\mathbb{T}}$.

Now introduce two sequences of probability kernels $M = (M_n)_{n\in\mathbb{T}}$ and $G = (G_n)_{n\in\mathbb{T}}$, with each $M_n : E \times \mathcal{B}(E) \to [0, 1]$ and $G_n : E \times \mathcal{B}(F) \to [0, 1]$. We assume that $G_n(x, dy) = g_n(x, y)\psi(dy)$ for some $g_n : E \times F \to [0, \infty[$ and ψ a σ-finite measure on $(F, \mathcal{B}(F))$.

Throughout Section 5, $\mathbf{P}$ is a probability measure on $(\Omega, \mathcal{F})$ under which (X, Y) is a hidden Markov model, constructed as follows. Fix some $\pi = (\pi_n)_{n\in\mathbb{T}} \in \Pi_M$. For each $n \in \mathbb{T}$ define a probability $\mathbf{P}^{(n)}$ on $(\mathcal{B}(E) \otimes \mathcal{B}(F))^{\otimes(|n|+1)}$ by

$$\mathbf{P}^{(n)}(A) = \int_A \pi_n(dx_n)G_n(x_n, dy_n) \prod_{k=n+1}^{0} M_k(x_{k-1}, dx_k)G_k(x_k, dy_k), \tag{5.1}$$

with the convention that the product is unity when $n = 0$. Since $\pi \in \Pi_M$, the $\mathbf{P}^{(n)}$ are consistent, giving rise via the usual extension argument to a probability measure $\mathbf{P}$ on $(\Omega, \mathcal{F})$. Expectation w.r.t. $\mathbf{P}$ is denoted by $\mathbf{E}$. We shall write $\mathbf{P} \circ Y^{-1}$ for the push-forward of $\mathbf{P}$ by Y, that is, $(\mathbf{P} \circ Y^{-1})(H) = \mathbf{P}(\{\omega \in \Omega : Y(\omega) \in H\})$, for $H \in \mathcal{B}(F)^{\otimes\mathbb{T}}$.

Let us now remark upon some details of this setup (the analogues of the following properties for an HMM on a nonnegative time horizon are well known and the arguments involved in establishing them depend only superficially on the direction of time). Under $\mathbf{P}$, the bivariate process $(X_n, Y_n)_{n\in\mathbb{T}}$ is Markov, which implies that the following holds $\mathbf{P}$-a.s.:

$$\mathbf{P}\big(X_n \in \cdot \mid \mathcal{F}^Y_{]-\infty,0]} \vee \mathcal{F}^X_{]-\infty,n-1]}\big) = \mathbf{P}\big(X_n \in \cdot \mid \mathcal{F}^Y_{]-\infty,0]} \vee \sigma(X_{n-1})\big). \tag{5.2}$$

Moreover, under $\mathbf{P}$, X is a Markov, with for each $n \in \mathbb{T}$, X_n distributed according to π_n and $X_n \mid X_{n-1} \sim M_n(X_{n-1}, \cdot)$. The observations Y are conditionally independent given X, and the conditional distribution of Y_n given X is $G_n(X_n, \cdot)$. It follows from this conditional-independence structure that the following holds $\mathbf{P}$-a.s.:

$$\mathbf{P}\big(X_n \in \cdot \mid \mathcal{F}^Y_{]-\infty,0]} \vee \sigma(X_{n-1})\big) = \mathbf{P}\big(X_n \in \cdot \mid \mathcal{F}^Y_{[n,0]} \vee \sigma(X_{n-1})\big). \tag{5.3}$$

5.2. *Transition kernels of the conditional signal process*. Define the sequence of functions $(\phi_n)_{n\in\mathbb{T}}$, each $\phi_n : E \times F^n \to [0,\infty[$, recursively as

$$\begin{aligned} \phi_0(x, y_0) &:= \sum_{x'\in E} M_0(x,x') g_0(x', y_0), \\ \phi_{n-1}(x, y_{n-1:0}) &:= \sum_{x'\in E} M_{n-1}(x,x') g_{n-1}(x', y_{n-1}) \phi_n(x', y_{n:0}). \end{aligned} \tag{5.4}$$

Now with $y = (y_n)_{n\in\mathbb{T}}$, introduce for each $n \in \mathbb{T}$,

$$\begin{aligned} &M_n^y(x,x') \\ &\quad := \begin{cases} \dfrac{M_n(x,x') g_n(x', y_n) \phi_{n+1}(x', y_{n+1:0})}{\phi_n(x, y_{n:0})}, & \phi_n(x, y_{n:0}) > 0, \\ M_n(x,x'), & \phi_n(x, y_{n:0}) = 0, \end{cases} \end{aligned} \tag{5.5}$$

with the convention that $\phi_1(x, y_{1:0}) \equiv 1$. Similar to (1.2), let

$$M_{k,k}^y := \mathrm{Id}, \qquad M_{n-1,k}^y(x,x') := \sum_{z\in E} M_n^y(x,z) M_{n,k}^y(z,x'), \qquad n \le k. \tag{5.6}$$

According to (5.5), for each y, $M_n^y(\cdot,\cdot)$ is clearly a Markov kernel on E. This kernel provides a version of the conditional probabilities in (5.2)–(5.3), in the sense of the following lemma, proof of which is given in the Appendix.

LEMMA 2. *For each $n \in \mathbb{T}$ and $x \in E$*,

$$\mathbf{P}(X_n = x \mid \mathcal{F}_{[n,0]}^Y \vee \sigma(X_{n-1})) = M_n^Y(X_{n-1}, x), \qquad \mathbf{P}\text{-}a.s.$$

We next establish the existence of a particular y-dependent sequence of absolute probabilities for the Markov kernels $M^y = (M_n^y)_{n\in\mathbb{T}}$.

LEMMA 3. *For each $n \in \mathbb{T}$, there exists a probability kernel $\mu_n^{\cdot}(\cdot) : \Omega^Y \times E \to [0,1]$, such that for all $x \in E$*,

$$\mathbf{P}(X_n = x \mid \mathcal{F}_{]-\infty,0]}^Y) = \mu_n^Y(x), \qquad \mathbf{P}\text{-}a.s.$$

and

$$\sum_{x'\in E} \mu_{n-1}^Y(x') M_n^Y(x', x) = \mu_n^Y(x), \qquad \mathbf{P}\text{-}a.s.$$

PROOF. Since E is a finite set, the existence for any n of a probability kernel $\mu_n^{\cdot}(\cdot) : \Omega^Y \times E \to [0,1]$ satisfying $\mathbf{P}(X_n = x \mid \mathcal{F}_{]-\infty,0]}^Y) = \mu_n^Y(x)$, $\mathbf{P}$-a.s. for all x, is immediate. Then by the tower property of conditional expectation, Lemma 2 and (5.3), the following equalities hold **P**-a.s.:

$$\begin{aligned}\mu_n^Y(x) &= \mathbf{P}(X_n = x \mid \mathcal{F}^Y_{]-\infty,0]}) \\ &= \mathbf{E}\big[\mathbf{E}\big[\mathbb{I}\{X_n = x\} | \mathcal{F}^Y_{]-\infty,0]} \vee \sigma(X_{n-1})\big] | \mathcal{F}^Y_{]-\infty,0]}\big] \\ &= \mathbf{E}\big[M_n^Y(X_{n-1}, x) \mid \mathcal{F}^Y_{]-\infty,0]}\big] \\ &= \sum_{x' \in E} \mu_{n-1}^Y(x') M_n^Y(x', x).\end{aligned}$$

□

5.3. *The coupling for the HMM.* Our next main objective is to apply the construction and results of Section 3 to derive and study a perfect sampling procedure associated with the Markov kernels M^y. The setup is as follows. With $\Omega^\xi = (E^s)^{\otimes \mathbb{T}}$ and $\mathcal{F}^\xi = (\mathcal{B}(E)^{\otimes s})^{\otimes \mathbb{T}}$, let $\mathbf{Q}^\cdot(\cdot) : \Omega^Y \times \mathcal{F}^\xi \to [0, 1]$ be a probability kernel such that for each $y \in \Omega^Y$ the coordinate projections $(\xi_n^x; x \in E; n \in \mathbb{T})$ are distributed under $\mathbf{Q}^y(\cdot)$ as

$$(\xi_n^x; x \in E, n \in \mathbb{T}) \text{ are independent,} \tag{5.7}$$

$$\mathbf{Q}^y(\xi_n^x = x') = M_n^y(x, x'), \qquad (n, x, x') \in \mathbb{T} \times E^2. \tag{5.8}$$

Thus with y fixed, $\mathbf{Q}^y(\cdot)$ may be regarded as an instance of the probability measure denoted $\mathbf{Q}(\cdot)$ in Section 3. Also, let the maps Φ_n, $\Phi_{n,k}$ and the coalescence times $(T_n)_{n \in \mathbb{T}}$ be defined exactly as in equations (3.4), (3.5) and (3.7) of Section 3.

