

A SUFFICIENT CONDITION FOR REDUCING RECURSIONS IN HIDDEN MARKOV MODELS

Yun S. Song*

*Department of Statistics, University of Oxford,
1 South Parks Road, Oxford, OX1 3TG, UK*

Abstract

In hidden Markov models, the probability of observing a set of strings can be computed using recursion relations. We construct a sufficient condition for simplifying the recursion relations for a certain class of hidden Markov models. If the condition is satisfied, then one can construct a reduced recursion where the dependence on Markov states completely disappears. We discuss a specific example – namely, statistical multiple alignment based on the TKF-model – in which the sufficient condition is satisfied.

Keywords: Hidden Markov model; recursions; reduction; Markov states; statistical alignment

1 Introduction

Hidden Markov models (HMMs) are probabilistic models with a wide range of important applications. For instance, HMMs have been extensively studied in speech recognition since the early 1970s (see [5] for a review on this subject). An equally important area of application is computational biology, where HMMs have been applied, with much success, to investigate such diverse problems as sequence alignment and gene finding [1].

A fundamental problem in HMMs is computing the probability or the likelihood of observing a set of strings of symbols. There exist general recursion-based algorithms for computing such probabilities, and, in principle, dynamic programming techniques can be employed to solve the recursions of interest. In practice, however, there is a major obstacle. Namely, when each step in a recursion involves many terms, exactly solving the recursion becomes infeasible. As we elaborate later in Section 3, the obvious recursion one can write down depends explicitly on Markov states and therefore may involve many terms, especially so when there are many Markov states in the model with the same emission pattern. The primary goal of this paper is to construct a sufficient condition for completely eliminating the dependence of recursions on Markov states for a certain class of HMMs. The condition is defined in terms of the quantities which arise in state-dependent recursions, and, when they exist, reduced recursions are defined in terms of the quantities which appear in the condition.

*Current Address: Department of Computer Science, University of California at Davis, Davis, CA 95616, U.S.A.
E-mail: yssong@cs.ucdavis.edu

Stochastic processes are widely used to model the evolution of biological sequences. For example, Thorne, Kishino and Felsenstein have constructed a reversible Markov model of sequence evolution incorporating three main types of mutation events, namely substitutions, insertions and deletions of nucleotides or amino acids [7]. This model is often referred to as the TKF model of statistical alignment. In [3], the TKF model on a general binary tree has been reformulated as a multi-state hidden Markov model. Furthermore, using a completely different approach from that discussed in the present paper, it has been recently shown that, under the TKF model, there exists a reduced recursion for the joint probability of observing a given set of sequences at the leaves of a given phylogenetic tree [4]. Although properties of the TKF model are heavily used in [4], it has not been considered what must happen in the corresponding multi-state hidden Markov model for there to be a reduction. In this paper, we consider the HMM proposed in [3] and explicitly check, for small number of sequences, that the sufficient condition we construct here is indeed satisfied by that model. This result is in agreement with the finding of [4] that there exists a reduced recursion for the joint probability.

This paper is organised as follows. In Section 2, we introduce some notations to be used throughout this paper. In Section 3, we define the probability of our interest and describe state-dependent recursions in HMMs. The main result of this paper is discussed in Section 4, where a sufficient condition for reduction is constructed. In Section 5, we consider the aforementioned statistical alignment example where our sufficient condition is satisfied. We conclude with some remarks in Section 6.

2 Notations

Throughout this paper, we adopt the following notational convention. Let \mathcal{A} denote a finite alphabet of letters and let $\mathbf{S} = (S_1, S_2, \dots, S_r)$ be an r -tuple of strings, the length of the string S_i being L_i ; that is, $S_i = S_{i,1}, S_{i,2}, \dots, S_{i,L_i}$, where $S_{i,j} \in \mathcal{A}$ for all $1 \leq i \leq r$ and $1 \leq j \leq L_i$. In what follows, we let $\mathbf{L} = (L_1, L_2, \dots, L_r)$.

Example 1 Consider a finite alphabet consisting of two letters, say H and T . Suppose that $\mathbf{S} = (S_1, S_2)$, where $S_1 = HT$ and $S_2 = H$. Then, $\mathbf{L} = (2, 1)$. In later examples, we use this simple example to illustrate our notation.

Given a pair of r -tuples $\mathbf{x} = (x_1, x_2, \dots, x_r)$ and $\mathbf{y} = (y_1, y_2, \dots, y_r)$ of non-negative integers, we say that $\mathbf{x} \leq \mathbf{y}$ if $x_i \leq y_i$ for all $1 \leq i \leq r$. Let $\mathbf{K} = (K_1, K_2, \dots, K_r)$ be an r -tuple of non-negative integers satisfying $\mathbf{K} \leq \mathbf{L}$ and let $\mathbf{u} = (u_1, u_2, \dots, u_r)$ be an r -tuple of 0s and 1s. We use $\mathbf{S}[\mathbf{K}, \mathbf{u}]$ to denote an r -tuple of symbols, each being either a letter in \mathcal{A} or the empty string \emptyset , such that

$$\text{the } i^{\text{th}} \text{ component of } \mathbf{S}[\mathbf{K}, \mathbf{u}] = \begin{cases} S_{i,K_i}, & \text{if } u_i = 1, \\ \emptyset, & \text{if } u_i = 0. \end{cases}$$

Example 2 Let $\mathbf{K} = (2, 1)$. Then, for the 2-tuple of strings defined in Example 1, $\mathbf{S}[\mathbf{K}, \mathbf{u}] = (T, \emptyset)$ if $\mathbf{u} = (1, 0)$, $\mathbf{S}[\mathbf{K}, \mathbf{u}] = (\emptyset, H)$ if $\mathbf{u} = (0, 1)$, and $\mathbf{S}[\mathbf{K}, \mathbf{u}] = (T, H)$ if $\mathbf{u} = (1, 1)$.

We use $K_i \circ S_i$ to denote the prefix of S_i of length K_i ; that is,

$$K_i \circ S_i := \begin{cases} \emptyset, & \text{if } K_i = 0, \\ S_{i,1}, S_{i,2}, \dots, S_{i,K_i}, & \text{if } K_i \geq 1. \end{cases}$$

We define $\mathbf{K} \circ \mathbf{S} := (K_1 \circ S_1, K_2 \circ S_2, \dots, K_r \circ S_r)$ and use $\mathcal{P}(\mathbf{K})$ to denote the probability of observing the prefixes $\mathbf{K} \circ \mathbf{S}$.

Example 3 Again, consider the strings defined in Example 1. If $\mathbf{K} = (1, 1)$, then $\mathbf{K} \circ \mathbf{S} = (H, H)$. If $\mathbf{K} = (2, 0)$, then $\mathbf{K} \circ \mathbf{S} = (HT, \emptyset)$.

A hidden Markov model generates strings \mathbf{S} with probability $\mathcal{P}(\mathbf{S})$ and satisfies $\sum_{\mathbf{S}} \mathcal{P}(\mathbf{S}) = 1$. The Markov state space is denoted by Ω . There are two special states in Ω , namely the initial state \mathcal{I} and the end state \mathcal{E} . Every Markov chain starts from the initial state \mathcal{I} and ends in the end state \mathcal{E} . We decompose the state space Ω as

$$\Omega = \{\mathcal{I}\} \cup \Psi \cup \{\mathcal{E}\}. \quad (1)$$

That is, Ψ contains all Markov states other than \mathcal{I} and \mathcal{E} . The probability of making a transition from a state $\eta \in \Omega$ to another state $\eta' \in \Omega$ is denoted by $p(\eta, \eta')$. In the present paper, we consider a class of HMMs in which a state emits at most one letter per string. For $\psi \in \Psi$, we define $\Delta(\psi) = (\delta_1(\psi), \delta_2(\psi), \dots, \delta_r(\psi))$, where

$$\delta_i(\psi) = \begin{cases} 1, & \text{if } \psi \text{ emits a letter } \in \mathcal{A} \text{ in the string } S_i, \\ 0, & \text{otherwise.} \end{cases}$$

In other words, $\Delta(\psi)$ denotes the *emission pattern* of the state ψ . In the context of alignment, each column in an alignment can be regarded as being generated by a certain state in Ψ . If ψ generates a column in an alignment and $\delta_i(\psi) = 0$, then that column will have a gap in the i^{th} sequence.

The probability of emitting $\mathbf{S}[\mathbf{K}, \Delta(\psi)]$ in the state ψ is denoted by

$$\pi(\mathbf{S}[\mathbf{K}, \Delta(\psi)] \mid \psi).$$

The set of *silent* states is defined as $\Phi := \{\psi \in \Psi \mid \Delta(\psi) = \mathbf{0}\}$. For every $\varphi \in \Phi$, we define

$$\pi(\mathbf{S}[\mathbf{K}, \Delta(\varphi)] \mid \varphi) = 1.$$

Example 4 Consider the HMM shown in Figure 1. The state space Ψ contains only three states—namely ψ_1, ψ_2 and φ —and their emission patterns are $\Delta(\psi_1) = (1, 0)$, $\Delta(\psi_2) = (0, 1)$ and $\Delta(\varphi) = (0, 0)$. Note that φ is a silent state. In this simple example, there is no state with $(1, 1)$ as its emission pattern and no two states have the same emission pattern. In general, there can be many states with the same emission pattern. We shall return to this toy HMM in Examples 5 and 6.

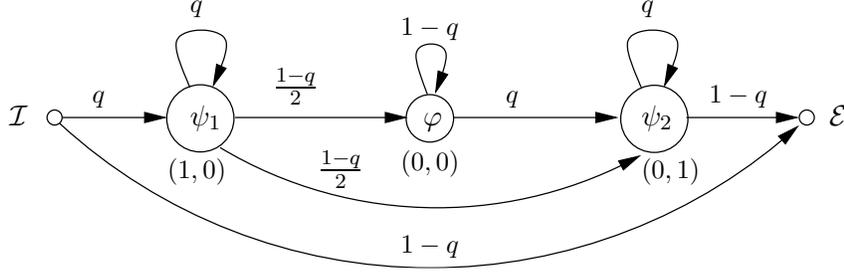


Figure 1: A toy HMM discussed in Examples 4,5 and 6. Possible transitions are indicated by arrows, and probabilities associated with transitions are shown above the arrows. Here, $0 < q < 1$. Emission patterns are shown below the states in Ψ .

3 State-dependent HMM recursions

Consider a Markov path $\Lambda = (\mathcal{I}, \lambda_1, \lambda_2, \dots, \lambda_k, \mathcal{E})$ —starting in the initial state \mathcal{I} , going through states $\lambda_i \in \Psi$, and finally terminating in the end state \mathcal{E} —such that $\sum_{i=1}^k \Delta(\lambda_i) = \mathbf{L}$. Then, the probability of generating \mathbf{S} while following the path Λ is given by

$$\mathcal{P}(\mathbf{S}, \Lambda) = p(\mathcal{I}, \lambda_1) \left[\prod_{j=1}^{k-1} p(\lambda_j, \lambda_{j+1}) \right] p(\lambda_k, \mathcal{E}) \left[\prod_{m=1}^k \pi(\mathbf{S}[\sum_{i=1}^m \Delta(\lambda_i), \Delta(\lambda_m)] | \lambda_m) \right].$$

The probability $\mathcal{P}(\mathbf{S})$ of observing the entire strings in $\mathbf{S} = (S_1, \dots, S_r)$ is formally defined as

$$\mathcal{P}(\mathbf{S}) = \sum_{\Lambda} \mathcal{P}(\mathbf{S}, \Lambda).$$

Here, the sum is over all possible Markov paths of arbitrary length such that $\sum_i \Delta(\lambda_i) = \mathbf{L}$. In practice, such a sum is usually impossible to carry out as it stands, the reason being that there are too many, often infinitely many, possible paths. As described in the following subsections, however, $\mathcal{P}(\mathbf{S})$ can be obtained from either

$$\mathcal{P}(\mathbf{S}) = \mathcal{F}(\mathbf{0} | \mathcal{I}) \quad \text{or} \quad \mathcal{P}(\mathbf{S}) = \mathcal{B}(\mathbf{L} | \mathcal{E}),$$

where the function $\mathcal{F}(\mathbf{0} | \mathcal{I})$ satisfies a forward recursion relation, whereas the function $\mathcal{B}(\mathbf{L} | \mathcal{E})$ satisfies a backward recursion relation.

