Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 130 (2020) 1388–1425

stochastic processes and their applications

www.elsevier.com/locate/spa

Existence of infinite Viterbi path for pairwise Markov models

Jüri Lember*, Joonas Sova

University of Tartu, Liivi 2 50409, Tartu, Estonia

Received 9 July 2018; received in revised form 16 April 2019; accepted 8 May 2019
Available online 15 May 2019

Abstract

For hidden Markov models one of the most popular estimates of the hidden chain is the Viterbi path — the path maximizing the posterior probability. We consider a more general setting, called the pairwise Markov model, where the joint process consisting of finite-state hidden regime and observation process is assumed to be a Markov chain. We prove that under some conditions it is possible to extend the Viterbi path to infinity for almost every observation sequence which in turn enables to define an infinite Viterbi decoding of the observation process, called the Viterbi process. This is done by constructing a block of observations, called a barrier, which ensures that the Viterbi path goes through a given state whenever this block occurs in the observation sequence.

1. Introduction and preliminaries

1.1. Introduction

We consider a Markov chain $Z = \{Z_k\}_{k\geq 1}$ with product state space $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{Y}$ is a finite set (state space) and $\mathcal{X}$ is an arbitrary separable metric space (observation space). Thus, the process Z decomposes as $Z = (X, Y)$, where $X = \{X_k\}_{k\geq 1}$ and $Y = \{Y_k\}_{k\geq 1}$ are random processes taking values in $\mathcal{X}$ and $\mathcal{Y}$, respectively. The process X is identified as an observation process and the process Y, sometimes called the *regime* or the *signal*, models the observations-driving hidden state sequence. Therefore our general model contains many well-known stochastic models as a special case: hidden Markov models (HMM), Markov switching

* Corresponding author.
E-mail addresses: jyril@ut.ee (J. Lember), joonas.sova@ut.ee (J. Sova).

https://doi.org/10.1016/j.spa.2019.05.004

models, hidden Markov models with dependent noise and many more. The *segmentation* or *path estimation* problem consists of estimating the realization of $(Y_1, \ldots, Y_n)$ given a realization $x_{1:n}$ of $(X_1, \ldots, X_n)$. We adopt the standard notation $a_{l:n}$ for a vector $(a_l, \ldots, a_n)$ with $l \leq n \leq \infty$. A standard estimate is any path $v_{1:n} \in \mathcal{Y}^n$ having maximum posterior probability:

$$v_{1:n} = \arg\max_{y_{1:n}} P(Y_{1:n} = y_{1:n} | X_{1:n} = x_{1:n}).$$

Any such path is called *Viterbi path* and we are interested in the behavior of $v_{1:n}$ as n grows. The study of asymptotics of Viterbi path is complicated by the fact that adding one more observation, x_{n+1} can change the whole path, and so it is not clear, whether there exists a limiting infinite Viterbi path. In fact, as we show in Example 1, for some models the Viterbi path keeps changing a.s. and so there is no infinite path. The goal of the present paper is to establish the conditions that ensure the existence of infinite Viterbi path, a.s. When this happens, one can define infinite Viterbi decoding of X-process called *Viterbi process*. In this paper, we construct the infinite Viterbi path using the barriers. A *barrier* is a fixed-sized block in the observations $x_{1:n}$ that fixes the Viterbi path up to itself: for every continuation of $x_{1:n}$, the Viterbi path up to the barrier remains unchanged. Therefore, if almost every realization $x_{1:\infty}$ of X-process contains infinitely many barriers, then the infinite Viterbi path exists a.s. The main task of the paper is to exhibit the conditions (in terms of the model) that guarantee the existence of infinite many barriers, a.s. Having infinitely many barriers is not necessary for existence of infinite Viterbi path (see Example 2), but the barrier-construction has several advantages. One of them is that it allows to construct the infinite path *piecewise*, meaning that to determine the first k elements $v_{1:k}$ of the infinite path it suffices to observe $x_{1:n}$ for n big enough. Another great advantage of the barriers is that under piecewise construction the Viterbi process is typically a regenerative process. The regenerativity allows to easily prove limit theorems to understand the asymptotic behavior of inferences based on Viterbi paths.

Our main construction theorems (Theorems 2.1 and 3.1) generalize the piecewise construction in [19,20], where the existence of Viterbi process was proven for HMM's. The important special case of HMM is analyzed in Section 4.1, but let us stress that generalization beyond the HMM is far from being straightforward. Moreover, we see that some assumptions of previous HMM-theorem in [19] can be relaxed and the statements can be strengthened.

The paper is organized as follows. In Section 1.2, we introduce our model and some necessary notation; in Section 1.3, the segmentation problem, infinite Viterbi path, barriers and many other concepts are introduced and defined. Also the idea of piecewise construction is explained in detail. This subsection also contains several examples like the above-mentioned example of an HMM with no infinite Viterbi path (Example 1). The subsection ends with the overview about the history of the problem. In Sections 2 and 3, the main barrier-construction theorems, Theorems 2.1 and 3.1, are stated and proven. In Section 4, these theorems are applied for several special cases and models: HMM's (Section 4.1, discrete $\mathcal{X}$ (Section 4.2), linear Markov switching model (Section 4.3) and a class of Gaussian PMM's (Section 4.4).

1.2. Pairwise Markov model

Let the observation-space $\mathcal{X}$ be a separable metric space equipped with its Borel σ-field $\mathcal{B}(\mathcal{X})$. Let the state-space be $\mathcal{Y} = \{1, 2, \ldots, |\mathcal{Y}|\}$, where $|\mathcal{Y}|$ is some positive integer. We denote $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, and equip $\mathcal{Z}$ with product topology $\tau \times 2^{\mathcal{Y}}$, where τ denotes the topology induced by the metrics of $\mathcal{X}$. Furthermore, $\mathcal{Z}$ is equipped with its Borel σ-field $\mathcal{B}(\mathcal{Z}) = \mathcal{B}(\mathcal{X}) \otimes 2^{\mathcal{Y}}$, which is the smallest σ-field containing sets of the form $A \times B$, where $A \in \mathcal{B}(\mathcal{X})$ and $B \in 2^{\mathcal{Y}}$. Let μ be a σ-finite measure on $\mathcal{B}(\mathcal{X})$ and let c be the counting measure on $2^{\mathcal{Y}}$. Finally, let

$$q\colon \mathcal{Z}^2 \to \mathbb{R}_{\geq 0}, \quad (z, z') \mapsto q(z|z')$$

be a such a measurable non-negative function that for each $z' \in \mathcal{Z}$ the function $z \mapsto q(z|z')$ is a density with respect to product measure $\mu \times c$.

We define random process $Z = \{Z_k\}_{k\geq 1} = \{(X_k, Y_k)\}_{k\geq 1}$ as a homogeneous Markov chain on the two-dimensional space $\mathcal{Z}$ having the transition kernel density $q(z|z')$. This means that the transition kernel of Z is defined as follows:

$$P(Z_2 \in A|Z_1 = z') = \int_A q(z|z')\, \mu \times c(dz), \quad z' \in \mathcal{Z}, \quad A \in \mathcal{B}(\mathcal{Z}).$$

The marginal processes $\{X_k\}_{k\geq 1}$ and $\{Y_k\}_{k\geq 1}$ will be denoted with X and Y, respectively. Following [5,6,24], we call the process Z a *pairwise Markov model* (PMM). It should be noted that even though Z is a Markov chain, this does not necessarily imply that either of the marginal processes X and Y are Markov chains, but is some special cases they can be. However, it is not difficult to see that conditionally, given Y, X is Markov chain, and vice-versa [24].

The letter p will be used to denote the various joint and conditional densities. By abuse of notation, the corresponding probability law is indicated by arguments of $p(\cdot)$, with lower-case x_k, y_k and z_k indicating random variables X_k, Y_k and Z_k, respectively. For example

$$p(x_{2:n}, y_{2:n}|x_1, y_1) = \prod_{k=2}^{n} q(x_k, y_k|x_{k-1}, y_{k-1}),$$

where $x_{2:n} = (x_2, \ldots, x_n)$ and $y_{2:n} = (y_2, \ldots, y_n)$. Sometimes it is convenient to use other symbols beside x_k, y_k, z_k as the arguments of some density; in that case we indicate the corresponding probability law using the equality sign, for example

$$p(x_{2:n}, y_{2:n}|x_1 = x, y_1 = i) = q(x_2, y_2|x, i) \prod_{k=3}^{n} q(x_k, y_k|x_{k-1}, y_{k-1}), \quad n \geq 3.$$

Also $p(z_1) = p(x_1, y_1)$ denotes the initial distribution density of Z with respect to measure $\mu_1 \times c$, where μ_1 is some σ-finite measure on $\mathcal{B}(\mathcal{X})$. Thus the joint density of $Z_{1:n}$ is $p(z_{1:n}) = p(z_1)p(z_{2:n}|z_1)$. For every $n \geq 2$ and $i, j \in \mathcal{Y}$ we also denote

$$p_{ij}(x_{1:n}) := \max_{y_{1:n}\colon y_1=i, y_n=j} \prod_{k=2}^{n} q(x_k, y_k|x_{k-1}, y_{k-1}), \quad x_{1:n} \in \mathcal{X}^n. \tag{1}$$

Thus

$$p_{ij}(x_{1:n}) = \max_{y_{1:n}\colon y_1=i, y_n=j} p(x_{2:n}, y_{2:n}|x_1, y_1).$$

If $p(y_2|x_1, y_1)$ does not depend on x_1, and $p(x_2|y_2, x_1, y_1)$ does not depend on neither x_1 nor y_1, then Z is called a *hidden Markov model* (HMM). In that case, denoting

$$p_{ij} := p(y_2 = j|y_1 = i), \quad f_j(x) := p(x_2 = x|y_2 = j),$$

the transition kernel density factorizes into

$$q(x, j|x', i) = p(x_2 = x|y_2 = j, x_1 = x', y_1 = i)p(y_2 = j|x_1 = x', y_1 = i) = p_{ij} f_j(x).$$

Density functions f_j are also called the *emission densities*. When $\mathcal{X}$ is discrete, then $f_j(x) = P(X_2 = x|Y_2 = j)$ is called the *emission probability* of x from state j.

If $p(y_2|x_1, y_1)$ does not depend on x_1, and $p(x_2|y_2, x_1, y_1)$ does not depend on y_1, then following [3] we call Z a *Markov switching model*. Thus HMM's constitute a sub-class of Markov switching models. In the case of Markov switching model, denoting

$$f_j(x|x') := p(x_2 = x|y_2 = j, x_1 = x'),$$

the transition kernel density becomes

$$q(x, j|x', i) = p_{ij} f_j(x|x').$$

It is easy to confirm that in case of Markov switching model (and therefore also in case of HMM) Y is a homogeneous Markov chain with transition matrix (p_{ij}). Most PMM's used in practice fall into the class of Markov switching models (see e.g. [3] and the references therein for the practical applications of Markov switching models). It should be noted that although PMM is a way more general model than HMM, it is not so much studied in the literature and therefore also the terminology and classification of dependence-structures is not unified. For example, a classification of PMM's according to their dependence structures can be found in [25], and a slightly different one is presented in [6].

We also admit that often in the literature X stands for the underlying process and Y is the observed one. Our notation is opposite and it is inherited from statistical learning, where the observations (sample) are typically denoted by $x_{1:n}$ and Y stands for the latent variables.

1.3. Viterbi path

The *segmentation problem* in general consists of guessing or estimating the unobserved realization of process $Y_{1:n}$ – the true path – given the realization $x_{1:n}$ of the observation process $X_{1:n}$. Since the true path cannot be exactly known, the segmentation procedure merely consists of finding the path that in some sense is the best approximation. Probably the most popular estimate is the path with maximum posterior probability. This path will be denoted with $v(x_{1:n})$ and also with $v_{1:n}$, when $x_{1:n}$ is assumed to be fixed:

$$v_{1:n} =: v(x_{1:n}) := \arg\max_{y_{1:n}} p(y_{1:n}, x_{1:n}) = \arg\max_{y_{1:n}} P(Y_{1:n} = y_{1:n}|X_{1:n} = x_{1:n}).$$

Typically $v_{1:n}$ is called Viterbi or MAP path (also Viterbi or MAP alignment). Clearly $v_{1:n}$ might not be unique. As it is well known, Viterbi path minimizes the average error over all possible paths, when the error between two sequences is zero if they are totally equal and one otherwise. On the other hand, Viterbi path is not in general the one that minimizes the expected number of errors, when the number of errors between two sequences are measured entry by entry (Hamming metric). For more detailed discussion about the segmentation problem and the properties of different estimates, we refer to [14–16,21,26]. Although these papers deal with HMM's only, the general theory applies for any model including PMM's.

The Viterbi path inherits its name by famous Viterbi algorithm that is used to find the Viterbi path in the case of HMM. It is easy to see that the algorithm also applies in the case of PMM. To see that, denote for every $y \in \mathcal{Y}$

$$\delta_1(y) := p(x_1, y_1 = y), \quad \delta_t(y) := \max_{y_{1:t}:\, y_t = y} p(x_{1:t}, y_{1:t}), \quad t \geq 2.$$

Clearly $\delta_t(y)$ also depends on $x_{1:t}$, but in our case the path $x_{1:n}$ is typically fixed and therefore $x_{1:t}$ is left out from the definition. The recursion behind the Viterbi algorithm is now

$$\delta_{t+1}(y) = \max_{y'} \delta_t(y') q(x_{t+1}, y|x_t, y'), \quad t = 2, \ldots, n. \tag{2}$$

At each time $t+1$ and state y the algorithm remembers the state y' achieving the maximum in (2) and by backtracking from the state $v_n = \arg\max_{y\in\mathcal{Y}} \delta_n(y)$, the Viterbi path can be found. To avoid the numerical underflow, the logarithmic or rescaled versions of the Viterbi recursion can be used, see e.g. [14].

Because Viterbi algorithm applies for PMM's as easily as for HMM's, using Viterbi path in segmentation is appealing computationally as well as conceptually. Therefore, to study the statistical properties of Viterbi path-based inferences, one has to know the long-run or typical behavior of random vectors $v(X_{1:n})$. As argued in [19], behavior of $v(X_{1:n})$ is not trivial since the $(n+1)^{\text{th}}$ observation can change the alignment based on the previous observations $x_{1:n}$. It might happen with a positive probability that the first entry of $v(x_{1:n+1})$ is different from corresponding entry of $v(x_{1:n})$. If this happens again and again, then the first element of $v(x_{1:n})$ keeps changing as n grows and there is not such thing as limiting Viterbi path. On the other hand, it is intuitively clear that there is a positive probability to observe $x_{1:n}$ such that regardless of the value of the $(n+1)^{\text{th}}$ observation (provided n is sufficiently large), the paths $v(x_{1:n})$ and $v(x_{1:n+1})$ agree on first u elements, where $u < n$. If this is true, then no matter what happens in the future, the first u elements of the paths remain constant. Provided there is an increasing unbounded sequence u_i ($u < u_1 < u_2 < \cdots$) such that the path up to u_i remains constant, one can define limiting or infinite Viterbi path. Let us formalize the idea. In the following definition $v(x_{1:n})$ is a Viterbi path and $v(x_{1:n})_{1:t}$ are the first t elements of the n-elemental vector $v(x_{1:n})$.

Definition 1.1. Let $x_{1:\infty}$ be a realization of X. The sequence $v_{1:\infty} \in \mathcal{Y}^\infty$ is called *infinite Viterbi path of* $x_{1:\infty}$ if for any $t \geq 1$ there exists $m(t) \geq t$ such that

$$v(x_{1:n})_{1:t} = v_{1:t}, \quad \forall n \geq m(t). \tag{3}$$

Hence $v_{1:\infty}$ is the infinite Viterbi path of $x_{1:\infty}$ if for any t, the first t elements of $v_{1:\infty}$ are the first t elements of a Viterbi path $v(x_{1:n})$ for all n big enough ($n \geq m(t)$). In other words, for every n big enough, there exists at least one Viterbi path so that $v(x_{1:n})_{1:t} = v_{1:t}$. Note that above-stated definition is equivalent to the following: for every $t \geq 1$,

$$\lim_n v(x_{1:n})_{1:t} \to v_{1:t}. \tag{4}$$

Indeed, since $\mathcal{Y}^t$ is finite, the convergence (4) holds if and only if $v(x_{1:n})_{1:t} = v_{1:t}$ eventually, and this is exactly (3). For infinite $\mathcal{Y}$, (3) is obviously much stronger than (4), and in this case, the infinite Viterbi path is defined via (4), see [4].

As we shall see, for many PMM's the infinite Viterbi path exists for almost every realization of X. However, the following counterexample shows that for some models the infinite Viterbi path exists for almost no realization of X.

Example 1. Let $p \in (\frac{1}{2}, 1)$; then there exists positive integer K such that taking $\epsilon = \frac{1}{2K}$, we have

$$\epsilon < 1 - p - \frac{\epsilon}{2} < \frac{1}{2} < p - \frac{\epsilon}{2} < 1. \tag{5}$$

We look at the model where $\mathcal{X} = \{1, 2\}$, $\mathcal{Y} = \{1, 2, \ldots, K + 2\}$ and the transmission matrix of Z is

$$\begin{array}{c} \\ (1,1) \\ (2,1) \\ (1,2) \\ (2,2) \\ (1,3) \\ (2,3) \\ \vdots \\ (1,K+2) \\ (2,K+2) \end{array}
\begin{array}{c} \begin{array}{ccccccccc} (1,1) & (2,1) & (1,2) & (2,2) & (1,3) & (2,3) & \ldots & (1,K+2) & (2,K+2) \end{array} \\
\begin{bmatrix}
p-\frac{\epsilon}{2} & 1-p-\frac{\epsilon}{2} & 0 & 0 & \epsilon^2 & \epsilon^2 & \ldots & \epsilon^2 & \epsilon^2 \\
p-\frac{\epsilon}{2} & 1-p-\frac{\epsilon}{2} & 0 & 0 & \epsilon^2 & \epsilon^2 & \ldots & \epsilon^2 & \epsilon^2 \\
0 & 0 & 1-p-\frac{\epsilon}{2} & p-\frac{\epsilon}{2} & \epsilon^2 & \epsilon^2 & \ldots & \epsilon^2 & \epsilon^2 \\
0 & 0 & 1-p-\frac{\epsilon}{2} & p-\frac{\epsilon}{2} & \epsilon^2 & \epsilon^2 & \ldots & \epsilon^2 & \epsilon^2 \\
0 & 0 & 0 & 0 & \epsilon & \epsilon & \ldots & \epsilon & \epsilon \\
0 & 0 & 0 & 0 & \epsilon & \epsilon & \ldots & \epsilon & \epsilon \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \epsilon & \epsilon & \ldots & \epsilon & \epsilon \\
0 & 0 & 0 & 0 & \epsilon & \epsilon & \ldots & \epsilon & \epsilon
\end{bmatrix} \end{array}.$$

We assume that the initial distribution of Z is such that

$$P(X_1 = 1) = P(X_1 = 2) = P(Y_1 = 1) = P(Y_1 = 2) = \frac{1}{2}$$

and X_1 and Y_1 are independent.