PROPOSITION 6. *Fix any $y \in \Omega^Y$. Any (and all) of Proposition 1 conditions 1–4 hold for the Markov kernels M^y, if and only if*

$$\mathbf{Q}^y(T_n > -\infty) = 1 \qquad \forall n \in \mathbb{T}. \tag{5.9}$$

*Furthermore, if (5.9) holds then for all $n \in \mathbb{T}$, $\mathbf{Q}^y(\Phi_{T_n,n}(x) \in \cdot) = \pi_n^y(\cdot)$ for all $x \in E$, where $\pi^y = (\pi_n^y)_{n \in \mathbb{T}}$ is the unique sequence of absolute probabilities for M^y. If (5.9) holds for y in a set of $\mathbf{P} \circ Y^{-1}$ probability 1, then for all $n \in \mathbb{T}$ and $x \in E$, $\mathbf{Q}^Y(\Phi_{T_n,n}(x) \in \cdot) = \mathbf{P}(X_n \in \cdot \mid \mathcal{F}^Y)$, **P**-a.s., and we call the coupling a.s. successful.*

PROOF. For fixed $y \in \Omega^Y$, the claimed equivalence between (5.9) and Proposition 1 conditions 1–4 holding for the Markov kernels M^y is an application of Theorem 1. So, too, is the equality $\mathbf{Q}^y(\Phi_{T_n,n}(x) \in \cdot) = \pi_n^y(\cdot)$. Since E is a finite set and $\mathbb{T}$ is countable, it follows from Lemma 3 that there exists $H \in \mathcal{F}^Y$ with $\mathbf{P}(H) = 1$ and such that for all $\omega \in H$, $n \in \mathbb{T}$ and $x \in E$, $\mathbf{P}(X_n = x \mid \mathcal{F}^Y)(\omega) = \mu_n^{Y(\omega)}(x)$ and $\sum_{x' \in E} \mu_{n-1}^{Y(\omega)}(x') M_n^{Y(\omega)}(x', x) = \mu_n^{Y(\omega)}(x)$. If, as hypothesized in the statement, there exists $\tilde{H} \in \mathcal{F}^{\tilde{Y}}$ such that for all $y \in \tilde{H}$, $\mathbf{Q}^y(T_n > -\infty) = 1$ for all $n \in \mathbb{T}$, then for all $\omega \in H \cap Y^{-1}(\tilde{H})$, $M^{Y(\omega)}$ admits a unique sequence of absolute probabilities, and so $\pi_n^{Y(\omega)}(x) = \mu_n^{Y(\omega)}(x) = \mathbf{P}(X_n = x \mid \mathcal{F}^Y)(\omega)$. □

Algorithm 1: Perfect sampling for the hidden Markov model

for each $x \in E$, set

$$\phi_0(x, y_0) = \sum_{x' \in E} M_0(x, x') g_0(x', y_0),$$

and by convention, $\phi_1(x, y_{1:0}) = 1$.

set $\Phi_{0,0} = \mathrm{Id}$

set $n = 0$

while card(image of $\Phi_{n,0}$) > 1

for each $x \in E$,

for each $x' \in E$, set

$$M_n^y(x, x') = \begin{cases} \dfrac{M_n(x, x') g_n(x', y_n) \phi_{n+1}(x', y_{n+1:0})}{\phi_n(x, y_{n:0})}, & \phi_n(x, y_{n:0}) > 0, \\ M_n(x, x'), & \phi_n(x, y_{n:0}) = 0, \end{cases}$$

sample $\xi_n^x \sim M_n^y(x, \cdot)$ and set $\Phi_n(x) = \xi_n^x$

set $\Phi_{n-1,0} = \Phi_{n,0} \circ \Phi_n$

set $n = n - 1$

for each $x \in E$, set

$$\phi_n(x, y_{n:0}) = \sum_{x' \in E} M_n(x, x') g_n(x', y_n) \phi_{n+1}(x', y_{n+1:0}).$$

return $\Phi_{n,0}(x)$, for any $x \in E$.

We present in Algorithm 1 some steps of the sampling procedure in order to emphasize the way that the observations enter into recursive computations. For simplicity of presentation, we consider the case of implementing the coupling until $T_0 = \sup\{n < 0 : \text{image of } \Phi_{n,0} \text{ is a singleton}\}$, thus upon termination in a.s. finite time of the below algorithm, the output value is a sample from μ_0^y. The important point here is that to run this algorithm one needs access to only the observations $y_0, \dots, y_{T_0}$.

5.4. *Successful coupling and conditional ergodicity.* With a little further technical work, we can relate the successful coupling in the sense of (5.9) to the conditional ergodicity properties of the HMM. Our next step is to perform some careful accounting of certain σ-algebras to help us transfer results backward and forward between the measurable space $(\Omega, \mathcal{F})$ underlying the HMM and the "marginal" space $(\Omega^X, \mathcal{F}^{\tilde{X}})$; the attentive reader will have noticed that under the definitions of Section 5.1, $\mathcal{F}_I^X$ consists of subsets of Ω, where as $\mathcal{F}_I^{\tilde{X}}$ consists of subsets Ω^X. On the other hand, $\mathcal{F}_I^{\tilde{X}}$ coincides with the object in Section 4 denoted there by

$\mathcal{F}_I^X$, and in terms of which Theorem 2 is phrased. The resolution of this issue is provided by the following technical lemma, the proof of which is given in the Appendix.

LEMMA 4. *With the definitions of Section 5.1 in force*,

$$\mathcal{F}_I^X = \{A \times \Omega^Y; A \in \mathcal{F}_I^{\tilde{X}}\} \qquad \forall I \subset \mathbb{T},$$

and,

$$\bigcap_{n \in \mathbb{T}} \mathcal{F}_{]-n,0]}^X = \left\{A \times \Omega^Y; A \in \bigcap_{n \in \mathbb{T}} \mathcal{F}_{]-n,0]}^{\tilde{X}}\right\}. \tag{5.10}$$

Lemma 4 allows us to set up correspondence between probabilities on $\mathcal{F}^{\tilde{X}}$ and $\mathcal{F}^X$, and in particular we have the following.

LEMMA 5. *There exist a probability kernel* $P^{\cdot}(\cdot) : \Omega^Y \times \mathcal{F}^{\tilde{X}} \to [0, 1]$ *and a set* $\tilde{H} \in \mathcal{F}^{\tilde{Y}}$ *of* $\mathbf{P} \circ Y^{-1}$ *probability* 1, *such that for all* $y \in \tilde{H}$

$$P^y(\{\tilde{X}_n = x_n, \dots, \tilde{X}_0 = x_0\}) = \mu_n^y(x_n) \prod_{k=n+1}^{0} M_k^y(x_{k-1}, x_k).$$

The function $\mathbf{P}^{\mathcal{F}^Y} : \Omega \times \mathcal{F}^X \to [0, 1]$ *defined by*

$$\mathbf{P}^{\mathcal{F}^Y}(\omega, A \times \Omega^Y) := P^{Y(\omega)}(A), \qquad A \in \mathcal{F}^{\tilde{X}},$$

is a probability kernel, *and for each* $A \in \mathcal{F}^{\tilde{X}}$,

$$\mathbf{P}^{\mathcal{F}^Y}(\omega, A \times \Omega^Y) = \mathbf{P}(A \times \Omega^Y \mid \mathcal{F}^Y)(\omega) \qquad \textit{for } \mathbf{P}\textit{-almost all } \omega \in \Omega. \tag{5.11}$$

The proof is in the Appendix. For any $y \in \Omega^Y$, we denote by Π_{M^y} the set of all sequences of absolute probabilities for $M^y = (M_n^y)_{n \in \mathbb{T}}$. The set Π_{M^y} is nonempty by Fact 1. For any $y \in \Omega^Y$ we shall write generically π^y for a member of Π_{M^y} (we do not claim measurable dependence of π_n^y on y except at least in the case of $\pi_n^y = \mu_n^y$ with the latter as in Lemma 3), and, by arguments only superficially different (we omit the details) to those used in the proof of Lemma 5, for any such $\pi^y \in \Pi_{M^y}$ there exists a probability measure P_{π^y} on $\mathcal{F}^{\tilde{X}}$ such that

$$P_{\pi^y}(\{\tilde{X}_n = x_n, \dots, \tilde{X}_0 = x_0\}) = \pi_n^y(x_n) \prod_{k=n+1}^{0} M_k^y(x_{k-1}, x_k), \tag{5.12}$$

and

$$\mathbf{P}_{\pi^y}(A \times \Omega^Y) := P_{\pi^y}(A), \qquad A \in \mathcal{F}^{\tilde{X}},$$

defines a probability measure on $\mathcal{F}^X$. We now have the technical and notational devices to state and prove the following theorem, which characterizes almost sure success of the coupling.

THEOREM 3. *The following are equivalent*:

1. $\mathbf{Q}^{Y(\omega)}(\bigcap_{n\in\mathbb{T}}\{T_n > -\infty\}) = 1$ *for* $\mathbf{P}$-*almost all* ω.
2. *There exists a set* $H \in \mathcal{F}$ *such that* $\mathbf{P}(H) = 1$ *and*

$$\mathbf{P}^{\mathcal{F}^Y}(\omega, A) = \mathbf{P}^{\mathcal{F}^Y}(\omega, A)^2 = \mathbf{P}_{\pi^{Y(\omega)}}(A),$$

for all $\omega \in H$, $A \in \bigcap_{n\in\mathbb{T}} \mathcal{F}^X_{]-\infty,n]}$ *and* $\pi^{Y(\omega)} \in \Pi_{M^{Y(\omega)}}$.