3.1 Backward recursion

If \mathbf{x} is an r -tuple of integers with at least one component $x_i < 0$, then we define $\mathcal{B}(\mathbf{x} | \psi) = 0$ for all $\psi \in \Psi$. Let \mathbf{K} be an r -tuple of non-negative integers satisfying $\mathbf{0} \leq \mathbf{K} \leq \mathbf{L}$ and $\mathbf{K} \neq \mathbf{0}$. For any state $\zeta \in \Psi \cup \{\mathcal{E}\}$, the backward recursion relation is given by

$$\mathcal{B}(\mathbf{K} | \zeta) = \sum_{\psi \in \Psi} \mathcal{B}(\mathbf{K} - \Delta(\psi) | \psi) \pi(\mathbf{S}[\mathbf{K}, \Delta(\psi)] | \psi) p(\psi, \zeta).$$

In the above equation, decompose the sum over Ψ as a sum over Φ (i.e. silent states) and a sum over $\Psi \setminus \Phi$ (i.e. non-silent states) as follows:

$$\mathcal{B}(\mathbf{K}|\zeta) = \sum_{\varphi \in \Phi} \mathcal{B}(\mathbf{K}|\varphi) p(\varphi, \zeta) + \sum_{\psi \in \Psi \setminus \Phi} \mathcal{B}(\mathbf{K} - \Delta(\psi) | \psi) \pi(\mathbf{S}[\mathbf{K}, \Delta(\psi)] | \psi) p(\psi, \zeta). \quad (2)$$

Let $\Phi = \{\varphi_1, \varphi_2, \dots, \varphi_h\}$ be the set of silent states. Then, substituting $\zeta = \varphi_i$ into (2) gives

$$\mathcal{B}(\mathbf{K}|\varphi_i) = \sum_{j=1}^h \mathcal{B}(\mathbf{K}|\varphi_j) p(\varphi_j, \varphi_i) + \sum_{\psi \in \Psi \setminus \Phi} \mathcal{B}(\mathbf{K} - \Delta(\psi) | \psi) \pi(\mathbf{S}[\mathbf{K}, \Delta(\psi)] | \psi) p(\psi, \varphi_i),$$

from which we can obtain

$$\begin{pmatrix} \mathcal{B}(\mathbf{K}|\varphi_1) \\ \mathcal{B}(\mathbf{K}|\varphi_2) \\ \vdots \\ \mathcal{B}(\mathbf{K}|\varphi_h) \end{pmatrix} = \sum_{\psi \in \Psi \setminus \Phi} \mathcal{B}(\mathbf{K} - \Delta(\psi) | \psi) \pi(\mathbf{S}[\mathbf{K}, \Delta(\psi)] | \psi) (\mathbb{I} - \mathbb{M}^t)^{-1} \begin{pmatrix} p(\psi, \varphi_1) \\ p(\psi, \varphi_2) \\ \vdots \\ p(\psi, \varphi_h) \end{pmatrix},$$

where \mathbb{I} denotes the $h \times h$ identity matrix and \mathbb{M} an $h \times h$ matrix with entries $[\mathbb{M}]_{i,j} = p(\varphi_i, \varphi_j)$, the probability of making a transition from a silent state $\varphi_i \in \Phi$ to another silent state $\varphi_j \in \Phi$. The matrix \mathbb{M}^t denotes the transpose of \mathbb{M} . The coefficient $\mathcal{B}(\mathbf{K} - \Delta(\psi) | \psi) \pi(\mathbf{S}[\mathbf{K}, \Delta(\psi)] | \psi)$ is a scalar quantity and every entry in the matrix to its right is supposed to be multiplied by it. In summary, the backward recursion relation can now be rewritten as

$$\mathcal{B}(\mathbf{K}|\zeta) = \sum_{\psi \in \Psi \setminus \Phi} \mathcal{B}(\mathbf{K} - \Delta(\psi) | \psi) \pi(\mathbf{S}[\mathbf{K}, \Delta(\psi)] | \psi) C(\psi, \zeta), \quad (3)$$

where

$$C(\psi, \zeta) := \begin{pmatrix} p(\psi, \varphi_1) \\ p(\psi, \varphi_2) \\ \vdots \\ p(\psi, \varphi_h) \end{pmatrix}^t (\mathbb{I} - \mathbb{M})^{-1} \begin{pmatrix} p(\varphi_1, \zeta) \\ p(\varphi_2, \zeta) \\ \vdots \\ p(\varphi_h, \zeta) \end{pmatrix} + p(\psi, \zeta). \quad (4)$$

We can therefore interpret the role of the silent states as giving rise to *effective* transition probabilities. For $\psi \in (\Psi \setminus \Phi) \cup \{\mathcal{I}\}$ and $\zeta \in (\Psi \setminus \Phi) \cup \{\mathcal{E}\}$, the effective transition probability $C(\psi, \zeta)$ includes, in addition to the *direct* transition probability $p(\psi, \zeta)$, the probability of going from the state ψ to one or more intermediate silent states $\varphi_i \in \Phi$ and then to the state ζ .

When $\mathbf{K} = \mathbf{0}$, the backward recursion for $\zeta \in (\Psi \setminus \Phi) \cup \{\mathcal{E}\}$ is

$$\mathcal{B}(\mathbf{0}|\zeta) = p(\mathcal{I}, \zeta) + \sum_{\varphi \in \Phi} \mathcal{B}(\mathbf{0}|\varphi) p(\varphi, \zeta),$$

which is none other than the effective transition probability $C(\mathcal{I}, \zeta)$. Thus, $\mathcal{B}(\mathbf{0}|\zeta) = C(\mathcal{I}, \zeta)$ serves as the boundary condition for the recursion given in (3).

3.2 Forward recursion

If \mathbf{x} is an r -tuple of integers with at least one component $x_i > L_i$, then we define $\mathcal{F}(\mathbf{x} | \psi) = 0$ for all $\psi \in \Psi$. Let \mathbf{K} be an r -tuple of non-negative integers satisfying $\mathbf{0} \leq \mathbf{K} \leq \mathbf{L}$ and $\mathbf{K} \neq \mathbf{L}$. For any state $\zeta \in \Psi \cup \{\mathcal{I}\}$, the forward recursion relation reads

$$\mathcal{F}(\mathbf{K} | \zeta) = \sum_{\psi \in \Psi} p(\zeta, \psi) \pi(\mathbf{S}[\mathbf{K} + \Delta(\psi), \Delta(\psi)] | \psi) \mathcal{F}(\mathbf{K} + \Delta(\psi) | \psi) .$$

For ζ a non-silent state, we can perform similar simplifying steps as in the previous subsection to obtain

$$\mathcal{F}(\mathbf{K} | \zeta) = \sum_{\psi \in \Psi \setminus \Phi} C(\zeta, \psi) \pi(\mathbf{S}[\mathbf{K} + \Delta(\psi), \Delta(\psi)] | \psi) \mathcal{F}(\mathbf{K} + \Delta(\psi) | \psi) , \quad (5)$$

where the effective transition probability $C(\zeta, \psi)$ is given by (4) with ψ and ζ interchanged.

When $\mathbf{K} = \mathbf{L}$, the forward recursion for $\zeta \in (\Psi \setminus \Phi) \cup \{\mathcal{I}\}$ satisfies

$$\mathcal{F}(\mathbf{L} | \zeta) = p(\zeta, \mathcal{E}) + \sum_{\varphi \in \Phi} p(\zeta, \varphi) \mathcal{F}(\mathbf{L} | \varphi) = C(\zeta, \mathcal{E}) , \quad (6)$$

which serves as the boundary condition for the forward recursion given in (5).

3.3 The toy HMM revisited

We now return to the toy HMM depicted in Figure 1 and demonstrate how either backward or forward recursion can be used to compute the joint probability $\mathcal{P}(\mathbf{S})$ of observing the two strings in $\mathbf{S} = (S_1, S_2)$. Since φ is the only silent state in this example, i.e. $\Phi = \{\varphi\}$, the effective transition probability (4) becomes

$$C(\eta, \zeta) = \frac{p(\eta, \varphi)p(\varphi, \zeta)}{1 - p(\varphi, \varphi)} + p(\eta, \zeta).$$

From Figure 1, it follows that non-vanishing effective transition probabilities are as follows.

$$\begin{aligned} C(\mathcal{I}, \psi_1) &= q, & C(\mathcal{I}, \mathcal{E}) &= 1 - q, \\ C(\psi_1, \psi_1) &= q, & C(\psi_1, \psi_2) &= 1 - q, \\ C(\psi_2, \psi_2) &= q, & C(\psi_2, \mathcal{E}) &= 1 - q. \end{aligned}$$

In the following two examples, we again consider the simple case where $S_1 = HT$ and $S_2 = H$.

Example 5 (Backward recursion) Note that $\Psi \setminus \Phi = \{\psi_1, \psi_2\}$. The backward recursion (3) can be used to compute $\mathcal{P}(\mathbf{S})$ as follows. First, recall that $\mathcal{P}(\mathbf{S}) = \mathcal{B}(\mathbf{L} | \mathcal{E})$. Since $\mathbf{L} = (2, 1)$ and $\Delta(\psi_2) = (0, 1)$, we have $\mathbf{S}[\mathbf{L}, \Delta(\psi_2)] = (\emptyset, H)$. Also, since $C(\psi_1, \mathcal{E}) = 0$, the backward recursion (3) implies

$$\mathcal{B}((2, 1) | \mathcal{E}) = \mathcal{B}((2, 0) | \psi_2) \pi((\emptyset, H) | \psi_2) C(\psi_2, \mathcal{E}).$$

Further applying the recursion to $\mathcal{B}((2, 0) | \psi_2)$ gives

$$\mathcal{B}((2, 0) | \psi_2) = \mathcal{B}((1, 0) | \psi_1) \pi((T, \emptyset) | \psi_1) C(\psi_1, \psi_2),$$

where we have used $\mathbf{S}[(2,0), (1,0)] = (T, \emptyset)$. Recall that $\mathcal{B}((2,-1) | \psi_2)$ is defined to be zero, so it does not appear above. Now, since $\mathbf{S}[(1,0), (1,0)] = (H, \emptyset)$, the recursion applied to $\mathcal{B}((1,0) | \psi_1)$ gives

$$\mathcal{B}((1,0) | \psi_1) = \mathcal{B}((0,0) | \psi_1) \pi((H, \emptyset) | \psi_1) C(\psi_1, \psi_1).$$

Finally, using the above results and the boundary condition $\mathcal{B}((0,0) | \psi_1) = C(\mathcal{I}, \psi_1)$, we obtain

$$\begin{aligned} \mathcal{B}((2,1) | \mathcal{E}) &= C(\mathcal{I}, \psi_1) C(\psi_1, \psi_1) C(\psi_1, \psi_2) C(\psi_2, \mathcal{E}) \pi((H, \emptyset) | \psi_1) \pi((T, \emptyset) | \psi_1) \pi((\emptyset, H) | \psi_2) \\ &= q^2(1-q)^2 \pi((H, \emptyset) | \psi_1) \pi((T, \emptyset) | \psi_1) \pi((\emptyset, H) | \psi_2), \end{aligned}$$

which determines $\mathcal{P}(\mathbf{S})$.