Let us see now what are the possible Viterbi paths for some observation sequence $x_{1:n}$. We have

$$\begin{aligned}
\max_{y_{1:n}} p(x_{1:n}, y_{1:n}) &= \frac{1}{4} \max_{y_1 \in \{1,2\},\ y_{2:n} \in \mathcal{Y}^{n-1}} \prod_{k=2}^{n} q(x_k, y_k | x_{k-1}, y_{k-1}) \\
&= \frac{1}{4} \max_{y_{1:n} \in \{1,2\}^n} \prod_{k=2}^{n} q(x_k, y_k | x_{k-1}, y_{k-1}) \\
&= \frac{1}{4} \max_{y_{1:n} \in \{(1,\ldots,1),(2,\ldots,2)\}} \prod_{k=2}^{n} q(x_k, y_k | x_{k-1}, y_{k-1}), \qquad (6)
\end{aligned}$$

where the second equality holds by (5). Let $n_i(x_{2:n})$ be the number of i-s in $x_{2:n}$. Since

$$\prod_{k=2}^{n} q(x_k, 1 | x_{k-1}, 1) = \left(p - \frac{\epsilon}{2}\right)^{n_1(x_{2:n})} \left(1 - p - \frac{\epsilon}{2}\right)^{n_2(x_{2:n})}$$

and

$$\prod_{k=2}^{n} q(x_k, 2 | x_{k-1}, 2) = \left(p - \frac{\epsilon}{2}\right)^{n_2(x_{2:n})} \left(1 - p - \frac{\epsilon}{2}\right)^{n_1(x_{2:n})}$$

we have by the fact that $p > 1 - p$ and (6), that Viterbi path of $x_{1:n}$ can be expressed as

$$v(x_{1:n}) = \begin{cases} (1, \ldots, 1), & \text{if } n_1(x_{2:n}) \geq n_2(x_{2:n}) \\ (2, \ldots, 2), & \text{else.} \end{cases} \qquad (7)$$

Now, let us take a closer look at the behavior of Z. Note that at some time, let us say at time T, Y moves to state space $\{3, \ldots, K + 2\}$. Time T is a random variable that is almost surely finite. Before time T, Y is constantly in the state 1 or constantly in the state 2 (both possibilities having probability $\frac{1}{2}$). After and at time T, Y is always in $\{3, \ldots, K + 2\}$. Note that $\{X_{T+k}\}_{k \geq 0}$ is an i.i.d. Bernoulli sequence with parameter $\frac{1}{2}$. Let

$$S_k = n_1(X_{2:k}) - n_2(X_{2:k}), \quad k \geq 2.$$

Random process $\{S_{T+k}\}_{k\geq 0}$ is a simple symmetric random walk with random starting point. Therefore the process $\{S_k\}_{k\geq 2}$ will almost surely fall below zero i.o. and rise above zero i.o. Together with (7) this implies that almost no realization of X has an infinite Viterbi path.

It is easy to confirm that this model is HMM with

$$p_{11} = p_{22} = 1-\epsilon, \quad p_{12} = p_{21} = 0,$$
$$p_{1k} = p_{2k} = \frac{\epsilon}{K}, \quad p_{k1} = p_{k2} = 0, \quad k = 3, \dots, K+2,$$
$$p_{kl} = \frac{1}{K}, \quad k, l \in \{3, \dots, K+2\},$$

and

$$f_2(2) = f_1(1) = \frac{1}{1-\epsilon}\left(p - \frac{\epsilon}{2}\right), \quad f_1(2) = f_2(1) = \frac{1}{1-\epsilon}\left(1 - p - \frac{\epsilon}{2}\right),$$
$$f_k(1) = f_k(2) = \frac{1}{2}, \quad k = 3, \dots, K+2.$$

Nodes. Suppose now $x_{1:\infty}$ is such that infinite Viterbi path exists. It means that for every time t, there exists time $m(t) \geq t$ such that the first t elements of $v(x_{1:n})$ are fixed as soon as $n \geq m$. Note that if $m(t)$ is such a time, then $m(t)+1$ is such a time too. Theoretically, the time m might depend on the whole sequence $x_{1:\infty}$. This means that after observing the sequence $x_{1:m}$, it is not yet clear, whether the first t elements of Viterbi path are now fixed (for any continuation of $x_{1:m}$) or not. In practice, one would not like to wait infinitely long, instead one prefers to realize that the time $m(t)$ is arrived right after observing $x_{1:m}$. In this case, the (random) time $m(t)$ is the stopping time with respect to the observation process. In particular, it means the following: for every possible continuation $x_{m+1:n}$ of $x_{1:m}$, the Viterbi path at time t passes the state v_t, let that state be i. This requirement is fulfilled, when the following holds: for every two states $j, k \in \mathcal{Y}$

$$\delta_t(i) p_{ij}(x_{t:m}) \geq \delta_t(k) p_{kj}(x_{t:m}), \tag{8}$$

where $p_{ij}(\cdot)$ is defined in (1). Indeed, there might be several states satisfying (8), but the ties can always be broken in favor of the state i, so that whenever $n \geq m$, there is at least one Viterbi path $v(x_{1:n})$ that passes the state i at time t. Therefore, if at time t, there is a state i satisfying (8), then m is the time $m(t)$ required in (3) and it depends on $x_{1:m}$ only.

Definition 1.2. Let $x_{1:m}$ be a vector of observations. If equalities (8) hold for any pair of states j and k, then the time t is called an *i-node of order $r = m - t$*. Time t is called a *strong i-node of order r*, if it is an i-node of order r, and the inequality (8) is strict for any j and $k \neq i$ for which the left side of the inequality is positive. We call t a node of order r if for some i, it is an i-node of order $r = m - t$.

The definition of node is a straightforward generalization of the corresponding definition in [13,19,20]. Note that when t is a node of order r, then $t-1$ is a node of order $r+1$.

Example 2. Following is an example of a model, for which infinite Viterbi path always exists, but no nodes ever occur. Let Z be a HMM with $\mathcal{Y} = \mathcal{X} = \{1, 2\}$, transition matrix of Y being identity. Suppose $p \in (\frac{1}{2}, 1)$ and let emission probabilities be

$$f_1(1) = f_2(2) = p, \quad f_1(2) = f_2(1) = 1 - p.$$

Let the initial distribution be uniform. This trivial model picks parameter p or $1-p$ with probability $\frac{1}{2}$ and then an i.i.d. Bernoulli sample with chosen probability. Let, for any $x_{1:t} \in \{0,1\}^t$, $n_i(x_{1:t})$ be the number of i-s in $x_{1:t}$. Now clearly for any $n \geq 1$

$$v(x_{1:n}) = \begin{cases} (1,\dots,1), & \text{if } n_1(x_{1:t}) \geq n_2(x_{1:t}) \\ (2,\dots,2), & \text{else.} \end{cases}$$

Since by SLLN

$$\frac{n_2(X_{1:n})}{n} \to \begin{cases} p, & \text{if } Y_1 = 2 \\ 1-p, & \text{if } Y_1 = 1 \end{cases} \quad \text{a.s.,}$$

we see that for almost every realization of X the infinite Viterbi path exists. This infinite path is constantly 1 if $Y_1 = 1$ and constantly 2 if $Y_1 = 2$. Surely, for any t, there exists $m(t)$ such that (3) holds, but in this case $m(t)$ depends on the whole sequence $x_{1:\infty}$, because for any $x_{1:m}$ one can find a continuation $x_{m+1:n}$ such that $v(x_{1:t}) \neq v(x_{1:n})_{1:t}$. This implies that there cannot be any nodes in any sequence $x_{1:\infty}$. Indeed, for any $x_{t+1:m}$, it holds $p_{12}(x_{t:m}) = p_{21}(x_{t:m}) = 0$, and so inequalities (8) cannot hold.

Barriers. The goal of the present paper is to find sufficient conditions for almost every realization of observation process to have infinitely many nodes. Whether a time t is a node of order r or not depends, in general, on the sequence $x_{1:t+r}$. Sometimes, however, there is some small block of observations that guarantees the existence of a node regardless of the other observations. Let us illustrate this by an example.

Example 3. Suppose that there exists a state $i \in \mathcal{Y}$ such that for any triplet $y_{t-1}, y_t, y_{t+1} \in \mathcal{Y}$

$$q(x_t, i|x_{t-1}, y_{t-1})q(x_{t+1}, y_{t+1}|x_t, i) \geq q(x_t, y_t|x_{t-1}, y_{t-1})q(x_{t+1}, y_{t+1}|x_t, y_t). \tag{9}$$

Then

$$\begin{aligned} \delta_t(i)q(x_{t+1}, y_{t+1}|x_t, i) &= \max_{y'} \delta_{t-1}(y')q(x_t, i|x_{t-1}, y')q(x_{t+1}, y_{t+1}|x_t, i) \\ &\geq \max_{y'} \delta_{t-1}(y')q(x_t, y_t|x_{t-1}, y')q(x_{t+1}, y_{t+1}|x_t, y_t) \\ &= \delta_t(y_t)q(x_{t+1}, y_{t+1}|x_t, y_t). \end{aligned}$$

We thus have that t is an i-node of order 1, because for every pair $j, k \in \mathcal{Y}$

$$\delta_t(i)p_{ij}(x_t, x_{t+1}) \geq \delta_t(k)p_{kj}(x_t, x_{t+1}).$$

Whether (9) holds or not, depends on triplet (x_{t-1}, x_t, x_{t+1}). In case of Markov switching model, (9) is

$$p_{y_{t-1}i} f_i(x_t|x_{t-1}) \cdot p_{iy_{t+1}} f_{y_{t+1}}(x_{t+1}|x_t) \geq p_{y_{t-1}y_t} f_{y_t}(x_t|x_{t-1}) \cdot p_{y_t y_{t+1}} f_{y_{t+1}}(x_{t+1}|x_t).$$

And in a more special case of HMM, (9) is equivalent to

$$p_{y_{t-1}i} f_i(x_t) \cdot p_{iy_{t+1}} \geq p_{y_{t-1}y_t} f_{y_t}(x_t) \cdot p_{y_t y_{t+1}}. \tag{10}$$

The inequalities (10) have very clear meaning — when the observation x_t has relatively big probability of being emitted from state i (in comparison of being emitted from any other state), then regardless of the observations before or after x_t, time t is a node. In particular, this is the case when the supports of the emission distributions are different and x_t can be emitted from one state, only. On the other hand, for many models, there are no such x_t possible, so (10) is rather an exception than a rule.

Definition 1.3. Given $i \in \mathcal{Y}$, $b_{1:M}$ is called an *(strong) i-barrier of order r and length M*, if, for any $x_{1:\infty}$ with $x_{m-M+1:m} = b_{1:M}$ for some $m \geq M$, $m - r$ is an (strong) i-node of order r.

Hence, if (9) holds, then the triplet (x_{t-1}, x_t, x_{t+1}) is an i-barrier of order 1 and length 3. In what follows, we give some sufficient conditions that guarantee the existence of infinitely many barriers in almost every realization of X. More closely, we construct a set $\mathcal{X}^* \subset \mathcal{X}^M$ such that every vector $x_{1:M}$ from $\mathcal{X}^*$ is i-barrier of order r for a given state $i \in \mathcal{Y}$, and $P(X \in \mathcal{X}^* \text{ i.o.}) = 1$, where

$$\{X \in \mathcal{X}^* \text{ i.o.}\} := \bigcap_{k=1}^{\infty} \bigcup_{l=k}^{\infty} \{X_{l:l+M-1} \in \mathcal{X}^*\}.$$

Since every barrier contains a r-order i-node, having infinitely many barriers in $x_{1:\infty}$ entails infinitely many i-nodes of order r, let the locations of these nodes be $u_1 < u_2 < \cdots$. Let $m \geq u_2 + r$. There must exist a Viterbi path $v(x_{1:m})$ passing state i at time u_1. There also exists a Viterbi path passing i at time u_2. If $v(x_{1:m})$ is unique, then the path passes i at both times, but if Viterbi path is not unique and u_1 and u_2 are too close to each other, then there might not be possible to break ties in favor of i at u_1 and u_2 simultaneously, as is shown in the following example.

Example 4. Let Z be HMM with $|\mathcal{Y}| \geq 4$ and let for some $\epsilon \in (0, 1)$

$$p_{ij} = \epsilon, \quad (i, j) \in 2^{\{1,2\}} = \{(1, 1), (1, 2), (2, 1), (2, 2)\},$$
$$p_{13} = p_{24} = p_{41} = p_{32} = \epsilon, \quad p_{14} = p_{23} = p_{42} = p_{31} = 0.$$

Also, let $\mathcal{X}$ be finite, and let $1, 2 \in \mathcal{X}$ be such that for some $p \in (0, 1)$

$$f_1(1) = f_2(1) = p; \quad f_k(1) = 0, \quad k \in \mathcal{Y} \setminus \{1, 2\};$$
$$f_3(2) = f_4(2) = p; \quad f_k(2) = 0, \quad k \in \mathcal{Y} \setminus \{3, 4\}.$$

Thus, whenever $X_t = 1$, then $Y_t \in \{1, 2\}$, and whenever $X_t = 2$, then $Y_t \in \{3, 4\}$ $(t \geq 2)$. Note that the word $(1, 1, 2, 1, 1)$ is 1- and 2-barrier of length 5 and of order 3. To see this, note that when for some t, $x_{t:t+4} = (1, 1, 2, 1, 1)$ then for all $k, j \in \mathcal{Y}$

$$\delta_{t+1}(k) p_{kj}(1, 2, 1, 1) = \delta_t(i^*) p^4 \epsilon^4 \cdot \mathbb{I}_{\{1,2\}}(k) \mathbb{I}_{\{1,2\}}(j),$$

where $i^* = \operatorname{argmax}_{l \in \{1,2\}} \delta_t(l)$ and $\mathbb{I}$ denotes the indicator function. Thus time $t+1$ is an i-node of order 3 for $i \in \{1, 2\}$: for all $k, j \in \mathcal{Y}$

$$\begin{aligned}
\delta_{t+1}(i) p_{ij}(1, 2, 1, 1) &= \delta_t(i^*) p^4 \epsilon^4 \cdot \mathbb{I}_{\{1,2\}}(i) \mathbb{I}_{\{1,2\}}(j) \\
&\geq \delta_t(i^*) p^4 \epsilon^4 \cdot \mathbb{I}_{\{1,2\}}(k) \mathbb{I}_{\{1,2\}}(j) \\
&= \delta_{t+1}(k) p_{kj}(1, 2, 1, 1).
\end{aligned}$$

This means that the word $(1, 1, 2, 1, 1)$ is indeed a 1- and 2-node of order 3.

Similarly, $(2, 1, 1, 1, 1)$ is also 1- and 2-barrier of length 5 and of order 3. Assuming that Y is irreducible, we have that the word $(1, 1, 2, 1, 1, 1, 1)$ occurs in X infinitely many times. Suppose now that t is such that $x_{t:t+6}$ is equal to that word. Then $t + 1$ and $t + 3$ are both 1- and 2-nodes. Now, breaking ties at these locations differently (to 1 at $t + 1$ and to 2 at $t + 3$, or vice-versa) is acceptable. But breaking ties to the same value (either both to 1 or both to 2) will result in a zero-likelihood path. Indeed, for $y_{t+1:t+3} \in \{1,2\} \times \mathcal{Y} \times \{1,2\}$

$$p(x_{t+2:t+3}, y_{t+2:t+3}|x_{t+1}, y_{t+1}) = p_{y_{t+1}y_{t+2}} f_{y_{t+2}}(2) \cdot p_{y_{t+2}y_{t+3}} f_{y_{t+3}}(1)$$
$$= \epsilon^2 \mathbb{I}_{\{(1,3,2)\}}(y_{t+1:t+3}) \cdot \mathbb{I}_{\{(2,4,1)\}}(y_{t+1:t+3}) \cdot p^2.$$

This problem does not occur, if the nodes are strong or if $u_2 \geq u_1 + r$. Indeed, since u_1 is an i-node of order r, then by definition of r-order node, between times u_1 and $u_1 + r + 1$, the ties can be broken so that whatever state the Viterbi path passes at time u_2, it passes i at time u_1. Thus, if the locations of nodes $u_1 < u_2 < \cdots$ are such that $u_k \geq u_{k-1} + r$ for all $k \geq 2$, it is possible to construct the infinite Viterbi path so that it passes the state i at every time u_k. In what follows, when the nodes u_k and u_{k-1} are such that $u_k \geq u_{k-1} + r$, then the nodes are called *separated*. Of course, there is no loss of generality in assuming that the nodes $u_1 < u_2 < \cdots$ are separated, because from any non-separated sequence of nodes it is possible to pick a separated subsequence. Another approach is to enlarge the barriers so that two barriers cannot overlap and, therefore, are separated. This is the way barriers are defined in [20].

Construction of infinite Viterbi path. Having infinitely many separated nodes $u_1 < u_2 < \cdots$ or order r, it is possible to construct the infinite Viterbi path *piecewise*. Indeed, we know that for every $n \geq u_k + r$, there is a Viterbi path $v_{1:n} = v(x_{1:n})$ such that $v_{u_j} = i$, $j = 1, \ldots, k$. Because of that property and by optimality principle clearly the piece $v_{u_{j-1}:u_j}$ depends on the observations $x_{u_{j-1}:u_j}$, only. Therefore $v_{1:\infty}$ can be constructed in the following way: first use the observations $x_{1:u_1}$ to find the first piece $v_{1:u_1}$ as follows:

$$v_{1:u_1} = \underset{y_{1:u_1}\,:\, y_{u_1}=i}{\arg\max}\; p(x_{1:u_1}, y_{1:u_1}).$$

Then use $x_{u_1:u_2}$ to find the second piece $v_{u_1:u_2}$ as follows:

$$v_{u_1:u_2} = \underset{y_{u_1:u_2}\,:\, y_{u_1}=y_{u_2}=i}{\arg\max}\; p(x_{u_1:u_2}, y_{u_1:u_2}),$$

and so on. Finally use $x_{u_k:n}$ to find the last piece $v_{u_k:n}$ as follows:

$$v_{u_k:n} = \underset{y_{u_k:n}\,:\, y_{u_k}=i}{\arg\max}\; p(x_{u_k:n}, y_{u_k:n}).$$

The last piece $v_{u_k:n}$ might change as n grows, but the rest of the Viterbi path is now fixed. Thus, if $x_{1:\infty}$ contains infinitely many nodes, the whole infinite path can be constructed piecewise.

If the nodes u_k are strong (not necessarily separated) then the piecewise construction detailed above is achieved when the Viterbi estimation is done by a lexicographic or co-lexicographic tie-breaking scheme induced by some ordering on $\mathcal{Y}$. Indeed, since the i-nodes u_k are strong, we know that regardless of tie-breaking scheme $v(x_{1:n})_{u_k} = i$ for all $k \geq 1$ and $n \geq u_k + r$. Therefore the lexicographic ordering ensures that for all $k \geq 1$ and $n \geq u_k + r$

$$v_{1:n} = \underset{y_{1:n}}{\arg\max}\, p(x_{1:n}, y_{1:n}) = (\underset{y_{1:u_k}}{\arg\max}\, p(x_{1:u_k}, y_{1:u_k}), \underset{y_{u_k+1:n}}{\arg\max}\, p(x_{u_k:n}, y_{u_k} = i, y_{u_k+1:n})).$$

This shows that $v(x_{1:n})_{1:u_k}$ is independent of $n \geq u_k + r$ for all $k \geq 1$ and so the infinite Viterbi path is well-defined.

Viterbi process. The notion of infinite Viterbi path of a fixed realization $x_{1:\infty}$ naturally carries over to an infinite Viterbi path of X, called the *Viterbi process*. Formally, this process is defined as follows.

Definition 1.4. A random process $V = \{V_k\}_{k\geq 1}$ on space $\mathcal{Y}$ is called a *Viterbi process*, if the event $\{V$ is not an infinite Viterbi path of $X\}$ is contained in a set of zero probability measure.

If there exists a barrier set $\mathcal{X}^*$ consisting of i-barriers of fixed order and satisfying $P(X \in \mathcal{X}^* \text{ i.o.}) = 1$, then the Viterbi process can be constructed by applying the piecewise construction detailed above to the process X. However, this construction has a serious weakness: it requires that the ties are broken in each piece of the piecewise path separately, which means that to obtain the correct Viterbi path (corresponding to the Viterbi process) one has to first identify the barriers in the observation sequence. In practice, this type of tie-breaking mechanism would complicate implementation of the Viterbi path estimation and add significantly to its computational cost. The solution to this problem is to allow only strong barriers, in which case, as we saw above, the piecewise tie-breaking can be replaced with lexicographic or co-lexicographic tie-breaking. Fortunately, as we will see later, the requirement of strong barriers as opposed to simply barriers does not seem to be restrictive.