PROOF. When 1 holds, there exists $H \in \mathcal{F}$ with $\mathbf{P}(H) = 1$ such that for all $\omega \in H$ the following hold: for all $n \in \mathbb{T}$, $\mathbf{Q}^{Y(\omega)}(T_n > -\infty) = 1$; then via Proposition 6 and Proposition 1, $\operatorname{card}(\Pi_{M^{Y(\omega)}}) = 1$; then by an application of Theorem 2 with the $\mathbf{P}_\pi$ appearing there taken to be $P^{Y(\omega)}(\cdot)$, and Lemma 4, we have $P^{Y(\omega)}(\tilde{A}) = P^{Y(\omega)}(\tilde{A})^2 = P_{\pi^{Y(\omega)}}(\tilde{A})$ for all $\tilde{A} \in \bigcap_{n\in\mathbb{T}} \mathcal{F}^{\tilde{X}}_{]-\infty,n]}$ and $\pi^{Y(\omega)} \in \Pi_{M^{Y(\omega)}}$. Lemmata 4 and 5 then give $\mathbf{P}^{\mathcal{F}^Y}(\omega, A) = \mathbf{P}^{\mathcal{F}^Y}(\omega, A)^2 = \mathbf{P}_{\pi^{Y(\omega)}}(A)$ for all $A \in \bigcap_{n\in\mathbb{T}} \mathcal{F}^X_{]-\infty,n]}$, which establishes 2.

If 2 holds, we apply this chain of reasoning in reverse to establish 1. The details are omitted in order to avoid repetition. □

6. Discussion. Throughout Section 6, the definitions and constructions of Section 5 are in force. In particular, it is timely to recall that the law of the HMM, $\mathbf{P}$, has the defining ingredients:

- $M = (M_n)_{n\in\mathbb{T}}$ a sequence of Markov kernels on E,
- $\pi = (\pi)_{n\in\mathbb{T}} \in \Pi_M$ a sequence of absolute probabilities for M,
- $G = (G_n)_{n\in\mathbb{T}}$ a sequence of probability kernels, each acting from E to F and such that for each n, $G_n(x, dy) = g_n(x, y)\psi(dy)$.

We shall consider various combinations of the following assumptions.

ASSUMPTION 1. Under $\mathbf{P}$ the signal process is ergodic, in that $\mathbf{P}(A) = \mathbf{P}(A)^2$ for all $A \in \bigcap_{n\in\mathbb{T}} \mathcal{F}^X_{]-\infty,n]}$.

ASSUMPTION 2. The observations are nondegenerate, in that $g_n(x, y) > 0$ for all $n \in \mathbb{T}$, $x \in E$ and $y \in F$.

ASSUMPTION 3. The signal transitions, absolute probabilities and observation kernels do not depend on time, in that $M_n = M_0$, $\pi_n = \pi_0$ and $G_n = G_0$ for all $n \in \mathbb{T}$.

Let us briefly comment on these assumptions.

Assumption 1 does not imply that the backward products of M are weakly ergodic: the latter is, by Proposition 1 and Theorem 2, equivalent to the *simultaneous* tail triviality of X under the probability measures over paths in Ω^X derived from *all* members of Π_M, whereas Assumption 1 involves only the particular $\pi \in \Pi_M$ used to construct $\mathbf{P}$.

Assumption 2 is the same type of assumption employed by van Handel (2009) and ensures that information from the observations cannot rule out with certainty any particular hidden state. We shall use several times the fact that when this assumption holds, $\phi_n(x, y_{n:0}) > 0$ for all n, x and $y_n, \dots, y_0$, which is established by a simple induction.

Assumption 3 sacrifices some of the generality of the HMM, but serves to simplify our discussions. Note that when this assumption holds the signal process X is stationary under $\mathbf{P}$ with $\mathbf{P}(X_n \in \cdot) = \pi_0(\cdot)$ for all $n \in \mathbb{T}$.

6.1. *The connection to filter stability.* Throughout Section 6.1, we adopt Assumption 3. Let (X^+, Y^+) be the time-reversal of (X, Y), that is, $X_n^+ = X_{-n}$, $Y_n^+ = Y_{-n}$, $n \in \mathbb{N}$. For some probability distribution $\bar{\pi}_0$, not necessarily an invariant distribution for M_0, but such that $\bar{\pi}_0 \ll \pi_0$, let $\bar{\boldsymbol{P}}$ be the probability measure on $(\Omega, \mathcal{F})$ under which (X^+, Y^+) has the same transition probabilities as under $\mathbf{P}$ but $X_0^+ \sim \bar{\pi}_0$, so, since $X_0^+ = X_0$,

$$\frac{d\bar{\boldsymbol{P}}}{d\mathbf{P}}(X, Y) = \frac{d\bar{\pi}_0}{d\pi_0}(X_0), \qquad \mathbf{P}\text{-a.s.}$$

For $n \in \mathbb{T}$, let ρ_n^Y and $\bar{\rho}_n^Y$ be respectively regular conditional probabilities of the form $\mathbf{P}(X_n \in \cdot \mid \mathcal{F}_{[n,0]}^Y)$ and $\bar{\mathbf{P}}(X_n \in \cdot \mid \mathcal{F}_{[n,0]}^Y)$, so ρ_n^Y (resp., $\bar{\rho}_n^Y$) is a filtering distribution under $\mathbf{P}$ (resp., $\bar{\mathbf{P}}$) for the time-reversed HMM (X^+, Y^+). Among various notions of forgetting associated with HMMs, *asymptotic filter stability* (in mean) is the phenomenon

$$\lim_{n \to -\infty} \bar{\mathbf{E}}[\|\rho_n^Y - \bar{\rho}_n^Y\|] = 0. \tag{6.1}$$

As discussed in Chigansky, Liptser and Van Handel (2011), van Handel (2009) and references therein, it is now well known that ergodicity of the signal as per Assumption 1 is, alone, not enough to establish filter stability (in various senses); see Section 6.2.1 for a counterexample. However, (6.1) does hold if

$$\bigcap_{n \in \mathbb{T}} \mathcal{F}_{]-\infty,0]}^Y \vee \mathcal{F}_{]-\infty,n]}^X = \mathcal{F}_{]-\infty,0]}^Y, \qquad \mathbf{P}\text{-a.s.} \tag{6.2}$$

[see Chigansky, Liptser and Van Handel (2011) for a proof] and using a result of von Weizsäcker (1983), when Assumption 1 holds a necessary and sufficient condition for (6.2) is

$$\bigcap_{n \in \mathbb{T}} \mathcal{F}_{]-\infty,n]}^X \text{ is } \mathbf{P}^{\mathcal{F}^Y}(\omega, \cdot)\text{-a.s. trivial, for } \mathbf{P}\text{-a.e. } \omega. \tag{6.3}$$

Here, $\mathbf{P}^{\mathcal{F}^Y}$ is the object defined in Lemma 5 and appearing in Theorem 3: it is a version of $\mathbf{P}(\cdot \mid \mathcal{F}^Y)$ as a regular conditional distribution over $\mathcal{F}^X$ given $\mathcal{F}^Y$.

Thus, we see that if the coupling for the HMM is a.s. successful, in the sense that condition 1 of Theorem 3 holds, then condition 2 of that Theorem holds, implying (6.3) and, therefore, (6.1). Thus, asymptotic filter stability for the reversed HMM is a necessary condition for a.s. successful coupling. However, as we shall discuss in Section 6.2.2, condition (6.3) is in general weaker than the *simultaneous* tail triviality in condition 2 of Theorem 3.

6.2. *Counterexamples to successful coupling.*

6.2.1. *Degenerate observations.* The purpose of this section is to show how consideration of unicity of absolute probabilities and a.s. success of the coupling bring a fresh perspective on a well-known counterexample to filter stability due to Baxendale, Chigansky and Liptser (2004). Our starting point is to observe that for any $y \in \tilde{H}$ where $\tilde{H}$ is as in Lemma 5, the tower property of conditional expectation and the Markov property of $\tilde{X}$ under P^y give

$$
\begin{aligned}
&P^y\biggl(\tilde{X}_k = x \biggm| \bigcap_{n\in\mathbb{T}} \mathcal{F}^{\tilde{X}}_{]-\infty,n]}\biggr) \\
&\quad= E^y\biggl[P^y\biggl(\tilde{X}_k = x \biggm| \sigma(\tilde{X}_{k-1}) \vee \bigcap_{n\in\mathbb{T}} \mathcal{F}^{\tilde{X}}_{]-\infty,n]}\biggr) \biggm| \bigcap_{n\in\mathbb{T}} \mathcal{F}^{\tilde{X}}_{]-\infty,n]}\biggr] \\
&\quad= E^y\biggl[M^y_k(\tilde{X}_{k-1}, x) \biggm| \bigcap_{n\in\mathbb{T}} \mathcal{F}^{\tilde{X}}_{]-\infty,n]}\biggr] \\
&\quad= \sum_{x'} P^y\biggl(\tilde{X}_{k-1} = x' \biggm| \bigcap_{n\in\mathbb{T}} \mathcal{F}^{\tilde{X}}_{]-\infty,n]}\biggr) M^y_k(x', x), \qquad P^y\text{-a.s.,}
\end{aligned}
\tag{6.4}
$$

where E^y denotes expectation w.r.t. P^y. So, if $\bigcap_{n\in\mathbb{T}} \mathcal{F}^{\tilde{X}}_{]-\infty,n]}$ is *not* P^y-a.s. trivial [cf. (6.3) via Lemmata 4 and 5], regular conditional probabilities of the form $P^y(\tilde{X}_n = x \mid \bigcap_{n\in\mathbb{T}} \mathcal{F}^{\tilde{X}}_{]-\infty,n]})$ give rise to absolute probabilities for M^y distinct from μ^y, hence $\mathrm{card}(\Pi_{M^y}) > 1$. The following is a concrete example of this phenomenon. Throughout the remainder of Section 6.2.1, Assumption 3 is in force.

