Tensor Rings for Learning Circular Hidden Markov Models

Mohammad Ali Javidian, Vaneet Aggarwal, and Zubin Jacob
School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN, 47907
{mjavidia, vaneet, zjacob}@purdue.edu

Abstract

In this paper, we propose circular Hidden Quantum Markov Models (c-HQMMs), which can be applied for modeling temporal data in quantum datasets (with classical datasets as a special case). We show that c-HQMMs are equivalent to a constrained tensor network (more precisely, circular Local Purified State with positive-semidefinite decomposition) model. This equivalence enables us to provide an efficient learning model for c-HQMMs. The proposed learning approach is evaluated on six real datasets and demonstrates the advantage of c-HQMMs on multiple datasets as compared to HQMMs, circular HMMs, and HMMs.

1 Introduction

Hidden Markov Models (HMMs) are commonly used for modeling temporal data, usually, in cases where the underlying probability distribution is unknown, but certain output observations are known [Rabiner and Juang, 1986, Zucchini and MacDonald, 2009]. Hidden Quantum Markov Models (HQMMs) [Monras et al., 2010, Clark et al., 2015] can be thought of a reformulation of HMMs in the language of quantum systems (see section 2 for the formal definition). It has been shown that quantum formalism allows for more efficient description of a given stochastic process as compared to the classical case [Monras et al., 2010, Clark et al., 2015, Adhikary et al., 2021]. In this paper, we propose circular HQMMs, and validate them to be more efficient than HQMMs.

We note that circular HMMs (c-HMMs) have been proposed to model HMMs, where the initial and terminal hidden states are connected through the state transition probability [Arica and Vural, 2000]. c-HMMs have found application in speech recognition [Zheng and Yuan, 1988, Shahin, 2006], biology and meteorology [Holzmann et al., 2006], shape recognition [Arica and Vural, 2000, Cai et al., 2007], biomedical engineering [Coast et al., 1990], among others. Given the improved performance of c-HMMs as compared to HMMs, it remains open if such an extension can be done for HQMMs, which is the focus of this paper.

Even though multiple algorithms for learning HQMMs have been studied, direct learning of the model parameters is inefficient and ends up to poor local-optimal points [Adhikary et al., 2020]. In order to deal with this challenge, a tensor network based approach is used to learn HQMMs [Adhikary et al., 2021], based on a result that HQMM is equivalent to uniform locally purified states (LPS) tensor network. The model in [Adhikary et al., 2021] deals with infinite horizon HQMM, which involves uniform Kraus operators and thus uniform LPS. In this paper, we model a finite sequence of random variables, which allows us to have different Kraus operators at each time instant. We further extend the finite-horizon HQMMs to circular HQMMs (c-HQMMs).

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

In order to train the parameters of the c-HQMM, we show the equivalence of c-HQMM with a restricted class of tensor networks. In order to do that, we first define a class of tensor networks, called circular LPS (c-LPS). Then, we show that c-HQMM is equivalent to c-LPS where certain matrices formed from the decomposition are positive semi-definite (p.s.d.). Finally, we propose an algorithm to train c-LPS with p.s.d. restrictions, thus providing an efficient algorithm for learning c-HQMMs.

The results in this paper show equivalence of finite-horizon HQMMs and c-HQMMs to the corresponding tensor networks. Further, we show that c-HMM is equivalent to a class of tensor networks (circular Matrix Product State (c-MPS)) with non-negative real entries. This allows for an alternate approach of learning c-HMMs, and may be of independent interest. The key contributions of this work are summarized as follows:

- We propose c-HQMM for modeling finite-horizon temporal data.
- We show that c-HQMMs are equivalent to c-LPS tensor networks with positive semi-definite matrix structure in the decomposition. Further, equivalent tensor structures for finite horizon HQMMs and c-HMMs are also provided.
- Learning algorithm for c-HQMM is provided using the tensor network equivalence.

In order to validate the proposed framework of c-HQMM and the proposed learning algorithm, we compare with standard HMMs (equivalent to MPS with non-negative real entries in decomposition), c-HMMs, and HQMMs. Evaluation on realistic datasets demonstrate the improved performance of c-HQMMs for modeling temporal data.

2 Related Work and Background

In this section, we briefly review the key related literature on hidden Markov models and tensor networks, with relevant definitions.