Proving the existence of the Viterbi process is the main motivation for barrier set construction. Once it is established that the Viterbi process $V = \{V_k\}_{k\geq 1}$ exists, the next step is to study its probabilistic properties. An important and very useful property is that the process $(Z, V) = \{(Z_k, V_k)\}_{k\geq 1}$ is regenerative. In case of HMM, this is achieved in [15,19] by, roughly speaking, constructing regeneration times for Z which are also nodes. When it is ensured that (Z, V) is regenerative, the standard theory for regenerative processes can be applied. See [7,12,15,19] for regeneration-based inferences of HMM. It is important to stress that the regenerativity of (Z, V) is possible due to the existence of barriers and that is an extra motivation of barrier construction studied in this paper. Note that the Viterbi process in Example 1.2 is not regenerative. Regenerativity of (Z, V) in case of general PMM's will be proven in a follow-up paper; in the present article we only deal with the existence of V.

History of the problem. To our best knowledge, so far the existence of Viterbi process has been proven in the case of HMM's only. The first attempts in that directions have been made by A. Caliebe and U. Rösler in [1,2]. They essentially define the concept of nodes and prove the existence of infinitely many nodes under rather restrictive assumptions like (10). For an overview of the main results in [1,2] as well as for the discussion about their assumption, see [19,20]. For HMM, the most general conditions for the existence of infinitely many barriers were given in Lemma 3.1 of [20] (the same lemma is also Lemma 3.1 in [19]). Let us now state that lemma.

Recall that in the case of HMM f_i are the emission densities with respect to measure μ. Denote

$$G_i := \{x \in \mathcal{X} \mid f_i(x) > 0\}, \quad i \in \mathcal{Y}. \tag{11}$$

A subset $C \subset \mathcal{Y}$ is called a *cluster*, if

$$\mu\left(\cap_{i\in C} G_i\right) > 0 \quad \text{and} \quad \mu\left[(\cap_{i\in C} G_i) \cap (\cup_{i\notin C} G_i)\right] = 0, \tag{12}$$

Distinct clusters need not be disjoint and a cluster can consist of a single state. In this latter case such a state is not hidden, since it is indicated by any observation it emits. When the number of states is two, then $\mathcal{Y}$ is the only cluster possible, since otherwise all observations would reveal their states and the underlying Markov chain would cease to be hidden.

Theorem 1.1 (*Lemma 3.1 in [20]*). *Suppose Z is stationary HMM satisfying the following conditions.*

(i) For each state $j \in \mathcal{Y}$

$$\mu\left(\left\{x \in \mathcal{X} \mid f_j(x)p_{\cdot j} > \max_{i\in\mathcal{Y},\ i\neq j} f_i(x)p_{\cdot i}\right\}\right) > 0, \quad \text{where } p_{\cdot j} := \max_{i\in\mathcal{Y}} p_{ij}. \tag{13}$$

(ii) There exists a cluster $C \subset \mathcal{Y}$ such that the sub-stochastic matrix $\mathbb{P}_C = (p_{ij})_{i,j\in C}$ is primitive, that is $\mathbb{P}_C^R$ has only positive elements for some positive integer R.

Also let Markov chain Y be irreducible and aperiodic. Then for $i \in \mathcal{Y}$ there exists a barrier set $\mathcal{X}^* \subset \mathcal{X}^M$, $M \geq 1$, consisting of i-barriers of fixed order and satisfying $P(X_{1:M} \in \mathcal{X}^*) > 0$.

Since stationary HMM with irreducible and aperiodic Y is ergodic, it immediately follows from this theorem that under the specified conditions almost every realization has infinitely many barriers and piecewise construction of Viterbi process is possible.

The assumptions (i) and (ii) are discussed in detail in [19,20]. Let us just mention that they are both natural and hold in the most models in practice. In particular, (ii) is much weaker than common assumptions of having all entries in transition matrix (p_{ij}) positive or all emission densities strictly positive and transition matrix primitive. It turns out that under (ii), it is possible to generalize the existing results of exponential forgetting properties of smoothing probabilities for HMM's [18]. This property has nothing to do with Viterbi paths so that (ii) is in a sense a natural and desirable property from many different aspects. For another application of (ii), see [15,16]. The generalization of (ii) in the case of PMM's is the condition **B1** in Theorem 3.1 and, just like (ii), also **B1** might be useful for proving many other properties of PMM besides the existence of infinite Viterbi path (filter stability, exponential forgetting on smoothing probabilities and many others). It also turns out that in the special case of 2-state HMM, both assumptions can be relaxed: namely an irreducible aperiodic 2-state Markov chain has always primitive transition matrix (but not necessarily having all entries positive as it is incorrectly stated in [13]), and as argued above, the cluster assumption (ii) trivially holds. It has been shown in [13], that for 2-state stationary HMM, almost every realization has infinitely many barriers if

$$\mu\left(\{x \in \mathcal{X} \mid f_1(x) \neq f_2(x)\}\right) > 0. \tag{14}$$

Obviously, the assumption (14) is most natural for any HMM, so essentially the result says that Viterbi process exists for any two-state stationary HMM.

Theorem 1.1 does have one weakness: it does not guarantee that the barrier set $\mathcal{X}^*$ consists of *strong* barriers. As we saw earlier, having infinitely many *strong* barriers (as opposed to simply barriers) is a very desirable property. We rectify this issue in Section 4.1, where we prove a generalized version of Theorem 1.1 (Corollary 4.1) which guarantees that the barrier set $\mathcal{X}^*$ consists of strong barriers. This corollary also relaxes the definition of cluster (12) by replacing it with so-called weak cluster assumption.

Finally we would like to add a few words on the stationarity assumption of Theorem 1.1. In case of HMM, this assumption is not very restrictive, since the stationary distribution of Z can easily be expressed through the stationary distribution of Y. More specifically, if (π_i) is the stationary distribution of Y, then the stationary density of Z is given by $p(x_1, y_1) = \pi_{y_1} f_{y_1}(x_1)$. For stationary HMM the stationary density can be easily calculated and hence the Viterbi algorithm is easy to implement. However, in general case of PMM we often do not have a way to calculate the stationary density (if it exists), and so the stationarity assumption becomes more restrictive. Therefore in the present paper we abandon this assumption altogether. This means that we can no longer rely on ergodicity of Z to ensure that $P(X \in \mathcal{X}^* \text{ i.o.}) = 1$ for the barrier set $\mathcal{X}^*$ — instead we apply the theory of *Harris recurrent* Markov chains.

When the hidden state space $\mathcal{Y}$ is infinite, then the infinite Viterbi path of $x_{1:\infty}$ is defined through convergences (4). The Viterbi process is then defined analogously to the case when $\mathcal{Y}$ is finite: it is the process on $\mathcal{Y}$ which is almost surely the infinite Viterbi path of X. In [4] P. Chigansky and Y. Ritov study the existence of such process in case of HMM with continuous hidden state space. The authors provide examples where the infinite Viterbi path does indeed exist, and moreover prove its existence under certain strong log-concavity conditions for transition and observation densities. They also provide an example where $\mathcal{Y}$ is countable and Markov chain Y is positive recurrent, but the infinite Viterbi path does not exist because it diverges coordinate-wise to infinity. This is similar to our Example 1 in the sense that both examples demonstrate a situation where the Viterbi process does not exist.

2. Barrier set construction theorem

Recall the definition of $p_{ij}(\cdot)$ in (1). We already saw that the inequalities (9) ensure that (x_{t-1}, x_t, x_{t+1}) is a barrier of order 1. We generalize this idea with:

Proposition 2.1. *Suppose $b_{1:M} \in \mathcal{X}^M$, $M \geq 3$, is such that for some $l \in \{2, \dots, M-1\}$*

$$p_{i1}(b_{1:l})p_{1j}(b_{l:M}) \geq p_{ik}(b_{1:l})p_{kj}(b_{l:M}), \quad \forall i, j, k \in \mathcal{Y}. \tag{15}$$

Then $b_{1:M}$ is a 1-barrier of order $M-l$. If inequalities (15) are strict for any i, j and any $k \neq 1$ for which the left side of the inequality is non-zero, then $b_{1:M}$ is a strong 1-barrier of order $M-l$.

Proof. Let $x_{1:\infty}$ be such a realization of X that $x_{t+1:t+M} = b_{1:M}$ for some t. Then for all $j, k \in \mathcal{Y}$

$$\begin{aligned} \delta_{t+l}(1)p_{1j}(x_{t+l:t+M}) &= \max_{i\in\mathcal{Y}} \delta_{t+1}(i)p_{i1}(x_{t+1:t+l})p_{1j}(x_{t+l:t+M}) \\ &\geq \max_{i\in\mathcal{Y}} \delta_{t+1}(i)p_{ik}(x_{t+1:t+l})p_{kj}(x_{t+l:t+M}) \\ &= \delta_{t+l}(k)p_{kj}(x_{t+l:t+M}), \end{aligned} \tag{16}$$

which shows that $x_{t+1:t+M} = b_{1:M}$ is indeed a 1-node of order $M-l$. Let now inequalities (15) be strict for any i, j and any $k \neq 1$ for which the left side of the inequality is non-zero. Then we have for every $j \in \mathcal{Y}$ for which $\delta_{t+l}(1)p_{1j}(x_{t+l:t+M}) > 0$ and for every $k \neq 1$ that the inequality (16) is strict, which makes $b_{1:M}$ a strong 1-barrier of order $M-l$. □

Proposition 2.1 allows us to derive conditions **A1–A3** detailed below, which ensure the existence of a barrier set. For any $n \geq 2$, define

$$\mathcal{Y}^+(x_{1:n}) := \{(i, j) \mid p_{ij}(x_{1:n}) > 0\}, \quad x_{1:n} \in \mathcal{X}^n. \tag{17}$$

For any set A consisting of vectors of length n we adopt the following notation:

$$A_{(k)} := \{x_k \mid x_{1:n} \in A\}, \quad 1 \leq k \leq n,$$
$$A_{(k,l)} := \{x_{k:l} \mid x_{1:n} \in A\}, \quad 1 \leq k \leq l \leq n.$$

Hence

$$\mathcal{Y}^+(x_{1:n})_{(1)} = \{i \mid \exists j(i) \text{ such that } p_{ij}(x_{1:n}) > 0\},$$

$$\mathcal{Y}^+(x_{1:n})_{(2)} = \{j \mid \exists i(j) \text{ such that } p_{ij}(x_{1:n}) > 0\}.$$

Observe that if $i \in \mathcal{Y}^+(x_{1:n})_{(1)}$ and $j \in \mathcal{Y}^+(x_{1:n})_{(2)}$, then not necessarily $(i, j) \in \mathcal{Y}^+(x_{1:n})$. The aforementioned conditions are the following.

A1 There exists $N \geq 2$, $n_1 < \cdots < n_{2N+2}$, set $\mathcal{X}^* \subset \mathcal{X}^{n_{2N+2}}$ and $\epsilon > 0$ such for all $k = 1, \ldots, 2N$ and all $x_{n_k:n_{k+1}} \in \mathcal{X}^*_{(n_k,n_{k+1})}$

$$p_{11}(x_{n_k:n_{k+1}}) \geq p_{i1}(x_{n_k:n_{k+1}}), \quad \forall i \in \mathcal{Y}, \tag{18}$$

$$p_{11}(x_{n_k:n_{k+1}}) \geq p_{1i}(x_{n_k:n_{k+1}}), \quad \forall i \in \mathcal{Y}, \tag{19}$$

$$p_{11}(x_{n_k:n_{k+1}})(1-\epsilon) > p_{ij}(x_{n_k:n_{k+1}}), \quad \forall i, j \in \mathcal{Y} \setminus \{1\}.$$

A2 There exist constants $0 < \delta \leq \Delta < \infty$ such that

$$p_{ij}(x_{1:n_1}), p_{ij}(x_{n_{2N+1}:n_{2N-2}}) \leq \Delta, \quad \forall i, j \in \mathcal{Y}, \quad \forall x_{1:n_1} \in \mathcal{X}^*_{(1,n_1)},$$
$$\forall x_{n_{2N+1}:n_{2N-2}} \in \mathcal{X}^*_{(n_{2N+1},n_{2N+2})},$$
$$\mathcal{Y}^+(x_{1:n_1}) \neq \emptyset, \quad p_{i1}(x_{1:n_1}) \geq \delta, \quad \forall i \in \mathcal{Y}^+(x_{1:n_1})_{(1)}, \quad \forall x_{1:n_1} \in \mathcal{X}^*_{(1,n_1)},$$
$$\mathcal{Y}^+(x_{n_{2N+1}:n_{2N+2}}) \neq \emptyset, \quad p_{1j}(x_{n_{2N+1}:n_{2N+2}}) \geq \delta, \quad \forall j \in \mathcal{Y}^+(x_{n_{2N+1}:n_{2N+2}})_{(2)},$$
$$\forall x_{n_{2N+1}:n_{2N+2}} \in \mathcal{X}^*_{(n_{2N+1},n_{2N+2})}.$$

A3 It holds

$$\frac{\Delta}{\delta}(1-\epsilon)^N < 1.$$

We also consider a strengthened version of **A1**:

A1′ The condition **A1** holds with either inequalities (18) or inequalities (19) being strict for all $i \neq 1$.

Theorem 2.1. *Suppose **A1–A3** are fulfilled. Then $\mathcal{X}^*$ consists of 1-barriers of order $n_{2N+2} - n_{N+1}$. Furthermore, if **A1′** holds instead of **A1**, then the 1-barriers are strong.*

Note how the condition **A1** concerns only the section $\mathcal{X}^*_{(n_1,n_{2N+1})}$ of $\mathcal{X}^*$ while the condition **A2** concerns the sections $\mathcal{X}^*_{(1,n_1)}$ and $\mathcal{X}^*_{(n_{2N+1},n_{2N+2})}$. This motivates:

Definition 2.1. If $\mathcal{X}^*$ satisfies **A1** (**A1′**), then $\mathcal{X}^*_{(n_1,n_{2N+1})}$ is called *a (strong) center part of a barrier set*.

Since a center part of a barrier set has a cyclic structure (consisting of $2N$ cycles), then it is natural that its construction is also cyclical. Consider for example the case when $\mathcal{X}$ is discrete and there exists a sequence $x_{1:n} \in \mathcal{X}^n$, $n \geq 2$, such that

$$x_1 = x_n \quad \text{and} \quad p_{11}(x_{1:n}) > p_{ij}(x_{1:n}), \quad \forall (i, j) \in \mathcal{Y}^2 \setminus \{(1, 1)\}.$$

Thus, denoting $x = x_{1:n-1}$, for any $N \geq 2$ we can take the strong center part to be

$$\{(\underbrace{x, x, \ldots, x}_{2N \text{ blocks of } x}, x_n)\}.$$

Since we can take N arbitrarily large, we can always ensure that **A3** holds when δ and Δ are fixed.

Let us consider now the case when $\mathcal{X}$ is uncountable. Then the situation is in general more complicated, since center part of a barrier set should typically contain uncountably many vectors for X to return to the corresponding barrier set $\mathcal{X}^*$ infinitely often. For any vector sets $A \subset \mathcal{X}^k$, $k \geq 1$, and $B \subset \mathcal{X}^l$, $l \geq 1$, we write

$$A \cdot B := \{x_{1:k+l-1} \mid x_{1:k} \in A,\ x_{k:k+l-1} \in B\}. \tag{20}$$

For a fixed $\epsilon > 0$ and $n \geq 1$ let

$$W = \{x_{1:n} \in \mathcal{X}^n \mid p_{11}(x_{1:n})(1-\epsilon) > p_{ij}(x_{1:n}),\ (i,j) \in \mathcal{Y}^2 \setminus \{(1,1)\}\}.$$

We can construct a strong center part of a barrier set by gluing together $2N$ instances of W:

$$\underbrace{W \cdots W}_{2N \text{ instances of } W}\quad . \tag{21}$$

Again, for a fixed δ and Δ, here N can be taken so large that **A3** holds. However, for X to enter the corresponding barrier set $\mathcal{X}^*$ infinitely often, the set (21) must have positive $\mu^{2N\cdot(n-1)+1}$ measure. This might be difficult to confirm for specific models. In fact, depending on ϵ, n and N, the set (21) might well be empty. But in many instances (21) does have a positive $\mu^{2N\cdot(n-1)+1}$-measure, regardless of the choice of N. The following example demonstrates this.

Example 5. Let the function $(x,x') \mapsto q(x,i|x',j)$ be continuous for all $i, j \in \mathcal{Y}$. Suppose there exists $x_{1:n} \in \mathcal{X}^n$, $n \geq 2$, such that

$$x_1 = x_n \quad \text{and} \quad p_{11}(x_{1:n}) > p_{ij}(x_{1:n}), \quad \forall (i,j) \in \mathcal{Y}^2 \setminus \{(1,1)\}.$$

Since $(x,x') \mapsto q(x,i|x',j)$ are continuous, then so must be maps $\mathcal{X}^n \ni x_{1:n} \mapsto p_{ij}(x_{1:n})$, and therefore there must exist open balls $B_1, \ldots, B_{n-1} \subset \mathcal{X}$ and $\epsilon > 0$ such that $x_{1:n} \in B_1 \times \cdots \times B_{n-1} \times B_1$ and for every $x_{1:n} \in B_1 \times \cdots \times B_{n-1} \times B_1$

$$p_{11}(x_{1:n})(1-\epsilon) > p_{ij}(x_{1:n}), \quad \forall (i,j) \in \mathcal{Y}^2 \setminus \{(1,1)\}.$$

Setting $B = B_1 \times \cdots \times B_{n-1}$, we have for arbitrary $N \geq 2$ that set

$$\underbrace{B \times \cdots \times B}_{2N \text{ blocks of } B} \times B_1$$

is a strong center part of a barrier set. Assuming that any open ball has positive μ-measure, this barrier set must have positive $\mu^{2N\cdot(n-1)+1}$-measure.