Let $E = \{0, 1, 2, 3\}$ and

$$M_0(x, x) = 1/2, \qquad M_0(x + 1 \bmod 4, x) = 1/2.$$

M_0 obviously has a unique invariant distribution and Assumption 1 holds. Let $Y_n = \mathbf{1}_{\{X_n \in \{1,3\}\}}$. Then the time-reversed model (X^+, Y^+) coincides up to a relabeling of states with Chigansky, Liptser and Van Handel (2011), Example 1.1, and as explained therein (6.2) does not hold, so neither does (6.3). It follows by Theorem 3 that the coupling is not a.s. successful. We can also verify that for this model $\mathrm{card}(\Pi_{M^y}) > 1$ for any $y \in \Omega^Y$ by direct calculation in connection with (6.4), so that the lack of successful coupling can also be deduced from Proposition 6.

Fix any $y = (y_n)_{n\in\mathbb{T}} \in \Omega^Y$. A simple induction argument provides that for any $n \in \mathbb{T}$ and $x \in E$,

$$\phi_n(x, y_{n:0}) = 2^{n-1},$$

and, therefore, M_n^y is given by

$$M_n^y(x, x') = \begin{cases} 2M_n(x, x'), & y_n = x' \bmod 2, \\ 0, & \text{otherwise.} \end{cases} \tag{6.5}$$

From this, we observe that for each $x \in E$ either $M_n^y(x, x) = 1$ or $M_n^y(x, x - 1 \bmod 4) = 1$, and in matrix form M_n^y is as follows, with the case $y_n = 0$ on the left and the case $y_n = 1$ on the right:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Let $A_n^y := \{x : y_n = x \bmod 2\}$. By (6.5), $x' \notin A_n^y \Rightarrow M_n^y(x, x') = 0$, so we observe that any $\pi^y \in \Pi_{M^y}$, that is, satisfying the equations

$$\pi_n^y(x) = \sum_{x'\in E} \pi_{n-1}^y(x') M_n^y(x', x) \qquad \forall x \in E, n \in \mathbb{T},$$

must be such that for all $n \in \mathbb{T}$,

$$\pi_n^y(x) = \begin{cases} \displaystyle\sum_{x'\in A_{n-1}^y} \pi_{n-1}^y(x') M_n^y(x', x), & x \in A_n^y, \\ 0, & x \notin A_n^y. \end{cases} \tag{6.6}$$

Via some simple manipulations, it then follows that if the values $\{\pi_n^y(x)\}_{x\in A_n^y}$ are fixed, (6.6) provides two equations which can be solved for the two values $\{\pi_{n-1}^y(x)\}_{x\in A_{n-1}^y}$. So, if for any $w \in [0, 1]$ we set $\pi_0^y(x) := w\mathbb{I}[x = y_0] + (1 - w)\mathbb{I}[x = y_0 + 2]$ and then recursively solve (6.6) for $(\pi_n^y)_{n\in\mathbb{T}\setminus\{0\}}$, we obtain by construction a sequence $(\pi_n^y)_{n\in\mathbb{T}} \in \Pi_{M^y}$ uniquely defined by w. Distinct values of w thus giving rise to distinct members of Π_{M^y}, we therefore have $\mathrm{card}(\Pi_{M^y}) > 1$. Moreover, when $w = 0$ or $w = 1$ the measure P_{π^y} defined as in (5.12) fixes all its mass on a single point in Ω^X, and π_n^y is a version of $P^y(\tilde{X}_n \in \cdot \mid \bigcap_{n\in\mathbb{T}} \mathcal{F}_{]-\infty,n]}^{\tilde{X}})$ as in (6.4).

6.2.2. *Reducible signal.* Throughout Section 6.2.2, Assumptions 2 and 3 are in force. The purpose of this example is to illustrate that condition 2 of Theorem 3 is strictly stronger than (6.3), and so the coupling may fail to be a.s. successful even when (6.3) holds. The main point here is that reducibility of M_0 is not ruled out by (6.3), and may compromise weak ergodicity of the backward products of M^y. Indeed, Let $E = \{0, 1, 2, 3\}$ and let M_0 be given by the matrix

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Obviously, $\pi_0 := [1/2 \quad 1/2 \quad 0 \quad 0]$ is an invariant distribution for M_0 and with this choice of π_0 Assumption 1 is satisfied, since under the probability measure $\mathbf{P}$, which is constructed using M_0 and π_0, the $(X_n)_{n\in\mathbb{T}}$ are then i.i.d. according to π_0.

With M_0 as given by the above matrix, $\phi_n(0, y_{n:0}) = \phi_n(1, y_{n:0})$ and $\phi_n(2, y_{n:0}) = \phi_n(3, y_{n:0})$ for all n and $y_n, \dots, y_0$, and M_n^y is given by the matrix

$$\begin{bmatrix} \frac{g_0(0, y_n)}{g_0(0, y_n) + g_0(1, y_n)} & \frac{g_0(1, y_n)}{g_0(0, y_n) + g_0(1, y_n)} & 0 & 0 \\ \frac{g_0(0, y_n)}{g_0(0, y_n) + g_0(1, y_n)} & \frac{g_0(1, y_n)}{g_0(0, y_n) + g_0(1, y_n)} & 0 & 0 \\ 0 & 0 & \frac{g_0(2, y_n)}{g_0(2, y_n) + g_0(3, y_n)} & \frac{g_0(3, y_n)}{g_0(2, y_n) + g_0(3, y_n)} \\ 0 & 0 & \frac{g_0(2, y_n)}{g_0(2, y_n) + g_0(3, y_n)} & \frac{g_0(3, y_n)}{g_0(2, y_n) + g_0(3, y_n)} \end{bmatrix}.$$

The equalities in the statement of Lemma 3 are satisfied if we take

$$\mu_n^y(x) = \frac{\pi_0(x) g_0(x, y_n)}{\sum_{z\in E} \pi_0(z) g_0(z, y_n)} = \frac{g_0(x, y_n)}{g_0(0, y_n) + g_0(1, y_n)} \mathbb{I}[x \in \{0, 1\}],$$

and then $(X_n)_{n\in\mathbb{T}}$ are independent under $\mathbf{P}^{\mathcal{F}^Y}(\omega, \cdot)$ for any $\omega \in \Omega$, so that (6.3) holds by the Kolmogorov 0–1 law. However, condition 2 of Theorem 3 does not hold: to see this first note that for any $y \in \Omega^Y$, $\pi^y = (\pi_n^y)_{n\in\mathbb{T}}$ with $\pi_n^y(x) \propto \mathbb{I}[x \in \{2, 3\}] g_0(x, y_n)$ defines a sequence of absolute probabilities for M^y, distinct from μ^y. Then, with $A := \{\omega : X_n(\omega) \in \{0, 1\} \text{ i.o.}\} \in \bigcap_n \mathcal{F}^X_{]-\infty,n]}$,

$$0 = \mathbf{P}_{\pi^{Y(\omega)}}(A) \neq \mathbf{P}^{\mathcal{F}^Y}(\omega, A) = 1 \qquad \forall \omega \in \Omega.$$

Thus, we conclude that condition 1 of Theorem 3 does not hold, that is, the coupling is not a.s. successful. Of course, this could have been verified more directly using Proposition 6; the backward products of M^y are clearly not weakly ergodic and in fact $\mathbf{Q}^y(\bigcap_{n\in\mathbb{T}}\{T_n = -\infty\}) = 1$.

6.3. *Verifiable conditions for successful coupling.* Our next aim is to present some sufficient conditions for condition 1 of Theorem 3 to hold. We shall make several uses of the following lemma, the proof of which is mostly technical and is given in the Appendix.

LEMMA 6. *Suppose that for some $n < 0$ there exists $k \in \{n+1, \dots, 0\}$, a probability distribution ν and constants $(\varepsilon^-, \varepsilon^+) \in]0, \infty[$ such that*

$$\varepsilon^- \nu(x') \le M_{n,k}(x, x') \le \varepsilon^+ \nu(x') \qquad \forall (x, x') \in E^2. \tag{6.7}$$

Then the following hold:

1. *If $k = n+1$ and $\phi_{n+1}(x, y_{n+1:0}) > 0$ for all x and $y_{n+1}, \dots, y_0$, then*

$$\sup_{y \in \Omega^Y} \beta(M^y_{n,k}) \le 1 - \frac{\varepsilon^-}{\varepsilon^+} < 1. \tag{6.8}$$

2. *If $k > n+1$,*

$$\phi_{k+1}(x, y_{k+1:0}) > 0 \qquad \forall (x, y_{k+1}, \dots, y_0) \in E \times F^k \tag{6.9}$$

and $g_j(x, y) > 0$ for all x, y and $j = n+1, \dots, k$, then

$$\beta(M^y_{n,k}) \le 1 - \frac{\varepsilon^-}{\varepsilon^+} \prod_{j=n+1}^{k} \frac{g_j^-(y_j)}{g_j^+(y_j)} < 1 \qquad \forall y = (y_n)_{n \in \mathbb{T}} \in \Omega^Y, \tag{6.10}$$

where $g_j^-(y) := \min_x g_j(x, y)$, $g_j^+(y) := \max_x g_j(x, y)$.