Example 6 (Forward recursion) We now wish to compute $\mathcal{P}(\mathbf{S})$ by solving for $\mathcal{F}((0,0) | \mathcal{I})$. For $\mathbf{K} = (0,0)$, we have $\mathbf{S}[\mathbf{K} + \Delta(\psi_1), \Delta(\psi_1)] = \mathbf{S}[(1,0), (1,0)] = (H, \emptyset)$. Hence, since $C(\mathcal{I}, \psi_2) = 0$, the forward recursion (5) implies

$$\mathcal{F}((0,0) | \mathcal{I}) = C(\mathcal{I}, \psi_1) \pi((H, \emptyset) | \psi_1) \mathcal{F}((1,0) | \psi_1).$$

Since $\mathbf{S}[(1,0) + \Delta(\psi_1), \Delta(\psi_1)] = \mathbf{S}[(2,0), (1,0)] = (T, \emptyset)$ and $\mathbf{S}[(1,0) + \Delta(\psi_2), \Delta(\psi_2)] = \mathbf{S}[(1,1), (0,1)] = (\emptyset, H)$, the forward recursion then implies

$$\mathcal{F}((1,0) | \psi_1) = C(\psi_1, \psi_1) \pi((T, \emptyset) | \psi_1) \mathcal{F}((2,0) | \psi_1) + C(\psi_1, \psi_2) \pi((\emptyset, H) | \psi_2) \mathcal{F}((1,1) | \psi_2).$$

It is straightforward to show that $\mathcal{F}((1,1) | \psi_2) = 0$, however, so we just need

$$\mathcal{F}((2,0) | \psi_1) = C(\psi_1, \psi_2) \pi((\emptyset, H) | \psi_2) \mathcal{F}((2,1) | \psi_2) = C(\psi_1, \psi_2) \pi((\emptyset, H) | \psi_2) C(\psi_2, \mathcal{E}),$$

where the boundary condition (6) has been used in the last equality. In summary, we obtain

$$\begin{aligned} \mathcal{F}((0,0) | \mathcal{I}) &= C(\mathcal{I}, \psi_1) C(\psi_1, \psi_1) C(\psi_1, \psi_2) C(\psi_2, \mathcal{E}) \pi((H, \emptyset) | \psi_1) \pi((T, \emptyset) | \psi_1) \pi((\emptyset, H) | \psi_2) \\ &= q^2(1-q)^2 \pi((H, \emptyset) | \psi_1) \pi((T, \emptyset) | \psi_1) \pi((\emptyset, H) | \psi_2). \end{aligned}$$

This is exactly what we obtained for $\mathcal{P}(\mathbf{S})$ in Example 5 using the backward recursion.

4 Reduced recursion

As we have seen in the previous section, both backward and forward recursions (c.f. (3) and (5)) explicitly depend on Markov states. As the number of Markov states can in general be quite large, with many states having the same emission pattern, it would be desirable if we could evaluate the probability $\mathcal{P}(\mathbf{S})$ without having to refer to the Markov states at every step of the recursion. In this section, we investigate when it is possible to eliminate the dependence on Markov states and thereby obtain a reduced recursion. We construct a sufficient condition for reduction by making a connection with the state-dependent HMM recursion described in Section 3.1.

Our main result is summarised at the end of this section, along with a brief description on how our findings can be applied in practice.

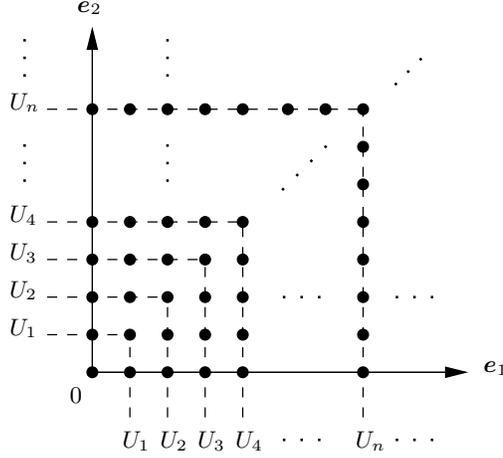


Figure 2: Stratification of M in two dimensions. Points connected by a dashed line form a single stratum.

4.1 More notations

We first lay out our notational convention to be used in this section. Recall that r denotes the number of strings in \mathcal{S} . Let $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_r\}$ be the standard basis of the rank- r Euclidean lattice \mathbb{Z}^r , and define M to be the subset of \mathbb{Z}^r generated by $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_r\}$ with non-negative integer coefficients, i.e.

$$M := \{a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \dots + a_r\mathbf{e}_r \mid a_1, a_2, \dots, a_r \in \mathbb{Z}_{\geq 0}\}.$$

Define a symmetric integral bilinear form $\langle \cdot, \cdot \rangle : \mathbb{Z}^r \times \mathbb{Z}^r \rightarrow \mathbb{Z}$, such that $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 1$ if $i = j$ or 0 if $i \neq j$. Hence, if $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \dots + x_r\mathbf{e}_r$ and $\mathbf{y} = y_1\mathbf{e}_1 + y_2\mathbf{e}_2 + \dots + y_r\mathbf{e}_r$, then $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^r x_i y_i$. The 1-norm of a vector $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \dots + x_r\mathbf{e}_r$ is defined as $|\mathbf{x}| := |x_1| + |x_2| + \dots + |x_r|$.

A *stratification* of M is given by

$$M = \bigcup_{n=0}^{\infty} U_n,$$

where a *stratum* U_n , for $n \in \mathbb{Z}_{\geq 0}$, is defined as

$$U_n := \left\{ a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \dots + a_r\mathbf{e}_r \in M \mid \begin{array}{l} 0 \leq a_i \leq n \text{ for every } i \in \{1, \dots, r\}, \text{ and} \\ a_j = n \text{ for at least one } j \in \{1, \dots, r\} \end{array} \right\}.$$

Note that $U_0 = \{\mathbf{0}\}$ and that U_i, U_j are disjoint if $i \neq j$. A two-dimensional example is shown in Figure 2, where points connected by a dashed line form a single stratum. The set of *emission vectors* is defined as

$$\Gamma_e = \{c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \dots + c_r\mathbf{e}_r \mid (c_1, c_2, \dots, c_r) \in \{0, 1\}^r, (c_1, c_2, \dots, c_r) \neq \mathbf{0}\}.$$

Note that $\Gamma_e = U_1$. For the two-dimensional example shown in Figure 2, $\Gamma_e = \{(1, 0), (0, 1), (1, 1)\}$.

Recall that we use $\mathcal{P}(\mathbf{K})$ to denote the probability of observing the prefixes $\mathbf{K} \circ \mathcal{S}$. We define $\mathcal{P}(\mathbf{x}) = 0$ if \mathbf{x} has at least one negative component.

4.2 Reduction for $\mathbf{K} \in U_1$

We first show that, for all $\mathbf{K} \in U_1$, there always exists a reduced recursion relation of the following form:

$$\mathcal{P}(\mathbf{K}) = \sum_{\substack{\mathbf{E} \in \Gamma_e, \\ \mathbf{0} \leq \mathbf{K} - \mathbf{E}}} G(\mathbf{E}; \mathbf{K}) \mathcal{P}(\mathbf{K} - \mathbf{E}). \quad (7)$$

Note that, unlike the usual recursions (3) and (5), this recursion has no explicit dependence on Markov states. In the reduced recursion, states with the same emission pattern do not need to be considered separately; we only need to consider distinct emission patterns. Our goal here is to construct the coefficients $G(\mathbf{E}; \mathbf{K})$ explicitly so that the above proposed recursion is satisfied.

Suppose that $\mathbf{K} = \mathbf{e}_i$. We show by construction that there exists a function $G(\mathbf{e}_i; \mathbf{e}_i)$ such that the following equation is true:

$$\mathcal{P}(\mathbf{e}_i) = \sum_{\substack{\mathbf{E} \in \Gamma_e, \\ \mathbf{0} \leq \mathbf{e}_i - \mathbf{E}}} G(\mathbf{E}; \mathbf{e}_i) \mathcal{P}(\mathbf{e}_i - \mathbf{E}) = G(\mathbf{e}_i; \mathbf{e}_i) \mathcal{P}(\mathbf{0}). \quad (8)$$

As mentioned before, our main idea is to draw a connection with the state-dependent recursion discussed in Section 3.1. First of all, note that $\mathcal{P}(\mathbf{0}) = (\mathcal{I}, \mathcal{E})$, which is the effective transition probability from the initial state \mathcal{I} to the end state \mathcal{E} (c.f. (4)). Secondly, using the backward recursion shown in (3), we obtain

$$\mathcal{P}(\mathbf{e}_i) = \mathcal{B}(\mathbf{e}_i | \mathcal{E}) = \left[\sum_{\psi \in \Psi_{\mathbf{e}_i}} C(\mathcal{I}, \psi) C(\psi, \mathcal{E}) \pi(\mathcal{S}[\mathbf{e}_i, \mathbf{e}_i] | \psi) \right],$$

where

$$\Psi_{\mathbf{y}} := \{\zeta \in \Psi \mid \Delta(\zeta) = \mathbf{y}\}. \quad (9)$$

It now follows that if we define

$$G(\mathbf{e}_i; \mathbf{e}_i) := \frac{1}{C(\mathcal{I}, \mathcal{E})} \left[\sum_{\psi \in \Psi_{\mathbf{e}_i}} C(\mathcal{I}, \psi) C(\psi, \mathcal{E}) \pi(\mathcal{S}[\mathbf{e}_i, \mathbf{e}_i] | \psi) \right], \quad (10)$$

then (8) is satisfied.

Now, suppose that $\mathbf{K} = \mathbf{e}_{i_1} + \mathbf{e}_{i_2}$ where \mathbf{e}_{i_1} and \mathbf{e}_{i_2} are distinct, i.e. $\langle \mathbf{e}_{i_1}, \mathbf{e}_{i_2} \rangle = 0$. Then, (7) implies

$$\begin{aligned} \mathcal{P}(\mathbf{e}_{i_1} + \mathbf{e}_{i_2}) &= \sum_{\substack{\mathbf{E} \in \Gamma_e, \\ \mathbf{0} \leq (\mathbf{e}_{i_1} + \mathbf{e}_{i_2}) - \mathbf{E}}} G(\mathbf{E}; \mathbf{e}_{i_1} + \mathbf{e}_{i_2}) \mathcal{P}(\mathbf{e}_{i_1} + \mathbf{e}_{i_2} - \mathbf{E}) \\ &= G(\mathbf{e}_{i_1}; \mathbf{e}_{i_1} + \mathbf{e}_{i_2}) \mathcal{P}(\mathbf{e}_{i_2}) + G(\mathbf{e}_{i_2}; \mathbf{e}_{i_1} + \mathbf{e}_{i_2}) \mathcal{P}(\mathbf{e}_{i_1}) + G(\mathbf{e}_{i_1} + \mathbf{e}_{i_2}; \mathbf{e}_{i_1} + \mathbf{e}_{i_2}) \mathcal{P}(\mathbf{0}). \end{aligned} \quad (11)$$

We wish to show that there exists $G(\mathbf{e}_{i_1} + \mathbf{e}_{i_2}; \mathbf{e}_{i_1} + \mathbf{e}_{i_2})$ such that the above equation is true. The probabilities $\mathcal{P}(\mathbf{e}_{i_1} + \mathbf{e}_{i_2})$, $\mathcal{P}(\mathbf{e}_{i_1})$, $\mathcal{P}(\mathbf{e}_{i_2})$, and $\mathcal{P}(\mathbf{0})$ can easily be determined using the state-dependent

recursion discussed in Section 3.1. Therefore, if we define $G(\mathbf{e}_i; \mathbf{e}_i + \mathbf{v}) = G(\mathbf{e}_i; \mathbf{e}_i)$ for all \mathbf{v} orthogonal to \mathbf{e}_i (i.e. $\langle \mathbf{v}, \mathbf{e}_i \rangle = 0$) and let

$$G(\mathbf{e}_{i_1} + \mathbf{e}_{i_2}; \mathbf{e}_{i_1} + \mathbf{e}_{i_2}) = \frac{\mathcal{P}(\mathbf{e}_{i_1} + \mathbf{e}_{i_2}) - G(\mathbf{e}_{i_1}; \mathbf{e}_{i_1})\mathcal{P}(\mathbf{e}_{i_2}) - G(\mathbf{e}_{i_2}; \mathbf{e}_{i_2})\mathcal{P}(\mathbf{e}_{i_1})}{\mathcal{P}(\mathbf{0})},$$

then (11) is satisfied.