Proof of Theorem 2.1. Fix $x_{1:n_{2N+2}} \in \mathcal{X}^*$. We will show that **A1**–**A3** imply inequalities

$$p_{i1}(x_{1:n_{N+1}}) \geq p_{ij}(x_{1:n_{N+1}}), \quad \forall i, j \in \mathcal{Y}, \tag{22}$$

$$p_{1i}(x_{n_{N+1}:n_{2N+2}}) \geq p_{ji}(x_{n_{N+1}:n_{2N+2}}), \quad \forall i, j \in \mathcal{Y}. \tag{23}$$

We also show that if **A1**$'$ holds instead of **A1**, then either inequalities (22) or (23) are strict for all i and $j \neq 1$ for which the left side of the inequality is non-zero. Then the statement follows from Proposition 2.1. Denote

$$a_{ij}(0) = p_{ij}(x_{1:n_1}), \quad a_{ij}(k) = p_{ij}(x_{n_k:n_{k+1}}), \quad k = 1, \ldots, N.$$

We start by proving (22). If $i \notin \mathcal{Y}^+(x_{1:n_1})_{(1)}$, then (22) holds, because for every $j \in \mathcal{Y}$

$$p_{ij}(x_{1:n_{N+1}}) = \max_{y \in \mathcal{Y}} p_{iy}(x_{1:n_1}) p_{yj}(x_{n_1:n_{N+1}}) = \max_{y \in \mathcal{Y}} 0 \cdot p_{yj}(x_{n_1:n_{N+1}}) = 0.$$

Consider now the case where $i \in \mathcal{Y}^+(x_{1:n_1})_{(1)}$. Let $M = M(i)$ be the set of all vectors $y_{1:N+1}$ which maximize the expression

$$a_{iy_1}(0)\prod_{k=1}^{N} a_{y_k y_{k+1}}(k). \tag{24}$$

Hence for every $y_{1:2N+1} \in M$, (24) is equal to $\max_{j\in\mathcal{Y}} p_{ij}(x_{1:n_{N+1}})$. First we will prove that for any $y_{1:N+1} \in M$, $y_{1:N}$ contains at least one 1, i.e.

$$M \cap (\mathcal{Y} \setminus \{1\})^N \times \mathcal{Y} = \emptyset. \tag{25}$$

Assuming on contrary, we would have

$$\begin{aligned}
\max_{j\in\mathcal{Y}} p_{ij}(x_{1:n_{N+1}}) &= \max_{y_1,\dots,y_N\in\mathcal{Y}\setminus\{1\},\, y_{N+1}\in\mathcal{Y}} a_{iy_1}(0)\prod_{k=1}^{N} a_{y_k y_{k+1}}(k) \\
&\overset{\mathbf{A1,A2}}{\leq} \frac{a_{i1}(0)}{\delta} \cdot \Delta \cdot (1-\epsilon)^N \cdot \prod_{k=1}^{N} a_{11}(k) \\
&\overset{\mathbf{A3}}{<} a_{i1}(0)\prod_{k=1}^{N} a_{11}(k) \\
&\leq \max_{j\in\mathcal{Y}} p_{ij}(x_{1:n_{N+1}})
\end{aligned}$$

— a contradiction. Fix $y'_{1:N+1} \in M$ arbitrarily; as we saw, there must exist $u \in \{1,\dots,N\}$ such that $y'_u = 1$. Since

$$\begin{aligned}
\max_{j\in\mathcal{Y}} p_{ij}(x_{1:n_{N+1}}) &= a_{iy'_1}(0)\prod_{k=1}^{N} a_{y'_k y'_{k+1}}(k) \\
&\leq a_{iy'_1}(0)a_{y'_1y'_2}(1)\cdots a_{y'_{u-1}1}(u-1)\cdot a_{11}(u)a_{11}(u+1)\cdots a_{11}(N) \\
&\leq p_{i1}(x_{1:n_{N+1}}),
\end{aligned}$$

then (22) holds. The proof of inequalities (23) is symmetrical.

Finally, we need to show that if $\mathbf{A1}'$ holds, then either inequalities (22) or (23) are strict for all i and $j \neq 1$ for which the left side of the inequality is non-zero. For this it suffices to prove the following two claims:

(i) if inequalities (19) are strict for all $i \neq 1$, then inequalities (22) are strict for all $i \in \mathcal{Y}^+(x_{1:n_1})_{(1)}$ and $j \neq 1$;
(ii) if inequalities (18) are strict for all $i \neq 1$, then inequalities (23) are strict for all $i \in \mathcal{Y}^+(x_{n_{2N+1}:n_{2N+2}})_{(2)}$ and $j \neq 1$.

We only prove the first claim; the proof for the second claim is symmetrical. Let the inequalities (19) be strict for all $i \neq 1$. Let now again $i \in \mathcal{Y}^+(x_{1:n_1})_{(1)}$ and let $j \neq 1$. Note that by **A1** and **A2** $p_{i1}(x_{1:n_{N+1}}) > 0$. We show now that assumption

$$p_{i1}(x_{1:n_1}) = p_{ij}(x_{1:n_1}) \tag{26}$$

leads to contradiction. Indeed, assuming (26), we have that there exists sequence $y'_{1:N+1}$ which belongs to set $M(i)$ and for which $y'_{N+1} = j \neq 1$. Therefore by (25) there must exist

$u \in \{1, \ldots, N\}$ such that $y'_u = 1$. Then

$$\begin{aligned} p_{ij}(x_{1:n_{N+1}}) &= a_{iy'_1}(0) \prod_{k=1}^{N} a_{y'_k y'_{k+1}}(k) \\ &< a_{iy'_1}(0) a_{y'_1 y'_2}(1) \cdots a_{y'_{u-1}1}(u-1) \cdot a_{11}(u) a_{11}(u+1) \cdots a_{11}(N) \\ &\le p_{i1}(x_{1:n_{N+1}}). \quad \square \end{aligned}$$

Theorem 2.1 gives conditions for constructing the barrier set $\mathcal{X}^*$, but we also need to ensure that X enters into $\mathcal{X}^*$ infinitely often a.s. For this we will use the following

Proposition 2.2. *Let $\mathcal{X}^* \subset \mathcal{X}^M$ for some $M \ge 1$. If for some $A \subset \mathcal{Z}$ and $\epsilon > 0$ it holds*

$$P(Z_k \in A \text{ i.o.}) = 1,$$
$$P(X_{1:M} \in \mathcal{X}^* | Z_1 = z) \ge \epsilon, \quad \forall z \in A,$$

then $P(X \in \mathcal{X}^* \text{ i.o.}) = 1$.

Proof. Take $B = \{((x_1, y_1), \ldots, (x_M, y_M)) \mid x_{1:M} \in \mathcal{X}^*, y_{1:M} \in \mathcal{Y}^M\}$. Thus

$$P(Z_{1:M} \in B | Z_1 = z) = P(X_{1:M} \in \mathcal{X}^* | Z_1 = z) \ge \epsilon, \quad \forall z \in A.$$

From Lemma A.1 it follows that $P(Z \in B \text{ i.o.}) = 1$ which implies $P(X \in \mathcal{X}^* \text{ i.o.}) = 1$. $\square$

Harris chains and reachable points. We will now introduce some general state space Markov chain terminology. Markov chain Z is called *φ-irreducible* for some σ-finite measure φ on $\mathcal{B}(\mathcal{Z})$, if $\varphi(A) > 0$ implies $\sum_{k=2}^{\infty} P(Z_k \in A | Z_1 = z) > 0$ for all $z \in \mathcal{Z}$. If Z is φ-irreducible, then there exists (see [23, Prop. 4.2.2.]) a *maximal irreducibility measure* ψ in the sense that for any other irreducibility measure φ' the measure ψ dominates φ', $\psi \succ \varphi'$. The symbol ψ will be reserved to denote the maximal irreducibility measure of Z. A point $z \in \mathcal{Z}$ is called *reachable* if for every open neighborhood O of z,

$$\sum_{k=2}^{\infty} P(Z_k \in O | Z_1 = z') > 0, \quad \forall z' \in \mathcal{Z}.$$

For ψ-irreducible Z, the point z is reachable if and only if it belongs to the support of ψ [23, Lemma 6.1.4]. Since we have equipped space $\mathcal{Z}$ with product topology $\tau \times 2^{\mathcal{Y}}$, where τ denotes the topology induced by the metrics of $\mathcal{X}$, the above-stated definition of reachable point is actually equivalent to the following: point $(x, i) \in \mathcal{Z}$ is called *reachable*, if for every open neighborhood O of x,

$$\sum_{k=2}^{\infty} P(Z_k \in O \times \{i\} | Z_1 = z) > 0, \quad \forall z \in \mathcal{Z}.$$

Chain Z is called *Harris recurrent*, if it is ψ-irreducible and $\psi(A) > 0$ implies $P(Z_k \in A \text{ i.o.} | Z_1 = z) = 1$ for all $z \in \mathcal{Z}$.

The following lemma links the conditions of Proposition 2.2 to the conditions **A1–A2** and Harris recurrence of Z.

Lemma 2.1. *Let $\mathcal{X}^* \subset \mathcal{X}^M$ satisfy **A1** and **A2** and let Z be Harris recurrent. Moreover, assume that there exists $i \in \mathcal{Y}$ such that $i \in \mathcal{Y}^+(x_{1:n_1})_{(1)}$ for every $x_{1:n_1} \in \mathcal{X}^*_{(1,n_1)}$. Denote*

$$\mathcal{X}^*(x_1) := \{x_{2:M} \mid x_{1:M} \in \mathcal{X}^*\}, \quad x_1 \in \mathcal{X}^*_{(1)}.$$

If

$$\mu^{M-1}(\mathcal{X}^*(x_1)) > 0, \quad \forall x_1 \in \mathcal{X}^*_{(1)}, \tag{27}$$

and $\psi(\mathcal{X}^*_{(1)} \times \{i\}) > 0$, *then* $P(X \in \mathcal{X}^* \text{ i.o.}) = 1$.

Proof. By **A1**, **A2** and (27) we have for every $x_1 \in \mathcal{X}^*_{(1)}$

$$\begin{aligned}
&P(X_{1:n_{2N+2}} \in \mathcal{X}^* | Z_1 = (x_1, i)) \\
&= \int_{\mathcal{X}^*(x_1)} \sum_{y_{1:M}:\, y_1 = i} p(x_{2:M}, y_{2:M} | x_1, y_1)\, \mu^{M-1}(dx_{2:M}) \\
&\geq \int_{\mathcal{X}^*(x_1)} p_{i1}(x_{1:n_1}) \left(\prod_{k=2}^{2N+1} p_{11}(x_{n_{k-1}:n_k}) \right) \max_{j \in \mathcal{Y}} p_{1j}(x_{n_{2N+1}:n_{2N+2}})\, \mu^{M-1}(dx_{2:M}) \\
&> 0.
\end{aligned}$$

Thus there must exist $A \subset \mathcal{X}^*_{(1)} \times \{i\}$ and $\epsilon' > 0$ such that $\psi(A) > 0$ and $P(X \in \mathcal{X}^* | Z_1 = z) \geq \epsilon'$ for all $z \in A$. Since Z is Harris recurrent, then $P(Z_k \in A \text{ i.o.}) = 1$ and the statement follows from Proposition 2.2. □

3. Barrier set construction with lower semi-continuous transition densities

In Section 4.1 we will show how Theorem 2.1 can be used to derive simple and general conditions for the existence of infinite Viterbi path in case of HMM. For non-HMM's the situation may be more complex and proving **A1**′, **A2** and **A3** might be difficult. In the present section we derive some conditions which are easier to handle by assuming lower semi-continuity and boundedness of functions $(x, x') \mapsto q(x, j|x', i)$. In what follows, the first assumption **B1** is closely related to the condition **A2** and the second assumption **B2** guarantees the existence of a strong center part of a barrier set.

- **B1** There exists an open set $E \subset \mathcal{X}^q$, $q \geq 2$, such that $\mathcal{Y}^+ := \mathcal{Y}^+(x_{1:q})$ is the same for every $x_{1:q} \in E$ and satisfies the following property: $(i, j) \in \mathcal{Y}^+$ for every $i \in \mathcal{Y}^+_{(1)}$ and $j \in \mathcal{Y}^+_{(2)}$. Furthermore, we assume that there exists a reachable point (x_E, i_E) in $E_{(1)} \times \mathcal{Y}^+_{(1)}$.
- **B2** For arbitrary $N \geq 2$ there exists a strong center part of a barrier set $\mathcal{X}^*_{(n_1, n_{2N+1})}$ which is open, non-empty and has $2N$ cycles. We assume that both set $\mathcal{X}^*_{(n_1)}$ and parameter ϵ of **A1** are independent of N, and there exists a compact set $K \subset \mathcal{X}$, which is independent of N, such that $\mathcal{X}^*_{(n_{2N+1})}$ is contained in K. Furthermore, we assume that there exists $x^* \in \mathcal{X}^*_{(n_1)}$ such that $(x^*, 1)$ is reachable.

Theorem 3.1. *Let μ be strictly positive*[1] *and let for every pair of states $i, j \in \mathcal{Y}$ function $(x, x') \mapsto q(x, i|x', j)$ be lower semi-continuous and bounded. If Z satisfies* **B1** *and* **B2**, *then there exists $\mathcal{X}^*$ satisfying* **A1**′, **A2** *and* **A3**. *Moreover, if Z is Harris recurrent, then $P(X \in \mathcal{X}^* \text{ i.o.}) = 1$.*

Before proving the theorem, let us briefly discuss its assumptions. Under ψ-irreducibility (that is implied by Harris recurrence) the existence of certain reachable points is not restrictive, because any point in the support of ψ is reachable. Thus **B1** is merely to guarantee that $\mathcal{Y}^+ = \mathcal{Y}^+_{(1)} \times \mathcal{Y}^+_{(2)}$. (we have already noted that this need not hold in general). To get some

[1] A measure is called *strictly positive* if it assigns a positive measure to all non-empty open sets.

insight into this assumption, consider an inhomogeneous Markov chain $Y = \{Y_t\}_{t\geq 1}$, with $\mathcal{Y}_t$ being the finite state space of Y_t. The canonical concepts of irreducibility and aperiodicity are not defined for such a Markov chain, but a natural generalization would be the following: for every time t, there exists a time $n > t$ such that $P(Y_n = j|Y_t = i) > 0$ for every $i \in \mathcal{Y}_t$ and $j \in \mathcal{Y}_n$. If Y is homogeneous, then this property implies that Y is irreducible and aperiodic, hence also geometrically ergodic. When we fix $t < n$, and define

$$\mathcal{Y}^+ = \{(i, j) : i \in \mathcal{Y}_t, j \in \mathcal{Y}_n^+, P(Y_n = j|Y_t = i) > 0\},$$

then the above-stated condition reads $\mathcal{Y}^+ = \mathcal{Y}_t \times \mathcal{Y}_n$. Suppose now that there exists a piece of observations $x_{1:q}$ such that when $x_{t:\infty}$ is a realization of $X_{t:\infty}$ beginning with $x_{1:q}$ (i.e. satisfying $x_{t:t+q-1} = x_{1:q}$), it holds that $\mathcal{Y}^+ = \mathcal{Y}^+_{(1)} \times \mathcal{Y}^+_{(2)}$, where $\mathcal{Y}^+ = \{(i, j) : P(Y_{t+q-1} = j|Y_t = i, X_{t:\infty} = x_{t:\infty}) > 0\}$ and

$$\mathcal{Y}^+_{(1)} = \{i : \exists j(i), \quad \text{such that } P(Y_{t+q-1} = j|Y_t = i, \quad X_{t:\infty} = x_{t:\infty}) > 0\}$$
$$\mathcal{Y}^+_{(2)} = \{j : \exists i(j) \quad \text{such that } P(Y_{t+q-1} = j|Y_t = i, \quad X_{t:\infty} = x_{t:\infty}) > 0\}.$$

The assumption **B1** ensures that a.e. realization of X contains infinitely many such pieces and therefore the above-described generalization of irreducibility and aperiodicity is satisfied for the conditional signal process $P(Y_{1:\infty} \in \cdot|X_{1:\infty} = x_{1:\infty})$.

It also turns out that for stationary Z assumption **B1** is closely related to *subpositivity* property of factor maps in ergodic theory. We shall return to that connection and also discuss the necessity of **B1** in Section 4.2.

Assumption **B2** provides some necessary assumptions for cycle-construction. Typically the center part of a barrier set is constructed by glueing together some fixed cycles and **B2** basically guarantees that no matter how many cycles are connected, the last one always ends in a fixed compact set K. As we shall see in the examples, this condition holds for many models.

Proof of Theorem 3.1 (*Lower Likelihood Bound for Vectors in E*)**.** We will show that with no loss of generality we may assume that there exist $\delta_0 > 0$ such that

$$p_{ij}(x_{1:q}) > \delta_0, \quad \forall x_{1:q} \in E, \quad \forall (i, j) \in \mathcal{Y}^+. \tag{28}$$

Let $x'_{1:q} \in E$ be such that $x'_1 = x_E$. Denote $\delta_0 = \frac{1}{2}\min_{(i,j)\in\mathcal{Y}^+} p_{ij}(x'_{1:q})$. Since $\mathcal{Y}^+(x'_{1:q}) = \mathcal{Y}^+$, then $\delta_0 > 0$. Denote

$$E' := \left\{x_{1:q} \in E \mid \min_{(i,j)\in\mathcal{Y}^+} p_{ij}(x_{1:q}) > \delta_0\right\}$$

It is not difficult to confirm by induction that lower semi-continuity and boundedness of $(x, x') \mapsto q(x, i|x', j)$ implies lower semi-continuity of functions $x_{1:n} \mapsto p(x_{2:n}, y_{2:n}|x_1, y_1)$ for all $n \geq 2$. Hence for all $i, j \in \mathcal{Y}$ the function $x_{1:q} \mapsto p_{ij}(x_{1:q})$ is lower semi-continuous since it expresses as a maximum over lower semi-continuous functions. Therefore the function $x_{1:q} \mapsto \min_{(i,j)\in\mathcal{Y}^+} p_{ij}(x_{1:q})$ must also be lower semi-continuous, and so E' must be open. Also $x_E \in E'_{(1)}$. Therefore E' could play the role of E and so there is no loss of generality in assuming that (28) holds true.

Construction of set D_1. Recall the element x_E from **B1**. Next we will show that there exist $l_1 > q$, $\delta_1 > 0$ and an open set $D_1 \subset \mathcal{X}^{l_1}$ such that $x_E \in D_{1(1)}$, $D_{1(l_1)} \subset \mathcal{X}^*_{(n_1)}$ and

$$\mathcal{Y}^+(x_{1:l_1}) = \mathcal{Y}^+, \quad \forall x_{1:l_1} \in D_1, \tag{29}$$

$$p_{i1}(x_{1:l_1}) \geq \delta_1, \quad \forall i \in \mathcal{Y}^+_{(1)}, \quad \forall x_{1:l_1} \in D_1. \tag{30}$$

The function of the set D_1 is to connect (the end of) the set E to (the beginning of) the strong center part $\mathcal{X}^*_{(n_1,n_{2N+1})}$.

Fix $x'_{1:q} \in E$ such that $x'_1 = x_E$. Also fix $j' \in \mathcal{Y}^+_{(2)}$. By **B2** set $\mathcal{X}^*_{(n_1)}$ is open (projection is an open map) and contains an element x^* such that $(x^*, 1)$ is reachable, so there must $k \geq 1$ such that $P(Z_{k+1} \in \mathcal{X}^*_{(n_1)} \times \{1\} | Z_1 = (x'_q, j')) > 0$. This implies that there exists $x'_{q+1:q+k} \in \mathcal{X}^k$ such that $x'_{q+k} \in \mathcal{X}^*_{(n_1)}$ and $\epsilon_0 := p_{j'1}(x'_{q:q+k}) > 0$. We define

$$D_1 = \left\{ x_{1:q+k} \mid \min_{i \in \mathcal{Y}^+_{(1)}} p_{i1}(x_{1:q+k}) > \delta_0 \epsilon_0,\ x_{1:q} \in E,\ x_{q+k} \in \mathcal{X}^*_{(n_1)} \right\},$$

where δ_0 is the constant from (28). Set D_1 is open by the fact that function $x_{1:q+k} \mapsto \min_{i \in \mathcal{Y}^+_{(1)}} p_{i1}(x_{1:q+k})$ is lower semi-continuous and both E and $\mathcal{X}^*_{(n_1)}$ are open. Note that inequality (30) is satisfied with $\delta_1 = \delta_0 \epsilon_0$. When $x_{1:q+k} \in D_1$, then

- $\mathcal{Y}^+(x_{1:q+k})_{(1)} \supset \mathcal{Y}^+_{(1)}$ by the fact that $\min_{i \in \mathcal{Y}^+_{(1)}} p_{i1}(x_{1:q+k}) > 0$;
- $\mathcal{Y}^+(x_{1:q+k})_{(1)} \subset \mathcal{Y}^+_{(1)}$, because by **B1** $\mathcal{Y}^+(x_{1:q}) = \mathcal{Y}^+$.

Hence (29) holds. By **B1** $(i, j') \in \mathcal{Y}^+$ for every $i \in \mathcal{Y}^+_{(1)}$. Consider the tuple $x'_{1:q+k}$, where $x'_{1:q}$ and $x'_{q+1:q+k}$ are defined as previously. By (28) we have for every $i \in \mathcal{Y}^+_{(1)}$ that

$$p_{i1}(x'_{1:q+k}) \geq p_{ij'}(x'_{1:q}) p_{j'1}(x'_{q:q+k}) > \delta_0 \epsilon_0.$$

This implies that $x'_{1:q+k} \in D_1$ and so $x_E = x'_1 \in D_{1(1)}$, as required.

