Infinite Viterbi alignments in the two state hidden Markov models

ALEXEY KOLOYDENKO AND JÜRI LEMBER

Abstract. Since the early days of digital communication, hidden Markov models (HMMs) have now been routinely used in speech recognition, processing of natural languages, images, and in bioinformatics. An HMM \((X_i, Y_i)_{i\geq 1}\) assumes observations \(X_1, X_2, \ldots\) to be conditionally independent given an “explanatory” Markov process \(Y_1, Y_2, \ldots\), which itself is not observed; moreover, the conditional distribution of \(X_i\) depends solely on \(Y_i\). Central to the theory and applications of HMM is the Viterbi algorithm to find a maximum a posteriori estimate \(q_1: n = (q_1, q_2, \ldots, q_n)\) of \(Y_{1:n}\) given the observed data \(x_{1:n}\). Maximum a posteriori paths are also called Viterbi paths or alignments. Recently, attempts have been made to study the behavior of Viterbi alignments of HMMs with two hidden states when \(n\) tends to infinity. It has indeed been shown that in some special cases a well-defined limiting Viterbi alignment exists. While innovative, these attempts have relied on rather strong assumptions. This work proves the existence of infinite Viterbi alignments for virtually any HMM with two hidden states.

1. Introduction

We consider hidden Markov models (HMM) \((Y, X)\) with two hidden states. Namely, \(Y\) represents the hidden process \(Y_1, Y_2, \ldots\), which is an irreducible aperiodic Markov chain with state space \(S = \{a, b\}\). In particular, the transition probabilities \(P = (p_{lm})\), \(l, m \in S\), are positive and the stationary distribution \(\pi = \pi P\) is unique. For technical convenience, \(Y_1\) is assumed to follow \(\pi\), however, the results of the paper hold for arbitrary initial distributions. To every state \(l \in S\) there corresponds an emission distribution \(P_l\) on \(X = \mathbb{R}^d\). Given a realization \(y_{1:\infty} \in S^\infty\) of \(Y\), the observations \(X_{1:\infty} := X_1, X_2, \ldots\)
are generated as follows. If $Y_i = a$ (resp. $Y_i = b$), then $X_i$ is distributed according to $P_a$ (resp. $P_b$) and independently of everything else. We refer to this model as the (general) 2-state HMM.

In Cappé et al. (2005), HMMs are called “one of the most successful statistical modelling ideas that have [emerged] in the last forty years”. Since their classical application to digital communication in 1960s (see further references in Cappé et al. (2005)), HMMs have had a defining impact on the mainstream research in speech recognition Huang et al. (1990); Jelinek (1976, 2001); McDermott and Hazen (2004); Ney et al. (1994); Padmanabhan and Picheny (2002); Rabiner and Juang (1993); Rabiner et al. (1986); Shu et al. (2003); Steinbiss et al. (1995); Ström et al. (1999), natural language models Ji and Bilmes (2006); Och and Ney (2000), and more recently computational biology Durbin et al. (1998); Eddy (2004); Lomsadze et al. (2005); Krogh (1998). Thus, for example, DNA regions can be labeled as $a$, “coding”, or $b$, “non-coding”, with $P_a$ and $P_b$ representing the respective distributions on the \{A,C,G,T\} alphabet.

Given observations $x_{1:n} := x_1, \ldots, x_n$, and treating the hidden states $y_{1:n} := y_1, \ldots, y_n$ as parameters, inference in HMMs typically involves $v(x_{1:n})$, a maximum a posteriori (MAP) estimate of $Y_{1:n}$. It has now been recognized that “[i]n spite of the theoretical and practical importance of the MAP path estimator, very little is known about its properties” Caliebe (2006). The same estimates are also known as Viterbi, or forced alignments and can be efficiently computed by a dynamic programming algorithm also bearing the name of Viterbi. When substituted for true $y_{1:n}$ in the likelihood function $\Lambda(y_{1:n}; x_{1:n}, \psi)$, Viterbi alignments can also be used to estimate $\psi$, any unknown free parameters of the model. Starting with an initial guess $\psi^{(0)}$ and alternating between maximization of the likelihood $\Lambda(y_{1:n}; x_{1:n}, \psi)$ in $y_{1:n}$ and $\psi$ is at the core of Viterbi training (VT), or extraction Jelinek (1976), also known as segmental K-means Juang and Rabiner (1990); Ephraim and Merhav (2002). Resulting estimates $\hat{\psi}_{VT}(x_{1:n}, \psi^{(0)})$ are known to be different from the maximum likelihood (ML) estimates $\hat{\psi}_{ML}(x_{1:n}, \psi^{(0)})$ which in this case are most commonly delivered by the EM procedure Baum and Petrie (1966); Ephraim and Merhav (2002); Bilmes (1998). Even if $\psi$ were known, Viterbi alignments $v(x_{1:n}; \psi)$ would typically differ from true paths $y_{1:n}$, and the long-run properties of $v(x_{1:n}; \psi)$ are not obvious Caliebe and Rösler (2002); Caliebe (2006); Lember and Koloydenko (2007, 2008); Koloydenko et al. (2007). Furthermore, Lember and Koloydenko (2007, 2008); Koloydenko et al. (2007) propose a hybrid of VT and EM which takes into account the asymptotic discrepancy between $\hat{\psi}_{ML}(x_{1:n}, \psi^{(0)})$ and $\hat{\psi}_{VT}(x_{1:n}, \psi^{(0)})$ in order to increase computational and statistical efficiencies of estimation of $\psi$ for $n$ large. Thus or otherwise, an important question is how to find the asymptotic properties of Viterbi alignments, given that $(n+1)^{th}$ observation can
in principle change the previous alignment entirely, i.e. $v(x_{1:n+1})_i \neq v(x_{1:n})_i$, $1 \leq i \leq n$? Do the Viterbi alignments then admit well-defined extensions? We answer this question positively in Lember and Koloydenko (2008) for general HMMs (in particular, allowing more than two hidden states) by constructing proper infinite Viterbi alignments. Generalizing and clarifying related results of Caliebe and Rösler (2002); Caliebe (2006), the approach in Lember and Koloydenko (2008) is to extend alignments piecewise, separating individual pieces by nodes (see §2 below). Although the construction is natural, a detailed formal proof of its correctness for general HMMs is rather long and requires certain mild technical assumptions. This paper, on the other hand, shows that in the special case of two state HMMs, the existence of infinite Viterbi alignments needs no special assumptions and can be proven considerably more easily. The results of this paper essentially complete and generalize those of Caliebe and Rösler (2002); Caliebe (2006).