6.3.1. *Almost surely successful coupling*. Throughout Section 6.3.1 Assumption 3 is in force. In the examples of Sections 6.2.1 and 6.2.2, it is, respectively, the issues of degeneracy of the observations and reducibility of M_0 which caused problems for successful coupling. Our next aim is to illustrate that once these two issues are ruled out, condition 1 of Theorem 3 holds.

Suppose that π_0 is the unique invariant distribution of M_0,

$$\begin{aligned} \pi_0(x) > 0 \qquad \forall x \in E \quad \text{and} \\ \lim_{n \to \infty} M_0^{(n)}(x, x') - \pi_0(x') = 0 \qquad \forall (x, x') \in E^2. \end{aligned} \tag{6.11}$$

It follows that there exists a probability distribution ν, $(\varepsilon^-, \varepsilon^+) \in]0, \infty[$ and $m \ge 1$ such that

$$\varepsilon^- \nu(x') \le M_0^{(m)}(x, x') \le \varepsilon^+ \nu(x') \qquad \forall (x, x') \in E^2. \tag{6.12}$$

We shall now argue that, if we adopt also Assumption 2, then for each $k \in \mathbb{T}$,

$$\lim_{n \to -\infty} \beta(M^Y_{n,k}) = 0, \qquad \mathbf{P}\text{-a.s.}, \tag{6.13}$$

which is, via Propositions 6 and 1, equivalent to condition 1 of Theorem 3.

Using the submultiplicativity of the Dobrushin coefficient and part 2 of Lemma 6, we have for any $y = (y_n)_{n \in \mathbb{T}}$, and $n < k \in \mathbb{T}$,

$$\begin{aligned} \beta(M^y_{n,k}) &\le \prod_{i=0}^{\lfloor (k-n)/m \rfloor - 1} \beta(M^y_{k-(i+1)m, k-im}) \\ &\le \prod_{i=0}^{\lfloor (k-n)/m \rfloor - 1} \bigl(1 - f(y_{k-(i+1)m+1}, \dots, y_{k-im})\bigr), \end{aligned}$$

where $f : F^m \to]0, 1]$ is given by

$$f(y_1, \dots, y_m) := \frac{\varepsilon^-}{\varepsilon^+} \prod_{j=1}^{m} \frac{g_0^-(y_j)}{g_0^+(y_j)}, \qquad (y_1, \dots, y_m) \in F^m.$$

Since for any sequence $(a_i)_{i\in\mathbb{N}}$ with values in $]0, 1]$, $\prod_{i=0}^{\infty}(1 - a_i) = 0 \Leftrightarrow \sum_{i=0}^{\infty} a_i = \infty$, in order to establish (6.13) it suffices to show

$$\sum_{i=0}^{\infty} f(Y_{k-(i+1)m+1}, \dots, Y_{k-im}) = \infty, \qquad \mathbf{P}\text{-a.s.} \tag{6.14}$$

To this end, let $Z_i^{(k)} = (X_{k-(i+1)m+1}, \dots, X_{k-im}, Y_{k-(i+1)m+1}, \dots, Y_{k-im})$. The time reversed bivariate process (X^+, Y^+) is a stationary Markov chain under $\mathbf{P}$, and it follows from (6.11) and the conditional independence structure of the HMM that the transition kernel of (X^+, Y^+) has a unique invariant distribution $\pi_0(dx)g_0(x, y)\psi(dy)$, and is uniformly ergodic, in the sense of Meyn and Tweedie (2009), Chapter 16. Some simple but tedious calculations show that under $\mathbf{P}$, $(Z_i^{(k)})_{i\in\mathbb{N}}$ is then also a stationary Markov chain, with transition kernel which admits a unique invariant distribution and which is uniformly ergodic. So by the strong law of large numbers for stationary and ergodic Markov chains,

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(Y_{k-(i+1)m+1}, \dots, Y_{k-im}) \\ &\quad = \mathbb{E}\big[f(Y_{-m+1}, \dots, Y_0)\big], \qquad \mathbf{P}\text{-a.s.} \end{aligned} \tag{6.15}$$

Since f is strictly positive the expectation in (6.15) is strictly positive, hence (6.14) holds, hence (6.13) holds.

6.3.2. *Surely successful coupling*. In practice, one is typically presented with an observation sequence which is not necessarily distributed according to $\mathbf{P}$. It may then be of some concern that even if one (and then both) of the conditions of Theorem 3 holds, the coupling may fail to be successful for y in a set of observation sequences which has zero probability under $\mathbf{P}$. In this section, we discuss some simple sufficient conditions for the stronger requirement that

$$\mathbf{Q}^y\Big(\bigcap_{n\in\mathbb{T}} \{T_n > -\infty\}\Big) = 1 \qquad \forall y \in \Omega^Y. \tag{6.16}$$

Suppose that

$$\inf_{n\in\mathbb{T}} \min_{(x,x')\in E^2} M_n(x, x') > 0 \tag{6.17}$$

and assume that for all $y = (y_n)_{n\in\mathbb{T}} \in \Omega^Y$

$$\forall n \in \mathbb{T}, \exists x : \qquad g_n(x, y_n) > 0. \tag{6.18}$$

An application of part 1 of Lemma 6 gives

$$\sup_{y\in\Omega^Y}\sup_{n\in\mathbb{N}}\beta(M_n^y)<1,$$

so for any $k\in\mathbb{T}$ and $y\in\Omega^Y$,

$$\lim_{n\to-\infty}\beta(M_{n,k}^y)\le\lim_{n\to-\infty}\prod_n^k\beta(M_j^y)=0,$$

which via Proposition 6 and Proposition 1 gives (6.16).

The reader can easily verify that part 2 of Lemma 6 can be used to show that (6.16) holds under conditions weaker than (6.17) and perhaps at the expense of strengthening (6.18).

6.4. *The case of finitely many observations*. In practice, typically only a finite number of observations are available, say $y_0,\dots,y_m$ for some $m\in\mathbb{T}$, and one aims to sample from conditional distributions of the form $\mathbf{P}(X_n\in\cdot\mid\mathcal{F}^Y_{[m,0]})$, for some $n\ge m$. There are a number of ways the coupling method can be applied in this situation.

As an example, fix $m\in\mathbb{T}$ and suppose for simplicity of exposition that M_n does not depend on n, and has unique invariant distribution π. Let the probability space of Section 5.3 be augmented so as to also support an E-valued random variable Z_m such that with y fixed, $\mathbf{Q}^y$ makes Z_m independent of $(\xi_n^x;x\in E;n\in\mathbb{T})$, and

$$\mathbf{Q}^y(Z_m=x)=\frac{\pi(x)g_m(x,y_m)\phi_{m+1}(x,y_{m+1:0})}{\sum_{x'\in E}\pi(x')g_m(x',y_m)\phi_{m+1}(x',y_{m+1:0})}=:\bar{\pi}_m^y(x).$$

Also introduce a random variable Z_0 such that on the event $\{T_0\ge m\}$, $Z_0:=\Phi_{T_0,0}(x)$ for an arbitrary $x\in E$; and on the event $\{T_0<m\}$, $Z_0:=\Phi_{m,0}(Z_m)$. Then using the fact that on the event $\{T_0\ge m\}$, $\Phi_{T_0,0}(x)=\Phi_{m,0}(x')$ for all x', we have

$$\begin{aligned}
&\mathbf{Q}^y(Z_0=z)\\
&\quad=\mathbf{Q}^y(\{Z_0=z\}\cap\{T_0\ge m\})+\mathbf{Q}^y(\{Z_0=z\}\cap\{T_0<m\})\\
&\quad=\mathbf{Q}^y(\{\Phi_{T_0,0}(x)=z\}\cap\{T_0\ge m\})+\mathbf{Q}^y(\{\Phi_{m,0}(Z_m)=z\}\cap\{T_0<m\})\\
&\quad=\mathbf{Q}^y(\{\Phi_{m,0}(Z_m)=z\}\cap\{T_0\ge m\})+\mathbf{Q}^y(\{\Phi_{m,0}(Z_m)=z\}\cap\{T_0<m\})\\
&\quad=\mathbf{Q}^y(\Phi_{m,0}(Z_m)=z)=\sum_{x'\in E}\bar{\pi}_m^y(x')M_{m,0}^y(x',z),
\end{aligned}$$

and it is easily checked that $\sum_{x'\in E}\bar{\pi}_m^Y(x')M_{m,0}^Y(x',z)=\mathbf{P}(X_0=z\mid\mathcal{F}^Y_{[m,0]})$, $\mathbf{P}$-a.s.

Some modifications of Algorithm 1 facilitate the sampling of Z_0. The "while" line is replaced by

while card(image of $\Phi_{n,0}$) >1 and $n>m$

and the "return" line is replaced by

$$
\begin{aligned}
&\text{if } \operatorname{card}(\text{image of } \Phi_{n,0}) = 1 \\
&\quad \text{return } Z_0 = \Phi_{n,0}(x), \ \text{for any } x \in E, \\
&\text{else} \\
&\quad \text{sample } Z_m \text{ from the distribution on } E \text{ with} \\
&\quad \operatorname{prob}(Z_m = x) \propto \pi(x) g_m(x, y_m) \phi_{m+1}(x, y_{m+1:0}) \\
&\quad \text{and return } Z_0 = \Phi_{m,0}(Z_m).
\end{aligned}
$$

The resulting procedure may be computationally cheaper than direct calculation and sampling from $\sum_{x \in E} \bar{\pi}_m^y(x) M_{m,0}^y(x, \cdot)$ if $T_0 \gg m$.