In a similar vein, we can recursively determine $G(\mathbf{E}; \mathbf{E})$ for all $\mathbf{E} \in \Gamma_e$. To be more specific, suppose we have determined $G(\mathbf{E}'; \mathbf{E}')$ for all $\mathbf{E}' \in \Gamma_e$ satisfying $|\mathbf{E}'| < n$. Then, consider $\mathbf{E} = \mathbf{e}_{i_1} + \mathbf{e}_{i_2} + \dots + \mathbf{e}_{i_n}$, where $\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \dots, \mathbf{e}_{i_n}$ are all distinct. The coefficient $G(\mathbf{E}; \mathbf{E})$ can be computed from

$$G(\mathbf{E}; \mathbf{E}) = \frac{\mathcal{P}(\mathbf{E})}{\mathcal{P}(\mathbf{0})} - \frac{1}{\mathcal{P}(\mathbf{0})} \times \sum_{\substack{\mathbf{E}' \in \Gamma_e, \\ \mathbf{E}' \neq \mathbf{E}, \mathbf{0} \leq \mathbf{E} - \mathbf{E}'}} G(\mathbf{E}'; \mathbf{E}') \mathcal{P}(\mathbf{E} - \mathbf{E}'),$$

where every term on the right hand side is known. The probabilities $\mathcal{P}(\mathbf{E})$ and $\mathcal{P}(\mathbf{E} - \mathbf{E}')$ appearing in the above equation can be evaluated using the state-dependent recursion (3); in practice, this evaluation can easily be carried out since $\mathbf{E} = \mathbf{e}_{i_1} + \mathbf{e}_{i_2} + \dots + \mathbf{e}_{i_n}$ is small enough for state-dependent recursions to handle. For all $\mathbf{E} \in \Gamma_e$ and $\mathbf{K} \in U_1$ satisfying $\mathbf{E} \leq \mathbf{K}$, we define $G(\mathbf{E}; \mathbf{K}) = G(\mathbf{E}; \mathbf{E})$. By construction, the coefficients $G(\mathbf{E}; \mathbf{K})$ thus determined are guaranteed to make (7) true.

Recall that we use $\mathcal{P}(\mathbf{K})$ to denote the probability of generating the prefixes $\mathbf{K} \circ \mathbf{S}$ starting from the initial state \mathcal{I} . More generally, we can define $\mathcal{P}(\mathbf{K} | \eta)$ as the probability of generating $\mathbf{K} \circ \mathbf{S}$ starting from the state $\eta \in (\Psi \setminus \Phi) \cup \{\mathcal{I}\}$; in the context of the backward recursion discussed in Section 3.1, this is equivalent to changing the boundary condition to $\mathcal{B}(\mathbf{0} | \zeta) = C(\eta, \zeta)$. In particular, $\mathcal{P}(\mathbf{0} | \eta) = \mathcal{B}(\mathbf{0} | \mathcal{E}) = C(\eta, \mathcal{E})$. As we have done for $\mathcal{P}(\mathbf{x})$, we define $\mathcal{P}(\mathbf{x} | \eta) = 0$ if \mathbf{x} contains at least one negative component. Let $\mathbf{E} \in \Gamma_e$ and define

$$\tilde{G}(\mathbf{E}; \mathbf{E}; \eta) := \frac{\mathcal{P}(\mathbf{E} | \eta)}{\mathcal{P}(\mathbf{0} | \eta)} - \frac{1}{\mathcal{P}(\mathbf{0} | \eta)} \times \sum_{\substack{\mathbf{E}' \in \Gamma_e, \\ \mathbf{E}' \neq \mathbf{E}, \mathbf{0} \leq \mathbf{E} - \mathbf{E}'}} G(\mathbf{E}'; \mathbf{E}') \mathcal{P}(\mathbf{E} - \mathbf{E}' | \eta). \quad (12)$$

As before, the state-dependent recursion (3) can be used to determine $\mathcal{P}(\mathbf{E} | \eta)$ and $\mathcal{P}(\mathbf{E} - \mathbf{E}' | \eta)$. For $\mathbf{E} = \mathbf{e}_i$, it leads to the following expression:

$$\tilde{G}(\mathbf{e}_i; \mathbf{e}_i; \eta) := \frac{1}{C(\eta, \mathcal{E})} \left[\sum_{\psi \in \Psi_{\mathbf{e}_i}} C(\eta, \psi) C(\psi, \mathcal{E}) \pi(\mathbf{S}[\mathbf{e}_i, \mathbf{e}_i] | \psi) \right]. \quad (13)$$

These general expressions $\tilde{G}(\mathbf{E}; \mathbf{E}; \eta)$ will appear later in our construction of the sufficient condition for reduction. As in $G(\mathbf{E}; \mathbf{K})$, we define $\tilde{G}(\mathbf{E}; \mathbf{K}; \eta) = \tilde{G}(\mathbf{E}; \mathbf{E}; \eta)$, for all $\mathbf{E} \in \Gamma_e$ and $\mathbf{K} \in U_1$ satisfying $\mathbf{E} \leq \mathbf{K}$. Also, note that $G(\mathbf{E}; \mathbf{K}) = \tilde{G}(\mathbf{E}; \mathbf{K}; \mathcal{I})$, where \mathcal{I} is the initial state.

4.3 Explicit formulae for $\tilde{G}(\mathbf{E}; \mathbf{K}; \eta)$

Let $\mathbf{E} = \mathbf{e}_{i_1} + \dots + \mathbf{e}_{i_n} \in \Gamma_e$ and $\mathbf{K} \in U_1$. After some algebraic simplification, the function $\tilde{G}(\mathbf{E}; \mathbf{K}; \eta)$ can be written in terms of effective transition probabilities $C(\psi, \psi')$, emission probabilities $\pi(\mathbf{S}[\mathbf{K}, \mathbf{E}'] | \psi)$,

where $\mathbf{E}' \leq \mathbf{E}$, and the functions $G(\mathbf{E}''; \mathbf{K})$, where $\mathbf{E}'' \neq \mathbf{E}$ and $\mathbf{E}'' \leq \mathbf{E}$. We will be referring back to this subsection when we discuss the proof of Proposition 1 in the next subsection.

In what follows, let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$ be an *ordered partition* of n , denoted by $\alpha \vdash n$. The length of the partition is m . Two length- m partitions $(\alpha_1, \alpha_2, \dots, \alpha_m)$ and $(\alpha'_1, \alpha'_2, \dots, \alpha'_m)$ are considered distinct as ordered partitions if $\alpha_i \neq \alpha'_i$ for some $i \in \{1, 2, \dots, m\}$. To simplify notation, we define $\omega_k := \sum_{i=1}^k \alpha_i$. The group of all permutations on a set of n letters is denoted by \mathfrak{S}_n . Then, the function $\tilde{G}(\mathbf{E}; \mathbf{K}; \eta)$ can be explicitly written as follows:

$$\tilde{G}(\mathbf{E}; \mathbf{K}; \eta) := \frac{\mathcal{B}(\mathbf{E}; \mathbf{K}; \eta) - \mathcal{G}(\mathbf{E}; \mathbf{K}; \eta)}{C(\eta, \mathcal{E})}, \quad (14)$$

where $\mathcal{B}(\mathbf{E}; \mathbf{K}; \eta)$ and $\mathcal{G}(\mathbf{E}; \mathbf{K}; \eta)$ are defined as

$$\begin{aligned} \mathcal{B}(\mathbf{E}; \mathbf{K}; \eta) &:= \sum_{(\alpha_1, \alpha_2, \dots, \alpha_m) \vdash n} \frac{1}{\alpha_1! \alpha_2! \cdots \alpha_m!} \\ &\times \sum_{\sigma \in \mathfrak{S}_n} \sum_{\psi_{j_1} \in \Psi_1(\sigma)} \cdots \sum_{\psi_{j_m} \in \Psi_m(\sigma)} \left[C(\eta, \psi_{j_1}) C(\psi_{j_1}, \psi_{j_2}) \cdots C(\psi_{j_{m-1}}, \psi_{j_m}) C(\psi_{j_m}, \mathcal{E}) \right. \\ &\left. \times \prod_{k=1}^m \pi(\mathcal{S}[\mathbf{K}, \mathbf{e}_{i_{\sigma(\omega_{k-1}+1)} + \mathbf{e}_{i_{\sigma(\omega_{k-1}+2)} + \cdots + \mathbf{e}_{i_{\sigma(\omega_{k-1}+\alpha_k)}}]} | \psi_{j_k}) \right] \end{aligned}$$

and

$$\begin{aligned} \mathcal{G}(\mathbf{E}; \mathbf{K}; \eta) &:= \sum_{\substack{(\alpha_1, \alpha_2, \dots, \alpha_m) \vdash n, \\ m \neq 1}} \frac{1}{\alpha_1! \alpha_2! \cdots \alpha_m!} \\ &\times \sum_{\sigma \in \mathfrak{S}_n} \sum_{\psi_{j_1} \in \Psi_1(\sigma)} \cdots \sum_{\psi_{j_{m-1}} \in \Psi_{m-1}(\sigma)} \left[C(\eta, \psi_{j_1}) C(\psi_{j_1}, \psi_{j_2}) \cdots C(\psi_{j_{m-2}}, \psi_{j_{m-1}}) C(\psi_{j_{m-1}}, \mathcal{E}) \right. \\ &\left. \times G(\mathbf{e}_{i_{\sigma(\omega_{m-1}+1)} + \cdots + \mathbf{e}_{i_{\sigma(\omega_{m-1}+\alpha_m)}}; \mathbf{K}) \prod_{k=1}^{m-1} \pi(\mathcal{S}[\mathbf{K}, \mathbf{e}_{i_{\sigma(\omega_{k-1}+1)} + \cdots + \mathbf{e}_{i_{\sigma(\omega_{k-1}+\alpha_k)}}]} | \psi_{j_k}) \right]. \end{aligned}$$

In the above formulae, we have defined

$$\Psi_k(\sigma) := \left\{ \psi \in \Psi \mid \Delta(\psi) = \mathbf{e}_{i_{\sigma(\omega_{k-1}+1)} + \mathbf{e}_{i_{\sigma(\omega_{k-1}+2)} + \cdots + \mathbf{e}_{i_{\sigma(\omega_{k-1}+\alpha_k)}}} \right\}.$$

Remarks:

- (i) In the expression for $\mathcal{G}(\mathbf{E}; \mathbf{K}; \eta)$, $m > 1$ and therefore $\alpha_m < n$. Hence, every function $G(\mathbf{E}'; \mathbf{K})$ appearing on the right hand side of (14) satisfies $\mathbf{E}' \neq \mathbf{E}$ and $\mathbf{E}' \leq \mathbf{E}$.
- (ii) In practice, dynamic programming can be employed to determine $\mathcal{P}(\mathbf{x} | \eta)$, for all $\mathbf{x} \in U_1$, and $\tilde{G}(\mathbf{E}; \mathbf{K}; \eta)$ can then be determined from (12). Hence, one does not need to deal with the above complicated expressions for \mathcal{B} and \mathcal{G} in actual computations.

4.4 A sufficient condition for extending reduction

The emission probabilities which appear in $\tilde{G}(\mathbf{E}; \mathbf{E}; \eta)$ are of the form $\pi(S[\mathbf{E}', \mathbf{E}'' | \psi])$ where $\mathbf{E}'' \leq \mathbf{E}' \leq \mathbf{E}$. For $\mathbf{K} \in U_n$, where $n \geq 1$ and $\mathbf{E} \leq \mathbf{K}$, we define $\tilde{G}(\mathbf{E}; \mathbf{K}; \eta)$ to be equal to $\tilde{G}(\mathbf{E}; \mathbf{E}; \eta)$ after replacing the emission probabilities $\pi(S[\mathbf{E}'; \mathbf{E}'' | \psi])$ with $\pi(S[\mathbf{K}; \mathbf{E}'' | \psi])$. For example, from (13), we obtain

$$\tilde{G}(\mathbf{e}_i; \mathbf{K}; \eta) := \frac{1}{C(\eta, \mathcal{E})} \left[\sum_{\psi \in \Psi_{\mathbf{e}_i}} C(\eta, \psi) C(\psi, \mathcal{E}) \pi(S[\mathbf{K}, \mathbf{e}_i | \psi]) \right].$$

Note that $\tilde{G}(\mathbf{E}; \mathbf{K}; \eta)$ does not depend on the components of \mathbf{K} orthogonal to \mathbf{E} . In a similar vein, for $\mathbf{K} \in U_n$, where $n \geq 1$ and $\mathbf{E} \leq \mathbf{K}$, we define $G(\mathbf{E}; \mathbf{K})$ to be equal to $G(\mathbf{E}; \mathbf{E})$ after replacing the emission probabilities $\pi(S[\mathbf{E}'; \mathbf{E}'' | \psi])$ with $\pi(S[\mathbf{K}; \mathbf{E}'' | \psi])$.