2. Preliminaries

Let $\lambda$ be a suitable $\sigma$-finite reference measure on $\mathbb{R}^d$ so that $P_a$ and $P_b$ have densities with respect to $\lambda$. For example, $\lambda$ can be a Lebesgue measure, or, as in the case of discrete observations, a counting measure. Thus, let $f_a$ and $f_b$ be the densities of $P_a$ and $P_b$, respectively. Throughout the rest of the paper, we assume that $P_a \neq P_b$ or, equivalently,

$$\lambda\{x \in X : f_a(x) \neq f_b(x)\} > 0. \quad (1)$$

Assumption (1) is natural since there would be no need to model the observation process by an HMM should the emission distributions coincide. Note also that unlike in the general case, the positivity of the transition probabilities is also a natural assumption for the two state HMMs. No more assumption on the HMM is made in this paper. In particular, unlike Caliebe (2006); Caliebe and Rösler (2002), we do not assume the square integrability of $\log(f_a/f_b)$, or equality of the supports of $P_a$ and $P_b$. However, the latter condition is not very restrictive, since for the two state HMMs with unequal supports the existence of infinite Viterbi alignments follows rather trivially (Corollary 2.1).

Thus, for any $n \geq 1$ and any $x_{1:n} \in X^n$ and $y_{1:n} \in S^n$, the likelihood $\Lambda_{x,y}(y_{1:n}; x_{1:n})$ is given by

$$P(Y_{1:n} = y_{1:n}) \prod_{i=1}^{n} f_{y_i}(x_i), \text{ where } P(Y_{1:n} = y_{1:n}) = \pi_{y_1} \prod_{i=2}^{n} p_{y_{i-1}y_i}. $$

Since estimation of $\psi$ is not a goal of this paper, the dependence on $\psi$ is suppressed. Decomposition (2) and recursion (3) below provide a basis for the Viterbi algorithm to compute alignments. Namely, for all $u \in \{1, 2, \ldots, n\}$,
\[ n - 1 \}, \]

\[
\max_{y_{1:n} \in S^n} \Lambda_\pi(y_{1:n}; x_{1:n}) = \max_{l \in S} \left[ \delta_u(l) \times \max_{y_{u+1,n} \in S^{n-u}} \Lambda_{\pi_l}(y_{u+1:n}; x_{u+1:n}) \right], \tag{2}
\]

where \((p_l)\) is the transition distribution given state \(l \in S\), and the scores

\[
\delta_u(l) := \max_{y_{1:n-1} \in S^{n-1}} \Lambda((y_{1:n-1}, l); x_{1:n}), \quad l = a, b,
\]

are defined for all \(u \geq 1\), and \(x_{1:n} \in \mathcal{X}^u\). Thus, \(\delta_u(l)\) is the maximum of the likelihood of the paths terminating at \(u\) in state \(l\). Note that \(\delta_1(l) = \pi_l f_1(x)\) and \(\delta_u(l)\) depends on \(x_{1:u}\). Further,

\[
\delta_{u+1}(a) = \max \{ \delta_u(a) p_{aa}, \delta_u(b) p_{ba} \} f_a(x_{u+1}), \quad \delta_{u+1}(b) = \max \{ \delta_u(a) p_{ab}, \delta_u(b) p_{bb} \} f_b(x_{u+1}), \quad u \geq 1.
\tag{3}
\]

**Example 2.1.** Let \(X_1, X_2, \ldots\) be i.i.d. following a mixture distribution \(\pi_a P_a + \pi_b P_b\) with density \(\pi_a f_a(x; \theta_a) + \pi_b f_b(x; \theta_b)\) and mixing weights \(\pi_a, \pi_b > 0\). Such a sequence is an HMM with the transition probabilities \(\pi_a = p_{aa} = p_{ba}, \pi_b = p_{bb} = p_{ab}\). In this special case the alignment is easy to exhibit. Indeed, in this case recursion (3) writes for any \(u \geq 1\)

\[
\delta_{u+1}(a) = c \pi_a f_a(x_{u+1}), \quad \delta_{u+1}(b) = c \pi_b f_b(x_{u+1}), \tag{4}
\]

where \(c = \max \{ \delta_u(a), \delta_u(b) \}\). Hence, the alignment \(v(x_{1:n})\) can be obtained pointwise as follows:

\[ v(x_{1:n}) = (v(x_1), \ldots, v(x_n)), \quad \text{where} \quad v(x) = \arg \max \{ \pi_a f_a(x), \pi_b f_b(x) \}. \]