6.5. *Numerical examples.*

6.5.1. *Sensitivity to model misspecification.* The purpose of this example is to numerically investigate an HMM for which the coupling is almost surely successful, that is, $\lim_{n \to -\infty} \beta(M_{n,k}^Y) = 0$, $\mathbf{P}$-a.s., but for which $\beta(M_{n,k}^Y)$ converges to zero very slowly or perhaps even remains bounded away from zero as $n \to -\infty$ when the HMM is misspecified, in the sense that Y is not distributed according to $\mathbf{P}$.

Consider $E = \{1, 2, 3\}$, $F = \mathbb{R}$, ψ Lebesgue measure, and a time-homogeneous HMM, that is, Assumption 3 holds, with for some $\delta \in (0, 1)$, M_0 written in matrix form:

$$
M_0 = \begin{bmatrix} 1-\delta & \delta & 0 \\ \delta/2 & 1-\delta & \delta/2 \\ 0 & \delta & 1-\delta \end{bmatrix},
$$

$$
g_0(1, y) = g_0(3, y) = \frac{e^{-y^2/2}}{\sqrt{2\pi}}, \qquad g_0(2, y) = \frac{e^{-(y-1)^2/2}}{\sqrt{2\pi}}.
$$

Clearly, Assumption 2 holds and (6.12) holds with $m = 2$. Hence, by the arguments of Section 6.3.1, for all $k \in \mathbb{T}$, $\lim_{n \to -\infty} \beta(M_{n,k}^Y) = 0$, $\mathbf{P}$-a.s.

Figure 1 illustrates $\beta(M_{n,0}^y)$ and histograms of the coupling time T_0 obtained from 10^4 independent runs of Algorithm 1, for three different data sequences y. The first, corresponding to the left column of plots, was drawn from $\mathbf{P}$, with the true sequence of hidden states also shown. The second, corresponding to the middle column, is a sample path of the process: $Y_0 = 0$ and $Y_n = Y_{n+1} + V_n$, where the V_n are i.i.d. $\mathcal{N}(0, 0.25)$. The third, corresponding to right column, is a realization of $Y_n = 0.003n + V_n$. To interpret these plots, note that for $x \in \{1, 3\}$, $\lim_{y_0 \to -\infty} g(2, y_0)/g(x, y_0) = 0$, and similarly, if it were true that the observation sequence were constant $y_n = y_{n+1} = \cdots = y_0$, elementary manipulations show that for $x \in \{1, 3\}$, $\lim_{y_0 \to -\infty} M_{n,0}^y(x, x) = 1$, hence $\lim_{y_0 \to -\infty} \beta(M_{n,0}^y) = 1$. The

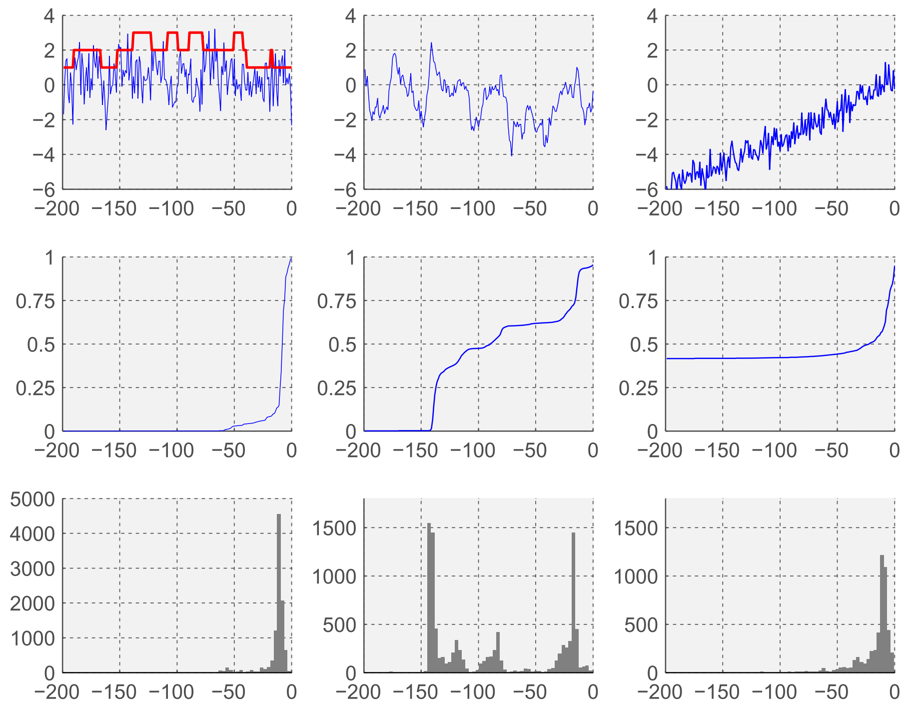

FIG. 1. *Top row*: y_n *vs.* n (*top left plot also shows true sequence of hidden states*). *Middle row*: $\beta(M^y_{n,0})$ *vs.* n. *Bottom row*: *histograms of* T_0 *obtained from* 10^4 *runs of the algorithm. First column corresponds to data simulated from the HMM*, *second and third columns correspond to data from misspecified models. In the first and second cases*, *all of the* 10^4 *realizations of* T_0 *were valued within* $\{-200, \dots, 0\}$, *for the third case*, *only* 50.3% *of the* 10^4 *realizations of* T_0 *were valued in* $\{-200, \dots, 0\}$, *the remaining realizations are not shown on the bottom-right histogram.*

plots in the second and third columns reflect a similar phenomenon, namely that long sequences of negative observations may slow down the convergence to zero of $\beta(M^y_{n,0})$ as $n \to -\infty$. In the case of the third column, it is notable that $\beta(M^y_{n,0})$ appears to be bounded away from zero, indeed the same phenomenon was observed in much longer runs of the algorithm (numerical results not shown).

6.5.2. *Simulating multiple samples*. Running Algorithm 1 several times in order to obtain multiple i.i.d. samples from π^y_0 may be prohibitively expensive. Consider the following procedure:

1. Fix n, compute $M^y_{n,0}$ and obtain an exact sample $X^\star_n$ from π^y_n using the perfect sampling scheme.
2. Given $X^\star_n$, drawn N conditionally independent samples,

$$X_0^{(i)} \sim M^y_{n,0}(X^\star_n, \cdot), \qquad i = 1, \dots, N.$$

TABLE 1

In the second row, $\lambda_n^y := \mathrm{Law}(X_0^{(1)}, X_0^{(2)})$, where $X_0^{(1)}, X_0^{(2)}$ are obtained from step 2 of the procedure, that is, $X_0^{(i)} \sim M_n^y(X_n^\star, \cdot)$. The third to fifth rows show mean percentage of overall CPU time spent on step 1, with estimated standard deviation in parentheses, obtained from 10^4 runs

		$n=5$	$n=10$	$n=25$	$n=50$	$n=100$
$\lVert\lambda_n^y - \pi_0^y \otimes \pi_0^y\rVert$		0.27	0.014	0.0033	3.3×10^{-4}	$<10^{-6}$
% time	$N=10^2$	23 (25)	26 (29)	41 (27)	43 (10)	47 (22)
step 1	$N=10^3$	4.2 (4.5)	4.8 (5.6)	9.2 (5.1)	9.8 (1.6)	11 (3.8)
	$N=10^4$	0.53 (0.57)	0.61 (0.72)	1.2 (0.66)	1.3 (0.20)	1.5 (0.49)

Since $X_n^\star \sim \pi_n^y$ and $\sum_x \pi_n^y(x) M_{n,0}^y(x, \cdot) = \pi_0^y(\cdot)$, the samples $(X_0^{(i)}; i = 1, \dots, N)$ each have marginal distribution π_0^y, but are not independent in general. Indeed writing λ_n^y for the joint distribution of $(X_0^{(1)}, X_0^{(2)})$,

$$\begin{aligned}
&\lVert\lambda_n^y - \pi_0^y \otimes \pi_0^y\rVert \\
&\quad= \frac{1}{2} \sum_{x,x'} \left| \sum_z \pi_n^y(z) M_{n,0}^y(z, x) M_{n,0}^y(z, x') - \pi_0^y(x) \pi_0^y(x') \right|.
\end{aligned}$$

Conditional ergodicity dictates this quantity converges to zero as $n \to -\infty$. Table 1 shows numerical values against n, for the HMM of Section 6.5.1 with the data sequence shown in the top-left plot of Figure 1 and with π_n^y and π_0^y approximated by $M_{-1000,n}^y(1, \cdot)$ and $M_{-1000,0}^y(1, \cdot)$. Also shown is the average percentage of CPU time spent on step 1 of the above two step procedure, obtained from 10^4 independent runs. The algorithms were implemented in Matlab on a 2.80 GHz desktop PC. These results illustrate that step 1, which amounts to guaranteeing that the marginal distribution of each $X_0^{(i)}$ is exactly π_0^y, is relatively cheap when N is large, even when n is large enough to make the pairwise dependence between $X_0^{(1)}$ and $X_0^{(2)}$ negligible.

