CrossMark

Quantum hidden Markov models based on transition operation matrices

Michał Cholewa[1] · Piotr Gawron[1] ·
Przemysław Głomb[1] · Dariusz Kurzyk[1,2] (iD)

Abstract In this work, we extend the idea of quantum Markov chains (Gudder in J Math Phys 49(7):072105 [3]) in order to propose quantum hidden Markov models (QHMMs). For that, we use the notions of transition operation matrices and vector states, which are an extension of classical stochastic matrices and probability distributions. Our main result is the Mealy QHMM formulation and proofs of algorithms needed for application of this model: Forward for general case and Vitterbi for a restricted class of QHMMs. We show the relations of the proposed model to other quantum HMM propositions and present an example of application.

Keywords Hidden Markov models · Open quantum walks · Transition operation matrices

✉ Dariusz Kurzyk
  dkurzyk@iitis.pl

  Michał Cholewa
  mcholewa@iitis.pl

  Piotr Gawron
  gawron@iitis.pl

  Przemysław Głomb
  przemg@iitis.pl

[1] Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Bałtycka 5,
  44-100 Gliwice, Poland

[2] Institute of Mathematics, Silesian University of Technology, Kaszubska 23, Gliwice 44-100, Poland

1 Introduction

The most basic Markov model is a Markov chain, which can be defined as a stochastic process with the Markov property. Formally, a Markov chain is a collection of random variables $\{n_t, t \geq 0\}$ having the property that $P(n_{t+1} = S_{k_{t+1}} | n_1 = S_{k_1}, n_2 = S_{k_2}, \ldots, n_t = S_{k_t}) = P(n_{t+1} = S_{k_{t+1}} | n_t = S_{k_t})$, where the values $\{S_1, \ldots, S_T\}$ of n_t are called states. They form the state space of the chain. According to the Markov property, the current state of a chain is only dependent on the previous state. Moreover, the state of a Markov chain is directly observed in each step. Any Markov chain can be described by a directed graph called the state diagram, where vertices are associated with states and each edge (i, j) is labeled by the probability of going from ith state to jth state. The information about Markov chain can be also represented by the initial state $\cdot$ and the stochastic matrix called transition matrix $P = [p_{ij}]$, such that $p_{ij} = P(n_{t+1} = S_j | n_t = S_i)$. If we consider a Markov chain where states are not observed directly, and these states generate symbols according to some random variables, then we obtain a hidden Markov model (HMM). Hence, in the case of a Markov chain, the states correspond with observations, but for a HMM, the states correspond with the random source of observations.

The classical hidden Markov model was introduced as a method of modeling signal sources observed in noise. It is now extensively used, e.g., in speech and gesture recognition or biological sequence analysis. Their popularity is a result of their versatile structure, which is able to model wide variety of problems, and effective algorithms that facilitate their application. The HMM is related to three fundamental: given a sequence of symbols of length T, $O = (o_1, o_2, \ldots, o_T)$, and a HMM parametrized by λ,

1. Compute the $P(O|\lambda)$, probability that the sequence O can be produced by a HMM λ.
2. Select the sequence of state indexes $N_T = (n_0, n_1, \ldots, n_T)$ that maximizes the probability $P(O|\lambda, N_T)$, in other words the most likely state sequence in HMM λ that produces O.
3. Adjust the model parameters λ to maximize $P(O|\lambda)$.

The above problems are solved, respectively, by the Forward, Vitterbi and Baum–Welch algorithms. The effectiveness of those algorithms is based on optimized procedure of computation, which uses a 'trellis': a two-dimensional lattice structure of observations and states. This formulation is based on the Markov property of model evolution and reduces the complexity from exponential $\mathcal{O}(T N^T)$ to polynomial $\mathcal{O}(N^2 T)$, where T is the number of observations and N is the number of model states [1].

Depending on the formulation, there are two definitions of a hidden Markov model: Mealy and Moore. In the former, the probability of next state being n_{t+1} depends both on the current state n_t and the generated output symbol o_t. In the latter, the symbol generation is independent from state switch, i.e., $P(n_{t+1} = S_i | o_{t+1} = o, n_t = S_j) = P(n_{t+1} = S_i | n_t = S_j)$. While the expressive power of Moore and Mealy models is the same, i.e., a process can be realized with Moore model if and only if it is realizable by Mealy model, the minimal model order for the realization is lower in Mealy models [2]. In this work, we focus only on Mealy models.

1.1 Related work

In this work, we follow the scheme proposed by Gudder in [3] and extend it in order to construct quantum hidden Markov models (QHMMs). Gudder introduced the notions of transition operation matrices and vector states, which give an elegant extension of classical stochastic matrices and probability distributions. These notions allow to define Markov processes that exhibit both classical and quantum behaviors.

Below we review two areas of research most closely related to our work: open quantum walks and Hidden quantum Markov models.

Open quantum walks　In recent years, a new subfield of quantum walks has emerged. In series of papers [4–9], Attal, Sabot, Sinayskiy and Petruccione introduced the notion of open quantum walks. Theorems for limit distributions of open quantum random walks were provided in [10]. In [11] the average position and the symmetry of distribution in the $SU(2)$ open quantum walk are studied. The notion of open quantum walks is generalized to quantum operations of any rank in [12] and analyzed in [13]. In first of these two papers, the notion of mean first passage time for a generalized quantum walk is introduced and studied for class of walks on Apollonian networks. In the second paper, a central limit theorem for reducible and irreducible open quantum walks is provided. In a recent paper [14], authors introduce the notion of hybrid quantum automaton—an object similar to quantum hidden Markov model. They use hybrid quantum automata and derived concepts in application to model checking.

Quantum hidden Markov models　Hidden quantum Markov models were introduced in [15]. The construction provided there by the authors is different from ours. In their work, the hidden quantum Markov model consists of a set of quantum operations associated with emission symbols. The evolution of the system is governed by the application of quantum operations on a quantum state. The sequence of emitted symbols defines the sequence of quantum operations being applied on the initial state of the hidden quantum Markov model.

1.2 Our contribution

In this work, we propose a quantum hidden Markov model formulation using the notions of transition operation matrices. We focus on Mealy models, for which we derive first the Forward algorithm in general case, then the Vitterbi algorithm, for models restricted to those in which sub-TOMs' elements are trace-monotonicity preserving quantum operations. Subsequently, we discuss the relationship between our model and model presented in [15]. The paper ends with the example of application of proposed model.

The paper is organized as follows: In Sect. 2, we collect the basic mathematical objects and their properties; in Sect. 3, we define quantum hidden Markov models and provide Forward and Viterbi algorithms for these models; in Sect. 4, we discuss the correspondences between proposed models and models described in [15]; Sect. 5 contains examples of application of our model; and finally in Sect. 6 we conclude.

2 Transition operation matrices

In what follows, we provide basic elements of quantum information theory and summarize definitions and properties of objects introduced by Gudder in [3].

2.1 Quantum theory

Let $\mathcal{H}$ be a complex finite Hilbert space and $\mathcal{L}(\mathcal{H})$ be the set of linear operators on $\mathcal{H}$. We also denote the set of positive operators on $\mathcal{H}$ as $\mathcal{P}^+(\mathcal{H})$ and the set of positive semi-definite operators on $\mathcal{H}$ as $\mathcal{P}(\mathcal{H})$.

Definition 1 (*Quantum state*) A linear operator $\rho \in \mathcal{P}(\mathcal{H})$ is called a quantum state if tr $\rho = 1$. Set of quantum states is denoted by $\Omega(\mathcal{H})$.

Definition 2 (*Sub-normalized quantum state*) A linear operator $\rho \in \mathcal{P}(\mathcal{H})$ is called sub-normalized [16] quantum state if tr $\rho \leq 1$. Set of sub-normalized quantum states is denoted by $\Omega_\leq(\mathcal{H})$.