Construction of sets $D_2(x)$. The function of D_2 is to connect (the end of) the strong center part $\mathcal{X}^*_{(n_1,n_{2N+1})}$ to (the beginning of) the set E.

Recall now the compact set K from **B2**. We will show that there exists a constant $\delta_2 > 0$ such that the following holds: for every $x \in K$ there exists $l_2(x) > q$ and a non-empty open set $D_2(x) \subset \mathcal{X}^{l_2}$ such that

$$\mathcal{Y}^+(x, x_{1:l_2})_{(2)} = \mathcal{Y}^+_{(2)}, \quad \forall x_{1:l_2} \in D_2(x), \tag{31}$$

$$p_{1j}(x, x_{1:l_2}) \geq \delta_2, \quad \forall j \in \mathcal{Y}^+_{(2)}, \quad \forall x_{1:l_2} \in D_2(x). \tag{32}$$

Denote for $s \geq 2$

$$p_s(x) := P(Z_s \in E_{(1)} \times \{i_E\} | Z_1 = (x, 1))$$

and $G_s = \{x \in \mathcal{X} \mid p_s(x) > 0\}$. Functions $x \mapsto p_s(x)$ are lower semi-continuous by Fatou's Lemma and lower semi-continuity of functions $x \mapsto p(x_{2:s}, y_{2:s} | x_1 = x, y_1 = 1)$, and so the sets G_s must be open. By **B1** set $E_{(1)}$ is open (projection is an open map) and contains an element x_E such that (x_E, i_E) is reachable for an $i_E \in \mathcal{Y}^+_{(1)}$. This implies that sets G_s form an open cover of compact set K. Hence there exists an $s_0 \geq 2$ such that $K \subset \cup_{s=2}^{s_0} G_s$.

Define now

$$h_s(x) := \sup_{x_{2:s} \in \mathcal{X}^{s-2} \times E_{(1)}} p(x_{2:s}, y_s = i_E | z_1 = (x, 1)),$$

$$s(x) := \arg\max_{s \in \{2,\dots,s_0\}} h_s(x)$$

and

$$\epsilon(x) := \frac{1}{2} \max_{s \in \{2,\dots,s_0\}} h_s(x).$$

Note that when $x \in K$, then $\epsilon(x) > 0$. Indeed, when $x \in K$ then there exists $s \in \{2, \dots, s_0\}$ such that $x \in G_s$. Hence $p_s(x) > 0$, and so $h_s(x) > 0$. This implies that $\epsilon(x) > 0$. Since $\epsilon(x)$ is lower semicontinuous and K is compact, we have that $\min_{x \in K} \epsilon(x) > 0$.

Denote

$$F(x) := \{x_{2:s(x)} \mid p(x_{2:s(x)}, y_{s(x)} = i_E | z_1 = (x, 1)) > \epsilon(x)\} \cap \mathcal{X}^{s(x)-2} \times E_{(1)}.$$

Functions $x_{2:s} \mapsto p(x_{2:s}, y_s = i_E | z_1 = (x, 1))$ must be lower semi-continuous, since they express as a finite sum of bounded lower semi-continuous functions. Therefore the sets $F(x)$ must be open. Also, when $x \in K$, then, as we saw, $\epsilon(x) > 0$, and so $F(x)$ is non-empty. We define

$$l_2(x) := s(x) + q - 2,$$
$$D_2(x) := F(x) \cdot E,$$

where operator $\cdot$ is defined in (20). Set $D_2(x)$ must be open, as it can be expressed as an intersection of two open sets:

$$D_2(x) = \mathcal{X}^{s(x)-2} \times E \cap F(x) \times \mathcal{X}^{q-1}.$$

Set $D_2(x)$ is also non-empty for all $x \in K$ by the fact that sets $F(x)$ are non-empty and by definition of sets $F(x)$ and $D_2(x)$.

Next, we prove the existence of $\delta_2 > 0$. Fix $x \in K$. Set $l = l_2(x)$ and $s = s(x)$ and let $x_{2:l+1} \in D_2(x)$. By definition of $F(x)$,

$$\sum_{y_{2:s-1}} p(x_{2:s}, y_{2:s-1}, y_s = i_E | z_1 = (x, 1)) > \epsilon(x),$$

and so

$$p_{1 i_E}(x, x_{2:s}) > |\mathcal{Y}|^{-s} \epsilon(x) \geq |\mathcal{Y}|^{-s_0} \epsilon(x) > 0. \tag{33}$$

Fix $j \in \mathcal{Y}^+_{(2)}$. Thus we have by **B1** that $(i_E, j) \in \mathcal{Y}^+$. Therefore by (33) and (28)

$$p_{1j}(x, x_{2:l+1}) \geq p_{1 i_E}(x, x_{2:s}) p_{i_E j}(x_{s:l+1}) \geq |\mathcal{Y}|^{-s_0} \epsilon(x) \delta_0 > 0. \tag{34}$$

Thus $\mathcal{Y}^+(x, x_{2:l+1})_{(2)} \supset \mathcal{Y}^+_{(2)}$; since also $\mathcal{Y}^+(x, x_{2:l+1})_{(2)} \subset \mathcal{Y}^+_{(2)}$, then (31) must hold. It follows from (34) that (32) holds with $\delta_2 = |\mathcal{Y}|^{-s_0} \min_{x \in K} \epsilon(x) \cdot \delta_0 > 0$.

Construction of $\mathcal{X}^*$. By the boundedness assumption there exists $\Delta_0 > 1$ such that $q(z|z') \leq \Delta_0$ for all $z, z' \in \mathcal{Z}$. Denote $l_{\max} := l_1 \vee (s_0 + q - 2)$, where $\vee$ denotes maximum. Hence $l_1 \vee \max_{x \in K} l_2(x) \leq l_{\max}$. We take $\Delta := \Delta_0^{l_{\max}}$ and $\delta := \delta_1 \wedge \delta_2$, where $\wedge$ denotes minimum. Take now $N \geq 2$ so large that **A3** holds — this is possible because according to **B2** set $\mathcal{X}^*_{(n_1)}$, ϵ and K are all independent of N.

We note that the set $D_1 \cdot \mathcal{X}^*_{(n_1, n_{2N+1})}$ is open and non-empty, since D_1 is open and non-empty by construction, $\mathcal{X}^*_{(n_1, n_{2N+1})}$ is open and non-empty by **B2** and $D_{1(l_1)} \subset \mathcal{X}^*_{(n_1)} \neq \emptyset$ by construction of D_1. Hence, taking $n_1 = l_1$, there exist open balls $(B_k)_{k=1}^{n_{2N+1}}$ in $\mathcal{X}$ such that $x_E \in B_1$ and

$$B_1 \times \cdots \times B_{n_{2N+1}} \subset D_1 \cdot \mathcal{X}^*_{(n_1, n_{2N+1})}. \tag{35}$$

Denote $\mathcal{X}(l) := \{x \in \mathcal{X} \mid l_2(x) = l\}$. The sets $(\mathcal{X}(l))_{l=1}^{l_{\max}}$ form a finite cover of $\mathcal{X}$. Therefore by the assumption that measure μ is strictly positive there must exist positive integer $l_2 \leq l_{\max}$ such that denoting $\mathcal{X}_0 := B_{n_{2N+1}} \cap \mathcal{X}(l_2)$, we have

$$\mu(\mathcal{X}_0) > 0. \tag{36}$$

Take $D_2 := \cup_{x\in\mathcal{X}_0}\{x\}\times D_2(x)$ and

$$\mathcal{X}^* := B_1 \times \cdots \times B_{n_{2N+1}-1} \times D_2.$$

Then **A1**′ is satisfied with $n_1 = l_1$ and $n_{2N+2} - n_{2N+1} = l_2$ by the fact that $\mathcal{X}^*_{(n_1,n_{2N+1})}$ is a strong center part of a barrier set (**B2**) and by (35). **A2** is satisfied by (29), (30) (31), (32) and (35).

Let now Z be Harris recurrent. To complete the proof it suffices to show that $P(X \in \mathcal{X}^* \text{ i.o.}) = 1$. To prove this, we show that the assumptions of Lemma 2.1 are fulfilled. We start with proving the assumption

$$\mu^{M-1}(\mathcal{X}^*(x_1)) > 0, \quad \forall x_1 \in \mathcal{X}^*_{(1)}, \tag{37}$$

where we define $M = n_{2N+2}$ and

$$\mathcal{X}^*(x_1) := \{x_{2:M} \mid x_{1:M} \in \mathcal{X}^*\}, \quad x_1 \in \mathcal{X}^*_{(1)}.$$

First we note that set D_2 is measurable. Indeed, setting $s = l_2 - q + 2$, set D_2 expresses as

$$\big(\cup_{x_1\in\mathcal{X}_0}\{x_1\}\times F(x_1)\big)\cdot E = \big(\{x_{1:s} \mid p(x_{2:s}, y_s = i_E|x_1, y_1 = 1) > \epsilon(x_1)\} \cap \mathcal{X}_0 \times \mathcal{X}^{s-1}\big)\cdot E.$$

The function $x_{1:s} \mapsto p(x_{2:s}, y_s = i_E|x_1, y_1 = 1) - \epsilon(x_1)$ is measurable, so D_2 must be measurable. Next, note that $\mu^{l_2}(D_2(x)) > 0$ for all $x \in \mathcal{X}_0 \subset K$ by the fact that sets $D_2(x)$ are by construction open and non-empty. Together with (36) the observations above imply that

$$\mu^{l_2+1}(D_2) = \int_{\mathcal{X}_0}\int_{D_2(x_1)} \mu^{l_2}(dx_{2:l_2+1})\,\mu(dx_1) > 0$$

which in turn implies (37).

Since Z is Harris recurrent, then it is by definition ψ-irreducible. To prove the rest of the assumptions of Lemma 2.1, it suffices to show that

$$i_E \in \mathcal{Y}^+(x_{1:n_1})_{(1)}, \quad \forall x_{1:n_1} \in \mathcal{X}^*_{(1,n_1)} \tag{38}$$

and

$$\psi(\mathcal{X}^*_{(1)} \times \{i_E\}) > 0. \tag{39}$$

By (29) $\mathcal{Y}^+_{(1)}(x_{1:n_1}) = \mathcal{Y}^+_{(1)}$ for every $x_{1:n_1} \in \mathcal{X}^*_{(1,n_1)} = \mathcal{X}^*_{(1,l_1)}$; also by **B1** $i_E \in \mathcal{Y}^+_{(1)}$, so (38) holds. Since point (x_E, i_E) is reachable by **B1**, then this point belongs to the support of measure ψ. Since $\mathcal{X}^*_{(1)} \times \{i_E\}$ is an open neighborhood of (x_E, i_E) (recall that $x_E \in \mathcal{X}^*_{(1)} = B_1$), then (39) holds by definition of measure support. □

4. Examples

4.1. Hidden Markov model

For HMM, Theorem 2.1 allows us to deduce a generalized version of Theorem 1.1. Recall the definitions of G_i(11). We introduce a new term obtained by weakening the cluster condition (12): a subset $C \subset \mathcal{Y}$ is called a *weak cluster*, if

$$\mu\,[(\cap_{i\in C} G_i) \setminus (\cup_{i\notin C} G_i)] > 0.$$

Observe that every state i belongs to at least one cluster. The result for HMM is the following:

Corollary 4.1. *Suppose Z is HMM satisfying the following conditions.*

(i) For each state $j \in \mathcal{Y}$

$$\mu\left(\left\{x \in \mathcal{X} \mid f_j(x)p_{\cdot j} > \max_{i\in\mathcal{Y},\ i\neq j} f_i(x)p_{\cdot i}\right\}\right) > 0, \quad \text{where } p_{\cdot j} := \max_{i\in\mathcal{Y}} p_{ij}.$$

(ii) There exists a weak *cluster $C \subset \mathcal{Y}$ such that the sub-stochastic matrix $\mathbb{P}_C = (p_{ij})_{i,j\in C}$ is primitive in the sense that $\mathbb{P}_C^R$ consists of only positive elements for some positive integer R.*

Also let Markov chain Y be irreducible. Then there exist $i \in \mathcal{Y}$ and a barrier set $\mathcal{X}^*$ consisting of strong i-barriers of fixed order and satisfying $P(X \in \mathcal{X}^* \text{ i.o.}) = 1$.

Compared to Theorem 1.1 we have removed the assumption of stationarity of Z and aperiodicity of Y, and replaced the assumption that C is cluster with a substantially weaker assumption that it is a weak cluster. Also, the result above guarantees the existence of infinitely many *strong* nodes, instead of just nodes like in Theorem 1.1.

Proof of Corollary 4.1. Fix $j_1 \in \mathcal{Y}$. Denote for every $k \geq 2$

$$j_k := \arg\max_{j\in\mathcal{Y}} p_{jj_{k-1}}.$$

There must exist integers u and v, $u < v$, such that $j_u = j_v$. We denote $n = v - u + 1$ and $i_{1:n} = (j_v, j_{v-1}, \dots, j_u)$. If needed, we will re-label the elements of $\mathcal{Y}$ so that $i_1 = i_n = 1$. By (i) there must exist $\epsilon > 0$ such that $\mu(A_j) > 0$ for each state $j \in \mathcal{Y}$, where

$$A_j := \left\{x \in \mathcal{X} \mid f_j(x)p_{\cdot j}(1-\epsilon) > \max_{i\in\mathcal{Y},\ i\neq j} f_i(x)p_{\cdot i}\right\}.$$

Denote $A := A_{i_2} \times A_{i_3} \times \dots \times A_{i_n}$. Next we show that for every $x_{1:n} \in \mathcal{X} \times A$

$$p_{11}(x_{1:n}) \geq p_{i1}(x_{1:n}), \quad \forall i \in \mathcal{Y}, \tag{40}$$

$$p_{11}(x_{1:n}) > p_{1i}(x_{1:n}), \quad \forall i \in \mathcal{Y}\setminus\{1\}, \tag{41}$$

$$p_{11}(x_{1:n})(1-\epsilon) > p_{ij}(x_{1:n}), \quad \forall i, j \in \mathcal{Y}\setminus\{1\}. \tag{42}$$

Indeed, by construction of $i_{1:n}$ and A, for any path $y_{1:n}$ for which $y_{2:n} \neq i_{2:n}$ and for any $x_{2:n} \in A$ we have

$$(1-\epsilon)\prod_{k=2}^{n} p_{i_{k-1}i_k} f_{i_k}(x_k) > \prod_{k=2}^{n} p_{y_{k-1}y_k} f_{y_k}(x_k).$$

On the other hand, for *any* path $y_{1:n}$ and any $x_{2:n} \in A$

$$\prod_{k=2}^{n} p_{i_{k-1}i_k} f_{i_k}(x_k) \geq \prod_{k=2}^{n} p_{y_{k-1}y_k} f_{y_k}(x_k).$$

Thus the inequalities (40), (41) and (42) must hold.

Note now that there must exist $0 < \delta_0 \leq \Delta_0 < \infty$ such that, defining $G_i^0 := \{x \in \mathcal{X} \mid \delta_0 \leq f_i(x);\ f_j(x) \leq \Delta_0, j \in \mathcal{Y}\}$, we have $\mu(G_i^0) > 0$ for every state $i \in \mathcal{Y}$. Furthermore, denoting $G := \left(\cap_{i\in C} G_i^0\right) \setminus (\cup_{i\notin C} G_i)$, by cluster assumption we may with no loss of generality assume that δ_0 is so small and Δ_0 is so large that $\mu(G) > 0$. Fix $j' \in C$. By irreducibility assumption there exists path $u_{1:K}$, $K \geq 2$, such that $u_1 = j'$, $u_K = 1$ and $p_{u_{k-1}u_k} > 0$ for all $k = 2, \dots, K$. Similarly, there exists path $v_{1:L}$, $L \geq 3$, such that $v_1 = 1$, $v_L = j'$ and $p_{v_{k-1}v_k} > 0$ for all $k = 2, \dots, L$. Denote $H_1 := G_{u_2}^0 \times \dots \times G_{u_K}^0$, $H_2 := G_{v_2}^0 \times \dots \times G_{v_L}^0$ and

$p^* := \min\{p_{ij} \mid p_{ij} > 0,\ i, j \in \mathcal{Y}\}$. With no loss of generality we may assume that $\delta_0 < 1$ and $\Delta_0 > 1$. Denote $M := K \vee L + R + 1$, where $\vee$ denotes maximum, and set $\delta := (p^*\delta_0)^M$, $\Delta := \Delta_0^M$ and $N \geq 2$ so big that **A3** holds. Take

$$\mathcal{X}^* := \mathcal{X} \times G^{R+1} \times H_1 \times A^{2N} \times H_2 \times G^R,$$

$n_1 = R + 1 + K$, $n_k = n_{k-1} + n - 1$ for $k = 2, \ldots, 2N + 1$, and $n_{2N+2} = n_{2N+1} + L - 1 + R$. By (40), (41) and (42) **A1**$'$ holds.

Next, we will prove **A2**. First note that by definition of sets G_i^0,

$$p_{ij}(x_{1:n_1}),\ p_{ij}(x_{n_{2N+1}:n_{2N+2}}) \leq \Delta, \quad \forall i, j \in \mathcal{Y}, \quad \forall x_{1:n_1} \in \mathcal{X}^*_{(1,n_1)},$$
$$\forall x_{n_{2N+1}:n_{2N+2}} \in \mathcal{X}^*_{(n_{2N+1},n_{2N+2})}.$$

Next, denote $\mathcal{Y}_C := \{i \in \mathcal{Y} | p_{ij} > 0,\ j \in C\}$. Note that by definition of set G, $\mathcal{Y}^+(x_{1:n_1})_{(1)} \subset \mathcal{Y}_C$ for all $x_{1:n_1} \in \mathcal{X}^*_{(1,n_1)}$. By the primitiveness of $\mathbb{P}_C$, we have for all $x_{1:n_1} \in \mathcal{X}^*_{(1,n_1)}$ and any $i \in \mathcal{Y}_C$

$$p_{i1}(x_{1:n_1}) \geq \max_{y_{1:n_1}\,:\, y_1 = i,\ y_{2:R+1} \in C^R,\ y_{R+2:n_1} = u_{1:K}} p(x_{1:n_1}, y_{1:n_1}) \geq \delta > 0.$$

Also note that by definition of sets G, $\mathcal{Y}^+(x_{n:n'})_{(2)} \subset C$ for all $x_{n:n'} \in \mathcal{X}^*_{(n_{2N+1},n_{2N+2})}$, where we denote $n := n_{2N+1}$ and $n' := n_{2N+2}$. By the primitiveness of $\mathbb{P}_C$ for all $x_{n:n'} \in \mathcal{X}^*_{(n,n')}$ and any $j \in C$, it holds

$$p_{1j}(x_{n:n'}) \geq \max_{y_{n:n'}\,:\, y_{n:n+L-1} = v_{1:L},\ y_{n+L:n'} \in C^R,\ y_{n'} = j} p(x_{n:n'}, y_{n:n'}) \geq \delta > 0.$$

The arguments above show that **A2** must hold and that

$$\mathcal{Y}^+(x_{1:n_1})_{(1)} = \mathcal{Y}_C, \quad \forall x_{1:n_1} \in \mathcal{X}^*_{(1,n_1)}. \tag{43}$$

From (43), Lemmas 2.1 and A.2 it follows that $P(X \in \mathcal{X}^* \text{ i.o.}) = 1$. □

4.2. Discrete $\mathcal{X}$

Consider the case where $\mathcal{X}$ is discrete (finite or countable) and Z is an irreducible and recurrent Markov chain with (discrete) state space $\mathcal{Z}' \subset \mathcal{X} \times \mathcal{Y}$. Here the state-space refers to the set of possible values of Z. Note that $\mathcal{Z}'$ can be a proper subset $\mathcal{X} \times \mathcal{Y}$. Also note: since the transition kernel $q(z|z')$ is defined on $\mathcal{Z}'$, the definition of $\mathcal{Y}^+(x_{1:q})$ immediately implies that $(i, x_1) \in \mathcal{Z}'$ for every $i \in \mathcal{Y}^+(x_{1:q})_{(1)}$. The following simple result can be derived from Theorem 3.1.