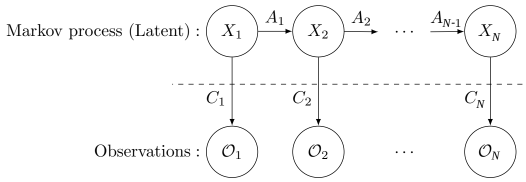
Hidden Markov Models (HMMs) [Rabiner and Juang, 1986, Zucchini and MacDonald, 2009] are a class of probabilistic graphical models that have found greatest use in problems that enjoy an inherent temporality. These problems consist of a process that unfolds in time, i.e., we have states at time t that are influenced directly by a state at $t-1$. HMMs have found application in such problems, for instance speech recognition [Juang and Rabiner, 1991], gesture recognition [Wilson and Bobick, 1999], face recognition [Nefian and Hayes, 1998], finance [Mamon and Elliott, 2007], computational biology [Siepel and Haussler, 2004, Koski, 2001, Krogh et al., 1994], among others. A finite-horizon hidden Markov model or HMM, as shown in Figure (1a), consists of a discrete-time, discrete-state Markov chain, with hidden states $X_t \in \{1, \cdots, d\}, t \in \{1, \cdots, N\}$[1], plus an observation model $p(o_t|x_t)$. The corresponding joint distribution has the form:

$$p(X_{1:N}, \mathcal{O}_{1:N}) = p(X_{1:N})p(\mathcal{O}_{1:N}|X_{1:N}) = \left[p(x_1)\Pi_{t=2}^{N} p(x_t|x_{t-1})\right]\left[\Pi_{t=1}^{N} p(o_t|x_t)\right] \quad (1)$$

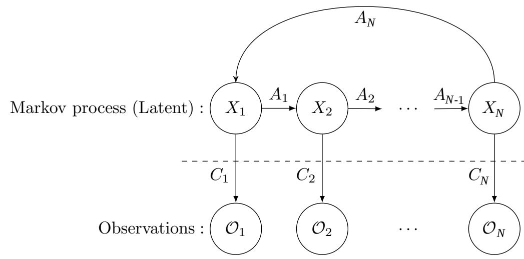
As shown in Figure (1a), evolution of hidden states is governed by column-stochastic matrices A_i's, called transition matrices, and emission matrices are column-stochastic matrices C_i's that determine the observation probabilities. Despite the fact that HMMs are a powerful and versatile tool for statistical modeling of complex time-series data and stochastic dynamic systems, many real-world problems include circular data (e.g., measurements in the form of angles or other periodic values), such as biology, climatology, oceanography, geophysics, and astronomy. In these problems, the periodic nature of the boundary requires a Hidden Markov topology which is both *temporal* (has a sequential order) and *ergodic* to allow the revisits of a state as the boundary returns to the starting point and repeats itself [Arica and Vural, 2000]. c-HMMs were proposed to address these problems. Formally, a c-HMM is a modification of HMM model, where the initial and terminal hidden states are connected through the state transition probability A_N, as shown in Figure (1b). The corresponding joint distribution has the form:

$$p(X_{1:N}, \mathcal{O}_{1:N}) = p(X_{1:N})p(\mathcal{O}_{1:N}|X_{1:N}) = \left[p(x_1|x_N)\Pi_{t=2}^{N} p(x_t|x_{t-1})\right]\left[\Pi_{t=1}^{N} p(o_t|x_t)\right] \quad (2)$$

[1] Finite horizon implies that N is finite, and thus the distributions can depend on the time-index. This paper focuses on finite horizon probability distributions.



(a) Representation of an HMM with observation variables $\mathcal{O}_i$, latent variables X_i, transition matrices A_i, and emission matrices C_i.



(b) Representation of a c-HMM with observation variables $\mathcal{O}_i$, latent variables X_i, transition matrices A_i, and emission matrices C_i.

Figure 1: Hidden Markov Model and Cyclic Hidden Markov Model.

Hidden Quantum Markov Model (HQMM) was introduced in [Monras et al., 2010] to model evolution from one quantum state to another, while generating classical output symbols. To produce an output symbol, a measurement or Kraus operation [Kraus et al., 1983] is performed on the internal state of the machine. To implement a Kraus operation, one can use an auxiliary quantum system, called *ancilla*. In every time step, the internal state of the HQMM interacts with its ancilla, which is then read out by a projective measurement. After every measurement, the ancilla is reset into its initial state, while the internal state of the HQMM remains hidden [Clark et al., 2015]. As in the classical case, an HQMM can be composed by the repeated application of the quantum sum rule (plays the role of transition matrices in HMMs) and quantum Bayes rule (plays the role of emission matrices in HMMs) [Adhikary et al., 2019] encoded using the sets of Kraus operators $\{K_{t,w}\}$ and $\{K_{t,x}\}$, respectively, for $t \in \{1, \cdots, N\}$:

$$\rho'_t = \sum_w K_{t,w}\rho_{t-1}K^\dagger_{t,w} \qquad \text{(quantum sum rule)}$$

$$\rho_t = \frac{K_{t,x}\rho'_t K^\dagger_x}{\mathbf{tr}(\sum_x K_{t,x}\rho'_t K^\dagger_{t,x})} \qquad \text{(quantum Bayes rule)}$$