Definition 3 (*Positive map*) A linear map $\Phi \in \mathcal{L}(\mathcal{L}(\mathcal{H}_1), \mathcal{L}(\mathcal{H}_2))$ is called positive map if, for every $\rho \in \mathcal{P}(\mathcal{H}_1)$, $\Phi(\rho) \in \mathcal{P}(\mathcal{H}_2)$.

Definition 4 (*Completely positive map*) A linear map $\Phi \in \mathcal{L}(\mathcal{L}(\mathcal{H}_1), \mathcal{L}(\mathcal{H}_2))$ is called completely positive (CP) if for any complex Hilbert space $\mathcal{H}_3$, the map $\Phi \otimes \mathbb{1} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_3), \mathcal{L}(\mathcal{H}_2 \otimes \mathcal{H}_3))$ is positive.

Definition 5 (*Trace preserving map*) A linear map $\Phi \in \mathcal{L}(\mathcal{L}(\mathcal{H}_1), \mathcal{L}(\mathcal{H}_2))$ is called trace preserving if $\text{tr}(\Phi(\rho)) = \text{tr } \rho$ for every $\rho \in \mathcal{L}(\mathcal{H}_1)$.

Definition 6 (*Trace non-increasing map*) A linear map $\Phi \in \mathcal{L}(\mathcal{L}(\mathcal{H}_1), \mathcal{L}(\mathcal{H}_2))$ is called trace non-increasing if $\text{tr}(\Phi(\rho)) \leq \text{tr } \rho = 1$ for every quantum state $\rho \in \Omega(\mathcal{H}_1)$.

Definition 7 (*Quantum operation*) A linear map $\Phi \in \mathcal{L}(\mathcal{L}(\mathcal{H}_1), \mathcal{L}(\mathcal{H}_2))$ is called a quantum operation if it is completely positive and trace non-increasing.

Definition 8 (*Quantum channel*) A linear map $\Phi \in \mathcal{L}(\mathcal{L}(\mathcal{H}_1), \mathcal{L}(\mathcal{H}_2))$ is called a quantum channel if it is completely positive and trace preserving.

Definition 9 (*Quantum measurement*) By quantum measurement we call a mapping from a finite set Θ of measurement outcomes to subset of set of measurement operators $\mu : \Theta \to \mathcal{P}(\mathcal{H})$ such that $\sum_{a \in \Theta} \mu(a) = \mathbb{1}$.

With each measurement μ, we associate a nonnegative functional $p : \Theta \to \mathbb{R}_+ \cup \{0\}$ which maps measurement outcome a for a given positive operator ρ and measurement μ to nonnegative real number in the following way $p(a)_\rho = \text{tr } \mu(a)\rho$. If tr $\rho = 1$, for given ρ and μ the value of p can be interpreted as probability of obtaining measurement outcome a in quantum state ρ.

If ρ is a sub-normalized state, the trivial measurement $\mu : a_e \mapsto \mathbb{1}$ measures the probability $p(a_e)_\rho = \text{tr } \rho$ that the state ρ exists. One should note that this kind of measurement commutes with any other measurement and thus does not disturb the quantum system.

2.2 Transition operation matrices

The core object of the Gudder's scheme is transition operation matrix (TOM) which generalizes the idea of stochastic matrix.

Definition 10 (*Transition operation matrix*) Let $\mathcal{H}_1, \mathcal{H}_2$ denote two finite-dimensional Hilbert spaces and $\Omega(\mathcal{H}_1), \Omega(\mathcal{H}_2)$ denote sets of quantum states acting on those spaces, respectively.

A TOM is a matrix in form $\mathcal{E} = \{\mathcal{E}_{ij}\}_{i,j=1}^{M,N}$, where $\mathcal{E}_{ij}$ is completely positive map in $\mathcal{L}(\mathcal{L}(\mathcal{H}_1), \mathcal{L}(\mathcal{H}_2))$ such that for every j and $\rho \in \Omega(\mathcal{H}_1)$ $\sum_i \mathcal{E}_{ij}(\rho) \in \Omega(\mathcal{H}_2)$.

Alternatively one can say that $\mathcal{E} = \{\mathcal{E}_{ij}\}_{i,j=1}^{M,N}$ is a TOM if and only if for every column j $\sum_i \mathcal{E}_{ij}$ is a quantum channel (completely positive trace preserving map). A simple implication of this definition is that each $\mathcal{E}_{ij}$ is CP-TNI mapping.

Note that in this definition TOM has four parameters:

 – size of matrix "output" (number of rows)—M,
 – size of matrix "input" (number of columns)—N,
 – "input" Hilbert space—$\mathcal{H}_1$,
 – "output" Hilbert space—$\mathcal{H}_2$.

The set of TOMs we will denote as $\Gamma^{M,N}(\mathcal{H}_1, \mathcal{H}_2)$.

Definition 11 (*Sub-transition operation matrix*) Let $\mathcal{H}_1, \mathcal{H}_2$ denote two finite-dimensional Hilbert spaces, $\Omega(\mathcal{H}_1)$ denotes set of quantum states acting on the first space, and $\Omega^{\leq}(\mathcal{H}_2)$ denotes set of sub-normalized quantum states acting on the second Hilbert space.

A sub-TOM is a matrix in the form $\mathcal{E} = \{\mathcal{E}_{ij}\}_{i,j=1}^{M,N}$, where $\mathcal{E}_{ij}$ is completely positive map in $\mathcal{L}(\mathcal{L}(\mathcal{H}_1), \mathcal{L}(\mathcal{H}_2))$ such that for every j and $\rho \in \Omega(\mathcal{H}_1)$. $\sum_i \mathcal{E}_{ij}(\rho) \in \Omega^{\leq}(\mathcal{H}_2)$.

The set of sub-TOMs we will denote as $\Gamma_{\leq}^{M,N}(\mathcal{H}_1, \mathcal{H}_2)$.

Definition 12 (*Quantum Markov chain*) Let a TOM $\mathcal{E} = \{\mathcal{E}_{ij}\}_{i,j=1}^{M,N}$ be given. Quantum Markov chain is a finite directed graph $G = (E, V)$ labeled by $\mathcal{E}_{ij}$ for $e \in E$ and by zero operator for $e \notin E$.

Definition 13 (*Vector state*) Vector state is a column vector $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N]^T$ such that $\alpha_i \in \Omega_{\leq}(\mathcal{H})$ are sub-normalized quantum states and $\sum_{i=1}^N \alpha_i \in \Omega(\mathcal{H})$. We will denote the set of vector states as $\Delta^N(\mathcal{H})$.

Definition 14 (*Sub-normalized vector state*) Sub-normalized vector state is a column vector $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N]^T$ such that $\alpha_i \in \Omega_{\leq}(\mathcal{H})$ are sub-normalized quantum states and $\sum_{i=1}^N \alpha_i \in \Omega_{\leq}(\mathcal{H})$. We will denote a set of sub-normalized vector states as $\Delta_{\leq}^N(\mathcal{H})$.