Equivalently (ignoring possible ties), using a generalized Voronoi partition \(\mathcal{X} = \mathcal{X}_a \cup \mathcal{X}_b\) with

\[ \mathcal{X}_a = \{ x \in \mathcal{X} : \pi_a f_a(x) \geq \pi_b f_b(x) \}, \quad \mathcal{X}_b = \{ x \in \mathcal{X} : \pi_b f_b(x) > \pi_a f_a(x) \}, \]

\[ v(x) = a \text{ if and only if } x \in \mathcal{X}_a, \quad \text{and otherwise (i.e. } x \in \mathcal{X}_b \text{) } v(x) = b. \]

Generally, it follows from (3) that, if

\[ \delta_u(a) p_{aa} > \delta_u(b) p_{ba}, \quad \delta_u(a) p_{ab} > \delta_u(b) p_{bb}, \tag{5} \]

for some \(u, 1 \leq u\), and some \(x_{1:u} \in \mathcal{X}^u\), then for any \(n > u\) and for any extension \(x_{u+1:n} \in \mathcal{X}^{n-u}\), the Viterbi alignment goes through state \(a\) at time \(u\). Namely, truncation \(v(x_{1:n})_{1:u}\) coincides with the Viterbi alignment \(v(x_{1:u})\) (indeed, (5) implies \(\delta_u(a) > \delta_u(b)\)). Thus, under condition (5), maximization of \(\Lambda_\pi((y_{1:n}, l); x_{1:n})\) can be reset at time \(u\) by clearing \(x_{1:u}\) from the memory, retaining \(v_{1:u}\), and replacing the initial distribution \(\pi\) by \((p_u)\) for further maximization of \(\Lambda_{\pi_u}(y_{u+1:n}; x_{u+1:n})\). Following Lember and Koloydenko (2008), if condition (5) holds, then \(x_u\) is called a strong a-node (of realization \(x_{1:n}, n > u\)), where “strong” refers to the inequalities in (5) being strict.

Suppose \(x_{1:u}\) contains infinitely many strong a-nodes at times \(u_1 < u_2 < \ldots\). Let \(v^1 = v(x_{1:u_1})\), and let \(v^k\) maximize \(\Lambda_{\pi_{u_k}}(y_{u_k+1:n}; x_{u_k+1:n})\), for
all \( k \geq 2 \). Then, concatenation \((v^1, v^2, v^3, \ldots)\) is naturally called the infinite piecewise Viterbi alignment Lember and Koloydenko (2008). Apparently, the almost sure existence of our infinite alignments directly depends on the existence of infinitely many (strong) nodes. At the same time, whether or not \( x_u \) is a node depends on \( x_{1:u} \) and hence is difficult to verify directly. Fortunately, in many cases \( x_u \) is guaranteed to be a node based on several preceding observations \( x_{u-m:u}, 1 \leq m < u \), ignoring the rest. Specifically, suppose for example that \( x \in \mathcal{X} \) is such that
\[
p_{ia} f_a(x)p_{aj} > p_{ib} f_b(x)p_{bj}, \quad \forall i,j \in S.
\] (6)

It is easy to check that for any \( u \geq 2 \), \( x_u = x \) is a strong \( a \)-node for any \( x_{1:u-1} \). Hence, if \( x_{1:\infty} \) contains infinitely many observations satisfying (6), then \( x_{1:\infty} \) also contains infinitely many strong nodes. This previous condition in its turn is met provided
\[
\lambda(\{x \in \mathcal{X} : p_{ia} f_a(x)p_{aj} > p_{ib} f_b(x)p_{bj}, \quad \forall i,j \in S\}) > 0.
\] (7)

Indeed, since our underlying Markov chain \( Y \) is ergodic, it is rather easy to see that \( \mathcal{X} \) is ergodic as well Ephraim and Merhav (2002); Genon-Catalot et al. (2000); Leroux (1992). Also, (7) implies that
\[
P_a (\{x \in \mathcal{X} : p_{ia} f_a(x)p_{aj} > p_{ib} f_b(x)p_{bj}, \quad \forall i,j \in S\}) > 0.
\]

Thus, it follows from ergodicity of \( \mathcal{X} \) that almost every realization of \( \mathcal{X} \) has infinitely many elements satisfying (6) and, hence infinitely many strong nodes. We have thus proved the following lemma.

**Lemma 2.1.** Assume that (7) holds. Then almost every sequence of observations \( x_{1:\infty} \) has infinitely many strong \( a \)-nodes.

(Clearly, interchanging \( a \) and \( b \) gives the same results in terms of \( b \)-nodes.) Lemma 2.1 is essentially Theorem 1 in Caliebe and Rösler (2002) (disregarding a misprint in the statement). Condition (7) holds for many two-state HMMs including the so-called additive Gaussian noise model Caliebe (2006), where the emission distributions are Gaussian. Another trivial example is the model with unequal supports of \( P_a \) and \( P_b \). Indeed, in that case (7) holds (at least up to swapping \( a \) and \( b \)). Hence, we have the following corollary.

**Corollary 2.1.** If the supports of \( P_a \) and \( P_b \) are not equal, then almost every sequence of observations has infinitely many strong nodes.

The goal of this work is essentially to remove condition (7) from Lemma 2.1.