We have already shown in Section 4.2 that a reduced recursion exists on the stratum U_1 . We now show that if the condition specified in the following proposition holds, then the reduced recursion can be extended to higher strata U_n , $n > 1$. More exactly, the proposition says the following: if the sufficient condition holds, then once $G(\mathbf{E}; \mathbf{K})$ have been determined for $\mathbf{K} \in U_1$, they can be used in the reduced recursion for all $\mathbf{K} \in M \setminus U_0$ after appropriately modifying the emission probabilities $\pi(\cdot)$.

Proposition 1 *Let $\tilde{G}(\mathbf{E}; \mathbf{E}; \eta)$ be defined as in (12). Recall that $G(\mathbf{E}; \mathbf{E}) = \tilde{G}(\mathbf{E}; \mathbf{E}; \mathcal{I})$, where \mathcal{I} is the initial state. Now, suppose that, for all r -tuple of strings and for all emission vectors $\mathbf{E} \in \Gamma_e$, we have*

$$\tilde{G}(\mathbf{E}; \mathbf{E}; \eta) = G(\mathbf{E}; \mathbf{E}) \tag{15}$$

for every state $\eta \in \Psi \setminus \Phi$ whose emission vector $\Delta(\eta)$ satisfies

$$\langle \Delta(\eta), \mathbf{E} \rangle \neq 0.$$

Then, for all $\mathbf{K} \in M \setminus U_0$, the following reduced recursion relation holds:

$$\mathcal{P}(\mathbf{K}) = \sum_{\substack{\mathbf{E} \in \Gamma_e, \\ \mathbf{0} \leq \mathbf{K} - \mathbf{E}}} G(\mathbf{E}; \mathbf{K}) \mathcal{P}(\mathbf{K} - \mathbf{E}). \tag{16}$$

(Note: The boundary condition for the recursion is $\mathcal{P}(\mathbf{0}) = C(\mathcal{I}, \mathcal{E})$, the effective transition probability from the initial state to the end state.)

Proof. Recall that both $\mathcal{P}(\mathbf{x})$ and $\mathcal{B}(\mathbf{x} | \psi)$, $\forall \psi \in \Psi$, are defined to be zero if \mathbf{x} contains at least one negative component. In this proof, we will use this fact to simplify our notation when writing summations, i.e. we omit writing restrictions like $\mathbf{0} \leq \mathbf{K} - \mathbf{E}$.

Let $\mathbf{J} \in U_n$, $n > 1$. Applying the proposed reduced recursion in (16) once and then subsequently applying the backward recursion in (3) gives

$$\begin{aligned} \mathcal{P}(\mathbf{J}) &= \sum_{\mathbf{E} \in \Gamma_e} G(\mathbf{E}; \mathbf{J}) \mathcal{P}(\mathbf{J} - \mathbf{E}) \\ &= \sum_{\mathbf{E} \in \Gamma_e} \sum_{\psi \in \Psi \setminus \Phi} G(\mathbf{E}; \mathbf{J}) \mathcal{B}(\mathbf{J} - \mathbf{E} - \Delta(\psi) | \psi) C(\psi, \mathcal{E}) \pi(S[\mathbf{J} - \mathbf{E}, \Delta(\psi)] | \psi). \end{aligned} \tag{17}$$

On the other hand, without using the proposed reduced recursion, we can obtain the following expression purely from applying the backward recursion in (3) twice:

$$\begin{aligned} \mathcal{P}(\mathbf{J}) = \mathcal{B}(\mathbf{J} | \mathcal{E}) &= \sum_{\psi_1, \psi_2 \in \Psi \setminus \Phi} \left\{ \mathcal{B}(\mathbf{J} - \Delta(\psi_1) - \Delta(\psi_2) | \psi_2) C(\psi_2, \psi_1) C(\psi_1, \mathcal{E}) \right. \\ &\quad \left. \times \pi(\mathbf{S}[\mathbf{J}, \Delta(\psi_1)] | \psi_1) \pi(\mathbf{S}[\mathbf{J} - \Delta(\psi_1), \Delta(\psi_2)] | \psi_2) \right\}. \end{aligned} \quad (18)$$

Our approach is to show that (16) holds for all $\mathbf{J} \in U_n, n > 1$, by showing that the expressions in (17) and (18) are equal if the condition shown in (15) is satisfied.

If ψ is a Markov state such that $\langle \mathbf{E}, \Delta(\psi) \rangle = 0$, then $\mathbf{E}' := \mathbf{E} + \Delta(\psi) \in \Gamma_e$. For every such a state appearing in (17), we decompose $\mathcal{B}(\mathbf{J} - \mathbf{E} - \Delta(\psi) | \psi)$ into two types as follows by further applying the backward recursion.

$$\begin{aligned} \mathcal{B}(\mathbf{J} - \mathbf{E}' | \psi) &= \sum_{\substack{\eta_1 \in \Psi \setminus \Phi, \\ \langle \mathbf{E}', \Delta(\eta_1) \rangle \neq 0}} \mathcal{B}(\mathbf{J} - \mathbf{E}' - \Delta(\eta_1) | \eta_1) C(\eta_1, \psi) \pi(\mathbf{S}[\mathbf{J} - \mathbf{E}', \Delta(\eta_1)] | \eta_1) \\ &+ \sum_{\substack{\eta_2 \in \Psi \setminus \Phi, \\ \langle \mathbf{E}', \Delta(\eta_2) \rangle = 0}} \mathcal{B}(\mathbf{J} - \mathbf{E}' - \Delta(\eta_2) | \eta_2) C(\eta_2, \psi) \pi(\mathbf{S}[\mathbf{J} - \mathbf{E}', \Delta(\eta_2)] | \eta_2). \end{aligned}$$

For every η_2 appearing in the second sum, we observe that $\mathbf{E}'' := \mathbf{E}' + \Delta(\eta_2) \in \Gamma_e$ and further apply the backward recursion to decompose $\mathcal{B}(\mathbf{J} - \mathbf{E}'' | \eta_2)$ into two sums, one over $\chi_1 \in \Psi \setminus \Phi$ such that $\langle \mathbf{E}'', \Delta(\chi_1) \rangle \neq 0$ and the other over $\chi_2 \in \Psi \setminus \Phi$ such that $\langle \mathbf{E}'', \Delta(\chi_2) \rangle = 0$. We repeatedly apply this process on the terms of the second type until every $\mathcal{B}(\mathbf{J} - \mathbf{x} - \Delta(\chi) | \chi)$ appearing on the right hand side of (17) has the property that $\mathbf{x} \in \Gamma_e$ and $\langle \mathbf{x}, \Delta(\chi) \rangle \neq 0$. We call the equation so obtained “modified (17).” Note that since Γ_e is of finite dimension, the process just described terminates after a finite number of steps.

In a similar vein, we construct “modified (18)” as follows. On the right hand side of (18), we decompose every $\mathcal{B}(\mathbf{J} - \Delta(\psi_1) - \Delta(\psi_2) | \psi_2)$ where $\langle \Delta(\psi_1), \Delta(\psi_2) \rangle = 0$ into two sums by further applying the backward recursion on those terms; the first sum is over $\zeta_1 \in \Psi \setminus \Phi$ such that $\langle \Delta(\psi_1) + \Delta(\psi_2), \Delta(\zeta_1) \rangle \neq 0$ and the second sum is over $\zeta_2 \in \Psi \setminus \Phi$ such that $\langle \Delta(\psi_1) + \Delta(\psi_2), \Delta(\zeta_2) \rangle = 0$. We repeat this process on the terms of the second type until every term is of the first type; i.e. at this point every $\mathcal{B}(\mathbf{J} - \mathbf{x} - \Delta(\chi) | \chi)$ appearing in the “modified (18)” would have the property that $\mathbf{x} \in \Gamma_e$ and $\langle \mathbf{x}, \Delta(\chi) \rangle \neq 0$.

We now define the difference between the two “modified” equations as

$$D := \text{“modified (17)”} - \text{“modified (18)”}, \quad (19)$$

which we want to show vanishes if the condition shown in (15) is satisfied. By construction, every term of the form $\mathcal{B}(\mathbf{J} - \mathbf{x} - \Delta(\chi) | \chi)$ appearing in (19) satisfies $\mathbf{x} \in \Gamma_e$ and $\langle \mathbf{x}, \Delta(\chi) \rangle \neq 0$. It is straightforward

but tedious to show that D can be written as

$$D = \sum_{\mathbf{E} \in \Gamma_e} \sum_{\substack{\eta \in \Psi \setminus \Phi, \\ \langle \mathbf{E}, \Delta(\eta) \rangle \neq 0}} \left\{ \left[G(\mathbf{E}; \mathbf{J}) - \tilde{G}(\mathbf{E}; \mathbf{J}; \eta) \right] \times \right. \\ \left. \mathcal{B}(\mathbf{J} - \mathbf{E} - \Delta(\eta) | \eta) C(\eta, \mathcal{E}) \pi(\mathbf{S}[\mathbf{J} - \mathbf{E}, \Delta(\eta)] | \eta) \right\}, \quad (20)$$

which arises as follows. For each term

$$\mathcal{B}(\mathbf{J} - \mathbf{E} - \Delta(\eta) | \eta) \pi(\mathbf{S}[\mathbf{J} - \mathbf{E}, \Delta(\eta)] | \eta)$$

appearing in the above equation, “modified (17)” contributes $G(\mathbf{E}; \mathbf{J})C(\eta; \mathcal{E}) + \mathcal{G}(\mathbf{E}; \mathbf{J}; \eta)$ to its coefficient, whereas “modified (18)” contributes $\mathcal{B}(\mathbf{E}; \mathbf{J}; \eta)$ (c.f. (14) and equations thereafter). Therefore, we have

$$G(\mathbf{E}; \mathbf{J})C(\eta; \mathcal{E}) + \mathcal{G}(\mathbf{E}; \mathbf{J}; \eta) - \mathcal{B}(\mathbf{E}; \mathbf{J}; \eta) = \left[G(\mathbf{E}; \mathbf{J}) - \frac{\mathcal{B}(\mathbf{E}; \mathbf{J}; \eta) - \mathcal{G}(\mathbf{E}; \mathbf{J}; \eta)}{C(\eta; \mathcal{E})} \right] C(\eta; \mathcal{E}),$$

which is none other than $\left[G(\mathbf{E}; \mathbf{J}) - \tilde{G}(\mathbf{E}; \mathbf{J}; \eta) \right] C(\eta; \mathcal{E})$.

Now, we observe that the only dependence of $\tilde{G}(\mathbf{E}; \mathbf{K}; \eta)$ on \mathbf{K} appears in emission probabilities of form the $\pi(\mathbf{S}[\mathbf{K}, \mathbf{E}'] | \psi)$, where $\mathbf{E}' \leq \mathbf{E}$. Therefore, if $\tilde{G}(\mathbf{E}; \mathbf{E}; \eta) = G(\mathbf{E}; \mathbf{E})$ for *all* r -tuples of strings \mathbf{S} , then

$$\tilde{G}(\mathbf{E}; \mathbf{J}; \eta) = G(\mathbf{E}; \mathbf{J}),$$

where $\mathbf{J} \in U_n, n > 1$, and $\mathbf{0} \leq \mathbf{J} - \mathbf{E}$. Hence, if the condition (15) is satisfied, then the coefficient of every $\mathcal{B}(\mathbf{J} - \mathbf{E} - \Delta(\eta) | \eta)$ appearing in (20) vanishes, and it follows that $D = 0$. We have thus shown that the expressions in (17) and (18) are, in fact, equal if the condition stated in (15) is satisfied. This completes our proof of the proposition. \blacksquare

4.5 Recapitulation of the main result

We now briefly summarise our main result and outline how Proposition 1 might be applied in practice to obtain a reduced recursion. Let $\mathbf{S} = (S_1, \dots, S_r)$ be an r -tuple of strings and let $\mathbf{L} = (L_1, \dots, L_r)$, where L_i denotes the length of the string S_i . As described in Section 3, the joint probability $\mathcal{P}(\mathbf{L})$ of observing the strings in \mathbf{S} can be computed, in principle, using standard recursions. For example, we can use $\mathcal{P}(\mathbf{L}) = \mathcal{B}(\mathbf{L} | \mathcal{E})$, where $\mathcal{B}(\mathbf{L} | \mathcal{E})$ satisfies the backward recursion

$$\mathcal{B}(\mathbf{K} | \zeta) = \sum_{\psi \in \Psi \setminus \Phi} \mathcal{B}(\mathbf{K} - \Delta(\psi) | \psi) \pi(\mathbf{S}[\mathbf{K}, \Delta(\psi)] | \psi) C(\psi, \zeta), \quad (21)$$

with the boundary condition $\mathcal{B}(\mathbf{0} | \zeta) = C(\mathcal{I}, \zeta)$. Here, $C(\psi, \zeta)$ are effective transition probabilities, defined in Section 3. Note that such a recursion explicitly depends on states, and that there can be many terms on the right hand side of (21) if the number of states in $\Psi \setminus \Phi$ is large. Therefore, it would

lead to significant computational acceleration if there existed a reduced recursion of the form

$$\mathcal{P}(\mathbf{K}) = \sum_{\substack{\mathbf{E} \in \Gamma_e, \\ \mathbf{0} \leq \mathbf{K} - \mathbf{E}}} G(\mathbf{E}; \mathbf{K}) \mathcal{P}(\mathbf{K} - \mathbf{E}), \quad (22)$$

with some well-defined boundary condition for $\mathcal{P}(\mathbf{0})$. If there are generally several states with the same emission pattern, the number of terms appearing on the right hand side of (22) can be considerably less than that of (21). In the reduced recursion (22), each emission pattern occurs only once.