Theorem 1 (Gudder [3]) *Applying TOM $\mathcal{E} \in \Gamma^{M,N}(\mathcal{H}_1, \mathcal{H}_2)$ on a vector state $\alpha \in \Delta^N(\mathcal{H}_1)$ produces vector state $\beta = \mathcal{E}(\alpha) \in \Delta^M(\mathcal{H}_2)$ where $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N]^T$, $\alpha_i \in \Omega^{\leq}(\mathcal{H}_1)$, where $\beta = [\beta_1, \beta_2, \ldots, \beta_M]^T$, $\beta_i \in \Omega^{\leq}(\mathcal{H}_2)$, and $\mathcal{E} \in \Gamma^{M,N}(\mathcal{H}_1, \mathcal{H}_2)$, and in the following way $\beta_i = \sum_{j=1}^N \mathcal{E}_{ij}(\alpha_j)$.*

Theorem 2 (Gudder [3]) *Product of TOM $\mathcal{A} \in \Gamma^{M,N}(\mathcal{H}_1, \mathcal{H}_2)$ and $\mathcal{B} \in \Gamma^{N,K}(\mathcal{H}_2, \mathcal{H}_3)$ is a TOM $\Gamma^{M,K}(\mathcal{H}_1, \mathcal{H}_3) \ni \mathcal{C} = \mathcal{B}\mathcal{A}$.*

Lemma 1 (Product of two sub-TOMs is a sub-TOM) *Product of sub-TOMs $\mathcal{A} \in \Gamma_{\leq}^{M,N}(\mathcal{H}_1, \mathcal{H}_2)$ and $\mathcal{B} \in \Gamma_{\leq}^{N,K}(\mathcal{H}_2, \mathcal{H}_3)$ is a sub-TOM $\Gamma_{\leq}^{M,K}(\mathcal{H}_1, \mathcal{H}_3) \ni \mathcal{C} = \mathcal{B}\mathcal{A}$.*

Proof of Lemma 1 According to proof of Lemma 2.2 in [3], $C_{ij} = \mathcal{B}_{ij}\mathcal{A}_{ij}$ is a completely positive map. For every $\rho \in \Omega(\mathcal{H}_1)$ and j we have that $\sigma = \sum_{i=1}^{M} \mathcal{A}_{ij}(\rho) \in \Omega^{\leq}(\mathcal{H}_2)$. If $\mathrm{tr}(\sigma) > 0$ then $\tilde{\sigma} = \sigma/\mathrm{tr}(\sigma) \in \Omega(\mathcal{H}_2)$ and

$$
\begin{aligned}
\mathrm{tr}\left(\sum_{i=1}^{M} \mathcal{B}_{ij}(\sigma)\right) &= \mathrm{tr}\left(\mathrm{tr}(\sigma)\sum_{i=1}^{M} \mathcal{B}_{ij}(\tilde{\sigma})\right) \\
&= \mathrm{tr}(\sigma)\,\mathrm{tr}\left(\sum_{i=1}^{M} \mathcal{B}_{ij}(\tilde{\sigma})\right) \leq 1.
\end{aligned}
\tag{1}
$$

In the case where $\mathrm{tr}(\sigma) = 0$, the σ is the zero operator and $\sum_{i=1}^{M} \mathcal{B}_{ij}(\sigma)$ is also the zero operator. Thus $\mathrm{tr}\left(\sum_{i=1}^{M} \mathcal{B}_{ij}(\sigma)\right) = 0$. Hence, $\sum_{i=1}^{M} C_{i,j}(\rho) \in \Omega^{\leq}(\mathcal{H}_3)$ and $\mathcal{C} \in \Gamma_{\leq}^{M,K}(\mathcal{H}_1, \mathcal{H}_3)$. $\qquad\square$

Product of (sub-)TOMs that have same dimensions is associative, i.e., $(\mathcal{E}\mathcal{F})\mathcal{G} = \mathcal{E}(\mathcal{F}\mathcal{G})$ and $(\mathcal{E}\mathcal{F})(\alpha) = \mathcal{E}(\mathcal{F}(\alpha))$.

3 Mealy quantum hidden Markov model

In order to explain the idea of QHMM, we can form following analogy. A QHMM might be understood as a system consisting of a particle that has an internal sub-normalized quantum state $\rho \in \Omega_{\leq}(\mathcal{H})$, and it occupies a classical state S_i. This particle hops from one classical state S_i into another state S_j passing trough a quantum operation associated with a sub-TOM element $\mathcal{P}_{S_j,S_i}^{V_k}$. With each transition, a symbol V_k is emitted from the system.

We will now define the classical and quantum version of the Mealy hidden Markov model.

Definition 15 (*Finite sequences*) Let

$$
\mathcal{V}^T = \underbrace{\mathcal{V} \times \mathcal{V} \times \cdots \times \mathcal{V}}_{T}
$$

defines the set of sequences of length T over alphabet $\mathcal{V}$.

Definition 16 (*Mealy hidden Markov model*) Let $\mathcal{S} = \{S_1, \ldots, S_N\}$ and $\mathcal{V} = \{V_1, \ldots, V_M\}$ be a set of states and an alphabet, respectively. The Mealy HMM is specified by a tuple $\lambda = (\mathcal{S}, \mathcal{V}, \Pi, \pi)$, where

- $\pi \in [0, 1]^N$ is a stochastic vector representing initial states, where π_i is the probability that the initial state is S_i;
- Π is a mapping $\mathcal{V} \ni V_i \mapsto \Pi^{V_i} \in \mathbb{R}^{N,N}$, where Π^{V_i} is sub-stochastic matrix, such that $\Pi^{\Sigma} := \sum_{i=1}^{M} \Pi^{V_i} \in \mathbb{R}^{N,N}$ is stochastic matrix and $\Pi_{j,k}^{V_i}$ is $p(n_{t+1} = S_k, o_{t+1} = V_i | n_t = S_j)$, that is probability of going from state j to k while generating the output V_i.

Let $O = o_1 o_2, \ldots o_T \in \mathcal{V}^T$ be a sequence of length T and $P : \mathcal{V}^T \to [0, 1]$ be string probabilities, defined as $P(O) = p(O(1) = o_1, O(2) = o_2, \ldots, O(T) = o_T)$. The concatenation of string O and o_{T+1} is denoted by Oo_{T+1}.

It is well known that for HMMs the function P satisfies

- $\sum_{O \in \mathcal{V}^T} P(O) = 1$ and
- $\sum_{o_{T+1} \in \mathcal{V}} P(Oo_{T+1}) = P(O)$, which follows from the law of total probability.

The string probabilities generated by Mealy HMM $\lambda = (\mathcal{S}, \mathcal{V}, \Pi, \pi)$ are given by

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_i,$$

where α_i is ith element of $\alpha = \Pi^{o_T} \Pi^{o_{T-1}} \ldots \Pi^{o_1} \pi$.

Definition 17 (*Mealy quantum hidden Markov model*) Let $\mathcal{S}$ and $\mathcal{V}$ be a set of states and an alphabet, respectively. Mealy QHMM is specified by a tuple $\lambda = (\mathcal{S}, \mathcal{V}, \mathcal{P}, \pi)$, where

- $\pi \in \Delta^N(\mathcal{H})$ is an initial vector state;
- $\mathcal{P}$ is a mapping $\mathcal{V} \to \Gamma_{\leq}^{N,N}(\mathcal{H}, \mathcal{H})$ such that $\mathcal{P}^S := \sum_{V_i \in \mathcal{V}} \mathcal{P}^{V_i} \in \Gamma^{N,N}(\mathcal{H}, \mathcal{H})$ is a TOM, with $\mathcal{P}^{V_i}$ being value of $\mathcal{P}$ for V_i.