6.6. *Outlook.* How much of all this can be generalized beyond the case in which E is a finite set? We believe: quite a lot, although the work involved is nontrivial. Of course if E is not a finite set, then we have to let go of Fact 1; without further assumption there is no guarantee that even a single sequence of absolute probabilities for M exists. The coupling we have specified in Section 3 relies heavily on the fact that E contains only finitely many points, but a generalization via the kind of mechanisms used for backward coupling of homogeneous chains [see, e.g., Foss and Tweedie (1998) and references therein] may be feasible, and that would be the starting point from which to investigate generalization of Theorem 1. Quite a few of the arguments used in the proof of Theorem 2 do not really rely on

E being a finite set. Regarding the application to HMMs, as soon as E contains infinitely many points, then with a few exceptions such as the linear-Gaussian state-space model, the functions $\phi_n(x, y_{n:0})$ are not available in closed form, so sampling from the kernels M_n^y becomes nontrivial and the perfect simulation algorithm may lose its practical relevance. Overall though, there are several possible avenues for further investigation.

APPENDIX

PROOF OF LEMMA 2. First, we claim that for any nonnegative measurable function f, the following holds $\mathbf{P}$-a.s.:

$$\begin{aligned}&\mathbf{E}\big[f(X_{n-1}, Y_n, \dots, Y_0) \mid \sigma(X_{n-1})\big]\\ &\quad= \int f(X_{n-1}, y_n, \dots, y_0)\phi_n(X_{n-1}, y_{n:0})\psi^{\otimes(|n|+1)}\big(d(y_n, \dots, y_0)\big).\end{aligned} \tag{A.1}$$

The right-hand side of (A.1) is clearly measurable w.r.t. $\sigma(X_{n-1})$, so to prove the claim it remains to check that for any $x_{n-1} \in E$,

$$\begin{aligned}&\int_{A(x_{n-1})}\int f(X_{n-1}, y_n, \dots, y_0)\phi_n(X_{n-1}, y_{n:0})\psi^{\otimes(|n|+1)}\big(d(y_n, \dots, y_0)\big)\,d\mathbf{P}\\ &\quad= \int_{A(x_{n-1})} f(X_{n-1}, Y_n, \dots, Y_0)\,d\mathbf{P},\end{aligned} \tag{A.2}$$

where $A(x_{n-1})$ is the event $\{X_{n-1} = x_{n-1}\}$. It follows from (5.1) and by writing out the definition of ϕ_n in (5.4) that the left-hand side of (A.2) is equal to

$$\begin{aligned}&\pi_{n-1}(x_{n-1})\int f(x_{n-1}, y_n, \dots, y_0)\phi_n(x_{n-1}, y_{n:0})\psi^{\otimes(|n|+1)}\big(d(y_n, \dots, y_0)\big)\\ &\quad= \int \pi_{n-1}(x_{n-1}) f(x_{n-1}, y_n, \dots, y_0) \sum_{(x_n,\dots,x_0)} \prod_{k=n}^{0} M_k(x_{k-1}, x_k)G_k(x_k, dy_k),\end{aligned}$$

which is also equal to the right-hand side (A.2), thus completing the proof of (A.1).

Next, note that $M_n^Y(X_{n-1}, x)$ is measurable w.r.t. $\mathcal{F}^Y_{[n,0]} \vee \sigma(X_{n-1})$, so in order to complete the proof of the lemma it remains, by a standard monotone class argument, to show

$$\begin{aligned}&\int_{\{X_{n-1}=x_{n-1}\}} \mathbb{I}\big[(Y_n, \dots, Y_0) \in A\big] M_n^Y(X_{n-1}, x)\,d\mathbf{P}\\ &\quad= \mathbf{P}\big(\{X_{n-1} = x_{n-1}\} \cap \big\{(Y_n, \dots, Y_0) \in A\big\} \cap \{X_n = x\}\big),\end{aligned} \tag{A.3}$$

for any $x_{n-1}, x \in E$ and $A \in \mathcal{B}(F)^{\otimes(|n|+1)}$. We proceed by fixing x and applying (A.1) with $f(x_{n-1}, y_n, \dots, y_0) = \mathbb{I}[(y_n, \dots, y_0) \in A]M_n^y(x_{n-1}, x)$, we have using the definitions of $M_n^y(x_{n-1}, x)$, ϕ_{n+1} and $\mathbf{P}$ that the following equalities hold $\mathbf{P}$-a.s.:

$$\begin{aligned}
&\mathbf{E}\bigl[\mathbb{I}\bigl[(Y_n, \dots, Y_0) \in A\bigr] M_n^Y(X_{n-1}, x) \mid \sigma(X_{n-1})\bigr] \\
&\quad = \int_{\{(y_n,\dots,y_0)\in A\}} M_n(X_{n-1}, x) g_n(x, y_n)\phi_{n+1} \\
&\qquad \times (x, y_{n+1:0})\psi^{\otimes(|n|+1)}\bigl(d(y_n, \dots, y_0)\bigr) \\
&\quad = \mathbf{P}\bigl(\bigl\{(Y_n, \dots, Y_0) \in A\bigr\} \cap \{X_n = x\} \mid \sigma(X_{n-1})\bigr),
\end{aligned}$$

using this identity and the tower property of conditional expectation, we can rewrite the left-hand side of (A.3) as

$$\begin{aligned}
&\int_{\{X_{n-1}=x_{n-1}\}} \mathbb{I}\bigl[(Y_n, \dots Y_0) \in A\bigr] M_n^Y(X_{n-1}, x)\, d\mathbf{P} \\
&\quad = \int_{\{X_{n-1}=x_{n-1}\}} \mathbf{E}\bigl[\mathbb{I}\bigl[(Y_n, \dots, Y_0) \in A\bigr] M_n^Y(X_{n-1}, x) \mid \sigma(X_{n-1})\bigr]\, d\mathbf{P} \\
&\quad = \int_{\{X_{n-1}=x_{n-1}\}} \mathbf{P}\bigl(\bigl\{(Y_n, \dots, Y_0) \in A\bigr\} \cap \{X_n = x\} \mid \sigma(X_{n-1})\bigr)\, d\mathbf{P} \\
&\quad = \mathbf{P}\bigl(\{X_{n-1} = x_{n-1}\} \cap \bigl\{(Y_n, \dots, Y_0) \in A\bigr\} \cap \{X_n = x\}\bigr).
\end{aligned}$$

Equality (A.3) therefore holds and this completes the proof of the lemma. □

REMARK 4. We note that the arguments of the above proof rely on the definition in (5.5) only through the values taken by $M_n^y(x, \cdot)$ on the support of ϕ_n.

PROOF OF LEMMA 4. To prove $\mathcal{F}_I^X = \{A \times \Omega^Y; A \in \mathcal{F}_I^{\tilde{X}}\}$, we need to show that $\mathcal{C}_I^X := \{A \times \Omega^Y; A \in \mathcal{F}_I^{\tilde{X}}\}$ is the smallest σ-algebra of subsets of Ω w.r.t. which all the $(X_n)_{n\in I}$ are measurable. We break this down into three steps: (i) show that $\mathcal{C}_I^X$ is a σ-algebra; (ii) show that every $(X_n)_{n\in I}$ is measurable w.r.t. $\mathcal{C}_I^X$; (iii) show that if any set is removed from $\mathcal{C}_I^X$ then the resulting collection of sets either does not contain $X_n^{-1}(A)$ for some $n \in I$ and $A \in \mathcal{B}(E)$, or is not a σ-algebra.

Step (i) is immediate since $\mathcal{F}_I^{\tilde{X}}$ is by definition a σ-algebra and Ω^Y is nonempty (because F is by definition nonempty). For step (ii), we have by definition of $\mathcal{F}_I^{\tilde{X}}$ that for any $n \in I$ and $A \in \mathcal{B}(E)$, $\tilde{X}_n^{-1}(A) \in \mathcal{F}_I^{\tilde{X}}$ and $X_n^{-1}(A) = \eta^{-1} \circ \tilde{X}_n^{-1}(A) = \tilde{X}_n^{-1}(A) \times \Omega^Y$, hence $X_n^{-1}(A) \in \mathcal{C}_I^X$. For step (iii), for an arbitrary $B \in \mathcal{F}_I^{\tilde{X}}$ let us remove the set $B \times \Omega^Y$ from $\mathcal{C}_I^X$, the resulting collection of sets being $\{A \times \Omega^Y; A \in \mathcal{F}_I^{\tilde{X}} \setminus B\} =: \mathcal{D}_I^X$. Since $\mathcal{F}_I^{\tilde{X}}$ is the smallest σ-algebra w.r.t. which all the $(\tilde{X}_n)_{n\in I}$ are measurable, either there exists some $n \in I$ and $A \in \mathcal{B}(E)$ such that $\tilde{X}_n^{-1}(A) \notin \mathcal{F}_I^{\tilde{X}} \setminus B$, or $\mathcal{F}_I^{\tilde{X}} \setminus B$ is not a σ-algebra. In the former case, $X_n^{-1}(A) = \eta^{-1} \circ \tilde{X}_n^{-1}(A) = \tilde{X}_n^{-1}(A) \times \Omega^Y \notin \mathcal{D}_I^X$, that is, X_n is not measurable w.r.t. $\mathcal{D}_I^X$.