Our main result, Proposition 1, gives a sufficient condition for the existence of a reduced recursion like (22). This result is applicable to HMMs in which a state emits at most one letter per string. If the sufficient condition is satisfied, then our result completely determines what the coefficients $G(\mathbf{E}; \mathbf{K})$ appearing in the reduced recursion should be. The sufficient condition involves checking whether some quantities are equal. The quantities one needs to compute are $\tilde{G}(\mathbf{E}; \mathbf{E}; \eta)$, where $\mathbf{E} \in \Gamma_e = U_1$. (For instance, in the two-dimensional example shown in Figure 2, U_1 consists of three points and one would need to check the sufficient condition for these three points.) More precisely, $\tilde{G}(\mathbf{E}; \mathbf{E}; \eta)$ are defined as follows. First, define $\mathcal{P}(\mathbf{K} | \eta) := \mathcal{B}(\mathbf{K} | \mathcal{E})$, where $\mathcal{B}(\mathbf{K} | \mathcal{E})$ still satisfies the same standard backward recursion (21) but has the new boundary condition $\mathcal{B}(\mathbf{0} | \zeta) = C(\eta, \zeta)$. In particular, $\mathcal{P}(\mathbf{0} | \eta) = C(\eta, \mathcal{E})$. Then, for $\mathbf{E} \in U_1$, $\tilde{G}(\mathbf{E}; \mathbf{E}; \eta)$ are defined as

$$\tilde{G}(\mathbf{E}; \mathbf{E}; \eta) := \frac{\mathcal{P}(\mathbf{E} | \eta)}{\mathcal{P}(\mathbf{0} | \eta)} - \frac{1}{\mathcal{P}(\mathbf{0} | \eta)} \times \sum_{\substack{\mathbf{E}' \in \Gamma_e, \\ \mathbf{E}' \neq \mathbf{E}, \mathbf{0} \leq \mathbf{E} - \mathbf{E}'}} G(\mathbf{E}'; \mathbf{E}') \mathcal{P}(\mathbf{E} - \mathbf{E}' | \eta),$$

where $G(\mathbf{E}'; \mathbf{E}') = \tilde{G}(\mathbf{E}'; \mathbf{E}'; \mathcal{I})$, with \mathcal{I} being the unique initial state. Note that computing $\mathcal{P}(\mathbf{0} | \eta)$ and $\mathcal{P}(\mathbf{K} | \eta)$, for all $\mathbf{K} \in U_1$, can be done efficiently using the standard recursion (21); the largest value of any entry in \mathbf{K} that needs to be considered here is one, so the recursion reaches its boundary condition very quickly.

The following procedure describes how one might apply our result to a specific HMM.

1. For all r -tuple of strings, check whether $\tilde{G}(\mathbf{E}; \mathbf{E}; \eta) = G(\mathbf{E}; \mathbf{E})$ for all emission vectors $\mathbf{E} \in \Gamma_e = U_1$ and non-silent states $\eta \in \Psi \setminus \Phi$ satisfying $\langle \Delta(\eta), \mathbf{E} \rangle \neq 0$.
2. If the above conditions are satisfied, then to compute the joint probability of observing a particular r -tuple of string \mathbf{S} , use the reduced recursion (22), where $G(\mathbf{E}; \mathbf{K})$ is obtained from replacing the emission probabilities $\pi(S[\mathbf{E}'; \mathbf{E}'' | \psi])$ in $G(\mathbf{E}; \mathbf{E})$ with $\pi(S[\mathbf{K}; \mathbf{E}'' | \psi])$.

5 An example: statistical multiple alignment

In this section, we consider a hidden Markov model for statistical alignment [3] and check whether the sufficient condition from the previous section is satisfied by that model.

5.1 Preliminaries

The model proposed by Thorne, Kishino and Felsenstein [7] is a reversible Markov model of sequence evolution, incorporating substitutions, insertions and deletions. In this model, insertions and deletions are modeled as a birth-death process. A deletion random variable is associated to each nucleotide and an insertion random variable, called a *mortal link*, is placed to the *right* of every nucleotide. If a nucleotide dies, then so does its associated mortal link. If a mortal link m gives rise to birth, a nucleotide is drawn from the stationary distribution of the substitution process, and that nucleotide and its associated mortal link are placed to the right of the mortal link m . To the left of every sequence is a special insertion random variable. It is not associated to any nucleotide and it never dies. As such, it is called an *immortal link*. Every random variable for the deletion process is exponentially distributed with parameter μ , and that associated to the insertion process is exponentially distributed with parameter $\lambda < \mu$.

In its original formulation [7], the TKF model was constructed for the case of two sequences. For more than two sequences, efficient methods have been developed recently for applying the TKF model to each edge in the rooted phylogenetic tree relating the sequences [2, 3, 6]. When there are r sequences, a leaf-labelled tree with r leaves is used. For $\mathbf{S} = (S_1, S_2, \dots, S_r)$, let $\mathcal{P}(\mathbf{S})$ denote the joint probability of observing the sequences S_1, S_2, \dots, S_r at the leaves of the underlying phylogenetic tree. In the HMM formulation of [3], to each path from the initial state \mathcal{I} to the final state \mathcal{E} consistent with the given data \mathbf{S} , there corresponds an alignment of the sequences S_1, S_2, \dots, S_r . The “likelihood” $\mathcal{P}(\mathbf{S})$ is defined as the sum of the probabilities of all such paths in the HMM. Backward and forward recursions, just like that discussed in Section 3, for computing $\mathcal{P}(\mathbf{S})$ have been constructed in the HMM formulation of [3]. If the sufficient condition in Proposition 1 is satisfied by that HMM, then it would be consistent with the findings of [4], where a reduced recursion for the likelihood $\mathcal{P}(\mathbf{S})$ has been constructed using some fundamental properties of the TKF model.

In what follows, we wish to check whether our sufficient condition is satisfied by the HMM of [3]. Before we proceed, we briefly describe how a state is defined in [3]. In general, a state consists of two things. First of all, to every state in Ψ (c.f. (1)), there corresponds a nucleotide birth event at some vertex v in the underlying rooted tree T , and its survival-death information is specified for the nodes in the descendant subtree t_v of v , if there is any. This determines the emission pattern at the leaves of T . Note that every non-silent state in the HMM (i.e. a state in $\Psi \setminus \Phi$) generates a column in an alignment. Secondly, a state may have some “history” information on the part of the tree T complementary to t_v . We refer the reader to [3] for the exact definition of “history” used in that work. For the purpose of checking the sufficient condition in Proposition 1, we only need to know the emission patterns of states and the transition probabilities between states.

Let τ_i denote the time parameter associated to the edge incident with the leaf labelled i and define

$$\beta_i := \frac{1 - \exp[(\lambda - \mu)\tau_i]}{\mu - \lambda \exp[(\lambda - \mu)\tau_i]}.$$

	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
S_1			×	×			×
S_2		×		×	×	×	

Table 1: A summary of the emission patterns of the Markov states for two sequences. If a state ψ_i emits a letter in the string S_k , then the corresponding entry in the above table is marked \times .

Henceforward, we sometimes use the following notations when writing probabilities:

$$\begin{aligned}
a_i &:= \frac{\mu\beta_i}{1 - \exp(-\mu\tau_i)} & , & \quad \bar{a}_i := 1 - a_i, \\
b_i &:= \lambda\beta_i & , & \quad \bar{b}_i := 1 - b_i, \\
c_i &:= \exp(-\mu\tau_i) & , & \quad \bar{c}_i := 1 - c_i, \\
\gamma &:= \lambda/\mu & , & \quad \bar{\gamma} := 1 - \gamma.
\end{aligned} \tag{23}$$

In the HMM formulation of the TKF model, emission probabilities satisfy the property that

$$\pi(\mathbf{S}[\mathbf{K}, \Delta(\psi)] | \psi) = \pi(\mathbf{S}[\mathbf{K}, \Delta(\zeta)] | \zeta)$$

if $\Delta(\psi) = \Delta(\zeta)$; that is, if two states have the same emission pattern, say Δ , then their corresponding probabilities of emitting $\mathbf{S}[\mathbf{K}, \Delta]$ are the same. So, in what follows, emission probabilities will be simply written as $\pi(\mathbf{S}[\mathbf{K}, \Delta])$.

5.2 Two sequences

Consider the case of two sequences, i.e. $r = 2$ and $\mathbf{S} = (S_1, S_2)$. The underlying tree is a two-leaved tree with one degree-2 vertex, the root. The set of emission vectors is

$$\Gamma_e = \{c_1\mathbf{e}_1 + c_2\mathbf{e}_2 \mid (c_1, c_2) \in \{0, 1\}^2, (c_1, c_2) \neq (0, 0)\} = \{(1, 0), (0, 1), (1, 1)\}.$$

Aside from the initial and the end states, there are 7 other Markov states in the case of two sequences [3]. The emission patterns of the states in Ψ are given in Table 1. As the table shows, there is only one silent state, ψ_1 , in this HMM and therefore the effective transition probability shown in (4) simplifies to the following expression:

$$C(\eta, \zeta) = \frac{p(\eta, \psi_1)p(\psi_1, \zeta)}{1 - p(\psi_1, \psi_1)} + p(\eta, \zeta). \tag{24}$$

Transition probabilities are summarised in Table 2, where notations shown in (23) have been employed for conciseness. The probability $p(\eta, \zeta)$ of making a transition from η to ζ is given in the (η, ζ) entry of the table. Note that some transitions are not allowed, i.e. have zero probabilities.