As an example we give a three-state two-symbol Mealy QHMM $\lambda = (\mathcal{S}, \mathcal{V}, \Pi, \pi)$, with

$$\mathcal{S} = \{S_1, S_2, S_3\},$$
$$\mathcal{V} = \{V_1, V_2\},$$
$$\Pi = \left\{ V_1 \mapsto \mathcal{P}^{V_1}, V_2 \mapsto \mathcal{P}^{V_2} \right\},$$
$$\pi = \begin{bmatrix} \pi_{S_1} \\ \pi_{S_2} \\ \pi_{S_3} \end{bmatrix},$$
$$\mathcal{P}^{V_1} = \begin{bmatrix} \mathcal{P}_{S_1 S_1}^{V_1} & \mathcal{P}_{S_1 S_2}^{V_1} & \mathcal{P}_{S_1 S_3}^{V_1} \\ \mathcal{P}_{S_2 S_1}^{V_1} & \mathcal{P}_{S_2 S_2}^{V_1} & \mathcal{P}_{S_2 S_3}^{V_1} \\ \mathcal{P}_{S_3 S_1}^{V_1} & \mathcal{P}_{S_3 S_2}^{V_1} & \mathcal{P}_{S_3 S_3}^{V_1} \end{bmatrix},$$
$$\mathcal{P}^{V_2} = \begin{bmatrix} \mathcal{P}_{S_1 S_1}^{V_2} & \mathcal{P}_{S_1 S_2}^{V_2} & \mathcal{P}_{S_1 S_3}^{V_2} \\ \mathcal{P}_{S_2 S_1}^{V_2} & \mathcal{P}_{S_2 S_2}^{V_2} & \mathcal{P}_{S_2 S_3}^{V_2} \\ \mathcal{P}_{S_3 S_1}^{V_2} & \mathcal{P}_{S_3 S_2}^{V_2} & \mathcal{P}_{S_3 S_3}^{V_2} \end{bmatrix}.$$
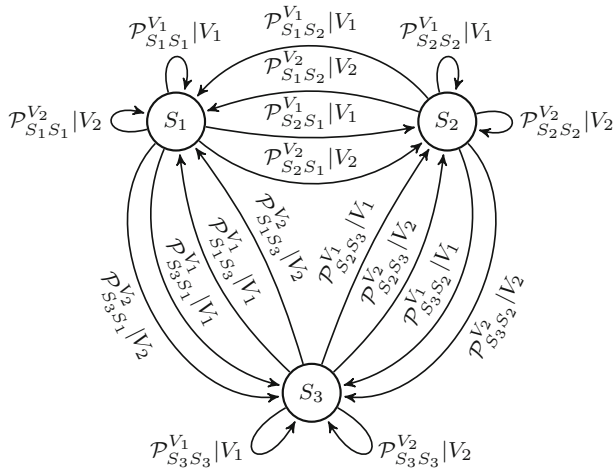
Fig. 1 Graphical representation of three-state Mealy QHMM λ, whose alphabet consists of two symbols V_1, V_2. The symbol $\mathcal{P}^{V_1}_{S_2 S_3}|V_1$ should be understood in the following way: when QHMM is in state S_3 and is being transformed to state S_2 while emitting symbol V_1, then the internal quantum sub-state is transformed by quantum operation $\mathcal{P}^{V_1}_{S_2 S_3}|V_1$

A graphical representation of this QHMM is presented in Fig. 1.

Remark 1 For dim $\mathcal{H} = 1$ QHMM reduces to classical HMM. In this case, TOMs reduce to stochastic matrices, sub-TOMs to sub-stochastic matrices, the vector states to probability vectors, sub-vector states to sub-normalized probability vectors.

3.1 Forward algorithm for Mealy QHMM

With each Mealy QHMM we can associate a mapping

$$\varrho : \mathcal{V}^* \to \Omega_{\leq}(\mathcal{H}),$$

where $\mathcal{V}^* = \bigcup_{T=1}^{\infty} \mathcal{V}^T$. Given a sequence $O = (o_1, o_2, \ldots, o_T)$ and a Mealy QHMM λ one can compute resulting sub-normalized quantum state $\rho_{O|\lambda}$.

Let us consider sub-normalized vector states

$$\alpha_T = \left[\alpha_{T,1}, \ldots, \alpha_{T,N}\right]^T \in \Delta^N_{\leq}(\mathcal{H})$$

such that

$$\alpha_T = \mathcal{P}^{o_T} \ldots \mathcal{P}^{o_2} \mathcal{P}^{o_1}(\pi), \tag{2}$$

then $\rho_{O|\lambda} := \varrho(O) = \sum_{i=1}^{N} \alpha_{T,i}$.

Equation (2) we call the Forward algorithm for QHMMs. Note that the result of this algorithm is a sub-normalized quantum state $\rho_{O|\lambda} \in \Omega_{\leq}(\mathcal{H})$. The sum of all

those states over all possible sequences of a given length forms a quantum state, as formulated in the following theorem.

Theorem 3 *For any QHMM λ we have $\sum_{O \in \mathcal{V}^T} \rho_{O|\lambda} \in \Omega(\mathcal{H})$.*

In order to prove this theorem, we will first prove the following lemma.

Lemma 2 *For any QHMM λ the following holds*

$$\sum_{O \in \mathcal{V}^T} \mathcal{P}^{o_T} \mathcal{P}^{o_{T-1}} \dots \mathcal{P}^{o_1}(\pi) \in \Delta^N(\mathcal{H}).$$

Proof Lemma 2 We will proceed by induction. For case $T = 1$, we have

$$\beta_1 = \sum_{o \in \mathcal{V}} \mathcal{P}^o(\pi) = \mathcal{P}^S(\pi) \in \Delta^N(\mathcal{H}).$$

For case $T = n + 1$

$$\begin{aligned}
\beta_T &= \sum_{O \in \mathcal{V}^{N+1}} \mathcal{P}^{o_{N+1}} \mathcal{P}^{o_N} \dots \mathcal{P}^{o_1}(\pi) \\
&= \sum_{o_{n+1} \in \mathcal{V}} \mathcal{P}^{o_{n+1}} \sum_{O \in \mathcal{V}^N} \mathcal{P}^{o_N} \mathcal{P}^{o_{N-1}} \dots \mathcal{P}^{o_1}(\pi) \\
&= \mathcal{P}^S \underbrace{\sum_{O \in \mathcal{V}^N} \mathcal{P}^{o_N} \mathcal{P}^{o_{N-1}} \dots \mathcal{P}^{o_1}(\pi)}_{\mathcal{X}}.
\end{aligned} \tag{3}$$

By inductive hypothesis $\mathcal{X}$ is a TOM. $\mathcal{P}^S$ is a TOM, therefore $\beta_T \in \Delta^N(\mathcal{H})$. □

Proof Theorem 3

$$\begin{aligned}
\sum_{O \in \mathcal{V}^T} &\rho_{O|\lambda} \\
&= \sum_{O \in \mathcal{V}^T} \sum_{i=1}^{N} \alpha_{T,i} \\
&= \sum_{O \in \mathcal{V}^T} \sum_{i=1}^{N} \left[\mathcal{P}^{o_T} \dots \mathcal{P}^{o_2} \mathcal{P}^{o_1}(\pi) \right]_i \\
&= \sum_{i=1}^{N} \left[\underbrace{\sum_{O \in \mathcal{V}^T} \mathcal{P}^{o_T} \dots \mathcal{P}^{o_2} \mathcal{P}^{o_1}(\pi)}_{\mathcal{X}} \right]_i \\
&= \sum_{i=1}^{N} [\mathcal{X}(\pi)]_i
\end{aligned} \tag{4}$$

Since by Lemma 2 $\mathcal{X}$ is a TOM, therefore $\sum_{O \in \mathcal{V}^T} \rho_{O|\lambda} \in \Omega(\mathcal{H})$. $\square$

Theorem 4 *Let $O = (o_1, o_2, \ldots, o_T)$ be a sequence of length T and Oo_{T+1} be a concatenation of O and o_{T+1}, then for any QHMM λ the following holds*

$$\sum_{o_{T+1} \in \mathcal{V}} \rho_{Oo_{T+1}|\lambda} = \rho_{O|\lambda}. \tag{5}$$

Proof Theorem 4

According to law of total probability for TOMs [3], we get

$$\sum_{o_{t+1} \in \mathcal{V}} \rho_{Oo_{t+1}|\lambda} = \sum_{o_{t+1} \in \mathcal{V}} \sum_{i=1}^{N} \mathcal{P}_i^{o_{t+1}}(\alpha_{T,i})$$

$$= \sum_{i=1}^{N} \sum_{o_{t+1} \in \mathcal{V}} \mathcal{P}_i^{o_{t+1}}(\alpha_{T,i}) \tag{6}$$

$$= \sum_{i=1}^{N} \alpha_{T,i} = \rho_{O|\lambda}.$$

$\square$

3.2 Viterbi algorithm for Mealy QHMM

We are given a QHMM λ with set of states $\mathcal{S} = \{S_1, S_2, \ldots, S_N\}$ and an alphabet of symbols $\mathcal{V} = \{V_1, V_2, \ldots, V_M\}$. We will denote $\mathcal{P}_{ij}^k := \mathcal{P}_{S_i S_j}^{V_k}$.