We can condense these two expressions into a single term for a given observation x by setting $K_{t,x,w} = K_{t,x}K_{t,w}$, for $t = 1, \cdots, N$:

$$\rho_t = \frac{\sum_w K_{t,x,w}\rho_{t-1}K^\dagger_{t,x,w}}{\mathbf{tr}(\sum_w K_{t,x,w}\rho_{t-1}K^\dagger_{t,x,w})} \qquad \text{(state update rule)}$$

We now formally define HQMMs using the Kraus operator-sum representation (the definition is modified from [Adhikary et al., 2020] to account for finite N).

Definition 1 (HQMM)**.** *An N-horizon d-dimensional Hidden Quantum Markov Model with a set of discrete observations $\mathcal{O}$ is a tuple $(\mathbb{C}^{d\times d}, \{K_{i,x,w_x}\})$, where the initial state $\rho_0 \in \mathbb{C}^{d\times d}$ and the Kraus operators $\{K_{i,x,w_x}\} \in \mathbb{C}^{d\times d}$, for all $x \in \mathcal{O}, i \in \{1, \cdots, N\}, w_x \in \mathbb{N}$, satisfy the following constraints:*

- *ρ_0 is a density matrix of arbitrary rank, and*
- *the full set of Kraus operators across all observables provide a quantum operation,[2] i.e., $\sum_{x,w_x} K^\dagger_{i,x,w_x}K_{i,x,w_x} = I$ for all $i \in \{1, \cdots, N\}$.*

The joint probability of a given sequence is given by:

$$p(x_1, \cdots, x_N) = \vec{I}^T \left(\sum_{w_{x_N}} K^\dagger_{N,x_N,w_{x_N}} \otimes K_{N,x_N,w_{x_N}}\right) \cdots \left(\sum_{w_{x_1}} K^\dagger_{1,x_1,w_{x_1}} \otimes K_{1,x_1,w_{x_1}}\right) \vec{\rho_0} \quad (3)$$

where $\vec{I}^T, \vec{\rho_0}$ indicate vectorization (column-first convention) of identity matrix and ρ_0, respectively. This model is illustrated in Figure (3a).

[2] If $\sum_w \mathcal{K}^\dagger_w \mathcal{K}_w = I$, then $\mathcal{K} = \sum_w \mathcal{K}_w \rho \mathcal{K}^\dagger_w$ is called a quantum channel. However, if $\sum_w \mathcal{K}^\dagger_w \mathcal{K}_w < I$, then $\mathcal{K}$ is called a stochastic quantum operation.

This representation was used in [Srinivasan et al., 2018] to show that any d dimensional HMM can be simulated as an equivalent d^2 dimensional HQMM [Srinivasan et al., 2018, Algorithm 1].

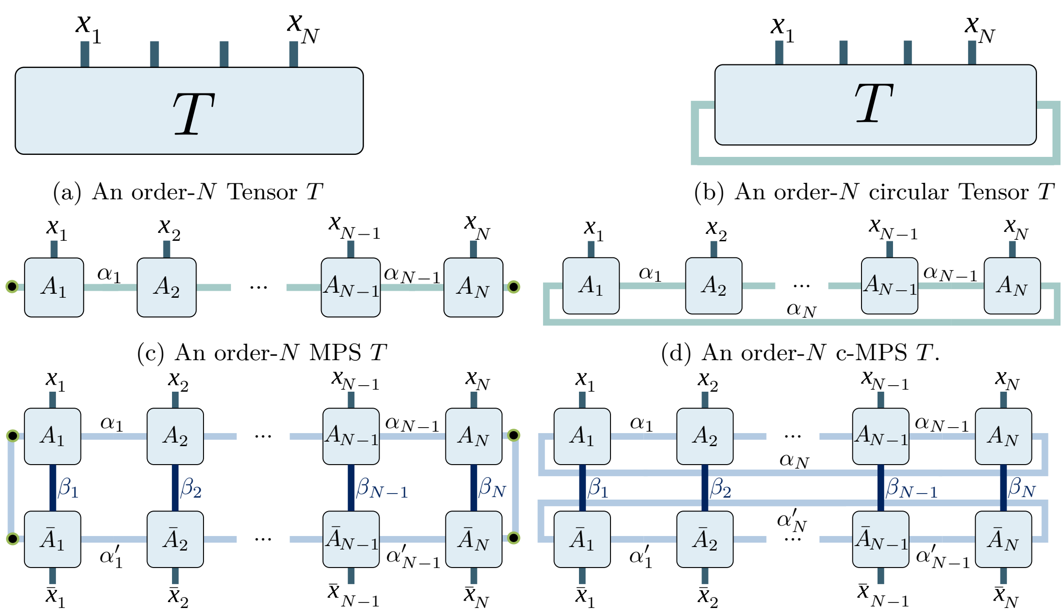
HQMMs enable us to generate more complex random output sequences than HMMs, even when using the same number of internal states [Clark et al., 2015, Srinivasan et al., 2018]. In other words, HQMMs are strictly more expressive than classical HMMs [Adhikary et al., 2021].