To this end, following Lember and Koloydenko (2008), we call an observation satisfying (6) an \( a \)-barrier of length 1. More generally, a block of observations \( z_{1:k} \in \mathcal{X}^k \) is called a (strong) barrier of length \( k \geq 1 \) if for every \( m \geq 0 \) and \( x_{1:m} \in \mathcal{X}^m \), \( z_{1:k} \) contains a (strong) node of realization \((x_{1:m}, z_{1:k})\). In Lember and Koloydenko (2008), we prove the existence of
infinitely many barriers for a very general class of HMMs. For the two-state HMMs, the conditions of our result in Lember and Koloydenko (2008) are given by (8) and (9) below:

\[
P_a \left( \{ x \in X : f_a(x) \max \{ p_{aa}, p_{ba} \} > f_b(x) \max \{ p_{bb}, p_{ab} \} \} \right) > 0, \tag{8}
\]

\[
P_b \left( \{ x \in X : f_b(x) \max \{ p_{bb}, p_{ab} \} > f_a(x) \max \{ p_{aa}, p_{ba} \} \} \right) > 0. \tag{9}
\]

To achieve our goal, we will first prove the same result for the two-state HMM under the relaxed assumption that (8) or (9) holds. As we shall see below (Lemma 3.1), in our two-state HMM one of these conditions is automatically satisfied and, moreover, all barriers are strong. Hence, occurrence of infinitely many strong barriers in this case will be shown (Theorem 4.1) to require no additional assumptions.

Finally, if a node is not strong and \( v(x_1:n) \) is not unique, an alignment might exist that does not go through this node. Such type of pathologies cause technical inconveniences in defining an infinite Viterbi alignment and are treated in Lember and Koloydenko (2008). Fortunately, unlike in the general case, in the case of two-state HMMs almost every realization has infinitely many strong nodes (Theorem 4.1). This allows for a simple resolution of the non-uniqueness in the case of two-state HMMs.

3. Main results

3.1. Three types of the two-state HMM. The following three cases exhaust all the possibilities:

- (1) \( p_{aa} > p_{ba} \) (\( \Leftrightarrow \) \( p_{bb} > p_{ab} \));
- (2) \( p_{aa} < p_{ba} \) (\( \Leftrightarrow \) \( p_{bb} < p_{ab} \));
- (3) \( p_{aa} = p_{ba} \) (\( \Leftrightarrow \) \( p_{bb} = p_{ab} \)).

From the definition of nodes, it follows that \( x_u \) is not a node only in one of the following two cases:

\[
(A) \left\{ \begin{array}{l}
\delta_u(a)p_{aa} > \delta_u(b)p_{ba} \\
\delta_u(b)p_{bb} > \delta_u(a)p_{ab}
\end{array} \right. \quad \text{or} \quad (B) \left\{ \begin{array}{l}
\delta_u(b)p_{ba} > \delta_u(a)p_{aa} \\
\delta_u(a)p_{ab} > \delta_u(b)p_{bb}
\end{array} \right.
\]

Case (A) is equivalent to

\[
\frac{p_{bb}}{p_{ab}} > \frac{\delta_u(a)}{\delta_u(b)} > \frac{p_{ba}}{p_{aa}}, \tag{10}
\]

and case (B) is equivalent to

\[
\frac{p_{bb}}{p_{ab}} < \frac{\delta_u(a)}{\delta_u(b)} < \frac{p_{ba}}{p_{aa}}. \tag{11}
\]

Thus, in case (A), we have \( \delta_{u+1}(a) = \delta_u(a)p_{aa}f_a(x_{u+1}) \) and \( \delta_{u+1}(b) = \delta_u(b)p_{bb}f_b(x_{u+1}) \), so that for any \( n > u \), the Viterbi alignment \( v(x_{1:n}) \) must satisfy \( v(x_{1:n})_u = v(x_{1:n})_{u+1} \). Similarly, in case (B), we have \( \delta_{u+1}(a) = \delta_u(b)p_{ba}f_a(x_{u+1}) \) and \( \delta_{u+1}(b) = \delta_u(a)p_{ab}f_b(x_{u+1}) \), i.e. \( v(x_{1:n})_u \neq v(x_{1:n})_{u+1} \).
Evidently, case 1 and case (B) are mutually exclusive, and so are case 2 and case (A). Therefore, if the transition matrix satisfies the conditions of case 1, then $x_n$ is not a node if and only if conditions (A) are fulfilled. This implies that in case 1, nodes are the only possibility for $v(x_{1:n})$ to change state. On the other hand, if the transition matrix satisfies the conditions of case 2, then $x_n$ is not a node if and only if (B) holds. Hence, in case 2 nodes are the only possibility for $v(x_{1:n})$ to remain in one state. Case 3 corresponds to the mixture model (see Example 2.1 above). Apparently (4), every observation is a node in this case (see also Figure 1 below).

Let us now examine conditions (8) and (9). From equation (1), it follows that

$$\lambda(\{x \in \mathcal{X} : f_a(x) > f_b(x)\}) > 0, \quad \lambda(\{x \in \mathcal{X} : f_a(x) < f_b(x)\}) > 0$$

(12)
and, for any $\alpha > \beta > 0$,
\[
\lambda \left( \{ x \in \mathcal{X} : \alpha f_a(x) > \beta f_b(x) \} \right) > 0 \quad \Leftrightarrow \quad P_a \left( \{ x \in \mathcal{X} : \alpha f_a(x) > \beta f_b(x) \} \right) > 0,
\]
\[
\lambda \left( \{ x \in \mathcal{X} : \alpha f_b(x) > \beta f_a(x) \} \right) > 0 \quad \Leftrightarrow \quad P_b \left( \{ x \in \mathcal{X} : \alpha f_b(x) > \beta f_a(x) \} \right) > 0.
\]  

Therefore, we have the following lemma.