We have a sequence of length T, $O = (o_1, o_2, \ldots, o_T)$, of symbols from alphabet $\mathcal{V}$, $o_i \in \mathcal{V}$.

A Mealy QHMM emits symbols on transition from one state to the next. For our sequence O, we index corresponding QHMM states by n_i, i.e., n_0 is the initial state (before the emission of the first symbol), and $n_i, i \geq 1$ is the state after emission of the symbol o_i. $n_i \in \mathcal{S}$.

The goal of the algorithm is to find most likely sequence of states conditioned on a sequence of emitted symbols O.

We denote the set of partial sequences of state indexes as $N_k = \{(n_0, n_1, \ldots, n_k) : n_j \in \mathcal{S}, j = 0, 1, \ldots, k\}$, where $k \leq T$. A set beginning with n_0 and ending after k steps with S_i we denote $N_k^{S_i} = \{(n_0, n_1, \ldots, n_{k-1}, n_k = S_i) : n_j \in \mathcal{S}, j = 0, 1, \ldots, k\} \subset N_k$.

Theorem 5 *Let O be a given sequence of emissions from $\mathcal{V}$. Let $\lambda = (\mathcal{S}, \mathcal{V}, \mathcal{P}, \pi)$ be a Mealy QHMM satisfying*

$$\forall_{n_i, n_j \in \mathcal{S}, o \in O} \forall_{\alpha, \beta \in \Omega_{\leq}(\mathcal{H})} \operatorname{tr} \alpha > \operatorname{tr} \beta$$
$$\implies \operatorname{tr} \mathcal{P}^o_{n_i, n_j}(\alpha) > \operatorname{tr} \mathcal{P}^o_{n_i, n_j}(\beta) \tag{7}$$

i.e., all sub-TOMs elements are trace-monotonicity preserving quantum operations.

We define $w \in N_k^{S_i}$ to be a sequence of k states ending with S_i. A sub-normalized state associated with w and sequence O is $B_w \in \Omega_{\leq}(\mathcal{H})$ defined as $B_w = \mathcal{P}^{o_k}_{n_k, n_{k-1}} \mathcal{P}^{o_{k-1}}_{n_{k-1}, n_{k-2}} \dots \mathcal{P}^{o_1}_{n_1, n_0}(\pi_{n_0})$. The sub-normalized state that maximizes trace over set of all B_ws with $w \in N_k^{S_i}$ is

$$A_{k, S_i} = \operatorname*{argmax}_{\left\{B_w : w \in N_k^{S_i}\right\}} \operatorname{tr} B_w. \tag{8}$$

Then the following holds

$$\operatorname{tr} A_{k, S_i} = \max_{n_{k-1} \in \mathcal{S}} \operatorname{tr} \mathcal{P}^{o_k}_{n_k = S_i, n_{k-1}}(A_{k-1, n_{k-1}}). \tag{9}$$

Proof Let us denote

$$w^*_{k, S_i} = (n_0^*, \dots, n_{k-1}^*, n_k^* = S_i) \in N_k^{S_i} \tag{10}$$

as the sequence of states maximizing trace of B_w, so that

$$\operatorname{tr} A_{k, S_i} = \operatorname{tr} B_{w^*_{k, S_i}}. \tag{11}$$

We now have

$$\operatorname{tr} A_{k, S_i} = \max_{w \in N_k^{S_i}} \operatorname{tr} B_w$$
$$= \max_{n_0, \dots, n_{k-1}, n_k = S_i} \operatorname{tr} \mathcal{P}^{o_k}_{n_k, n_{k-1}} \mathcal{P}^{o_{k-1}}_{n_{k-1}, n_{k-2}} \dots \mathcal{P}^{o_1}_{n_1, n_0}(\pi_{n_0}) \tag{12}$$

Obviously

$$\operatorname{tr} A_{k, S_i} = \operatorname{tr} \mathcal{P}^{o_k}_{n_k^*, n_{k-1}^*} \mathcal{P}^{o_{k-1}}_{n_{k-1}^*, n_{k-2}^*} \dots \mathcal{P}^{o_1}_{n_1^*, n_0^*}(\pi_{n_0^*}) \tag{13}$$

We will now prove that for $n_k^* = S_i$

$$w^*_{k, S_i} = (n_0^*, \dots, n_{k-1}^*, n_k^*)$$
$$\implies w^*_{k-1, n_{k-1}^*} = (n_0^*, \dots, n_{k-1}^*). \tag{14}$$

Let us assume that it is not true. That would mean that

$$w^*_{k-1, n_{k-1}^*} = (l_0^*, \dots, l_{k-2}^*, n_{k-1}^*) \neq (n_0^*, \dots, n_{k-2}^*, n_{k-1}^*).$$

Of course

$$\text{tr } B_{(l_0^*,...,l_{k-2}^*,n_{k-1}^*)} > \text{tr } B_{(n_0^*,...,n_{k-2}^*,n_{k-1}^*)}$$

From this, and (7), we have

$$\forall_{n_k \in \mathcal{S}, y_k \in \{1,2,...,M\}} \text{ tr } \mathcal{P}^{o_k}_{n_k,n_{k-1}^*}(B_{(l_0^*,...,l_{k-2}^*,n_{k-1}^*)})$$
$$> \text{tr } \mathcal{P}^{o_k}_{n_k,n_{k-1}^*}(B_{(n_0^*,...,n_{k-2}^*,n_{k-1}^*)}),$$

that leads to

$$w_{k,S_i}^* \neq (n_0^*, \ldots, n_{k-1}^*, n_k^*)$$

which is a contradiction. That proves that implication (14) holds.

Then, for $n_k = S_i$

$$
\begin{aligned}
\text{tr } A_{k,S_i} &= \text{tr } B_{w_{k,S_i}^*} \\
&= \text{tr } \mathcal{P}^{o_k}_{n_k^*,n_{k-1}^*}(B_{w_{k,n_{k-1}^*}^*}) \\
&= \text{tr } \mathcal{P}^{o_k}_{n_k^*,n_{k-1}^*}(A_{k-1,n_{k-1}^*}) \\
&= \max_{n_{k-1} \in \mathcal{S}} \text{tr } \mathcal{P}^{o_k}_{n_k=S_i,n_{k-1}}(A_{k-1,n_{k-1}}).
\end{aligned}
\tag{15}
$$

$\square$

Remark 2 It can be easily seen that Theorem 7 holds iff quantum operation $\mathcal{P}^y_{n_j,n_i}$ is of form $c \cdot \Phi$, where $c \in [0, 1)$ and Φ is a quantum channel (CP-TP map).

From Theorem 5, we immediately derive the Viterbi algorithm for Mealy QHMMs conditioned with (7) that computes most likely sequence of states for a given sequence O.

Initialization:

$$A_{0,S_i} = \pi_{S_i} \tag{16}$$

Computation for step number k:

$$
\begin{aligned}
\forall_{S_i \in \mathcal{S}, k \in \{1,...,T\}} \; n_{k-1}^*(S_i) \\
= \underset{n_{k-1} \in \mathcal{S}}{\text{argmax}} \text{ tr } \mathcal{P}^{o_k}_{S_i,n_{k-1}}(A_{k-1,n_{k-1}})
\end{aligned}
\tag{17}
$$

$$\forall_{S_i \in \mathcal{S}, k \in \{1,...,T\}} \; A_{k,S_i} = \mathcal{P}^{o_k}_{S_i,n_{k-1}^*(S_i)}(A_{k-1,n_{k-1}^*(S_i)}), \tag{18}$$

Termination:

$$n_T^* = \underset{S_i \in \mathcal{S}}{\text{argmax}} \text{ tr } A_{T,S_i}. \tag{19}$$

The most probable state sequence is $(n_0^*, \ldots, n_T^*)$, with resulting state being A_{T,n_T^*} with probability given by tr A_{T,n_T^*}.