In the latter case, we claim that $\mathcal{D}_I^X$ is not a σ-algebra. To prove this claim, we shall argue to the contrapositive that for $\tilde{\mathcal{C}}$ any collection of subsets of $E^{\mathbb{T}}$, if $\mathcal{D} := \{A \times \Omega^Y; A \in \tilde{\mathcal{C}}\}$ is a σ-algebra, then $\tilde{\mathcal{C}}$ is a σ-algebra. To this end, observe: if $\mathcal{D}$ contains Ω, then $\tilde{\mathcal{C}}$ contains Ω^X; if $\mathcal{D}$ is closed under complements, then $A \in \tilde{\mathcal{C}} \Rightarrow A \times \Omega^Y \in \mathcal{D} \Rightarrow (A \times \Omega^Y)^c \in \mathcal{D} \Rightarrow A^c \times \Omega^Y \in \mathcal{D} \Rightarrow A^c \in \tilde{\mathcal{C}}$, that is, $\tilde{\mathcal{C}}$ is closed under complements; if $\mathcal{D}$ is closed under countable unions, $A_n \in \tilde{\mathcal{C}} \Rightarrow A_n \times \Omega^Y \in \mathcal{D} \Rightarrow \bigcup_{n\in\mathbb{N}}(A_n \times \Omega^Y) \in \mathcal{D} \Rightarrow (\bigcup_{n\in\mathbb{N}} A_n) \times \Omega^Y \in \mathcal{D} \Rightarrow \bigcup_n A_n \in \tilde{\mathcal{C}}$, that is, $\tilde{\mathcal{C}}$ is closed under countable unions. This completes the proof of $\mathcal{F}_I^X = \{A \times \Omega^Y; A \in \mathcal{F}_I^{\tilde{X}}\}$, from which (5.10) follows directly. $\square$

PROOF OF LEMMA 5. As a consequence of Lemma 3, there exists $H \in \mathcal{F}^Y$ with $\mathbf{P}(H) = 1$ such that for all $\omega \in H$, $\sum_{z\in E} \mu_{n-1}^{Y(\omega)}(z) M_n^{Y(\omega)}(z, x) = \mu_n^{Y(\omega)}(x)$ for all n and x. Set $\tilde{H} = Y(H)$. For $y \in \tilde{H}$, we are assured by the usual extension argument of the existence of $P^y(\cdot)$ a measure with the desired properties. For $y \notin \tilde{H}$ set $P^y(\cdot)$ to an arbitrary probability. We thus obtain the desired kernel. It follows from Lemma 4 that every set in $\mathcal{F}^X$ is of the form $A \times \Omega^Y$ for some $A \in \mathcal{F}^{\tilde{X}}$, and then $\mathbf{P}^{\mathcal{F}^Y}$ is a probability kernel because P is. In order to establish (5.11), we argue as follows. With $\mathbf{P} \circ Y^{-1}$ the push-forward of $\mathbf{P}$ by Y, define $\tilde{\mathbf{P}}(A) := \int_{\Omega^X \times \Omega^Y} \mathbb{I}[(x, y) \in A] P^y(dx)(\mathbf{P} \circ Y)(dy)$, which is a probability measure on $(\Omega, \mathcal{F})$ and by construction $\tilde{\mathbf{P}}(A \times \Omega^Y \mid \mathcal{F}^Y)(\omega) = \mathbf{P}^{\mathcal{F}^Y}(\omega, A \times F)$, $\tilde{\mathbf{P}}$-a.s., for each $A \in \mathcal{F}^{\tilde{X}}$. The proof of (5.11) will be complete if we can show that $\tilde{\mathbf{P}} = \mathbf{P}$, since then $\tilde{\mathbf{P}}(\cdot \mid \mathcal{F}^Y) = \mathbf{P}(\cdot \mid \mathcal{F}^Y)$. For $\tilde{\mathbf{P}} = \mathbf{P}$, it is sufficient that for each $n \in \mathbb{T}$ and $A \in \mathcal{F}_n$, $\tilde{\mathbf{P}}(A) = \mathbf{P}(A)$, and the latter holds since, using (5.2), (5.3), Lemmata 2 and 3,

$$\begin{aligned}
\tilde{\mathbf{P}}(A) &= \int \sum_{(x_n,\dots,x_0)\in E^{n+1}} \mathbb{I}[(x, y) \in A] \mu_n^y(x_n) \prod_{k=n+1}^{0} M_k^y(x_{k-1}, x_k)(\mathbf{P} \circ Y)(dy) \\
&= \mathbf{E}[\mathbf{E}[\mathbf{E}[\cdots \mathbf{E}[\mathbb{I}[(X, Y) \in A] \mid \mathcal{F}^Y \vee \mathcal{F}^X_{]-\infty,0]}] \cdots \mid \mathcal{F}^Y \vee \mathcal{F}^X_{]-\infty,n]}] \mid \mathcal{F}^Y]] \\
&= \mathbf{P}(A).
\end{aligned}$$

The proof of the lemma is complete. $\square$

PROOF OF LEMMA 6. Throughout the proof, fix $y \in \Omega^Y$. For 1, first note that by (5.4) and (6.7),

$$\varepsilon^- \nu(g\phi)_{n+2}^y \le \phi_{n+1}(x, y_{n+1:0}) \le \varepsilon^+ \nu(g\phi)_{n+2}^y, \tag{A.4}$$

where

$$\nu(g\phi)_{n+2}^y := \sum_z \nu(z) g_{n+1}(z, y_{n+1}) \phi_{n+2}(z, y_{n+2:0}) > 0,$$

the positivity being due to the hypotheses of 1 combined with (A.4) and $\varepsilon^+ > 0$. It follows from (5.6), (5.5), the hypothesis of 1, (6.7) and (A.4) that

$$\begin{aligned} M^y_{n,k}(x,x') &= M^y_{n+1}(x,x') \\ &= \frac{M_{n+1}(x,x')g_{n+1}(x',y_{n+1})\phi_{n+2}(x',y_{n+2:0})}{\phi_{n+1}(x,y_{n+1:0})} \\ &\geq \frac{\varepsilon^-}{\varepsilon_+}\frac{\nu(x')g_{n+1}(x',y_{n+1})\phi_{n+2}(x',y_{n+2:0})}{\nu(g\phi)^y_{n+2}}, \end{aligned}$$

thus there exists a probability distribution $\tilde{\nu}^y_{n+1}$ such that $M^y_{n+1}(x,x') \geq \frac{\varepsilon^-}{\varepsilon_+}\tilde{\nu}^y_{n+1}(x')$. Combining this fact with the expression for $\beta(\cdot)$ in (2.1) gives (6.8).

A simple induction shows that when (6.9) and the hypotheses of 2 hold, $\phi_j(x, y_{j:0}) > 0$ for all x and $j = k+1, k, \dots, n+1$. Combining this fact with (5.6), (5.5), (5.4) and (6.7),

$$\begin{aligned} &M^y_{n,k}(x_n,x_k) \\ &\quad= \frac{\sum_{(x_{n+1},\dots,x_{k-1})}(\prod_{j=n+1}^k M_j(x_{j-1},x_j)g_j(x_j,y_j))\phi_{k+1}(x_k,y_{k+1:0})}{\phi_{n+1}(x_n,y_{n+1:0})} \\ &\quad\geq \frac{M_{n,k}(x_n,x_k)\phi_{k+1}(x_k,y_{k+1:0})}{\sum_z M_{n,k}(x_n,z)\phi_{k+1}(z,y_{k+1:0})}\prod_{j=n+1}^k \frac{g_j^-(y_j)}{g_j^+(y_j)} \\ &\quad\geq \frac{\varepsilon^-}{\varepsilon^+}\frac{\nu(x_k)\phi_{k+1}(x_k,y_{k+1:0})}{\sum_z \nu(z)\phi_{k+1}(z,y_{k+1:0})}\prod_{j=n+1}^k \frac{g_j^-(y_j)}{g_j^+(y_j)}, \end{aligned}$$

and the final denominator is strictly positive due to (6.9). Using again (2.1) gives (6.10). □

REFERENCES

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

CHIGANSKY, P., LIPTSER, R. and VAN HANDEL, R. (2011). Intrinsic methods in filter stability. In *The Oxford Handbook of Nonlinear Filtering* 319–351. Oxford Univ. Press, Oxford. MR2884601

DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York. MR0058896

FOSS, S. G. and TWEEDIE, R. L. (1998). Perfect simulation and backward coupling. *Comm. Statist. Stochastic Models* **14** 187–203. MR1617572

GLYNN, P. W. and THORISSON, H. (2001). Two-sided taboo limits for Markov processes and associated perfect simulation. *Stochastic Process. Appl.* **91** 1–20. MR1807359

HÄGGSTRÖM, O. (2002). *Finite Markov Chains and Algorithmic Applications*. Cambridge Univ. Press, Cambridge. MR1908939

LINDVALL, T. (2002). *Lectures on the Coupling Method*. Dover Publications, Inc., Mineola, NY. MR1924231

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

MURDOCH, D. J. and GREEN, P. J. (1998). Exact sampling from a continuous state space. *Scand*. *J*. *Stat*. **25** 483–502. MR1650023

PROPP, J. G. and WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9** 223–252.

SENETA, E. (2006). *Non-negative Matrices and Markov Chains*. Springer, New York. MR2209438

STENFLO, Ö. (2008). Perfect sampling from the limit of deterministic products of stochastic matrices. *Electron*. *Commun*. *Probab*. **13** 474–481. MR2438818

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

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

SCHOOL OF MATHEMATICS
UNIVERSITY OF BRISTOL
UNIVERSITY WALK
BRISTOL BS8 1TW
UNITED KINGDOM
E-MAIL: nick.whiteley@bristol.ac.uk

DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL
UNITED KINGDOM
E-MAIL: anthony.lee@warwick.ac.uk