Tensor Network is a set of tensors (high-dimensional arrays), where some or all of its indices are contracted according to some pattern [Oseledets, 2011, Orús, 2014]. They have been used in the study of many-body quantum systems [Orús, 2019, Montangero et al., 2018]. Further, they have been adopted for supervised learning in large-scale machine learning [Stoudenmire and Schwab, 2016, Wang et al., 2017, 2018]. Some of the classes of tensor networks we use in this work include variants of Matrix Product States (MPSs) and Locally Purified States (LPSs).

One class of tensor networks is Matrix Product State (MPS), where an order-N tensor $T_{d\times\cdots\times d}$, with rank r has entry $(x_1,\cdots,x_N)$ $(x_i \in \{1,\cdots,d)$ given as

$$T_{x_1,\cdots,x_N} = \sum_{\{\alpha_i\}_{i=0}^{N}=1}^{r} A_0^{\alpha_0} A_{1,x_1}^{\alpha_0,\alpha_1} A_{2,x_2}^{\alpha_1,\alpha_2} \cdots A_{N-1,x_{N-1}}^{\alpha_{N-2},\alpha_{N-1}} A_{N,x_N}^{\alpha_{N-1},\alpha_N} A_N^{\alpha_N} \tag{4}$$

where $A_k, k \in \{0, N+1\}$, is a vector of dimension r, where element (α_k) is denoted as $A_k^{\alpha_k}$. Further, $A_k, k \in \{1,\cdots,N\}$ is an order-3 tensor of dimension $d \times r \times r$, where element (x, α_L, α_R) is denoted as $A_{k,x}^{\alpha_L,\alpha_R}$.



(a) An order-N Tensor T (b) An order-N circular Tensor T

(c) An order-N MPS T (d) An order-N c-MPS T.

(e) An order-N LPS T (f) An order-N c-LPS T

Figure 2: Tensor diagrams corresponding to different tensor networks. Black end dots indicate boundary vectors.

Another class of tensor network that is studied in this paper is the Locally Purified State (LPS). An order-N MPS with d-dimensional indices, admits an LPS representation of puri-rank r and purification dimension μ when the entries of T can be written as:

$$\begin{aligned} T_{x_1,\cdots,x_N} = \sum_{\{\alpha_i,\alpha_i'\}_{i=0}^{N}=1}^{r} \sum_{\{\beta_i\}_{i=1}^{N}=1}^{\mu} & A_0^{\alpha_0,\alpha_0'} A_{1,x_1}^{\beta_1,\alpha_0,\alpha_1} \overline{A_{1,x_1}^{\beta_1,\alpha_0',\alpha_1'}} A_{2,x_2}^{\beta_2,\alpha_1,\alpha_2} \overline{A_{2,x_2}^{\beta_2,\alpha_1',\alpha_2'}} \\ & \cdots A_{N-1,x_{N-1}}^{\beta_{N-1},\alpha_{N-2},\alpha_{N-1}} \overline{A_{N-1,x_{N-1}}^{\beta_{N-1},\alpha_{N-2},\alpha_{N-1}}} A_{N,x_N}^{\beta_N,\alpha_{N-1},\alpha_N} \overline{A_{N,x_N}^{\beta_N,\alpha_{N-1}',\alpha_N'}} A_{N+1}^{\alpha_N,\alpha_N'} \end{aligned} \tag{5}$$

where $A_k, k \in \{0, N+1\}$, is an $r \times r$ matrix, where the element (α_k, α'_k) is denoted as $A_k^{\alpha_k, \alpha'_k}$. Further, $A_k, k \in \{1, \cdots, N\}$, is an order-4 tensor of dimension $d \times \mu \times r \times r$, where the element $(x, \beta, \alpha_L, \alpha_R)$ is denoted as $A_{k,x}^{\beta, \alpha_L, \alpha_R}$, and elements belong to $\mathbb{R}$ or $\mathbb{C}$, as defined based on the context.