In case when (7) does not apply, one can resort to exhaustive search over all state sequences. As a result of the multitude of possible quantum operations the behavior of the quantum hidden Markov model can be markedly different than its classical counterpart. This is similar to the relation of quantum and classical Markov models [12].

4 Relation with model proposed by Monras et al.

In [15] hidden quantum Markov model is defined, by Monras et al., as a tuple consisting of: a d-level quantum system with an initial state ρ_0, alphabet $\mathcal{V} = \{V_i\}$, a set of quantum operations (CP-TNI maps) $\{\mathcal{K}^{V_i}\}$ such that $\sum_i \mathcal{K}^{V_i}$ is a quantum channel (CP-TP map). The system evolves in discrete time steps and subsequently generates symbols $O = (o_1, o_2, \ldots, o_T)$ from alphabet $\mathcal{V}$ with probability $P(o_t) = \text{tr}\, \mathcal{K}^{V_i=o_t}(\rho_t)$ in every time step. After generation of the symbol o_t, the sub-normalized quantum state is updated to $\rho_t = \mathcal{K}^{V_i=o_t}(\rho_{t-1})$. Moreover, $\mathcal{K}^{V_i}$ can be represented by Kraus operators $\{K_j^{V_i}\}$. It means that $\mathcal{K}^{V_i}(\rho) = \sum_j K_j^{V_i} \rho (K_j^{V_i})^\dagger$ and $\rho_t = \sum_j K_j^{V_i=o_t} \rho_{t-1} (K_j^{V_i=o_t})^\dagger$, where $\sum_j (K_j^{V_i})^\dagger (K_j^{V_i}) \leq \mathbb{1}$. Here we omit the normalization factor; therefore, with every sequence O a sub-normalized quantum state is associated.

In the case of Monras et al. model, the number of internal states is equal to dimension of quantum system. In our case, the states are divided into two distinct classes: "internal" quantum states and "external" classical states. Our model can be reduced to the model presented by Monras et al. by performing the following transformation. First, we need to extend the alphabet $\mathcal{V}$ with the symbol \$. Second, we concatenate every sequence O with the symbol \$. Third, we associate symbol \$ with operation of partial trace over the classical system: $\mathcal{K}^{\$}(\rho) = \text{tr}_{\mathcal{H}_2} \rho$. Fourth, we express (sub-)vector states α as block diagonal (sub-)normalized quantum states and sub-TOMs $\mathcal{P}$ as quantum operations.

According to the above, we can notice that $\mathcal{K}^{V_i}$ corresponds to $\mathcal{P}^{V_i}$, whose elements $\mathcal{P}_{k,l}^{V_i}$ are represented by Kraus operators $\{E_{k,l,j}^{V_i}\}$, hence $\mathcal{P}_{k,l}^{V_i}(\rho) = \sum_j E_{k,l,j}^{V_i} \rho (E_{k,l,j}^{V_i})^\dagger$. Let us construct the set of operators $\{\hat{E}_{k,l,j}^{V_i}\}$ in the form $\hat{E}_{k,l,j}^{V_i} = E_{k,l,j}^{V_i} \otimes |k\rangle\langle l|$, then similarly as in [12], it can be proved that

$$\sum_{j,k,l} (\hat{E}_{k,l,j}^{V_i})^\dagger \hat{E}_{k,l,j}^{V_i} \leq \mathbb{1}.$$

Now, consider vector state $\alpha_T = \mathcal{P}^{o_T} \ldots \mathcal{P}^{o_2} \mathcal{P}^{o_1}(\pi) = [\alpha_1, \alpha_2, \ldots, \alpha_N]^T$ with associated a block diagonal quantum state $\rho_\alpha = \sum_i^N \alpha_i \otimes |i\rangle\langle i| \in \Omega(\mathcal{H}_1 \otimes \mathcal{H}_2)$, then

$$\rho_{O\$} = \text{tr}_{\mathcal{H}_2} \sum_{j,k,l} \hat{E}_{k,l,j}^{V_i=o_T} \cdots \hat{E}_{k,l,j}^{V_i=o_2} \hat{E}_{k,l,j}^{V_i=o_1} \rho_\alpha \left(\hat{E}_{k,l,j}^{V_i=o_1}\right)^\dagger \left(\hat{E}_{k,l,j}^{V_i=o_2}\right)^\dagger \cdots \left(\hat{E}_{k,l,j}^{V_i=o_T}\right)^\dagger.$$

(20)

Thus, our model can be expressed in the language proposed by Monras et al. However, formalism proposed in this paper has three notable advantages. First, it presents a hybrid quantum-classical model similar to the one presented in [14] therefore has similar field of applications. Our model intuitively generalizes both classical and quantum models. Second, this model allows us to propose a generalized version of Viterbi algorithm. Third, the use of TOM and vector states formalism reduces the amount of memory required to numerically simulate hybrid quantum-classical Markov models.

5 Examples of application

5.1 Example 1

Let us consider alphabet $\mathcal{V} = \{a, b, c\}$. We define a set of sequences $\mathcal{O} \subset \mathcal{V}^T$ of length T and having $O_i = a$ for odd i, and $O_i \in \{b, c\}$ for even i, i.e., $aba, abaca, acacabaca$.

Let $T = 3$. Our objective is to build a model able to differentiate sequences in O from all other sequences. In classical case, our model could be given by a HMM parametrized by $\lambda_1^c = (\mathcal{S}, \mathcal{V}, \Pi, \pi)$, where

$$
\mathcal{S} = \{s_1, s_2\}, \quad \pi = \begin{bmatrix} 0 \\ 1 \end{bmatrix},
$$
$$
\Pi = \left\{ \Pi^a = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \Pi^b = \begin{bmatrix} 0 & 0 \\ \frac{1}{2} & 0 \end{bmatrix}, \Pi^c = \begin{bmatrix} 0 & 0 \\ \frac{1}{2} & 0 \end{bmatrix} \right\}.
\tag{21}
$$

It is obvious that $p(aba|\lambda_1^c) = p(aca|\lambda_1^c) = \frac{1}{2}$, whereas for other possible sequences we get $\sum_{O \in \mathcal{V}^T \setminus \mathcal{O}} p(O|\lambda_1^c) = 0$.

If we are interested in further differentiating aba from aca, we could either construct two HMMs, one for each sequence, i.e., for aba parametrized by $\lambda_2^c = (\mathcal{S}, \mathcal{V}, \Pi, \pi)$, where

$$
\Pi = \left\{ \Pi^a = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \Pi^b = \begin{bmatrix} 0 & 0 \\ \frac{1}{2} & 0 \end{bmatrix} \right\}
\tag{22}
$$

and similarly for aca, or by building a three-state HMM $\lambda_3^c = (\mathcal{S}, \mathcal{V}, \Pi, \pi)$

$$
\mathcal{S} = \{s_1, s_2, s_3\}, \quad \pi = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},
$$
$$
\Pi = \left\{ \Pi^a = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \Pi^b = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \Pi^c = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix} \right\}
\tag{23}
$$

and recognize the sequences—aba from aca—based on the output of Vitterbi algorithm.