**Lemma 3.1.** Any two state HMM satisfies at least one of the conditions (8) and (9).

**Proof.** In case 1, (8) and (9) are equivalent to
\[
P_a \left( \{ x \in \mathcal{X} : f_a(x)p_{aa} > f_b(x)p_{bb} \} \right) = P_a \left( \left\{ x \in \mathcal{X} : \frac{f_b(x)p_{ab}}{f_a(x)p_{aa}} < 1 \right\} \right) > 0,
\]
\[
P_b \left( \{ x \in \mathcal{X} : f_b(x)p_{bb} > f_a(x)p_{aa} \} \right) = P_b \left( \left\{ x \in \mathcal{X} : \frac{f_a(x)p_{ba}}{f_b(x)p_{bb}} < 1 \right\} \right) > 0,
\]
respectively. If $p_{aa} = p_{bb}$, then (12) implies that both (15) and (16) are satisfied, and hence both (8) and (9) hold. If $p_{aa} > p_{bb}$, then (15), and subsequently (8), follow from (13). If $p_{aa} < p_{bb}$, then (16), and subsequently (9), follow from (14). Hence, at least one of the assumptions (8), (9) is always guaranteed to hold.

In case 2, (8) and (9) are equivalent to
\[
P_a \left( \{ x \in \mathcal{X} : f_a(x)p_{ba} > f_b(x)p_{ab} \} \right) = P_a \left( \left\{ x \in \mathcal{X} : \frac{f_b(x)p_{ab}}{f_a(x)p_{ba}} < 1 \right\} \right) > 0,
\]
\[
P_b \left( \{ x \in \mathcal{X} : f_b(x)p_{ab} > f_a(x)p_{ba} \} \right) = P_b \left( \left\{ x \in \mathcal{X} : \frac{f_a(x)p_{ba}}{f_b(x)p_{ab}} < 1 \right\} \right) > 0,
\]
respectively. Again, if $p_{aa} = p_{bb}$, then (17) and (18) both hold without further assumptions. If $p_{aa} > p_{bb}$, then (17) is automatically satisfied. Likewise, (18) holds if $p_{aa} < p_{bb}$. Hence, one of the assumptions (8), (9) is always guaranteed to hold.

In case 3, (8) and (9) write
\[
P_a \left( \{ x \in \mathcal{X} : f_a(x)\pi_a > f_b(x)\pi_b \} \right) > 0,
\]
\[
P_b \left( \{ x \in \mathcal{X} : f_b(x)\pi_b > f_a(x)\pi_a \} \right) > 0.
\]
Assume $\pi_a \geq \pi_b$. Then, (12) implies $\lambda \left( \{ x \in \mathcal{X} : \pi_a f_a(x) > \pi_b f_b(x) \} \right) > 0$, which in turn implies (19). □
Finally, we state and prove the main results for each of the three cases.

### 3.2. Case 1.

First, note that condition (7) in this case is equivalent to

\[ \lambda \left( \{ x \in \mathcal{X} : p_{ba} f_a(x) p_{ab} > p_{bb} f_b(x) p_{bb} \} \right) > 0. \]  

(21)

As mentioned in §2, condition (7) need not hold in general. Nonetheless, for the two-state HMM, we have the following lemma.

**Lemma 3.2.** In case 1, almost every realization of the two-state HMM has infinitely many strong barriers.

**Proof.** Without loss of generality, assume \( p_{aa} \geq p_{bb} \). Then (15) holds implying that there exists \( \epsilon > 0 \) such that

\[ P_a( \mathcal{X}_a ) > 0, \quad \text{where} \quad \mathcal{X}_a := \left\{ x \in \mathcal{X} : \frac{f_b(x) p_{ba}}{f_a(x) p_{aa}} < 1 - \epsilon \right\}. \]

Let integer \( k \) be sufficiently large for \((1 - \epsilon)^k < p_{ab} p_{ba} / (p_{aa} p_{bb})\) to hold. Then every sequence \( z_{1:k} \in \mathcal{X}_a^k \) satisfies

\[ \prod_{j=1}^{k} \frac{f_b(z_j) p_{bb}}{f_a(z_j) p_{aa}} < (1 - \epsilon)^k < \frac{p_{ab} p_{ba}}{p_{aa} p_{bb}}. \]  

(22)

Let \( u > k \) be arbitrary and let \( z_{0:k} \in \mathcal{X}_a^{k+1} \) be the last \( k + 1 \) observations in a generic sequence \( x_{1:u} \in \mathcal{X}_a^{u-k-1} \times \mathcal{X}_a^{k+1} \). To shorten the notation, we write \( d_j(z_i) \) for \( \delta_{u-k+1}(j) \) for every \( i = 0, 1, \ldots, k \), \( j = a, b \). Next, we show that \( x_{u-k:u} \) contains at least one strong node, and consequently, \( z_{0:k} \) is a strong barrier. Indeed, if none of the observations \( x_{u-k:u} \) were a strong a-node, then we would have

\[ d_b(z_k) = d_b(z_0) \prod_{j=1}^{k} f_b(z_j) p_{bb}. \]

Similarly, if none among the observations \( x_{u-k+1:u} \) were a strong b-node, we would have

\[ \delta_u(a) \geq \delta_{u-k}(b) p_{ba} \left( \prod_{j=1}^{k} f_a(z_j) \right) p_{aa}^{k-1}. \]

Hence,

\[ \frac{\delta_u(b)}{\delta_u(a)} \leq \frac{\delta_{u-k}(b) p_{bb} \left( \prod_{j=1}^{k} f_b(z_j) \right) p_{bb}^{k-1}}{\delta_{u-k}(b) p_{ba} \left( \prod_{j=1}^{k} f_a(z_j) \right) p_{aa}^{k-1}} = \frac{\prod_{j=1}^{k} (f_b(z_j) p_{bb}) p_{aa}}{\prod_{j=1}^{k} (f_a(z_j) p_{aa}) p_{ba}} \]

and by (22)

\[ \frac{\delta_u(b)}{\delta_u(a)} < \frac{p_{ab}}{p_{bb}}. \]
that contradicts (10). Thus, at least one of $x_{u-k:n}$ must be a strong node. Since $P_a(X_a) > 0$, by ergodicity of HMM, almost every realization has infinitely many barriers $z_{0:k} \in X_a^{k+1}$, implying also that every realization has infinitely many strong nodes.