The tensor networks can be represented using tensor diagrams, where tensors are represented by boxes, and indices in the tensors are represented by lines emerging from the boxes. The lines connecting tensors between each other correspond to contracted indices, whereas lines that do not go from one tensor to another correspond to open indices Orús [2014]. The tensor diagrams corresponding to tensor networks MPS and LPS can be seen in Figure (2c) and (2e), respectively.

Relation between HMMs and Tensor Networks: As shown recently in [Glasser et al., 2019, Adhikary et al., 2021], tensor networks have direct correspondence with HMMs. In particular, non-negative matrix product states (MPS) are HMMs [Glasser et al., 2019], and uniform locally purified states are QHMMs [Adhikary et al., 2021]. Note that equivalence assumes that the tensor networks are normalized as the probabilities, while we will not explicitly normalize the tensor networks in the proofs, while will be accounted in the learning.

Learning of HQMM: Two state-of-the-art algorithms for learning HQMMs were proposed in [Srinivasan et al., 2018, Adhikary et al., 2020]. Both algorithms use an iterative maximum-likelihood algorithm to learn Kraus operators to model sequential data using an HQMM. The proposed algorithm in [Srinivasan et al., 2018] is slow and there is no theoretical gaurantee that the algorithm steps towards the optimum at every iteration [Adhikary et al., 2020]. The proposed algorithm in [Adhikary et al., 2020], however, uses a gradient-based algorithm. Although, the proposed algorithm in [Adhikary et al., 2020] is able to learn an HQMM that outperforms the corresponding HMM, this comes at the cost of a rapid scaling in the number of parameters. In order to deal with this issue, equivalence between HQMMs and Tensor Networks have been considered to achieve efficient learning [Glasser et al., 2019, Adhikary et al., 2021].

3 Proposed c-HQMM

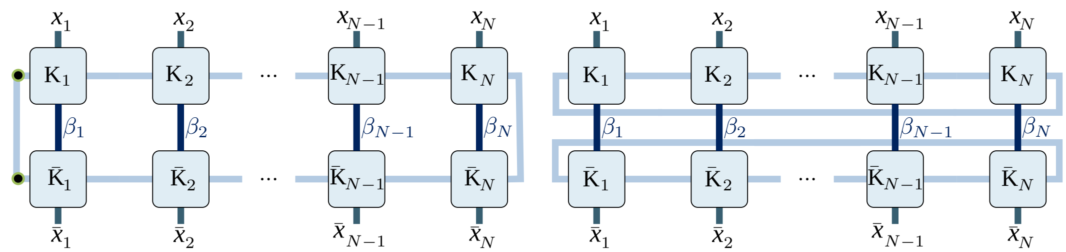
In this section, we propose Circular HQMM (c-HQMM) for modeling temporal data.

Definition 2 (c-HQMM)**.** *An N-horizon d-dimensional circular Hidden Quantum Markov Model (c-HQMM) with a set of discrete observations $\mathcal{O}$ is a tuple $(\mathbb{C}^{d\times d}, \{K_{i,x,w_x}\}, \boldsymbol{tr}(\cdot))$, where the Kraus operators are given as $\{K_{i,x,w_x}\} \in \mathbb{C}^{d\times d}$, for all $x \in \mathcal{O}, i \in \{1, \cdots, N\}, w_x \in \mathbb{N}$. The full set of Kraus operators across all observables provide a quantum operation, i.e., $\sum_{i,x,w_x} K_{i,x,w_x}^\dagger K_{x,w_x} = I$, for all $i \in \{1, \cdots, N\}$. The joint probability of a given sequence is given by:*

$$p(x_1, \cdots, x_N) = \boldsymbol{tr}\Bigg(\Bigg(\sum_{w_{x_N}} \bar{K}_{N,x_N,w_{x_N}} \otimes K_{N,x_N,w_{x_N}}\Bigg) \cdots \Bigg(\sum_{w_{x_1}} \bar{K}_{1,x_1,w_{x_1}} \otimes K_{1,x_1,w_{x_1}}\Bigg)\Bigg) \tag{6}$$

where $\boldsymbol{tr}(\cdot)$ indicates the trace of the resulting matrix. This model is illustrated in Figure (3b)*.*

We note that this representation along with the algorithm proposed in [Srinivasan et al., 2018, Algorithm 1] can be used to show that any d dimensional circular HMM can be simulated as an equivalent d^2 dimensional c-HQMM.



(a) A Hidden Quantum Markov Model. The rightmost connecting line at the boundary represents the application of the identity. Black end dots indicate boundary vectors.

(b) A Circular Hidden Quantum Markov Model.

Figure 3: Hidden Quantum Markov Model and circular Hidden Quantum Markov Model: with observation variables x_i, where K_i are $K_{i,x_i,w_{x_i}}$ and $\beta_i = |w_{x_i}|$ is determined by the Kraus-rank.