Corollary 4.2. *Let $\mathcal{X}$ be discrete and let Z be an irreducible and recurrent Markov chain with the state-space $\mathcal{Z}' \subset \mathcal{X} \times \mathcal{Y}$. Then the following conditions ensure that there exists a barrier set $\mathcal{X}^*$ consisting of strong 1-barriers of fixed order and satisfying $P(X \in \mathcal{X}^* \text{ i.o.}) = 1$.*

- *(i) There exists $q \geq 2$ and a sequence $x_{1:q} \in \mathcal{X}^q$ such that $\mathcal{Y}^+(x_{1:q})_{(1)}$ is non-empty and $(i, j) \in \mathcal{Y}^+(x_{1:q})$ for every $i \in \mathcal{Y}^+(x_{1:q})_{(1)}$ and $j \in \mathcal{Y}^+(x_{1:q})_{(2)}$.*
- *(ii) There exists $n \geq 2$ and $x^*_{1:n} \in \mathcal{X}^n$ such that $(x^*_1, 1) \in \mathcal{Z}'$ and*
  1. *it holds*

$$x^*_1 = x^*_n \quad \text{and} \quad p_{11}(x^*_{1:n}) > p_{ij}(x^*_{1:n}), \quad \forall i, j \in \mathcal{Y} \setminus \{1\};$$

2. *it holds*

$$p_{11}(x^*_{1:n}) > p_{i1}(x^*_{1:n}), \quad \forall i \in \mathcal{Y}, \tag{44}$$

$$p_{11}(x^*_{1:n}) > p_{1i}(x^*_{1:n}), \quad \forall i \in \mathcal{Y}, \tag{45}$$

where either inequalities (44) or inequalities (45) could be non-strict.

Proof. The proof is straightforward application of Theorem 3.1. To formally apply Theorem 3.1, Z should be viewed as a Markov chain on product space $\mathcal{Z} = \mathcal{Z} \times \mathcal{Y}$. In that perspective Z may no longer be irreducible. However with no loss of generality we may assume that Z is ψ-irreducible and Harris recurrent, where support of ψ is $\mathcal{Z}'$. Indeed, this can be achieved by fixing $(x', i') \in \mathcal{Z}'$ and taking $q(x', i'|z) = 1$ for all $z \in \mathcal{Z} \setminus \mathcal{Z}'$. Then all elements of $\mathcal{Z}'$ are reachable. Also, assuming with no loss of generality that $x' \neq x_2$, we have that $\mathcal{Y}^+(x_{1:q})$ is the same regardless if it is defined on the product space $\mathcal{Z}$ or subspace $\mathcal{Z}'$. Next, simply take $E = \{x_{1:q}\}$ so that **B1** holds with $E_{(1)} = \{x_1\}$ and i_E being any element of $\mathcal{Y}^+(x_{1:q})_{(1)}$. To see that **B2** holds, denote $x^* = x^*_{1:n-1}$ and note that for arbitrary $N \geq 2$ the strong center part of the barrier set can be taken to be

$$\mathcal{X}^*_{(n_1:n_{2N+1})} = \{(\underbrace{x^*, x^*, \ldots, x^*}_{2N \text{ blocks of } x^*}, x^*_n)\}.$$

In the discrete case measure μ is counting measure on $2^{\mathcal{X}}$, which is strictly positive, and the functions $(x', x) \mapsto q(x, j|x', i)$ are always continuous and bounded. Therefore Theorem 3.1 applies. □

Remarks about the condition (i).

1. If Z is stationary MC, then the set $\mathcal{Y}^+(x_{1:q})_{(1)}$ consists of states i satisfying the following property: there exists $y_{1:q} \in \mathcal{Y}^q$ such that $y_1 = i$ and $p(x_{1:q}, y_{1:q}) > 0$. Similarly $\mathcal{Y}^+(x_{1:q})_{(2)}$ consists of states j satisfying the following property: there exists $y_{1:q} \in \mathcal{Y}^q$ such that $y_q = j$ and $p(x_{1:q}, y_{1:q}) > 0$. However, given $i \in \mathcal{Y}^+(x_{1:q})_{(1)}$ and $j \in \mathcal{Y}^+(x_{1:q})_{(2)}$, there need not necessary be any path $y_{1:q}$ beginning with i (i.e. $y_1 = i$) and ending with j (i.e. $y_q = j$) such that $p(x_{1:q}, y_{1:q}) > 0$. The condition (i) ensures that for every pair $i \in \mathcal{Y}^+(x_{1:q})_{(1)}$ and $j \in \mathcal{Y}^+(x_{1:q})_{(2)}$ such a path exists and then $(i, j) \in \mathcal{Y}^+(x_{1:q})$. Interestingly, in ergodic theory, this property is the same as the *subpositivity* of the word $x_{1:q}$ for *factor map* $\pi : \mathcal{Z} \to \mathcal{X}, \pi(x, y) = x$, see ([27], Def 3.1). Thus (i) ensures that a.e. realization of X process has infinitely many subpositive words.
2. Let us now argue that for stationary Z, the subpositivity is also very close to be a necessary property of a barrier. Indeed, if $x_{k:l}$ $(1 < k < l < n)$ is a barrier containing a strong 1-node, then for any Viterbi path $v(x_{1:n})$, $(v_k, v_l) \in \mathcal{Y}^+(x_{k:l})$. Suppose now there exists another words of observations $x'_{1:k-1}$ and $x'_{l+1:n}$ such that the corresponding Viterbi path $v' = v(x'_{1:k-1}, x_{k:l}, x'_{l+1:n})$ satisfies: $v'_k \neq v_k$ and $v'_l \neq v_l$. Then also $(v'_k, v'_l) \in \mathcal{Y}^+(x_{k:l})$. Take now $v'_k \in \mathcal{Y}^+(x_{k:l})_{(1)}$ and $v_l \in \mathcal{Y}^+(x_{k:l})_{(2)}$ and ask: does $(v'_k, v_l) \in \mathcal{Y}^+(x_{k:l})$? Since $x_{k:l}$ is a barrier containing a strong 1-node, then by piecewise construction there exists a Viterbi path $w = v(x'_{1:k-1}, x_{k:n})$ such that $w_k = v'_k$ and $w_l = v_l$ and so $(v'_k, v_l) \in \mathcal{Y}^+(x_{k:l})$. We have seen that if $i \in \mathcal{Y}^+(x_{k:l})_{(1)}$ is such that for some $x'_{1:k-1}$, $v_k(x'_{1:k-1}, x_{k:n}) = i$ and if $j \in \mathcal{Y}^+(x_{k:l})_{(2)}$ is such that for some $x'_{l+1:n}$, $v_l(x_{1:l}, x'_{l+1:n}) = j$, then $(i, j) \in \mathcal{Y}^+(x_{k:l})$. Therefore, if every $i \in \mathcal{Y}^+(x_{k:l})_{(1)}$ and every

$j \in \mathcal{Y}^+(x_{k:l})_{(2)}$ satisfies above-stated property of being included into a Viterbi path (and often this is the case), then (i) and also **B1** is a necessary property of a barrier.

Example 6. Let $\mathcal{X} = \mathcal{Y} = \{1, 2\}$, and assume that X and Y are Markov chains both having the transition matrix $\begin{pmatrix} p & 1-p \\ q & 1-q \end{pmatrix}$, where $p, q \in (0, 1)$. Then, as is shown in [22], the transition matrix of Z has the form

$$\mathbb{Q} = \begin{array}{c} \\ (1,1) \\ (1,2) \\ (2,1) \\ (2,2) \end{array} \begin{array}{c} \begin{array}{cccc} (1,1) & (1,2) & (2,1) & (2,2) \end{array} \\ \left[\begin{array}{cccc} p\lambda_1 & p(1-\lambda_1) & p(1-\lambda_1) & 1 + p\lambda_1 - 2p \\ p\lambda_2 & p(1-\lambda_2) & q - p\lambda_2 & 1 + p\lambda_2 - q - p \\ q\mu_1 & q(1-\mu_1) & p - q\mu_1 & 1 + q\mu_1 - p - q \\ q\mu_2 & q(1-\mu_2) & q(1-\mu_2) & 1 + q\mu_2 - 2q \end{array} \right] \end{array},$$

where

$$\lambda_1 \in \left[\frac{2p-1}{p} \vee 0, 1 \right], \quad \lambda_2 \in \left[\frac{q+p-1}{p} \vee 0, \frac{q}{p} \wedge 1 \right], \tag{46}$$

$$\mu_1 \in \left[\frac{p+q-1}{q} \vee 0, \frac{p}{q} \wedge 1 \right], \quad \mu_2 \in \left[\frac{2q-1}{q} \vee 0, 1 \right]. \tag{47}$$

Here $\vee$ and $\wedge$ denote the maximum and minimum, respectively. When Z is stationary, then X and Y are independent if and only if

$$\lambda_1 = \mu_1 = p, \quad \lambda_2 = \mu_2 = q.$$

Assume now that λ_i and μ_i are not allowed to have the extreme values of the constraints (46) and (47). Then the elements of $\mathbb{Q}$ are positive, which implies that $\mathcal{Z}' = \mathcal{Z}$ and Z is irreducible and recurrent. Also (i) of Corollary 4.2 trivially holds for any $x_{1:q} \in \{1, 2\}^q$, where $q \geq 2$. Thus the existence of infinitely many strong nodes for almost every realization of X is guaranteed if there exists $x^*_{1:n} \in \{1, 2\}^n$, $n \geq 2$, such that

$$x^*_1 = x^*_n \quad \text{and} \quad p_{11}(x^*_{1:n}) > p_{ij}(x^*_{1:n}), \quad \forall (i, j) \in \mathcal{Y}^2 \setminus \{(1, 1)\}. \tag{48}$$

Taking $x^*_{1:n} = (1, 1)$, we have that (48) holds whenever

$$\begin{aligned} & p\lambda_1 > \max\{p(1-\lambda_1), p\lambda_2, p(1-\lambda_2)\} \\ \Leftrightarrow \quad & \lambda_1 > \max\{1-\lambda_1, \lambda_2, 1-\lambda_2\} \\ \Leftrightarrow \quad & \lambda_1 > \lambda_2 \vee (1-\lambda_2). \end{aligned}$$

Taking $x^*_{1:n} = (2, 2)$, we have that (48) holds when

$$p - q\mu_1 > \max\{1 + q\mu_1 - p - q, q(1-\mu_2), 1 + q\mu_2 - 2q\}.$$

Switch now the labels of $\mathcal{Y}$. Taking $x^*_{1:n} = (1, 1)$, we obtain that (48) holds when

$$\begin{aligned} & p(1-\lambda_2) > \max\{p\lambda_1, p(1-\lambda_1), p\lambda_2\} \\ \Leftrightarrow \quad & 1-\lambda_2 > \max\{\lambda_1, 1-\lambda_1, \lambda_2\} \\ \Leftrightarrow \quad & -\lambda_2 > \max\{\lambda_1 - 1, -\lambda_1, \lambda_2 - 1\} \\ \Leftrightarrow \quad & \lambda_2 < (1-\lambda_1) \wedge \lambda_1. \end{aligned}$$

Taking $x^*_{1:n} = (2, 2)$, we have that (48) holds when

$$1 + qp_2 - 2q > \max\{p - q\mu_1, 1 + q\mu_1 - p - q, q(1-\mu_2)\}.$$

Further conditions can be found with $n = 3, 4, \ldots$

4.3. Linear Markov switching model

Let $\mathcal{X} = \mathbb{R}^d$ for some $d \geq 1$ and for each state $i \in \mathcal{Y}$ let $\{\xi_k(i)\}_{k\geq 2}$ be an i.i.d. sequence of random variables on $\mathcal{X}$ with $\xi_2(i)$ having density h_i with respect to Lebesgue measure on $\mathbb{R}^d$. We consider the *linear Markov switching model*, where X is defined recursively by

$$X_k = F(Y_k)X_{k-1} + \xi_k(Y_k), \quad k \geq 2. \tag{49}$$

Here $F(i)$ are some $d \times d$ matrices, $Y = \{Y_k\}_{k\geq 1}$ is a Markov chain with transition matrix (p_{ij}), X_1 is some random variable on $\mathcal{X}$, and random variables $\{\xi_k(i)\}_{k\geq 2,\, i\in\mathcal{Y}}$ are assumed to be independent and independent of X_1 and Y. Recall that for Markov switching model, the transition density expresses as $q(x, j|x', i) = p_{ij} f_j(x|x')$. For the current model measure μ is Lebesgue measure on $\mathbb{R}^d$ and $f_j(x|x') = h_j(x - F(j)x')$. When $F(i)$ are zero-matrices, then the linear Markov switching model simply becomes HMM with h_i being the emission densities. When $F(i) = F$ for every $i \in \mathcal{Y}$, then the model reduces to an autoregressive model with correlated noise. When $d = 1$, we obtain the *switching linear autoregression of order 1*. The switching linear autoregressions are popular in econometric modeling, see e.g. [3,9–11] and the references therein.

We will now apply Theorem 3.1 to the linear Markov switching model. The requirement (of Theorem 3.1) that μ must be strictly positive is trivially fulfilled in the case where μ is Lebesgue measure. The requirement that functions $(x', x) \mapsto q(x, i|x', j)$ must be lower semi-continuous and bounded is fulfilled when h_j are lower semi-continuous and bounded (composition of lower semi-continuous function with continuous function is lower semi-continuous). Deriving simple conditions which ensure **B1** is also quite easy. In what follows, let $\|\cdot\|$ denote the 2-norm on $\mathcal{X} = \mathbb{R}^d$, and for any $x \in \mathcal{X}$ and $r > 0$ let $B(x, r)$ denote an open ball in $\mathcal{X}$ with respect to 2-norm with center point x and radius $r > 0$.

Lemma 4.1. *Let Z be the linear Markov switching model. If the following conditions are fulfilled, then Z satisfies* **B1**.

(i) There exists set $C \subset \mathcal{Y}$ and $r > 0$ such that the following two conditions are satisfied:

1. *for $x \in B(0, r)$, $h_i(x) > 0$ if and only if $i \in C$;*
2. *the sub-stochastic matrix $\mathbb{P}_C = (p_{ij})_{i,j\in C}$ is primitive, i.e. there exists $R \geq 1$ such that matrix $\mathbb{P}_C^R$ has only positive elements.*

(ii) Denote $\mathcal{Y}_C = \{i \in \mathcal{Y} \mid p_{ij} > 0,\ j \in C\}$. There exists $i_E \in \mathcal{Y}_C$ such that $(0, i_E)$ is reachable.

Conditions (i) and (ii) are not very restrictive. For example, when all the elements of $\mathbb{P}$ are positive, then (i) is fulfilled if densities h_i are either positive around 0 or zero around 0 and there exists at least one $j \in \mathcal{Y}$ such that h_j is positive around 0. If densities h_i are all positive around 0, then (i) is fulfilled when $\mathbb{P}$ is primitive with $C = \mathcal{Y}$. If h_i are positive everywhere and Y is irreducible, then all points in $\mathcal{Z}$ are reachable and so (ii) trivially holds.

Proof of Lemma 4.1. There must exist $r_0 > 0$ such that

$$\|x - F(j)x'\| < r, \quad \forall j \in \mathcal{Y}, \quad \forall x, x' \in B(0, r_0). \tag{50}$$

By (i) there exists $R \geq 1$ such that $\mathbb{P}_C^R$ contains only positive elements. We take $E = B(0, r_0)^{R+2}$. Fixing $x_{1:R+2} \in E$, we have for any $i, j \in \mathcal{Y}$

$$p_{ij}(x_{1:R+2}) = \max_{y_{1:R+2}:\,(y_1, y_{R+2})=(i,j)} \prod_{k=2}^{R+2} p_{y_{k-1}y_k} h_{y_k}(x_k - F(y_k)x_{k-1}).$$

Together with (50) and (i) this implies that $p_{ij}(x_{1:R+2}) > 0$ if and only if $i \in \mathcal{Y}_C$ and $j \in C$. Hence $\mathcal{Y}^+(x) = \mathcal{Y}_C \times C$ for every $x \in E$. Together with (ii) this implies that **B1** holds with $x_E = 0$. □

As for condition **B2**, the following lemma provides one possible way to construct the center part of the barrier set.

Lemma 4.2. *Let Z be the linear Markov switching model. If the following condition is fulfilled, then Z satisfies* **B2**: *there exists $x^* \in \mathcal{X}$ such that*

(i) $p_{11} = \max_{i \in \mathcal{Y}} p_{i1}$;
(ii) $(x^*, 1)$ *is reachable;*
(iii) h_i *is continuous at* $x^* - F(i)x^*$ *for all* $i \in \mathcal{Y}$, *and*

$$p_{11}h_1(x^* - F(1)x^*) > p_{ij}h_j(x^* - F(j)x^*), \quad \forall i \in \mathcal{Y}, \quad \forall j \in \mathcal{Y} \setminus \{1\}.$$

Proof. By (iii) there must exist $\epsilon > 0$ and $r > 0$ such that

$$\begin{aligned} &p_{11}h_1(x)(1-\epsilon) > p_{ij}h_j(x'), \quad \forall i \in \mathcal{Y}, \quad \forall j \in \mathcal{Y} \setminus \{1\}, \\ &\quad \forall x \in B(x^* - F(1)x^*, r), \quad \forall x' \in B(x^* - F(j)x^*, r). \end{aligned} \tag{51}$$

Also there must exist $r' > 0$ such that

$$\|x - F(j)x' - (x^* - F(j)x^*)\| < r, \quad \forall j \in \mathcal{Y}, \quad \forall x, x' \in B(x^*, r'). \tag{52}$$

For some $N \geq 2$ we define the center part of the barrier set by

$$\mathcal{X}^*_{(n_1, n_{2N+1})} = B(x^*, r')^{2N+1}$$

We confirm that $\mathcal{X}^*_{(n_1, n_{2N+1})}$ is indeed a strong center part of a barrier set, i.e. that it satisfies **A1**$'$. We take $n_{k+1} - n_k = 1$ for all $k = 1, \ldots, 2N$. Let $x', x \in B(x^*, r')$. We have for all $i \in \mathcal{Y}$ and $j \in \mathcal{Y} \setminus \{1\}$

$$p_{11}(x', x)(1-\epsilon) = p_{11}h_1(x - F(1)x')(1-\epsilon) > p_{ij}h_j(x - F(j)x') = p_{ij}(x', x).$$

Here the inequality follows from (52) and (51). On the other hand we have by (i) for all $i \in \mathcal{Y}$

$$p_{11}(x', x) = p_{11}h_1(x - F(1)x') \geq p_{i1}h_1(x - F(1)x') = p_{i1}(x', x).$$

The arguments above show that **A1**$'$ does indeed hold. Hence by (ii) **B2** holds. □

For the sake of simplicity Lemma 4.2 uses only cycles of length 2 in the construction of barrier set, but this could easily be generalized to include cycles of arbitrary length.

Remark. In the proofs of Lemmas 4.1 and 4.2 the specific structure of the linear Markov switching model has not played a very big role, so a natural question is, if analogous results could be proven for more general models. More specifically, we can consider a Markov switching model, where instead of recursion (49) X is more generally defined by

$$X_k = G(Y_k, X_{k-1}) + \xi_k(Y_k), \quad k \geq 2, \tag{53}$$

where $G(i,\cdot)\colon \mathbb{R}^d \to \mathbb{R}^d$ are some continuous functions. For this model, the transition kernel density expresses as $q(x, j|x', i) = p_{ij} h_j(x - G(j, x'))$. The statement of Lemma 4.2 indeed holds for this model, if we replace the condition (iii) with the following generalized version: h_i is continuous at $x^* - G(i, x^*)$ for all $i \in \mathcal{Y}$, and

$$p_{11}h_1(x^* - G(1, x^*)) > p_{ij}h_j(x^* - G(j, x^*)), \quad \forall i \in \mathcal{Y}, \quad \forall j \in \mathcal{Y} \setminus \{1\}.$$

The statement of Lemma 4.1 also holds for model (53), if we demand that the $G(i,\cdot)$ satisfy the following additional condition:

$$G(i, 0) = 0, \quad \forall i \in \mathcal{Y}. \tag{54}$$

If (54) is too restrictive, a different approach is needed to prove **B1**. In any case, if h_i are everywhere positive and $\mathbb{P}$ is primitive, then, as it is easy to verify, **B1** holds regardless of whether (54) holds or not.