A general expression for $G(\mathbf{e}_i; \mathbf{e}_i)$ is given in (10). We can use (24) and the transition probabilities in Table 2 to obtain the following results for our present example:

$$G(\mathbf{e}_1; \mathbf{e}_1) = \frac{\lambda\beta_2 + \lambda\beta_1[1 - (\lambda + \mu)\beta_2]}{1 - \lambda\mu\beta_1\beta_2} \pi(\mathbf{S}[\mathbf{e}_1, \mathbf{e}_1]), \tag{25}$$

$$G(\mathbf{e}_2; \mathbf{e}_2) = \frac{\lambda\beta_1 + \lambda\beta_2[1 - (\lambda + \mu)\beta_1]}{1 - \lambda\mu\beta_1\beta_2} \pi(\mathbf{S}[\mathbf{e}_2, \mathbf{e}_2]). \tag{26}$$

	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7	\mathcal{E}
\mathcal{I}	$\gamma \bar{b}_1 \bar{b}_2 \bar{c}_1 \bar{c}_2$	$\gamma \bar{b}_1 \bar{b}_2 \bar{c}_1 c_2$	$\gamma \bar{b}_1 \bar{b}_2 \bar{c}_2 c_1$	$\gamma \bar{b}_1 \bar{b}_2 c_1 c_2$	0	b_2	$\bar{b}_2 b_1$	$\bar{\gamma} \bar{b}_1 \bar{b}_2$
ψ_1	$\gamma a_1 a_2 \bar{c}_1 \bar{c}_2$	$\gamma a_1 a_2 \bar{c}_1 c_2$	$\gamma a_1 a_2 \bar{c}_2 c_1$	$\gamma a_1 a_2 c_1 c_2$	\bar{a}_2	0	$a_2 \bar{a}_1$	$\bar{\gamma} a_1 a_2$
ψ_2	$\gamma a_1 \bar{b}_2 \bar{c}_1 \bar{c}_2$	$\gamma a_1 \bar{b}_2 \bar{c}_1 c_2$	$\gamma a_1 \bar{b}_2 \bar{c}_2 c_1$	$\gamma a_1 \bar{b}_2 c_1 c_2$	b_2	0	$\bar{a}_1 \bar{b}_2$	$\bar{\gamma} a_1 \bar{b}_2$
ψ_3	$\gamma a_2 \bar{b}_1 \bar{c}_1 \bar{c}_2$	$\gamma a_2 \bar{b}_1 \bar{c}_1 c_2$	$\gamma a_2 \bar{b}_1 \bar{c}_2 c_1$	$\gamma a_2 \bar{b}_1 c_1 c_2$	0	\bar{a}_2	$a_2 b_1$	$\bar{\gamma} a_2 \bar{b}_1$
ψ_4	$\gamma \bar{b}_1 \bar{b}_2 \bar{c}_1 \bar{c}_2$	$\gamma \bar{b}_1 \bar{b}_2 \bar{c}_1 c_2$	$\gamma \bar{b}_1 \bar{b}_2 \bar{c}_2 c_1$	$\gamma \bar{b}_1 \bar{b}_2 c_1 c_2$	0	b_2	$\bar{b}_2 b_1$	$\bar{\gamma} \bar{b}_1 \bar{b}_2$
ψ_5	$\gamma a_1 \bar{b}_2 \bar{c}_1 \bar{c}_2$	$\gamma a_1 \bar{b}_2 \bar{c}_1 c_2$	$\gamma a_1 \bar{b}_2 \bar{c}_2 c_1$	$\gamma a_1 \bar{b}_2 c_1 c_2$	b_2	0	$\bar{a}_1 \bar{b}_2$	$\bar{\gamma} a_1 \bar{b}_2$
ψ_6	$\gamma \bar{b}_1 \bar{b}_2 \bar{c}_1 \bar{c}_2$	$\gamma \bar{b}_1 \bar{b}_2 \bar{c}_1 c_2$	$\gamma \bar{b}_1 \bar{b}_2 \bar{c}_2 c_1$	$\gamma \bar{b}_1 \bar{b}_2 c_1 c_2$	0	b_2	$\bar{b}_2 b_1$	$\bar{\gamma} \bar{b}_1 \bar{b}_2$
ψ_7	$\gamma \bar{b}_1 \bar{c}_1 \bar{c}_2$	$\gamma \bar{b}_1 \bar{c}_1 c_2$	$\gamma \bar{b}_1 \bar{c}_2 c_1$	$\gamma \bar{b}_1 c_1 c_2$	0	0	b_1	$\bar{\gamma} \bar{b}_1$

Table 2: A summary of transition probabilities in the hidden Markov model for two sequences. The probability $p(\eta, \zeta)$ of the transition $\eta \rightarrow \zeta$ is shown in the (η, ζ) entry of the above table.

Now, consider $\mathbf{E} = \mathbf{e}_1 + \mathbf{e}_2$. An explicit expression for $\tilde{G}(\mathbf{E}; \mathbf{E}; \eta)$ can be obtained using either (12) or (14) and the equations thereafter; it can be written as

$$\begin{aligned}
\tilde{G}(\mathbf{E}; \mathbf{E}; \eta) &= \frac{1}{C(\eta, \mathcal{E})} \left[\pi(\mathbf{S}[\mathbf{E}, \mathbf{E}]) \sum_{\zeta \in \Psi_{\mathbf{e}_1 + \mathbf{e}_2}} C(\eta, \zeta) C(\zeta, \mathcal{E}) \right. \\
&+ \sum_{\zeta_1 \in \Psi_{\mathbf{e}_1}} \sum_{\zeta_2 \in \Psi_{\mathbf{e}_2}} \pi(\mathbf{S}[\mathbf{E}, \mathbf{e}_1]) \pi(\mathbf{S}[\mathbf{E}, \mathbf{e}_2]) C(\eta, \zeta_2) C(\zeta_2, \zeta_1) C(\zeta_1, \mathcal{E}) \\
&+ \sum_{\zeta_1 \in \Psi_{\mathbf{e}_2}} \sum_{\zeta_2 \in \Psi_{\mathbf{e}_1}} \pi(\mathbf{S}[\mathbf{E}, \mathbf{e}_1]) \pi(\mathbf{S}[\mathbf{E}, \mathbf{e}_2]) C(\eta, \zeta_2) C(\zeta_2, \zeta_1) C(\zeta_1, \mathcal{E}) \\
&- G(\mathbf{e}_1; \mathbf{E}) \pi(\mathbf{S}[\mathbf{E}, \mathbf{e}_2]) \sum_{\zeta \in \Psi_{\mathbf{e}_2}} C(\eta, \zeta) C(\zeta, \mathcal{E}) \\
&\left. - G(\mathbf{e}_2; \mathbf{E}) \pi(\mathbf{S}[\mathbf{E}, \mathbf{e}_1]) \sum_{\zeta \in \Psi_{\mathbf{e}_1}} C(\eta, \zeta) C(\zeta, \mathcal{E}) \right], \tag{27}
\end{aligned}$$

where $\Psi_{\mathbf{y}}$ is defined as in (9). In our present example, the above formula implies

$$\begin{aligned}
G(\mathbf{E}; \mathbf{E}) &= \frac{\lambda}{\mu(1 - \lambda\mu\beta_1\beta_2)} \left\{ e^{-\mu(\tau_1 + \tau_2)} (1 - \lambda\beta_1)(1 - \lambda\beta_2) \pi(\mathbf{S}[\mathbf{E}, \mathbf{E}]) + \right. \\
&\pi(\mathbf{S}[\mathbf{E}, \mathbf{e}_1]) \pi(\mathbf{S}[\mathbf{E}, \mathbf{e}_2]) \left[(1 - e^{-\mu(\tau_1 + \tau_2)})(1 - \lambda\beta_1)(1 - \lambda\beta_2) + \right. \\
&\left. \left. (1 - \mu\beta_1)(1 - \mu\beta_2) - (1 - \lambda\mu\beta_1\beta_2) \right] \right\},
\end{aligned}$$

where we have used $G(\mathbf{E}; \mathbf{E}) = \tilde{G}(\mathbf{E}; \mathbf{E}; \mathcal{I})$.

We now check the sufficient condition stated in Proposition 1. As shown in Table 1, the states satisfying $\langle \mathbf{e}_1, \psi_a \rangle \neq 0$ are ψ_3, ψ_4 and ψ_7 . Using (13) and (24), one can verify that

$$\tilde{G}(\mathbf{e}_1; \mathbf{e}_1; \psi_3) = \tilde{G}(\mathbf{e}_1; \mathbf{e}_1; \psi_4) = \tilde{G}(\mathbf{e}_1; \mathbf{e}_1; \psi_7) = G(\mathbf{e}_1; \mathbf{e}_1),$$

where $G(\mathbf{e}_1; \mathbf{e}_1)$ is shown in (25). Likewise, the states satisfying $\langle \mathbf{e}_2, \psi_a \rangle \neq 0$ are ψ_2, ψ_4, ψ_5 and ψ_6 , and explicit computation shows that

$$\tilde{G}(\mathbf{e}_2; \mathbf{e}_2; \psi_2) = \tilde{G}(\mathbf{e}_2; \mathbf{e}_2; \psi_4) = \tilde{G}(\mathbf{e}_2; \mathbf{e}_2; \psi_5) = \tilde{G}(\mathbf{e}_2; \mathbf{e}_2; \psi_6) = G(\mathbf{e}_2; \mathbf{e}_2),$$

	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7	ψ_8	ψ_9	ψ_{10}	ψ_{11}	ψ_{12}	ψ_{13}	ψ_{14}	ψ_{15}
S_1					×	×	×	×							×
S_2			×	×			×	×					×	×	
S_3	×			×		×		×	×	×	×	×			

Table 3: A summary of the emission patterns of the Markov states for three sequences. If a state ψ_i emits a letter in the string S_k , then the corresponding entry in the above table is marked \times .

where $G(\mathbf{e}_2; \mathbf{e}_2)$ is shown in (26). Lastly, the general expression in (27) can be used to show that $\tilde{G}(\mathbf{e}_1 + \mathbf{e}_2; \mathbf{e}_1 + \mathbf{e}_2; \psi_a) = G(\mathbf{e}_1 + \mathbf{e}_2; \mathbf{e}_1 + \mathbf{e}_2)$, for all $a \in \{2, 3, \dots, 7\}$. Thus, the sufficient condition in Proposition 1 is satisfied by the present example.

5.3 Three sequences

According to the formulation of [3], there are 15 Markov states in Ψ in the case of three sequences. The underlying tree in this case is a three-leaved tree with one degree-3 vertex, which serves as the root. As in the previous example, there is exactly one silent state, and therefore the simple form of effective transition probability shown in (24) still applies to this case. The emission patterns of the 15 states are shown in Table 3, whereas transition probabilities are summarised in Table 4.

As we have done for two sequences, we can explicitly compute $G(\mathbf{E}; \mathbf{E})$, where \mathbf{E} is an emission vector in

$$\Gamma_e = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1), (1, 1, 1)\}.$$

For instance, we can obtain

$$G(\mathbf{e}_i; \mathbf{e}_i) = \frac{\lambda[\beta_i - \mu^2\beta_i\beta_j\beta_k + \mu\beta_j\beta_k(1 - \lambda\beta_i)]}{1 - \lambda\mu^2\beta_i\beta_j\beta_k} \pi(\mathbf{S}[e_i, e_i]),$$

where i, j, k denote three distinct integers in $\{1, 2, 3\}$, and show that

$$\begin{aligned} \tilde{G}(\mathbf{e}_1; \mathbf{e}_1; \psi_a) &= G(\mathbf{e}_1; \mathbf{e}_1), & \text{for } a \in \{5, 6, 7, 8, 15\}, \\ \tilde{G}(\mathbf{e}_2; \mathbf{e}_2; \psi_a) &= G(\mathbf{e}_2; \mathbf{e}_2), & \text{for } a \in \{3, 4, 7, 8, 13, 14\}, \\ \tilde{G}(\mathbf{e}_3; \mathbf{e}_3; \psi_a) &= G(\mathbf{e}_3; \mathbf{e}_3), & \text{for } a \in \{2, 4, 6, 8, 9, 10, 11, 12\}. \end{aligned}$$

We have used *Mathematica* to check explicitly that the condition shown in (15) holds for all other $\mathbf{E} \in \Gamma_e$ as well.

5.4 Four and higher number of sequences

For four and higher number of sequences, i.e. $r \geq 4$, there is more than one inequivalent tree that can be used as the underlying phylogenetic tree, and the number of states in the HMM of [3] depends on the tree being used. An r -leaved star-shaped tree is a tree with exactly one internal vertex of degree- r and r degree-1 vertices. An r -leaved binary tree is a tree with $r - 2$ degree-3 vertices and r degree-1 vertices.