4 Proposed c-LPS Model

In this section, we propose circular LPS (c-LPS) which is an extension of LPS. For this purpose, we first briefly review circular MPS (c-MPS) model. Circular MPS (c-MPS) is an extension of MPS, where an order-N c-MPS T, with d-dimensional indices and rank r has the entries given as:

$$T_{x_1,\cdots,x_N} = \sum_{\{\alpha_i\}_{i=1}^{N}=1}^{r} A_{1,x_1}^{\alpha_N,\alpha_1} A_{2,x_2}^{\alpha_1,\alpha_2} \cdots A_{N,x_N}^{\alpha_{N-1},\alpha_N} \tag{7}$$

where $A_k, k \in \{1,\cdots,N\}$, is an order-3 tensors of dimension $d \times r \times r$, as shown in Figure (2d), where element (x, α_L, α_R) is denoted as $A_{k,x}^{\alpha_L,\alpha_R}$. c-MPS are studied in the literature as tensor rings. Tensor rings [Zhao et al., 2016, Mickelin and Karaman, 2020] have found application in compression of convolutional neural networks [Wang et al., 2018], image and video compression [Zhao et al., 2019], data completion [Wang et al., 2017], among others.

Now, we introduce circular LPS (c-LPS) as a tensor ring extension of LPS, where an order-N with d-dimensional indices, puri-rank r, and purification dimension μ has entries given as:

$$T_{x_1,\cdots,x_N} = \sum_{\{\alpha_i,\alpha_i'\}_{i=1}^{N}=1}^{r} \sum_{\{\beta_i\}_{i=1}^{N}=1}^{\mu} A_{1,x_1}^{\beta_1,\alpha_N,\alpha_1} \overline{A_{1,x_1}^{\beta_1,\alpha_N',\alpha_1'}} \cdots A_{N,x_N}^{\beta_N,\alpha_{N-1},\alpha_N} \overline{A_{N,x_N}^{\beta_N,\alpha_{N-1}',\alpha_N'}} \tag{8}$$

where $A_k, k \in \{1,\cdots,N\}$, is an order-4 tensor of dimension $d \times \mu \times r \times r$, as shown in Figure (2f), where the element $(x, \beta, \alpha_L, \alpha_R)$ is denoted as $A_{k,x}^{\beta,\alpha_L,\alpha_R}$.[3]

5 c-HQMM are c-LPS with Positive Semi-Definite Matrix Structure

We first note that non-negative MPS (denoting by $\text{MPS}_{\mathbb{R}_{\geq 0}}$) are HMM [Glasser et al., 2019]. In other words, any HMM can be mapped to an MPS with non-negative elements, and any $\text{MPS}_{\mathbb{R}_{\geq 0}}$ can be mapped to a HMM. Similarly, local quantum circuits with ancillas are locally purified states [Glasser et al., 2019]. The authors of Adhikary et al. [2021] recently considered an infinite time model of HQMM, where Kraus operators do not depend on time, and showed the equivalence of these HQMMs with the uniform LPS with a positive definite matrix structure. However, our work considers a non-uniform finite-time structure by having Kraus operators depend on time. We note that the equivalent tensor structure corresponding to c-HMM and c-HQMM are open, which is studied in this section.

The next result describes the relation between c-HQMM and c-LPS:

[3] Since Born machines (BMs) are LPS of purification dimension $\mu = 1$, we can similarly define circular BMs.

Theorem 1. *c-HQMM model is equivalent to a c-LPS structure where the decomposition entries $A_{i,x}^{b,a_1,a_2}$ are complex, and the $r \times r$ matrices formed by $A_{i,x}^{b,\cdot,\cdot}$ for all i, x, b are positive semi-definite (p.s.d.).*

Proof. See [Javidian et al., 2021] for the proof. □

The proof structure can be directly specialized to HQMM, where we can obtain the following result:

Lemma 1. *HQMM model is equivalent to a LPS structure where the decomposition entries $A_{i,x}^{b,a_1,a_2}$ are complex, and the $r \times r$ matrices formed by $A_{i,x}^{b,\cdot,\cdot}$ for all i, x, b are positive semi-definite (p.s.d.) for $i \in \{1, \cdots, N\}$. Further, $r \times r$ matrix A_0 is p.s.d. and $r \times r$ matrix A_{N+1} is the identity matrix.*

Proof. See [Javidian et al., 2021] for the proof. □

Remark 1. *Non-terminating uniform LPS (uLPS) are equivalent to HQMM [Adhikary et al., 2021]. In uLPS boundary vectors originates from density matrices of arbitrary rank. As shown in [Adhikary et al., 2021], the evaluation functional of uLPS can be rescaled and transformed into a one that will converge to $\vec{I}^T$ when $N \to \infty$. To have the equivalency of finite-horizon HQMM and LPS, we need to restrict LPS models to the evaluation functional $\vec{I}^T$, as stated in Lemma 1.*

Further, we note that the prior works do not relate c-HMM to tensor networks, to the best of our knowledge. In the following result, we relate the c-HMM to c-MPS.