The next theorem refines the previous result.

**Theorem 3.1.** Suppose the (transition matrix of the) two-state HMM meets the condition of case 1. If $p_{aa} \geq p_{bb}$, then almost every realization has infinitely many strong $a$-barriers. (If $p_{aa} \leq p_{bb}$, then almost every realization has infinitely many strong $b$-barriers.)

**Proof.** Let $p_{aa} \geq p_{bb}$ and use the notation of the proof of Lemma 3.2. First, we show that none of the observations $x_{k-u+1:n}$ is a $b$-node. Indeed, since

$$d_b(z_1) = \max\{d_a(z_0)p_{ab}, d_b(z_0)p_{bb}\}f_b(z_1),$$

at least one of the following two inequalities must hold:

$$p_{ab}f_b(z_1)p_{ba} \geq p_{aa}f_a(z_1)p_{aa}, \quad p_{bb}f_b(z_1)p_{ba} \geq p_{ba}f_a(z_1)p_{aa}$$

(23)

in order for $x_{u-k+1}$ to be a $b$-node. However, (15) implies that $p_{ba}f_a(z_1)p_{aa} > p_{bb}f_b(z_1)p_{ba}$ and, since $p_{bb} > p_{ab}$, we have $p_{bb}f_b(z_1)p_{ba} > p_{ba}f_a(z_1)p_{aa}$. Hence, neither of the two inequalities (23) holds. Thus, $x_{u-k+1}$ cannot be a $b$-node, and the same argument shows that none of the subsequent observations $x_{u-k+2}, \ldots, x_u$ can be a $b$-node either.

The argument of the proof of Lemma 3.2 then shows that one of the observations in $x_{u-k:n}$ is a strong $a$-node and therefore $z_{0:k}$ is a strong $a$-barrier. The ergodic argument finishes the proof. (The same argument with $a$ and $b$ swapped establishes the second part of the theorem.)

Note that the condition $p_{bb} \geq p_{aa}$ is sufficient but not necessary for (16) to hold. In fact, for many 2-state HMMs, such as the one with additive white Gaussian noise, both (15) and (16) hold for any (positive) values of $p_{aa}$ and $p_{bb}$. On the other hand, it might happen that one of the conditions (15) and (16), say (16), fails. This would mean $P_b\left(\{x \in \mathcal{X} : p_{bb}f_b(x) > p_{aa}f_a(x)\}\right) = 0$, or, equivalently,

$$\lambda(\{x \in \mathcal{X} : p_{bb}f_b(x) > p_{aa}f_a(x)\}) = 0.$$  

(24)

**Corollary 3.1.** In case 1, equation (24) implies that almost every sequence of observations has infinitely many strong $a$-barriers and no strong $b$-nodes. Furthermore, equation (24) in case 1 implies that for almost every realization, if a $b$-node does occur, it occurs before the first $a$-node.

**Proof.** From the proof of Theorem 3.1, it follows that no observation $x \in \mathcal{X}$ such that $p_{bb}f_b(x) \leq p_{aa}f_a(x)$ (i.e. from the complement of the set in (24)) can be a strong $b$-node; a closer inspection of the proof actually shows that even a weak (i.e. not strong) $b$-node cannot occur after an $a$-node (since in
Corollary 3.1 in its turn implies that starting with the first strong $a$-node onward, the Viterbi alignment $v(x_{1:n})$ stays in state $a$. As we have already mentioned, Viterbi alignments need not be unique (see Lember and Koloydenko (2008)), i.e. ties are possible in general, and in this case, in particular, they are possible up until the first strong $a$-node. However, the impossibility of strong $b$-nodes in this case implies that the ties can be broken in favor of $a$, resulting in the constant all $a$ alignment.

Theorem 3.1 is a generalization of Theorem 7 in Caliebe (2006), which basically states that in case 1, if (15) and (16) hold, then under some additional assumptions (equal supports of $P_a$ and $P_b$ and further conditions $A2$), almost every realization has infinitely many nodes. Thus, Caliebe (2006) stops short of realizing that in case 1 conditions (15) and (16) alone ensure the existence of $a$- and $b$-nodes. This results in Caliebe (2006) invoking Theorem 2 of Caliebe and Rösler (2002) to prove the existence of nodes, hence superfluous assumptions $A1$, $A2$. Also the proof of Theorem 7 in Caliebe and Rösler (2002) could be simplified and shortened with the help of the notions of nodes and barriers. Finally, Corollary 3.1 generalizes Theorems 8 and 9 of Caliebe (2006).

3.3. Case 2. Recall that we have been proving the existence of barriers without condition (7). Note that in case 2, condition (7) becomes

$$\lambda \left\{ x \in X : p_{aa} f_a(x) p_{aa} > p_{ab} f_b(x) p_{ba} \right\} > 0.$$ 

Recall (§2) also that interchanging $a$ with $b$ gives a similar condition for strong $b$-nodes to occur infinitely often in almost every realization.