It remains to address the issue of Harris recurrence of the linear Markov switching model. In what follows, for $x \in \mathcal{X}$ we denote with $\|x\|_1$ the 1-norm of x, and for a $d \times d$ matrix A we denote with $\|A\|_1$ the 1-norm of matrix A, that is $\|A\|_1$ is the maximum absolute column sum of A.

Lemma 4.3. *Let Z be the linear Markov switching model. If the following conditions are fulfilled, then Z is Harris recurrent:*

(i) *Z is ψ-irreducible and support of ψ has non-empty interior;*
(ii) $\mathbb{E}\|\xi_2(i)\|_1 < \infty$ *for all* $i \in \mathcal{Y}$;
(iii) $\max_{i\in\mathcal{Y}} \sum_{j\in\mathcal{Y}} p_{ij}\|F(j)\|_1 < 1$.

Proof of this statement is given in Appendix B.

Applying the results above to the case where h_i are Gaussian yields

Corollary 4.3. *Let Z be the linear Markov switching model, with densities h_i being Gaussian with respective mean vectors μ_i and positive definite covariance matrices Σ_i. If the following conditions are fulfilled, then there exist a barrier set $\mathcal{X}^*$ consisting of strong 1-barriers of fixed order and satisfying $P(X \in \mathcal{X}^* \text{ i.o.}) = 1$.*

(i) *Matrix $\mathbb{P} = (p_{ij})$ is primitive, i.e. there exists R such that $\mathbb{P}^R$ consists of only positive elements.*
(ii) *It holds* $p_{11} = \max_{i\in\mathcal{Y}} p_{i1}$.
(iii) *Matrix $\mathbb{I}_d - F(1)$, where $\mathbb{I}_d$ denotes the identity matrix of dimension d, is non-singular, and for all $i \in \mathcal{Y}$ and $j \in \mathcal{Y} \setminus \{1\}$*

$$(\mathbb{I}_d - F(j))(\mathbb{I}_d - F(1))^{-1}\mu_1 \in \mathbb{R}^d \setminus H_{ij},$$

where

$$H_{ij} := \begin{cases} \emptyset, & \text{if } p_{ij} = 0 \text{ or } \dfrac{p_{11}\sqrt{|\Sigma_j|}}{p_{ij}\sqrt{|\Sigma_1|}} > 1, \\ \left\{x \in \mathbb{R}^d \mid (x-\mu_j)^\top \Sigma_j^{-1}(x-\mu_j) \right. \\ \quad \left. \leq -2\ln\left(\dfrac{p_{11}\sqrt{|\Sigma_j|}}{p_{ij}\sqrt{|\Sigma_1|}}\right)\right\}, & \text{else}. \end{cases}$$

(iv) It holds $\max_{i\in\mathcal{Y}}\sum_{j\in\mathcal{Y}} p_{ij}\|F(j)\|_1 < 1$.

Proof. By (i) Y is irreducible. This together with the fact that densities h_i are positive on the whole space $\mathbb{R}^d$ implies that all elements in $\mathcal{Z} = \mathbb{R}^d \times \mathcal{Y}$ are reachable and that Z is $\mu \times c$-irreducible, where μ denotes the Lebesgue measure on $\mathbb{R}^d$ and c denotes the counting measure on $\mathcal{Y}$. It follows from (i) and Lemma 4.1 that **B1** holds. Take now $x^* = (\mathbb{I}_d - F(1))^{-1}\mu_1$ (then $x^* - F(1)x^* = \mu_1$ and so $x^* - F(1)x^*$ maximizes h_1). Condition (iii) implies

$$x^* - F(j)x^* \notin H_{ij}, \quad \forall i \in \mathcal{Y}, \quad \forall j \in \mathcal{Y}\setminus\{1\}. \tag{55}$$

Some calculation reveals that $\{x \in \mathbb{R}^d \mid p_{ij}h_j(x) \geq p_{11}h_1(x^* - F(1)x^*)\} = H_{ij}$ and so (55) implies

$$p_{ij}h_j(x^* - F(j)x^*) < p_{11}h_1(x^* - F(1)x^*), \quad \forall i \in \mathcal{Y}, \quad \forall j \in \mathcal{Y}\setminus\{1\}.$$

This together with assumption (ii) and Lemma 4.2 implies that **B2** holds. By (iv) and Lemma 4.3 Z is Harris recurrent, so the statement follows from Theorems 2.1 and 3.1. □

In some cases the condition (ii) of Corollary 4.3 can be rather restrictive, particularly when the diagonal entries of $\mathbb{P} = (p_{ij})$ are small and so there are not many (or none at all) diagonal entries of $\mathbb{P}$ which dominate their column (i.e. are larger than or equal to other column entries). In that case one possible solution is to group the elements of Z to pairs, that is consider the model $Z' = \{((X_{2k-1}, X_{2k}), (Y_{2k-1}, Y_{2k}))\}_{k\geq 1}$ instead of $Z = \{(X_k, Y_k)\}_{k\geq 1}$. The transition kernel density of chain Z' is simply

$$q'(z_3, z_4|z_1, z_2)\colon ((z_1, z_2), (z_3, z_4)) \mapsto p(z_3, z_4|z_2) = q(z_3|z_2)q(z_4|z_3).$$

Let X' and Y' denote the marginals of Z': $X' = \{(X_{2k-1}, X_{2k})\}_{k\geq 1}$ and $Y' = \{(Y_{2k-1}, Y_{2k})\}_{k\geq 1}$. The existence of Viterbi process for X' implies the existence of Viterbi process for X under appropriate tie-breaking rules. Indeed, consider the case where the tie-breaking scheme corresponding to X' is lexicographic, induced by the following ordering on $\mathcal{Y}^2$:

$$(1,1) \succ (1,2) \succ (1,3) \succ \cdots \succ (2,1) \succ (2,2) \succ (2,3) \succ \cdots \succ (|\mathcal{Y}|, |\mathcal{Y}|).$$

We also assume that the ordering on $\mathcal{Y}$ is

$$1 \succ 2 \succ \cdots \succ |\mathcal{Y}|.$$

Thus $\mathcal{Y}^2$ is equipped with lexicographic ordering induced by the ordering on $\mathcal{Y}$. We assume that the tie-breaking scheme corresponding to X is lexicographic as well. Then, if $\{(V_{2k-1}, V_{2k})\}_{k\geq 1}$ is the Viterbi process of X', $\{V_k\}_{k\geq 1}$ is the Viterbi process of X.

Chain Z' is a linear Markov switching model on space $\mathbb{R}^{2d} \times \mathcal{Y}'$, where $\mathcal{Y}' := \{(i,j) \in \mathcal{Y}^2 \mid p_{ij} > 0\}$. To see this, note that

$$\begin{pmatrix} X_{2k-1} \\ X_{2k} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & F(Y_{2k-1}) \\ \mathbf{0} & F(Y_{2k})F(Y_{2k-1}) \end{pmatrix} \begin{pmatrix} X_{2k-3} \\ X_{2k-2} \end{pmatrix} + \xi'_k(Y_{2k-1}, Y_{2k}), \quad k \geq 2,$$

where

$$\xi'_k(i,j) := \begin{pmatrix} \xi_{2k-1}(i) \\ F(j)\xi_{2k-1}(i) + \xi_{2k}(j) \end{pmatrix} = \begin{pmatrix} \mathbb{I}_d & \mathbf{0} \\ F(j) & \mathbb{I}_d \end{pmatrix} \begin{pmatrix} \xi_{2k-1}(i) \\ \xi_{2k}(j) \end{pmatrix}.$$

The matrix $B := \begin{pmatrix} \mathbb{I}_d & \mathbf{0} \\ F(j) & \mathbb{I}_d \end{pmatrix}$ has full rank, so assuming that $\xi_k(i)$ are non-degenerate Gaussian with respective mean vectors μ_i and covariance matrices Σ_i, then random vectors $\xi'_k(i,j)$ are non-degenerate Gaussian with respective mean values $\begin{pmatrix} \mu_i \\ F(j)\mu_i + \mu_j \end{pmatrix}$ and covariance matrices $B\begin{pmatrix} \Sigma_i & \mathbf{0} \\ \mathbf{0} & \Sigma_j \end{pmatrix}B^\top$. Therefore Corollary 4.3 applies to both Z and Z'.

Markov chain Y', the hidden process of model Z', has transition matrix $\mathbb{P}' := (p_{jk}p_{kl})_{(i,j),(k,l)\in\mathcal{Y}'}$. Matrix $\mathbb{P}'$ might have diagonal entries which dominate their column, even if $\mathbb{P}$ does not have such entries. As a simple example consider the case where $\mathbb{P} = \begin{pmatrix} \epsilon & 1-\epsilon \\ 1-\epsilon' & \epsilon' \end{pmatrix}$, where $\epsilon, \epsilon' \in (0, \frac{1}{2})$. Then

$$\mathbb{P}' = \begin{array}{c} (1,1) \\ (1,2) \\ (2,1) \\ (2,2) \end{array} \overset{\begin{array}{cccc} (1,1) & (1,2) & (2,1) & (2,2) \end{array}}{\begin{bmatrix} \epsilon^2 & \epsilon(1-\epsilon) & (1-\epsilon)(1-\epsilon') & (1-\epsilon)(\epsilon')^2 \\ (1-\epsilon')\epsilon & (1-\epsilon')(1-\epsilon) & \epsilon'(1-\epsilon') & (\epsilon')^2 \\ \epsilon^2 & \epsilon(1-\epsilon) & (1-\epsilon)(1-\epsilon') & (1-\epsilon)(\epsilon')^2 \\ (1-\epsilon')\epsilon & (1-\epsilon')(1-\epsilon) & \epsilon'(1-\epsilon') & (\epsilon')^2 \end{bmatrix}},$$

and so the second and the third diagonal entry of $\mathbb{P}'$ dominates its column.

In general primitiveness of $\mathbb{P}$ does not imply primitiveness of $\mathbb{P}'$, but the reader can easily verify that if there exists an *odd* positive integer R such that $\mathbb{P}^R$ contains only positive entries, then $\mathbb{P}'^{(R+1)/2}$ contains only positive entries and so $\mathbb{P}'$ is primitive. We also note that the approach described above can easily be generalized to include groupings of triplets, quadruplets, etc.

4.4. A class of Gaussian models

In this section, we consider a class of Gaussian PMM's introduced, studied and applied in a series of papers by W. Pieczynski et al. [5,6,8,17,24]. These models are defined by the density $p(z_1, z_2)$ of (Z_1, Z_2) as follows: with $z_1 = (x', i)$ and $z_2 = (x, j)$

$$p(z_1, z_2) = p(y_1 = i, y_2 = j)p(x', x|y_1 = i, y_2 = j) = p(i, j)f_{ij}(x', x),$$

where $p(i, j) := p(y_1 = i, y_2 = j)$ and $f_{ij}(x', x)$ stands for the bivariate conditional density $p(x', x|y_1 = i, y_2 = j)$. We assume that for every $i, j \in \mathcal{Y}$, $p(i, j) > 0$ and all conditional densities f_{ij} are bivariate normal with parameters

$$\mu_{ij} = \begin{pmatrix} \mu_{ij}(1) \\ \mu_{ij}(2) \end{pmatrix}, \quad \Sigma_{ij} = \begin{pmatrix} \sigma^2_{ij}(1) & \rho_{ij}\sigma_{ij}(1)\sigma_{ij}(2) \\ \rho_{ij}\sigma_{ij}(1)\sigma_{ij}(2) & \sigma^2_{ij}(2) \end{pmatrix}.$$

Thus the transition kernel has the following density with respect to $\mu \times c$, where μ is the Lebesgue measure:

$$q(x, j|x', i) = \frac{f_{ij}(x', x)p(i, j)}{\sum_{k\in\mathcal{Y}} f_{ik}(x')p(i, k)}.$$

Here $f_{ik}(x')$ stands for the marginal density: $f_{ij}(x') = \int f_{ij}(x', x)\mu(dx)$. Since $p(i, j) > 0$ for every i, j and all densities are Gaussian, it follows that Z is $\mu \times c$-irreducible and so all elements of $\mathcal{Z} = \mathbb{R} \times \mathcal{Y}$ are reachable. For this model, Y is not necessarily a Markov chain. If Z is stationary, then Y is Markov if and only if the marginal distribution $p(x_1 = x|y_1 = i, y_2 = j)$ is independent of j, in terms of densities $f_{ij}(x') = f_i(x')$, for every $i, j \in \mathcal{Y}$ [5,17,24]. In practice often the matrix $(p(i, j))$ is symmetric, i.e. $p(i, j) = p(j, i)$ for every i, j and densities f_{ij} are symmetric in the following sense:

$$f_{ij}(x', x) = f_{ji}(x, x'), \quad i, j \in \mathcal{Y}, \quad x', x \in \mathcal{X}. \tag{56}$$

For Gaussian densities (56) simply means that for every $i, j \in \mathcal{Y}$, $\mu_{ij}(1) = \mu_{ji}(2)$, $\mu_{ij}(2) = \mu_{ji}(1)$ and $\sigma_{ij}(1) = \sigma_{ji}(2)$, $\sigma_{ij}(2) = \sigma_{ji}(2)$ and $\rho_{ij} = \rho_{ji}$. The symmetric matrix $(p(i, j))$ and (56) ensure that $p(z_1, z_2) = p(z_2, z_1)$ so that Z is reversible, hence stationary.

The following lemma establishes sufficient conditions for Z to be Harris recurrent.

Lemma 4.4. *Suppose that there exist $\epsilon > 0$ and $M < \infty$ so that*

$$\sum_{j\in\mathcal{Y}} b_{ij}\, p(i, j) f_{ij}(x') \le (1-\epsilon) \sum_{j\in\mathcal{Y}} p(i, j) f_{ij}(x'), \quad \forall i \in \mathcal{Y}, \quad \forall x' \in [-M, M]^c, \tag{57}$$

where

$$b_{ij} := \frac{\sigma_{ij}(2)}{\sigma_{ij}(1)} \rho_{ij}.$$

Then Z is Harris recurrent.

The proof of Lemma 4.4 is given in Appendix C. Observe that (57) is satisfied for any x' when $\max_{ij} b_{ij} < 1$, and this holds when all correlation coefficients ρ_{ij} are all sufficiently small. In particular, (57) holds when $\rho_{ij} = 0$ for every i, j so that X_1, X_2 are conditionally independent (that does not necessarily imply HMM). On the other hand, as we shall see in the examples below, the condition (57) can also hold when for some i, j, $b_{ij} > 1$.

Let us now discuss the assumptions **B1** and **B2**. Observe

$$p_{ij}(x_1, x_2) = \frac{p(i, j) f_{ij}(x_1, x_2)}{\sum_k p(i, k) f_{ik}(x_1, x_2)}. \tag{58}$$

Since f_{ij} are Gaussian and $p(i, j) > 0$ it follows that for every pair x_1, x_2, $\mathcal{Y}^+(x_1, x_2) = \mathcal{Y} \times \mathcal{Y} = \mathcal{Y}_{(1)} \times \mathcal{Y}_{(2)}$. Thus **B1** trivially holds with $q = 2$, $E = \mathcal{X}^2$ (recall that every point $(x, i) \in \mathcal{X} \times \mathcal{Y} = E_{(1)} \times \mathcal{Y}_{(1)}$ is reachable).

For checking **B2** we apply Example 5, which is possible because the mappings $(x'x) \mapsto q(x, i|x, j)$ are continuous. To show the existence of an open strong center part of a barrier, it suffices to find a n and a vector $x_{1:n}$ so that

$$x_1 = x_n \quad \text{and} \quad p_{11}(x_{1:n}) > p_{ij}(x_{1:n}), \quad \forall (i, j) \in \mathcal{Y}^2 \setminus \{(1, 1)\}. \tag{59}$$

In Example 5, it is shown that when such $x_{1:n}$ exists, then for any N, there exists an open strong center part $\mathcal{X}^*_{(1,2N(n-1)+1)}$ consisting of $2N$ cycles. According to the construction, $\mathcal{X}^*_{(1)} = \mathcal{X}^*_{(2N(n-1)+1)} = B_1$ is an open ball, independent of N. Thus the closure of B_1 is the compact set K required in **B2** . Since any pair $(x, 1)$ is reachable, we see that **B2** holds whenever there exists $x_{1:n}$ that satisfies (59).

The vector $x_{1:n}$ satisfying (59), when it exists, depends on the model. For simplicity, we shall more closely consider the case $K = 2$ (often in met in practice) and we look for the shortest possible vector, i.e. $n = 2$. Thus, we look for a $x \in \mathcal{X}$ such that all three inequalities simultaneously hold

$$p_{11}(x, x) > p_{22}(x, x), \quad p_{11}(x, x) > p_{12}(x, x), \quad p_{11}(x, x) > p_{21}(x, x).$$

These three conditions are equivalent to the following

$$\begin{aligned} &p(1, 1) f_{11}(x, x) \cdot p(2, 1) f_{21}(x, x) > p(2, 2) f_{22}(x, x) \cdot p(1, 2) f_{12}(x, x) \\ &p(1, 1) f_{11}(x, x) > p(1, 2) f_{12}(x, x) \\ &p(1, 1) f_{11}(x, x) \cdot p(2, 2) f_{22}(x, x) > p(1, 2) f_{12}(x, x) \cdot p(2, 1) f_{21}(x, x). \end{aligned} \tag{60}$$

If $p(1,2) = p(2,1)$ and (56) holds, then the first inequality simplifies $p(1,1)f_{11}(x,x) > p(2,2)f_{22}(x,x)$. If, in addition $p(1,1) \geq p(1,2)$ and $p(2,2) \geq p(1,2)$, then the inequalities (60) hold when

$$f_{11}(x,x) > f_{22}(x,x), \quad f_{11}(x,x) > f_{12}(x,x), \quad f_{11}(x,x)f_{22}(x,x) > f_{12}^2(x,x). \tag{61}$$

Surely $x = \mu_{11}(1) = \mu_{11}(2)$ satisfies the first two inequalities, but the third one needs checking. Let us proceed with numerical examples.

In [24] the following model was considered: $K = 2$,

$$\mu_{11} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \mu_{12} = \begin{pmatrix} 2 \\ 8 \end{pmatrix}, \quad \mu_{12} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}, \quad \mu_{21} = \begin{pmatrix} 10 \\ 10 \end{pmatrix}$$

and

$$\sigma_{11}(1) = \sigma_{11}(2) = \sigma_{12}(1) = \sigma_{21}(2) = 14, \ \sigma_{12}(2) = \sigma_{21}(1) = \sigma_{22}(1) = \sigma_{22}(2) = 20,$$
$$\rho_{11} = \rho_{22} = 0.9, \ \rho_{12} = \rho_{21} = 0.1.$$

Thus the densities f_{ij} are such that (56) holds. The probabilities $p(i,j)$ were taken

$$1)\ p(i,j) = 0.25, \quad i,j = 1,2; \quad 2)\ p(1,1) = p(2,2) = 0.48,$$
$$p(1,2) = p(2,1) = 0.02.$$

Thus in the both cases the matrix $(p(i,j))$ is symmetric satisfying the inequalities $p(1,1) \geq p(1,2)$ and $p(2,2) \geq p(1,2)$. Hence it suffices to find a single x satisfying (61). One can check that $x = -10$ is such an element, hence **B2** holds (for both cases).