\mathcal{E}	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	
\mathcal{I}	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_1c_2c_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_2\bar{c}_3c_1$
ψ_1	$\bar{\gamma}a_1a_2a_3$	$\bar{\gamma}a_1a_2a_3\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}a_1a_2a_3\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}a_1a_2a_3\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}a_1a_2a_3\bar{c}_1c_2c_3$	$\bar{\gamma}a_1a_2a_3\bar{c}_2\bar{c}_3c_1$
ψ_2	$\bar{\gamma}a_1a_2\bar{b}_3$	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_1c_2c_3$	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_2\bar{c}_3c_1$
ψ_3	$\bar{\gamma}a_1a_3\bar{b}_2$	$\bar{\gamma}a_1a_3\bar{b}_2\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}a_1a_3\bar{b}_2\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}a_1a_3\bar{b}_2\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}a_1a_3\bar{b}_2\bar{c}_1c_2c_3$	$\bar{\gamma}a_1a_3\bar{b}_2\bar{c}_2\bar{c}_3c_1$
ψ_4	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_1c_2c_3$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_2\bar{c}_3c_1$
ψ_5	$\bar{\gamma}a_2a_3\bar{b}_1$	$\bar{\gamma}a_2a_3\bar{b}_1\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}a_2a_3\bar{b}_1\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}a_2a_3\bar{b}_1\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}a_2a_3\bar{b}_1\bar{c}_1c_2c_3$	$\bar{\gamma}a_2a_3\bar{b}_1\bar{c}_2\bar{c}_3c_1$
ψ_6	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_1c_2c_3$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_2\bar{c}_3c_1$
ψ_7	$\bar{\gamma}a_3\bar{b}_1\bar{b}_2$	$\bar{\gamma}a_3\bar{b}_1\bar{b}_2\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}a_3\bar{b}_1\bar{b}_2\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}a_3\bar{b}_1\bar{b}_2\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}a_3\bar{b}_1\bar{b}_2\bar{c}_1c_2c_3$	$\bar{\gamma}a_3\bar{b}_1\bar{b}_2\bar{c}_2\bar{c}_3c_1$
ψ_8	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_1c_2c_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_2\bar{c}_3c_1$
ψ_9	$\bar{\gamma}a_1a_2\bar{b}_3$	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_1c_2c_3$	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_2\bar{c}_3c_1$
ψ_{10}	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_1c_2c_3$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_2\bar{c}_3c_1$
ψ_{11}	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_1c_2c_3$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_2\bar{c}_3c_1$
ψ_{12}	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_1c_2c_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_2\bar{c}_3c_1$
ψ_{13}	$\bar{\gamma}a_1\bar{b}_2$	$\bar{\gamma}a_1\bar{b}_2\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}a_1\bar{b}_2\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}a_1\bar{b}_2\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}a_1\bar{b}_2\bar{c}_1c_2c_3$	$\bar{\gamma}a_1\bar{b}_2\bar{c}_2\bar{c}_3c_1$
ψ_{14}	$\bar{\gamma}\bar{b}_1\bar{b}_2$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{c}_1c_2c_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{c}_2\bar{c}_3c_1$
ψ_{15}	$\bar{\gamma}\bar{b}_1$	$\bar{\gamma}\bar{b}_1\bar{c}_1\bar{c}_2\bar{c}_3$	$\bar{\gamma}\bar{b}_1\bar{c}_1\bar{c}_2c_3$	$\bar{\gamma}\bar{b}_1\bar{c}_1\bar{c}_3c_2$	$\bar{\gamma}\bar{b}_1\bar{c}_1c_2c_3$	$\bar{\gamma}\bar{b}_1\bar{c}_2\bar{c}_3c_1$

	ψ_6	ψ_7	ψ_8	ψ_9	ψ_{10}	ψ_{11}	ψ_{12}	ψ_{13}	ψ_{14}	ψ_{15}
\mathcal{I}	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_2c_1c_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_3c_1c_2$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3c_1c_2c_3$	0	0	0	b_3	0	\bar{b}_3b_2	$\bar{b}_2\bar{b}_3b_1$
ψ_1	$\bar{\gamma}a_1a_2a_3\bar{c}_2c_1c_3$	$\bar{\gamma}a_1a_2a_3\bar{c}_3c_1c_2$	$\bar{\gamma}a_1a_2a_3c_1c_2c_3$	\bar{a}_3	0	0	0	$a_3\bar{a}_2$	0	$a_2a_3\bar{a}_1$
ψ_2	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_2c_1c_3$	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_3c_1c_2$	$\bar{\gamma}a_1a_2\bar{b}_3c_1c_2c_3$	b_3	0	0	0	$\bar{a}_2\bar{b}_3$	0	$a_2\bar{a}_1\bar{b}_3$
ψ_3	$\bar{\gamma}a_1a_3\bar{b}_2\bar{c}_2c_1c_3$	$\bar{\gamma}a_1a_3\bar{b}_2\bar{c}_3c_1c_2$	$\bar{\gamma}a_1a_3\bar{b}_2c_1c_2c_3$	0	\bar{a}_3	0	0	a_3b_2	0	$a_3\bar{a}_1\bar{b}_2$
ψ_4	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_2c_1c_3$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_3c_1c_2$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3c_1c_2c_3$	0	b_3	0	0	\bar{b}_3b_2	0	$\bar{a}_1\bar{b}_2\bar{b}_3$
ψ_5	$\bar{\gamma}a_2a_3\bar{b}_1\bar{c}_2c_1c_3$	$\bar{\gamma}a_2a_3\bar{b}_1\bar{c}_3c_1c_2$	$\bar{\gamma}a_2a_3\bar{b}_1c_1c_2c_3$	0	0	\bar{a}_3	0	0	$a_3\bar{a}_2$	$a_2a_3b_1$
ψ_6	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_2c_1c_3$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_3c_1c_2$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3c_1c_2c_3$	0	0	b_3	0	0	$\bar{a}_2\bar{b}_3$	$a_2\bar{b}_3b_1$
ψ_7	$\bar{\gamma}a_3\bar{b}_1\bar{b}_2\bar{c}_2c_1c_3$	$\bar{\gamma}a_3\bar{b}_1\bar{b}_2\bar{c}_3c_1c_2$	$\bar{\gamma}a_3\bar{b}_1\bar{b}_2c_1c_2c_3$	0	0	0	\bar{a}_3	0	a_3b_2	$a_3\bar{b}_2b_1$
ψ_8	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_2c_1c_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_3c_1c_2$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3c_1c_2c_3$	0	0	0	b_3	0	\bar{b}_3b_2	$\bar{b}_2\bar{b}_3b_1$
ψ_9	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_2c_1c_3$	$\bar{\gamma}a_1a_2\bar{b}_3\bar{c}_3c_1c_2$	$\bar{\gamma}a_1a_2\bar{b}_3c_1c_2c_3$	b_3	0	0	0	$\bar{a}_2\bar{b}_3$	0	$a_2\bar{a}_1\bar{b}_3$
ψ_{10}	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_2c_1c_3$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3\bar{c}_3c_1c_2$	$\bar{\gamma}a_1\bar{b}_2\bar{b}_3c_1c_2c_3$	0	b_3	0	0	\bar{b}_3b_2	0	$\bar{a}_1\bar{b}_2\bar{b}_3$
ψ_{11}	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_2c_1c_3$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3\bar{c}_3c_1c_2$	$\bar{\gamma}a_2\bar{b}_1\bar{b}_3c_1c_2c_3$	0	0	b_3	0	0	$\bar{a}_2\bar{b}_3$	$a_2\bar{b}_3b_1$
ψ_{12}	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_2c_1c_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3\bar{c}_3c_1c_2$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{b}_3c_1c_2c_3$	0	0	0	b_3	0	\bar{b}_3b_2	$\bar{b}_2\bar{b}_3b_1$
ψ_{13}	$\bar{\gamma}a_1\bar{b}_2\bar{c}_2c_1c_3$	$\bar{\gamma}a_1\bar{b}_2\bar{c}_3c_1c_2$	$\bar{\gamma}a_1\bar{b}_2c_1c_2c_3$	0	0	0	0	b_2	0	$\bar{a}_1\bar{b}_2$
ψ_{14}	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{c}_2c_1c_3$	$\bar{\gamma}\bar{b}_1\bar{b}_2\bar{c}_3c_1c_2$	$\bar{\gamma}\bar{b}_1\bar{b}_2c_1c_2c_3$	0	0	0	0	0	b_2	\bar{b}_2b_1
ψ_{15}	$\bar{\gamma}\bar{b}_1\bar{c}_2c_1c_3$	$\bar{\gamma}\bar{b}_1\bar{c}_3c_1c_2$	$\bar{\gamma}\bar{b}_1c_1c_2c_3$	0	0	0	0	0	0	b_1

Table 4: A summary of transition probabilities in the hidden Markov model for three sequences. The probability $p(\eta, \zeta)$ of making a transition from η to ζ is shown in the (η, ζ) entry of the above tables.

Consider the case $r = 4$. If a star-shaped tree is used, then there are 31 states in Ψ . Of these, exactly one state is silent and therefore the simple expression in (24) for effective transition probabilities can be used. Hence, if a star-shaped tree is used, it is not difficult to check the sufficient condition of Proposition 1 analytically as we have done in the previous two subsections. If a binary tree is used, however, then there are 45 states in Ψ , of which 3 are silent. In that case, the general form of effective transition probabilities $C(\psi, \zeta)$ has to be used. As shown in (4), the formula for $C(\psi, \zeta)$ involves a matrix inversion, and so in general one ends up with a complicated expression for $C(\psi, \zeta)$ when there is more than one silent state. Hence, in general it is impractical to check the sufficient condition of Proposition 1 analytically if a binary tree is used. It is possible, however, to check the sufficient condition numerically using arbitrary parameter values. In the example of 4 sequences involving a 4-leaved binary tree, we have tried different sets of parameter values and have numerically verified that the sufficient condition holds in each case.

Using our approach, we are not able to prove that the sufficient condition of Proposition 1 holds for arbitrary number of sequences. Nevertheless, as mentioned before, it has been shown in [4] that there exists a reduced recursion for statistical alignment of arbitrary number of sequences on a binary tree. It thus seems likely that the sufficient condition in Proposition 1 also holds for $r > 4$.

6 Concluding remarks

In this paper, we have constructed a sufficient condition for reducing state-dependent recursions in a certain class of hidden Markov models. If the condition is satisfied, then the dependence of the recursions on Markov states can be eliminated. When an HMM contains several states with the same emission pattern, a corresponding reduced recursion, if it exists, can considerably accelerate the computation of likelihoods, possibly making a computationally intractable problem tractable.

In the case of few sequences, we have explicitly shown that a TKF-based hidden Markov model for statistical alignment satisfies the sufficient condition, but it still remains unclear why it actually does so. It would be interesting to investigate what kind of underlying structure is required for the condition to hold in general. Finding more affirmative examples through which we can gain more intuition would help unravel that puzzle. If testing the condition analytically is too cumbersome or impractical, then one could try to test the condition numerically as we have done in the statistical alignment example with 4 sequences. If the sufficient condition is satisfied numerically, then numerical values of likelihoods can be computed using the reduced recursion shown in Proposition 1.

In our work, we have assumed that every state emits at most one letter per string. Suppose that k is some non-negative integer and that every string in $\mathcal{S} = (S_1, \dots, S_r)$ has length divisible by k . If every state in the HMM emits either zero or k letters per string, then the result obtained in this paper should extend to that case as well; one can rescale all string lengths by dividing by k and regard \mathcal{A}^3 as the new

alphabet. It would be interesting to investigate whether a sufficient condition for reduction exists for more general cases.

In closing, we mention that a simpler version of the sufficient condition in Proposition 1 may exist. The condition (15) encodes certain constraints on the structure of an HMM, and it might be possible to translate these constraints into more transparent graphical statements regarding HMM architecture diagrams.

Acknowledgments

The author gratefully acknowledges A. Hobolth, G.A. Lunter, I. Miklós and J. Hein for useful discussions. Also, two anonymous referees are thanked for helpful suggestions on improving the presentation of this manuscript. This research is supported by EPSRC under grant HAMJW and by MRC under grant HAMKA.

References

- [1] DURBIN, R., EDDY, S. R., KROGH, A., AND MITCHISON, G. (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge.
- [2] HEIN, J. (2001). An algorithm for statistical alignment of sequences related by a binary tree. In *Proceedings of the Pacific Symposium on Biocomputing*, eds. R. B. Altman, A. K. Dunker, L. Hunter, K. Lauderdale, and T. E. Klein. World Scientific, pp. 179–190,
- [3] HEIN, J., JENSEN, J. L. AND PEDERSEN, C. N. S. (2003). Recursions for statistical multiple alignment. *PNAS* **100**, 14960–14965.
- [4] LUNTER, G. A., MIKLÓS, I., SONG, Y. S. AND HEIN, J. (2003). An efficient algorithm for statistical multiple alignment. *J. Comput. Biol.* **10**, 869–889.
- [5] RABINER, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77**, 257–286.
- [6] STEEL, M. AND HEIN, J. (2001). Applying the Thorne-Kishino-Felsenstein model to sequence evolution on a star-shaped tree. *Appl. Math. Let.* **14**, 679–684.
- [7] THORNE, J. L., KISHINO, H. AND FELSENSTEIN, J. (1991). An evolutionary model for maximum likelihood alignment of DNA sequences. *J. Mol. Evol.* **33**, 114–124.