It follows from (12) that for some $\epsilon > 0$, the sets

$$X_a := \{ x \in X : f_a(x)(1-\epsilon) > f_b(x) \}, \quad X_b := \{ x \in X : f_a(x) < f_b(x)(1-\epsilon) \}$$

both have positive $\lambda$-measure. Hence $P_a(X_a) > 0$ and $P_b(X_b) > 0$. Then, for $x_{1:2} \in X_a \times X_b$, the following holds:

$$\frac{f_b(x_1) f_a(x_2)}{f_a(x_1) f_b(x_2)} < (1-\epsilon)^2.$$ (25)

**Lemma 3.3.** In case 2, almost every realization has infinitely many strong barriers.

**Proof.** Let $X_a$ and $X_b$ be as above. Choose $k$ sufficiently large for

$$(1-\epsilon)^{2k} < \frac{P_{aa} P_{bb}}{P_{ba} P_{ab}}$$

to hold. Next, consider a sequence $z_{0:2k} \in X^{2k+1}$, where $z_0, z_{2i} \in X_a$, $z_{2i-1} \in X_b$, for every $i = 1, \ldots, k$. We show that for every $u > 2k$, every sequence
of observations $x_{1u} \in \mathcal{X}^u$ such that $x_{u-2k:u} = z_{0:2k}$, contains a strong node, making $z_{0:2k}$ a strong barrier.

The choice of $k$ and $z_{0:2k}$ implies
\[
\prod_{i=1}^k p_{ba} f_a(z_{2i-1}) p_{ab} f_b(z_{2i}) < (1 - \epsilon)^{2k} \frac{p_{bb} p_{aa}}{p_{ba} p_{ab}}. \tag{26}
\]

If there is no strong node among $x_{u-2k:u}$, then
\[
d_b(z_{2k}) = d_b(z_0) \prod_{i=1}^k p_{ba} f_a(z_{2i-1}) p_{ab} f_b(z_{2i})
\]
and
\[
d_a(z_{2k}) \geq d_b(z_0) \frac{p_{bb}}{p_{ab}} \prod_{i=1}^k p_{ab} f_b(z_{2i-1}) p_{ba} f_a(z_{2i}).
\]
Hence, by (26)
\[
d_b(z_{2k}) \leq \frac{d_b(z_0)}{d_a(z_{2k})} \prod_{i=1}^k p_{ba} f_a(z_{2i-1}) p_{ab} f_b(z_{2i}) < p_{aa} p_{ba} p_{bb}\prod_{i=1}^k p_{ab} f_b(z_{2i-1}) p_{ba} f_a(z_{2i})< p_{aa} p_{ba} p_{bb},\tag{29}
\]
which contradicts (11). \hfill \square

Next, we refine this result. Without loss of generality assume $p_{ba} \geq p_{ab}$. Therefore
\[
p_{ab} p_{aa} \geq p_{ba} p_{bb}, \tag{27}
\]
and also, for every $x \in \mathcal{X}_a$,
\[
p_{ba} f_a(x) > p_{ab} f_b(x). \tag{28}
\]
Hence, (17) holds. We multiply the right side of (28) by $p_{ba} p_{bb}$ and the left side by $p_{ab} p_{aa}$, and use (27) to obtain
\[
f_a(x) p_{aa} > f_b(x) p_{bb} \tag{29}.
\]
Finally, for $x \in \mathcal{X}_b$, we have
\[
f_a(x) < f_b(x). \tag{30}
\]

We will need the following lemma.

**Lemma 3.4.** Assume (in addition to being in case 2) that $p_{ab} \leq p_{ba}$.

a) In any pair of observations $z_{1:2} \in \mathcal{X}_a \times \mathcal{X}_b$, $z_1$ is not a $b$-node.

b) In any pair of observations $z_{2:3} \in \mathcal{X}_b \times \mathcal{X}_a$, if $z_2$ is a $b$-node, then $z_3$ is a strong $a$-node.

**Proof.** Assume that $p_{ab} \leq p_{ba}$, and consider a). First note that since we are in case 2, $z_1$ is a $b$-node if and only if
\[
d_b(z_1) p_{bb} \geq d_a(z_1) p_{ab}. \tag{31}
\]
Suppose first that $z_0$ is not a node, in which case $d_b(z_1) = d_a(z_0)p_{ab}f_b(z_1)$ and $d_a(z_1) = d_b(z_0)p_{ba}f_a(z_1)$. Then

$$d_a(z_1)p_{ab} = d_b(z_0)p_{ba}f_a(z_1)p_{ab} \geq d_a(z_0)p_{aa}f_a(z_1)p_{ab}$$

$$> d_a(z_0)p_{ab}f_b(z_1)p_{bb} = d_a(z_0)p_{ab}f_b(z_1)p_{bb} = d_b(z_1)p_{bb}.$$  

The first inequality above follows from the recursion property (3) of scores $\delta$, whereas the second one follows from (29). Thus, when $z_0$ is not a node, $z_1$ cannot be a $b$-node. Similarly, supposing that $z_0$ is an $a$-node, we obtain that $z_1$ is not a $b$-node. Suppose finally that $z_0$ is a $b$-node. Then $d_b(z_1) = d_b(z_0)p_{bb}f_b(z_1)$ and $d_a(z_1) = d_b(z_0)p_{ba}f_a(z_1)$. Applying consecutively $p_{bb} < p_{ab}$, (28) and $p_{bb} < p_{ab}$ again, we obtain: $p_{bb}f_b(z_1)p_{bb} < p_{bb}f_b(z_1)p_{bb} \leq p_{ba}f_a(z_1)p_{bb} < p_{ba}f_a(z_1)p_{ab}$. Thus, contrary to (31),

$$d_b(z_1)p_{bb} = d_b(z_0)p_{bb}f_b(z_1)p_{bb} < d_b(z_0)p_{ba}f_a(z_1)p_{ab} = d_a(z_1)p_{ab},$$

that is, $z_1$ is not a $b$-node in this case either.