In [5] the following models were considered: the mean vectors are fixed

$$\mu_{11} = \begin{pmatrix} -5 \\ -5 \end{pmatrix}, \quad \mu_{12} = \begin{pmatrix} -3 \\ 3 \end{pmatrix}, \quad \mu_{12} = \begin{pmatrix} 3 \\ -3 \end{pmatrix}, \quad \mu_{21} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}$$

and the covariation matrices are such that

$$\sigma_{11}(1) = \sigma_{11}(2) = 14, \quad \sigma_{12}(1) = \sigma_{21}(2) = 7, \ \sigma_{12}(2) = \sigma_{21}(1) = 9,$$
$$\sigma_{22}(1) = \sigma_{22}(2) = 20$$

but the correlation coefficients vary

$$(1) \quad \rho_{11} = \rho_{22} = 0.9, \quad \rho_{12} = \rho_{21} = 0.1; \quad (2) \quad \rho_{11} = \rho_{22} = 0.1, \quad \rho_{12} = \rho_{21} = 0.9;$$
$$(3) \quad \rho_{11} = \rho_{22} = \rho_{12} = \rho_{21} = 0.1; \quad (4) \quad \rho_{11} = \rho_{22} = \rho_{12} = \rho_{21} = 0.9.$$

Again, in all four cases the emissions satisfy (56). The probabilities $p(i,j)$ satisfy the inequalities $p(1,1) \geq p(1,2)$ and $p(2,2) \geq p(1,2)$, so again it suffices to find x satisfying (61). One can check that suitable values for x are: (1) -10, (2) -28, (3) -19, (4) -35. Thus, for all models considered in [5,24] the condition **B2** holds.

To conclude the section, let us study the Harris recurrency of the considered models. In the models studied in [24], the coefficients b_{ij} are as follows: $b_{11} = 0.9, b_{12} = 1/7, b_{21} = 7/100, b_{22} = 0.9$. We see that $b_{ij} < 1$ for every i, j and so by Lemma 4.4 the chain is Harris. For the models in [5], we see that

$$(1) \quad b_{11} = b_{22} = 0.9, \ b_{12} = 9/70, \ b_{21} = 7/90; \quad (2) \quad b_{11} = b_{22} = 0.1, \ b_{12} = 81/70, \ b_{21} = 63/90;$$
$$(3) \quad b_{11} = b_{22} = 0.1, \ b_{12} = 9/70, \ b_{21} = 7/90; \quad (4) \quad b_{11} = b_{22} = 0.9, \ b_{12} = 81/70, \ b_{21} = 63/90.$$

We see that for cases (1) and (3) clearly the condition (57) holds, because $b_{ij} < 1$ for $i, j = 1, 2$. For case (2) and (4) observe that for $b_{1j} < 1$, so that for $i = 1$ the condition (57) holds for any x'. Thus it suffices to show that when $|x'|$ is big enough, then

$$\frac{b_{21}p(2,1)f_{21}(x') + b_{22}p(2,2)f_{22}(x')}{p(2,1)f_{21}(x') + p(2,2)f_{22}(x')} < 1. \tag{62}$$

Observe that in both cases $f_{21}(x')$ is the Gaussian density with standard deviation 9 and $f_{22}(x')$ is the Gaussian density with standard deviation 20. Therefore

$$\lim_{|x'|\to\infty} \frac{f_{21}(x')}{f_{22}(x')} = 0$$

and so the ratio in (62) tends to $b_{22} < 1$ as $|x'| \to \infty$. Thus (57) holds in all cases and all models considered in [5,24] are Harris recurrent.

Acknowledgment

The research is supported by Estonian institutional research funding IUT34-5.

Appendix A. Supporting results

Lemma A.1. *Suppose there exist sets $A \subset \mathcal{Z}$ and $B \subset \mathcal{Z}^M$, $M \geq 1$, and $\epsilon > 0$ such that*

$$P(Z_k \in A \text{ i.o.}) = 1,$$
$$P(Z_{1:M} \in B | Z_1 = z) \geq \epsilon, \quad \forall z \in A.$$

Then

$$P(Z \in B \text{ i.o.}) = 1.$$

Proof. The proof is just a slightly modified version of the proof of [23, Th. 9.1.3]. It suffices to show that

$$P(Z \in A \text{ i.o.}) \leq P(Z \in B \text{ i.o.}). \tag{63}$$

Define

$$E_n = \{Z_{n:n+M-1} \in B\}, \quad n \geq 1.$$

For each $n \geq 1$ let $\mathcal{F}_n$ be a σ-field generated by $\{Z_1, \ldots, Z_n\}$. First we show that as $n \to \infty$

$$P\left(\bigcup_{i=n}^{\infty} E_i | \mathcal{F}_n\right) \to \mathbb{I}\left(\bigcap_{k=1}^{\infty}\bigcup_{i=k}^{\infty} E_i\right), \quad \text{a.s.,} \tag{64}$$

where $\mathbb{I}(\cdot)$ is the indicator function. To see this, note that for fixed $l \leq n$

$$P\left(\bigcup_{i=l}^{\infty} E_i | \mathcal{F}_n\right) \geq P\left(\bigcup_{i=n}^{\infty} E_i | \mathcal{F}_n\right) \geq P\left(\bigcap_{k=1}^{\infty}\bigcup_{i=k}^{\infty} E_i | \mathcal{F}_n\right). \tag{65}$$

Applying the Martingale Convergence Theorem to the extreme elements of the inequalities (65), we obtain

$$\mathbb{I}(\cup_{i=l}^{\infty} E_i) \geq \limsup_n P(\cup_{i=n}^{\infty} E_i | \mathcal{F}_n) \geq \liminf_n P(\cup_{i=n}^{\infty} E_i | \mathcal{F}_n) \geq \mathbb{I}(\cap_{k=1}^{\infty} \cup_{i=k}^{\infty} E_i). \tag{66}$$

As $l \to \infty$, the two extremes in (66) converge, which shows that the convergence (64) holds as required.

Next, define

$$L(z) = P(\cup_{i=1}^{\infty} E_i | Z_1 = z), \quad z \in \mathcal{Z}.$$

Clearly $L(z) \geq \epsilon$ for every $z \in A$. Also by Markov property $P(\cup_{i=n}^{\infty} E_i | \mathcal{F}_n) = L(Z_n)$ a.s. Thus, using (64), we have almost surely

$$\mathbb{I}(\cap_{k=1}^{\infty} \cup_{i=k}^{\infty} \{Z_i \in A\}) \leq \mathbb{I}(\limsup_n L(Z_n) \geq \epsilon) = \mathbb{I}(\lim_n L(Z_n) = 1) = \mathbb{I}(\cap_{k=1}^{\infty} \cup_{i=k}^{\infty} E_i).$$

This implies (63). □

Lemma A.2. *Let Z be HMM and define $G_i = \{x \in \mathcal{X} \mid f_i(x) > 0\}$ for $i \in \mathcal{Y}$. If Markov chain Y is irreducible then Z is ψ-irreducible, where*

$$\psi(A \times \{i\}) = \mu(A \cap G_i), \quad A \in \mathcal{B}(\mathcal{X}), \quad i \in \mathcal{Y},$$

and Harris recurrent.

Proof. That Z is ψ-irreducible follows directly from the irreducibility of Y and definition of HMM. Because for every $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{Z})$

$$P(Z_2 \in A | Z_1 = (x, 1)) = P(Z_2 \in A | Y_1 = 1),$$

then set $\mathcal{X} \times \{1\}$ is small (see definition in [23]). Harris recurrence now follows from [23, Prop. 9.1.7]. □

Appendix B. Proof of Lemma 4.3

We start with the following auxiliary lemma:

Lemma B.1. *Let Z be the linear Markov switching model defined by (49). Suppose Z is ψ-irreducible, support of ψ has non-empty interior, and there exist a compact set $C \subset \mathbb{R}^d \times \mathcal{Y}$ and a positive measurable function $V : \mathcal{Z} \to \mathbb{R}_{>0}$ satisfying*

$$\mathbb{E}[V(Z_2) | Z_1 = z] - V(z) \leq 0, \quad \forall z \in \mathcal{Z} \setminus C. \tag{67}$$

If set $\{z \mid V(z) \leq k\}$ is contained in a compact set for every $k < \infty$, then Z is Harris recurrent.

Proof. According to our assumption space $\mathcal{Z}$ is equipped with product topology $\tau \times 2^{\mathcal{Y}}$, where τ is the topology on the Euclidean space $\mathcal{X} = \mathbb{R}^d$. In this topology saying that some function $h : \mathcal{Z} \to \mathbb{R}$ is continuous means the following: for every $x_0 \in \mathcal{X}$ and $i \in \mathcal{Y}$

$$\lim_{x \to x_0} h(x, i) = h(x_0, i).$$

First we show that for any bounded and continuous function $h : \mathcal{Z} \to \mathbb{R}$, the function

$$z \mapsto \mathbb{E}[h(Z_2) | Z_1 = z] \tag{68}$$

is also bounded and continuous. Indeed, we have

$$\begin{aligned}\mathbb{E}[h(Z_2)|Z_1=(x',i)] &= \int_{\mathcal{Z}} h(x,j)\,q(x,j|x',i)\mu\times c(d(x,j))\\ &= \int_{\mathcal{Z}} h(x,j)p_{ij}f_j(x|x')\,\mu\times c(d(x,j))\\ &= \sum_{j\in\mathcal{Y}}\int_{\mathcal{X}} h(x,j)p_{ij}f_j(x|x')\,\mu(dx)\\ &= \sum_{j\in\mathcal{Y}}\int_{\mathcal{X}} h(x,j)p_{ij}h_j(x-F(j)x')\,\mu(dx)\\ &= \sum_{j\in\mathcal{Y}}\int_{\mathcal{X}} h(x+F(j)x',j)p_{ij}h_j(x)\,\mu(dx).\end{aligned}$$

Thus (68) is bounded, and also continuous by Dominated Convergence Theorem. In what follows, the definitions for the terms in italic can be found from [23]. By [23, Prop. 6.1.1(i)] Z is *weak Feller*. Hence by [23, Prop. 6.2.8] every compact set in $\mathcal{Z}$ is *petite*. Thus the statement follows from [23, Th. 9.1.8]. □

We take $V(x,i)=\|x\|_1+1$; then

$$\begin{aligned}\mathbb{E}[V(Z_2)|Z_1=(x',i)]-V(x',i) &= \sum_{j\in\mathcal{Y}} p_{ij}\int \|x\|_1 h_j(x-F(j)x')\,\mu(dx)-\|x'\|_1\\ &= \sum_{j\in\mathcal{Y}} p_{ij}\int \|x+F(j)x'\|_1 h_j(x)\,\mu(dx)-\|x'\|_1\\ &\le \sum_{j\in\mathcal{Y}} p_{ij}\int (\|x\|_1+\|F(j)\|_1\|x'\|_1)h_j(x)\,\mu(dx)-\|x'\|_1\\ &\le \sum_{j\in\mathcal{Y}} p_{ij}\mathbb{E}\|\xi_2(j)\|_1+\|x'\|_1\sum_{j\in\mathcal{Y}} p_{ij}\|F(j)\|_1-\|x'\|_1.\end{aligned}$$

Thus by the assumptions that the expectations $\mathbb{E}\|\xi(i)\|_1$ are finite and $\max_{i\in\mathcal{Y}}\sum_{j\in\mathcal{Y}} p_{ij}\|F(j)\|_1 < 1$, we have that (69) holds with $C=[-n,n]^d\times\mathcal{Y}$, when n is sufficiently large. Also set $\{z\,|\,V(z)\le k\}$ is contained in a compact set $[-k,k]^d\times\mathcal{Y}$ for every $k<\infty$. Hence Lemma B.1 applies.

Appendix C. Proof of Lemma 4.4

Lemma C.1. *Let Z be the Gaussian PMM as in Section 4.4. Suppose there exist a compact set $C\subset\mathbb{R}\times\mathcal{Y}$ and a positive measurable function $V:\mathcal{Z}\to\mathbb{R}_{>0}$ satisfying*

$$\mathbb{E}[V(Z_2)|Z_1=z]-V(z)\le 0,\quad \forall z\in\mathcal{Z}\setminus C. \tag{69}$$

If set $\{z\,|\,V(z)\le k\}$ is contained in a compact set for every $k<\infty$, then Z is Harris recurrent.

Proof. The proof is the same as the one of Lemma B.1: it is easy to check that also for this model (68) is bounded and continuous, thus every compact set is petite and so [23, Th. 9.1.8] applies. □

We take $V(x,i) = |x| + 1$, then for every $i \in \mathcal{Y}$

$$\begin{aligned}\mathbb{E}[V(Z_2)|Z_1 = (x',i)] - V(x',i) &= \sum_{j\in\mathcal{Y}} \int |x| q(x,j|x',i)\mu(dx) - |x'| \\ &= \frac{\sum_{j\in\mathcal{Y}} \int |x| f_{ij}(x',x)\mu(dx)p(i,j)}{\sum_{k\in\mathcal{Y}} f_{ik}(x')p(i,k)} - |x'| \\ &= \frac{\sum_{j\in\mathcal{Y}} \int |x| f_{ij}(x|x') f_{ij}(x')\mu(dx)p(i,j)}{\sum_{k\in\mathcal{Y}} f_{ik}(x')p(i,k)} - |x'|, \end{aligned} \tag{70}$$

where

$$f_{ij}(x|x') := \frac{f_{ij}(x',x)}{f_{ij}(x')}$$

is the conditional density. Let

$$\mu_{ij}(x') := \int x f_{ij}(x|x')\mu(dx) = \mu_{ij}(2) + b_{ij}(x' - \mu_{ij}(1)).$$

Then

$$\int |x| f_{ij}(x|x')\mu(dx) \le \int |x - \mu_{ij}(x')| f_{ij}(x|x')\mu(dx) + |\mu_{ij}(x')| = a_{ij} + |\mu_{ij}(2) - b_{ij}\mu_{ij}(1) + b_{ij}x'|,$$

where

$$a_{ij} = \int |x - \mu_{ij}(x')| f_{ij}(x|x')\mu(dx)$$

is independent of x'. Thus, with $c_{ij} := a_{ij} + |\mu_{ij}(2) - b_{ij}\mu_{ij}(1)|$, we obtain that

$$\int |x| f_{ij}(x|x')\mu(dx) \le c_{ij} + b_{ij}|x'|$$

and so, with $c := \max_{ij} c_{ij}$, we have

$$\begin{aligned}\sum_{j\in\mathcal{Y}} \int |x| f_{ij}(x|x')\mu(dx) f_{ij}(x')p(i,j) &\le \sum_{j\in\mathcal{Y}} \big(c_{ij} + b_{ij}|x'|\big) f_{ij}(x')p(i,j) \\ &\le \sum_{j\in\mathcal{Y}} \Big(\frac{c}{|x'|} + b_{ij}\Big)|x'| f_{ij}(x')p(i,j).\end{aligned}$$

Let now $M_1 \ge M$ be such that $c/M_1 < \epsilon$. Then, for every x' such that $|x'| \ge M_1$, it holds $c/|x'| < \epsilon$ and then by (57)

$$\sum_{j\in\mathcal{Y}} \Big(\frac{c}{|x'|} + b_{ij}\Big)|x'| f_{ij}(x')p(i,j) \le |x'| \sum_{j\in\mathcal{Y}} f_{ij}(x')p(i,j),$$

and so the right hand side of (70) is less than 0. Thus with $C = [-M_1, M_1] \times \mathcal{Y}$, the condition (69) holds. Lemma C.1 finishes the proof.

References

[1] A. Caliebe, Properties of the maximum a posteriori path estimator in hidden Markov models, IEEE Trans. Inform. Theory 52 (1) (2006) 41–51.

[2] A. Caliebe, U. Rösler, Convergence of the maximum a posteriori path estimator in hidden Markov models, IEEE Trans. Inform. Theory 48 (7) (2002) 1750–1758.

[3] O. Cappé, E. Moulines, T. Rydén, Inference in Hidden Markov Models, Springer, 2005.

[4] P. Chigansky, Y. Ritov, On the Viterbi process with continuous state space, Bernoulli 17 (2) (2011) 609–627.

[5] S. Derrode, W. Pieczynski, Signal and image segmentation using pairwise Markov chains, IEEE Trans. Signal Process. 52 (9) (2004) 2477–2489.

[6] S. Derrode, W. Pieczynski, Unsupervised data classification using pairwise Markov chains with automatic copula selection, Comput. Statist. Data Anal. 63 (2013) 81–98.

[7] A. Ghosh, E. Kleiman, A. Roitershtein, Large deviation bounds for functionals of Viterbi paths, IEEE Trans. Inform. Theory 57 (6) (2011) 3932–3937.

[8] I. Gorynin, H. Gangloff, E. Monfrini, W. Pieczynski, Assessing the segmentation performance of pairwise and triplet Markov models, Signal Process. 145 (2018) 183–192.

[9] J. Hamilton, A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica (1989) 357–384.

[10] J. Hamilton, Analysis of time series subject to changes in regime, J. Econometrics 45 (1–2) (1990) 39–70.

[11] J. Hamilton, Regime switching models, in: Macroeconometrics and Time Series Analysis, Springer, 2010, pp. 202–209.

[12] A. Koloydenko, M. Käärik, J. Lember, On adjusted Viterbi training, Acta Appl. Math. 96 (1) (2007) 309–326.

[13] A. Koloydenko, J. Lember, Infinite Viterbi alignments in the two state hidden Markov models, Acta Comment. Univ. Tartu. Math. 12 (2008) 109–124.

[14] A. Koloydenko, J. Lember, Bridging Viterbi and posterior decoding: A generalized risk approach to hidden path inference based on hidden Markov models, J. Mach. Learn. Res. 15 (2014) 1–58.

[15] K. Kuljus, J. Lember, Asymptotic risks of Viterbi segmentation, Stochastic Process. Appl. 122 (9) (2012) 3312–3341.

[16] K. Kuljus, J. Lember, On the accuracy of the MAP inference in HMMs, Methodol. Comput. Appl. Probab. 18 (3) (2016) 597–627.

[17] P. Lanchantin, J. Lapuyade-Lahorgue, W. Pieczynski, Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise, Signal Process. (2011) 163–175.

[18] J. Lember, On approximation of smoothing probabilities for hidden Markov models, Stat. Probab. Lett. 81 (2) (2011) 310–316.

[19] J. Lember, A. Koloydenko, The adjusted Viterbi training for hidden Markov models, Bernoulli 14 (1) (2008) 180–206.

[20] J. Lember, A. Koloydenko, A constructive proof of the existence of Viterbi processes, IEEE Trans. Inform. Theory 56 (4) (2010) 2017–2033.

[21] J. Lember, K. Kuljus, A. Koloydenko, Theory of segmentation, in: P. Dymarsky (Ed.), Hidden Markov Models, Theory and Applications, InTech, 2011, pp. 51–84.

[22] J. Lember, H. Matzinger, J. Sova, F. Zucca, Lower bounds for moments of global scores of pairwise Markov chains, Stochastic Process. Appl. 128 (5) (2018) 1678–1710.

[23] S.P. Meyn, R. Tweedie, Markov Chains and Stochastic Stability, Cambridge University Press, 2009.

[24] W. Pieczynski, Pairwise Markov chains, IEEE Trans. Pattern Anal. Mach. Intell. 25 (5) (2003) 634–639.

[25] X.T. Tong, R. van Handel, Ergodicity and stability of the conditional distributions of nondegenerate Markov chains, Ann. Appl. Probab. 22 (4) (2012) 1495–1540.

[26] C. Yau, C. Holmes, A decision-theoretic approach for segmental classification, Ann. Appl. Stat. 7 (3) (2013) 1814–1835.

[27] J. Yoo, On factor maps that send Markov measures to Gibbs measures, J. Stat. Phys. 141 (6) (2010) 1055–1070.