We can solve the problem of discrimination by using QHMM $\lambda_1^q = (\mathcal{S}, \mathcal{V}, \mathcal{P}, \pi)$, with the following parameters

$$
\mathcal{S} = \{s_1, s_2\}, \quad \pi = \begin{bmatrix} 0_2 \\ |0\rangle\langle 0| \end{bmatrix},
$$

$$
\Pi = \left\{ \mathcal{P}^a = \begin{bmatrix} 0_4 & \mathbb{1}_4 \\ 0_4 & 0_4 \end{bmatrix}, \quad \mathcal{P}^b = \begin{bmatrix} 0_4 & 0_4 \\ \frac{1}{2}\Phi_U & 0_4 \end{bmatrix}, \quad \mathcal{P}^c = \begin{bmatrix} 0_4 & 0_4 \\ \frac{1}{2}\mathbb{1}_4 & 0_4 \end{bmatrix} \right\},
$$

(24)

where $\Phi_U(\cdot) = U \cdot U^\dagger$ is unitary channel, such that $U = \begin{bmatrix} \cos\frac{\pi}{2} & -\sin\frac{\pi}{2} \\ \sin\frac{\pi}{2} & \cos\frac{\pi}{2} \end{bmatrix}$ and $0_n, \mathbb{1}_n$ are zero and identity operators over vector space of dimension n, respectively.

Moreover, let $\mu : \{b \mapsto |1\rangle\langle 1|, c \mapsto |0\rangle\langle 0|\}$ be a measurement.

Let us consider the application of quantum Forward algorithm on sequence aba. Initial vector state of the algorithm is $\alpha_0 = \begin{bmatrix} 0_2 \\ |0\rangle\langle 0| \end{bmatrix}$, and the final state is $\alpha_3 = \begin{bmatrix} \frac{1}{2}U|0\rangle\langle 0|U^\dagger \\ 0_2 \end{bmatrix}$. The associate sub-normalized quantum state is $\rho_{aba|\lambda_1^q} = \frac{1}{2}U|0\rangle\langle 0|U^\dagger$; therefore, the resulting sequence of probabilities is given by

$$
(\mathrm{tr}\, \rho_{aba|\lambda_1^q}\mu(b), \mathrm{tr}\, \rho_{aba|\lambda_1^q}\mu(c)) = \left(\frac{1}{2}, 0 \right).
$$

(25)

It is obvious that application of quantum Forward algorithm on sequence aca gives result $(0, \frac{1}{2})$ and $\sum_{O \in \mathcal{V}^T \setminus \mathcal{O}} \mathrm{tr}\, \rho_{O|\lambda_1^q} = 0$.

We have shown that it is possible to construct two-state QHMM that fulfills the same task as pair of two-state HMMs or three-state HMM.

5.2 Example 2

Let us consider language A consisting of the sequences $a^{k_1} b^{k_2} a^{k_3} \ldots$, where $k_1, k_2, k_3, \ldots$ are nonnegative odd integers and a, b are symbols from alphabet $\mathcal{V} = \{a, b\}$. In other words, language A contains these sentences in which odd length subsequences of letters a and b alternate.

Classically, sequences from this language can be generated by four-state HMM $\lambda^c = (\mathcal{S}, \mathcal{V}, \Pi, \pi)$ presented in Fig. 2a, where

$$
\mathcal{S} = \{s_1, s_2, s_3, s_4\}, \quad \pi = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},
$$

(26)

$$
\Pi = \left\{ \Pi^a = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \end{bmatrix}, \quad \Pi^b = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{bmatrix} \right\}.
$$

It is easy to check that for any sequence $a^{k_1} b^{k_2} a^{k_3} \cdots$ from the language A, probability $p(a^{k_1} b^{k_2} a^{k_3} \cdots | \lambda^c)$ is nonzero and equals $p(a^{k_1} b^{k_2} a^{k_3} \cdots | \lambda^c) = (\frac{1}{2})^{\frac{k_1+1}{2}} (\frac{1}{2})^{\frac{k_2+1}{2}}$ $(\frac{1}{2})^{\frac{k_3+1}{2}} \cdots$. Moreover, if any k_i is even, then $p(a^{k_1} b^{k_2} a^{k_3} \cdots | \lambda^c) = 0$.

Let us consider matrix of probabilities $p(a^{k_1} b^{k_2} a^{k_3} \cdots | \lambda^c)$ given as

$$
\tilde{H} = \begin{bmatrix}
1 & p(a|\lambda^c) & p(b|\lambda^c) & p(aa|\lambda^c) & p(ab|\lambda^c) & p(ba|\lambda^c) & \cdots \\
p(a|\lambda^c) & p(aa|\lambda^c) & p(ba|\lambda^c) & p(aaa|\lambda^c) & p(aba|\lambda^c) & p(baa|\lambda^c) & \cdots \\
p(b|\lambda^c) & p(ab|\lambda^c) & p(bb|\lambda^c) & p(aab|\lambda^c) & p(abb|\lambda^c) & p(bab|\lambda^c) & \cdots \\
p(aa|\lambda^c) & p(aaa|\lambda^c) & p(baa|\lambda^c) & p(aaaa|\lambda^c) & p(abaa|\lambda^c) & p(baaa|\lambda^c) & \cdots \\
p(ab|\lambda^c) & p(aab|\lambda^c) & p(bab|\lambda^c) & p(aaab|\lambda^c) & p(abab|\lambda^c) & p(baab|\lambda^c) & \cdots \\
p(ba|\lambda^c) & p(aba|\lambda^c) & p(bba|\lambda^c) & p(aaba|\lambda^c) & p(abba|\lambda^c) & p(baba|\lambda^c) & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & & \ddots
\end{bmatrix}.
$$

(27)

Notice that any upper-left corner of matrix $\tilde{H}$ is known as the Hankel matrix. Denote by $\tilde{H}_d$ a upper-left d-size sub-matrix of matrix $\tilde{H}$. Subsequently, let us notice that

$$
\text{rank}(\tilde{H}_{11}) = \text{rank} \begin{bmatrix}
1 & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{8} & \frac{1}{4} \\
\frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{8} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{8} & 0 \\
0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{8} & 0 & \frac{1}{16} & \frac{1}{8} \\
\frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{8} & 0 & 0 & \frac{1}{8} & 0 & \frac{1}{16} & 0 \\
\frac{1}{4} & 0 & 0 & \frac{1}{8} & \frac{1}{16} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{8} & 0 & 0 & 0 & 0 & 0 & \frac{1}{16} & 0 & \frac{1}{32} & \frac{1}{16} \\
0 & \frac{1}{4} & 0 & 0 & \frac{1}{8} & 0 & 0 & \frac{1}{8} & 0 & \frac{1}{16} & \frac{1}{8} \\
\frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{8} & \frac{1}{16} & 0 & 0 & \frac{1}{8} & 0 & \frac{1}{16} & 0 \\
0 & \frac{1}{8} & 0 & 0 & 0 & 0 & 0 & \frac{1}{16} & 0 & \frac{1}{32} & 0 \\
\frac{1}{8} & 0 & 0 & \frac{1}{16} & \frac{1}{32} & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{4} & 0 & 0 & \frac{1}{8} & \frac{1}{16} & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} = 4. \quad (28)
$$

Since $\text{rank}(\tilde{H}_{11}) = 4$, four-state HMM $\lambda^c = (\mathcal{S}, \mathcal{V}, \Pi, \pi)$ cannot be reduced to HMM with smaller number of states [17,18].