Let us now prove b). If $z_2$ is a $b$-node, then $d_a(z_3) = d_b(z_2)p_{ba}f_a(z_3)$ and $d_b(z_3) = d_b(z_2)p_{bb}f_b(z_3)$. By (29), we now have

$$d_a(z_3)p_{aa} = d_b(z_2)p_{ba}f_a(z_3)p_{aa} > d_b(z_2)p_{bb}f_b(z_3)p_{ba} = d_b(z_3)p_{ba}.$$  

Similarly to the argument regarding $b$-nodes guaranteed by (31) above, we now have $d_a(z_3) > d_b(z_3)$, implying $d_a(z_3)p_{ab} > d_b(z_3)p_{bb}$. Thus $z_3$ is a strong $a$-node.

**Theorem 3.2.** If $p_{ba} \geq p_{ab}$, then almost every realization has infinitely many strong $a$-nodes. If $p_{ba} \leq p_{ab}$, then almost every realization has infinitely many strong $b$-nodes.

**Proof.** Assume again that $p_{ba} \geq p_{ab}$. Let $z_{0:2k}$ be as in the proof of Lemma 3.3 and attach one more element $z_{2k+1} \in X_b$ to the end. Thus, $z_{2i} \in X_a$ and $z_{2i+1} \in X_b$, $i = 0, 1, \ldots, k$.

From (the proof of) Lemma 3.3 we know that $z_{0:2k}$ contains at least one strong node. If this is an $a$-node, then the theorem is proven. Otherwise this is a $b$-node, which, according to part a) of Lemma 3.4, can only be among $z_1, z_3, \ldots, z_{2k-1}$. Applying part b) of Lemma 3.4 shows that there must also be a strong $a$-node $z_2, z_4, \ldots, z_{2k}$. Invoking ergodicity again finishes the proof.

Clearly, swapping $a$ and $b$ in the above discussion following the proof of Lemma 3.3, establishes the other part of the theorem.  

Inequality (27) guarantees (17). Often, the model is such that in addition to (17), (18) also holds. However, to apply the previous proof (i.e. of Theorem 3.2) to guarantee the simultaneous existence of infinitely many strong $a$ and $b$-nodes, we would need the following counterpart of (29):

$$P_b(\{x \in X : f_b(x)p_{ab} > f_a(x)p_{ba}, f_b(x)p_{bb} > f_a(x)p_{aa}\}) > 0,$$

which is stronger than (18). However, this previous condition is indeed often met,
resulting in infinitely many strong a- and b-nodes (in almost every realization $x_{1:\infty}$).

Lemma 3.3 appears without proof as Theorem 10 in Caliebe (2006). The author of Caliebe (2006) actually suggests that Theorem 10 and other results for case 2 are analogous to the corresponding results for case 1, mainly Theorem 7 (of the same work). It is further stated in Caliebe (2006) that the proofs of those results are not given as they “are very similar” to the corresponding proofs in case 1. Our present workings actually show that case 2 is quite dissimilar to case 1 (due to the fluctuating nature of the typical Viterbi alignment) and in particular requires a more careful treatment. Note that, even if Theorem 10 in Caliebe (2006) assumed (8) and (9) (as Theorem 7 in Caliebe (2006) does) to help one to prove this theorem by analogy to Theorem 7, it is still not clear how the two proofs could be very similar.

3.4. Case 3 (the mixture model). Recall that every observation in this case is a (not necessarily strong) node. Furthermore, every observation from \( \{ x \in X : \pi_a f_a(x) > \pi_b f_b(x) \} \) is a strong a node. Thus, we have the following counterpart of Theorems 3.1 and 3.2.

Theorem 3.3. If \( \pi_a \geq \pi_b \), then almost every realization has infinitely many strong a-nodes. If \( \pi_a \leq \pi_b \), then almost every realization has infinitely many strong b-nodes.

4. Conclusion

In summary, we have proved Theorem 4.1 stated below and providing a basis for the piecewise construction and asymptotic analysis of the Viterbi alignments of two-state HMMs.

Theorem 4.1. Almost every realization of the two-state HMM has infinitely many strong barriers. Furthermore

a) if the transition probabilities satisfy \( p_{aa} \geq p_{ba} \) then (almost every realization of) the chain has infinitely many strong s-barriers where s is such that \( p_{ss} = \max\{p_{aa}, p_{bb}\} \),

b) otherwise (i.e. if \( p_{aa} < p_{ba} \)) (almost every realization of) the chain has infinitely many strong s-barriers where s is such that \( p_{ts} = \max\{p_{ab}, p_{ba}\} \) (for some \( t \in S \)).

References


REFERENCES


**Mathematics Department, Royal Holloway University of London, Egham, TW20 0EX, United Kingdom**

*E-mail address: alexey.koloydenko@rhul.ac.uk*

**Institute of Mathematical Statistics, University of Tartu, Liivi 2, 50409 Tartu, Estonia**

*E-mail address: jyril@ut.ee*