The application of the QHMM for the generation of sequences from A can reduce the number of the states to three. Let us consider QHMM $\lambda^q = (\mathcal{S}, \mathcal{V}, \Pi, \pi)$ presented in Fig. 2b, with
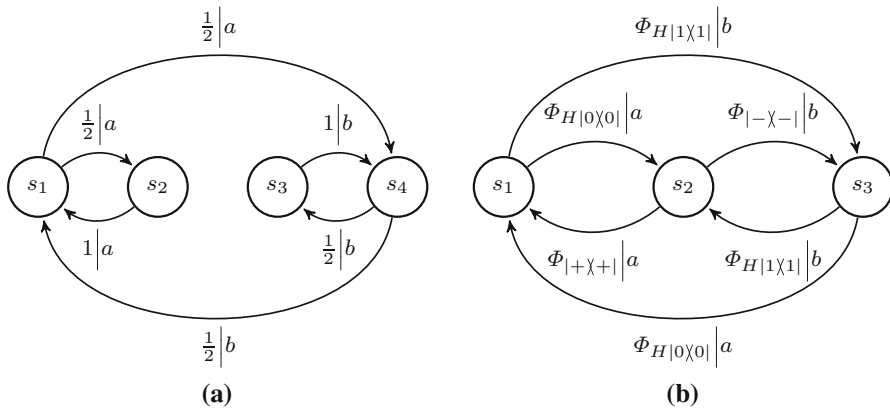
Fig. 2 Examples of HMM (**a**) and QHMM (**b**) generating with nonzero probabilities sequences $a^{k_1}b^{k_2}a^{k_3}\cdots$, where $k_1, k_2, k_3, \ldots$ are nonnegative odd integers

$$S = \{s_1, s_2, s_3\}, \quad \pi = \begin{bmatrix} 0_2 \\ 0_2 \\ |0\rangle\langle 0| \end{bmatrix},$$

$$\Pi = \left\{ \mathcal{P}^a = \begin{bmatrix} 0_4 & \Phi_{|+\rangle\langle+|} & \Phi_{H|0\rangle\langle0|} \\ \Phi_{H|0\rangle\langle0|} & 0_4 & 0_4 \\ 0_4 & 0_4 & 0_4 \end{bmatrix}, \quad \mathcal{P}^b = \begin{bmatrix} 0_4 & 0_4 & 0_4 \\ 0_4 & 0_4 & \Phi_{H|1\rangle\langle1|} \\ \Phi_{H|1\rangle\langle1|} & \Phi_{|-\rangle\langle-|} & 0_4 \end{bmatrix} \right\},$$

(29)

where $\Phi_X(\cdot) = X \cdot X^\dagger$ and $X \in \{|+\rangle\langle+|, |-\rangle\langle-|, H|0\rangle\langle0|, H|1\rangle\langle1|\}$.

Notice that, for any sequence $a^{k_1}b^{k_2}a^{k_3}\cdots$, where $k_1, k_2, k_3, \ldots$ are nonnegative odd integers, the final state is given as

$$\alpha_{k_1 k_2 k_3\ldots} = \begin{bmatrix} \left(\frac{1}{2}\right)^{\frac{k_1+1}{2}} \left(\frac{1}{2}\right)^{\frac{k_2+1}{2}} \left(\frac{1}{2}\right)^{\frac{k_3+1}{2}} \cdots \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\ 0_2 \\ 0_2 \end{bmatrix}$$

or

$$\alpha_{k_1 k_2 k_3\ldots} = \begin{bmatrix} 0_2 \\ 0_2 \\ \left(\frac{1}{2}\right)^{\frac{k_1+1}{2}} \left(\frac{1}{2}\right)^{\frac{k_2+1}{2}} \left(\frac{1}{2}\right)^{\frac{k_3+1}{2}} \cdots \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \end{bmatrix}.$$

Moreover, if any k_i is even, then $\alpha_{k_1 k_2 k_3\ldots} = \begin{bmatrix} 0_2 \\ 0_2 \\ 0_2 \end{bmatrix}$. Therefore, we have shown that it is possible to construct thee-state QHMM generating sequences from A with the same probabilities like its classical four-state counterpart. Those probabilities

tr $\rho_{a^{k_1}b^{k_2}a^{k_3}\dots|\lambda^q}$ are obtained from trivial measurements of sub-normalized quantum states $\rho_{a^{k_1}b^{k_2}a^{k_3}|\lambda^q\dots}$.

6 Conclusions

We have introduced a new model of quantum hidden Markov models based on the notions of transition operation matrices and vector states. We have shown that for a subclass of QHMMs and emission sequences the modified Viterbi algorithm can be used to calculate the most likely sequence of internal states that lead to a given emission sequence. Because of the fact that the structure of quantum hidden Markov models is more complicated than their classical counterparts, in general case the most likely sequence of states leading to a given emissions sequence has to be calculated using extensive search. We have also proposed a formulation of the Forward algorithm that is applicable for general QHMMs.

For given a sequence of symbols of length T, $O = (o_1, o_2, \dots, o_T)$, a sequence of states $N_T = (n_0, n_1, \dots, n_T)$ and a classical Mealy HMM with parameters λ, the joint probability distribution $P(N_T, O)$ can be factored into

$$P(N_T, O) = P(n_0) \prod_{t=1}^{T} P(n_t|o_t, n_{t-1}) P(o_t|n_{t-1}). \tag{30}$$

As in the case of classical Moore HMM [19], the above factorization can be considered as a simple dynamic Bayesian network. Hence, the concept of QHMM proposed in this manuscript gives basis to quantum generalization of dynamic Bayesian networks.

We believe that proposed model can find applications in modeling systems that posses both quantum and classical features.

References

1. Rabiner, L.: A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE **77**(2), 257–286 (1989)
2. Vanluyten, B., Willems, J.C., De Moor, B.: Equivalence of state representations for hidden Markov models. Syst. Control Lett. **57**(5), 410–419 (2008)
3. Gudder, S.: Quantum Markov chains. J. Math. Phys. **49**(7), 072105 (2008)
4. Petruccione, F., Sinayskiy, I.: Open quantum random walks and the open quantum Pascal triangle. In: International Conference on Quantum Information, page QMH2. Optical Society of America (2011)

5. Attal, S., Petruccione, F., Sinayskiy, I.: Open quantum walks on graphs. Phys. Lett. A **376**(18), 1545–1548 (2012)
6. Sinayskiy, I., Petruccione, F.: Efficiency of open quantum walk implementation of dissipative quantum computing algorithms. Quantum Inf. Process. **11**(5), 1301–1309 (2012)
7. Sinayskiy, I., Petruccione, F.: Properties of open quantum walks on $\mathbb{Z}$. Phys. Scr. **2012**(T151), 014077 (2012)
8. Attal, S., Petruccione, F., Sabot, C., Sinayskiy, I.: Open quantum random walks. J. Stat. Phys. **147**(4), 832–852 (2012)
9. Sinayskiy, I., Petruccione, F.: Open quantum walks: a short introduction. In: Journal of Physics: Conference Series, vol. 442, p. 012003. IOP Publishing (2013)
10. Konno, N., Yoo, H.J.: Limit theorems for open quantum random walks. J. Stat. Phys. **150**(2), 299–319 (2013)
11. Ampadu, C.: Averaging in SU(2) open quantum random walk. Chin. Phys. B **23**(3), 030302 (2014)
12. Pawela, Ł., Gawron, P., Miszczak, J.A., Sadowski, P.: Generalized open quantum walks on apollonian networks. PLoS ONE **10**(7), e0130967 (2015)
13. Sadowski, P., Pawela, Ł.: Central limit theorem for reducible and irreducible open quantum walks. Quantum Inf. Proc. **15**(7), 1–19 (2016)
14. Li, L., Feng, Y.: Quantum Markov chains: of hybrid systems, decidability of equivalence, and model checking linear-time properties. Inf. Comput. **244**, 229–244 (2015)
15. Monras, A., Beige, A., Wiesner, K.: Hidden quantum Markov models and non-adaptive read-out of many-body states. Appl. Math. Comput. Sci. **3**, 93 (2011)
16. Cappellini, V., Sommers, H.-J., Życzkowski, K.: Subnormalized states and trace-nonincreasing maps. J. Math. Phys. **48**(5), 052110 (2007)
17. Vidyasagar, M.: The complete realization problem for hidden Markov models: a survey and some new results. Math. Control Signals Syst. **23**(1–3), 1–65 (2011)
18. Huang, Q., Ge, R., Kakade, S., Dahleh, M.: Minimal realization problem for hidden markov models. In: Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on, pp. 4–11. IEEE (2014)
19. Ghahramani, Z.: An introduction to hidden Markov models and Bayesian networks. Int. J. Pattern Recognit. Artif. Intell. **15**(01), 9–42 (2001)