Lemma 2. *c-HMM model is equivalent to a c-MPS model, where the entries of each decomposition are real and non-negative.*

Proof. See [Javidian et al., 2021] for the proof. □

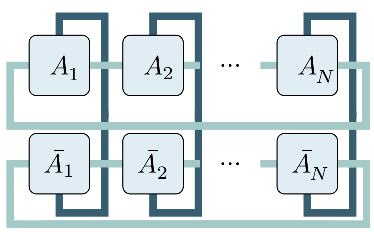


Figure 4: Contraction of tensor ring to compute: $Z_T = \sum_{X_1,\cdots,X_N} T_{X_1,\cdots,X_N}$.

6 Learning Algorithm for Circular LPS Models

In this section we propose an algorithm for learning c-LPS Models as in Theorem 1 via a maximum likelihood estimation (MLE) approach. The proposed algorithm is a modification of the algorithm proposed in [Glasser et al., 2019] for learning LPS models, except that we take into account the positive semi-definite nature of the decomposition and the cyclic structure.

Problem 1 (MLE for Distribution Approximation)**.** *Assume that $\{\boldsymbol{x}_i = (x_1^i, \cdots, x_N^i)\}_{i=1}^n$ is a sample of size n from an experiment with N discrete random variables. To estimate this discrete multivariate distribution, we use c-LPS model as defined in section 4. So, we have:*

$$p(x_1, \cdots, x_N) \approxeq \sum_{\{\alpha_i,\alpha_i'\}_{i=1}^N}^{r} \sum_{\{\beta_i\}_{i=1}^N}^{\mu} A_{1,x_1}^{\beta_1,\alpha_N,\alpha_1} \overline{A_{1,x_1}^{\beta_1,\alpha_N',\alpha_1'}} \cdots A_{N,x_N}^{\beta_N,\alpha_{N-1},\alpha_N} \overline{A_{N,x_N}^{\beta_N,\alpha_{N-1}',\alpha_N'}},$$

where the tensor decomposition entries follow the structure in Theorem 1. Our objective here is to estimate tensor elements of the c-LPS, i.e., $w = A_{i,x_i}^{\beta_i,\alpha_{L_i},\alpha_{R_i}}$*, for* $i = 1, \cdots, N$*. For this purpose, we minimize the negative log-likelihood:*

$$L = -\sum_i \log \frac{T_{\boldsymbol{x}_i}}{Z_T} \tag{9}$$

where $T_{\boldsymbol{x}_i}$ *is given by the contraction of c-LPS, and* $Z_T = \sum_{\boldsymbol{x}_i} T_{\boldsymbol{x}_i}$ *is a normalization factor.*

To find the optimal solution, we calculate the derivative of the log-likelihood with respect to w as follows:

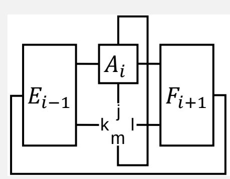
$$\partial_w L = -\sum_i \frac{\partial_w T_{\mathbf{x}_i}}{T_{\mathbf{x}_i}} - \frac{\partial_w Z_T}{Z_T} \tag{10}$$

We use a mini-batch gradient-descent algorithm to minimize the negative log-likelihood. At each step of the optimization, the sum is computed over a batch of training instances. The parameters in the tensor network are then updated by a small step in the inverse direction of the gradient. To satisfy the condition in Theorem 1, we project the $r \times r$ matrices formed by $A_{i,x}^{b,\cdot,\cdot}$ for all i, x, b to the positive semi-definite (p.s.d.) matrices using the standard Singular Value Decomposition (SVD) method [Dattorro, 2010]. Note that we use Wirtinger derivatives with respect to the conjugated tensor elements. Now, we explain how to compute $\partial_w Z_T$, Z_T, $\partial_w T_{\mathbf{x}_i}$, and $T_{\mathbf{x}_i}$ in equation (10). For a c-LPS of puri-rank r, the normalization Z_T can be computed by contracting the tensor network:

$$Z_T = \sum_{x_1,\cdots,x_N} T_{x_1,\cdots,x_N} \tag{11}$$

This contraction is performed, as shown in Figure (4), from left to right by contracting at each step the two vertical indices (corresponding to d_i and $\bar{d}_i$ with respect to the supports of X_i and $\bar{X}_i$) and then each of the two horizontal indices (with respect to α_is and α_i's, respectively). Finally, we trace out the indices corresponding to the rings. In this contraction, intermediate results from the contraction of the first i tensors are stored in E_i, and the same procedure is repeated from the right with intermediate results of the contraction of the last $N - i$ tensors stored in F_{i+1}. The derivatives of the normalization for each tensor are then computed as

$$\frac{\partial_w Z_T}{\partial_w \bar{A}_{i,m}^{j,k,l}} =$$



Computing $T_{\mathbf{x}_i}$ for a training sample and its derivative is done in the same way, except that the contracted index corresponding to an observed variable is now fixed to its observed value. We note that a similar approach to learn the model can be used for HQMM, HMM, and c-HMM structures.

7 Numerical Evaluations: Maximum Likelihood Estimation on Real Data

To evaluate the performance of the proposed algorithm for learning c-HQMMs, we used the same datasets as used in [Glasser et al., 2019] and learn HMM, c-HMM, HQMM, and c-HQMM using their respective tensor representations. HMM is equivalent to $\mathrm{MPS}_{\mathbb{R}_{>0}}$, and is a baseline for other structures. We note that equivalence of both c-HMMs and c-HQMMs to tensor networks first appear in this paper. Further, the equivalence of LPS and HQMM for finite N with non-uniform Kraus operators is also studied for the first time in this paper. We compare the performance of training HMM, c-HMM, HQMM, and c-HQMM using equivalent tensor representations on six different real data of categorical variables, where following parameters are used:

- Bond dimension/rank of the tensor networks: $r = 2, 3, 4, 5$ and 6.
- Learning rate was chosen using a grid search on powers of 10 going from 10^{-5} to 10^{5}.
- Batch size, i.e., the number of training samples per minibatch, was set to 20.
- Number of iterations was set to a maximum of 1000.
- The dimension of the purification index, i.e., μ for LPS and c-LPS was set to 2.

Each data point reported here is the lowest negative log-likelihood obtained from 10 trials with different initialization of tensors.

Results: The obtained results, summarized in Figure 5, show that (1) The tensor representations can be used to learn different HMMs. (2) We observe that despite the different algorithm choice, on almost all data sets, c-LPS and LPS lead to better modeling of the data distribution for the same rank as compared to MPS. (3) The results indicate that c-HQMM outperforms HQMM, HMM, and c-HMM. (4) In many cases, the performance difference between LPS and c-LPS for rank 5 and 6 is more significant than the cases with the rank 2 and 3. Further, the improvement depends on the dataset. We also note that we plot negative of log likelihoods, so the gap in the likelihoods is larger. The results suggest that in generic settings HQMM and c-HQMM should be preferred over both HMM and c-HMM models, respectively. Further, c-HQMM gives the best performance among the considered models.

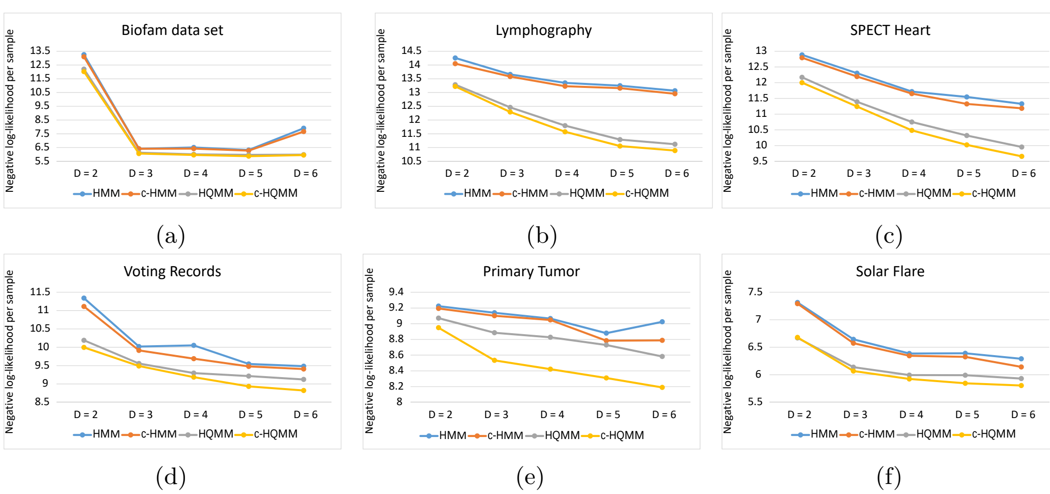


Figure 5: Maximum likelihood estimation with tensor rings MPS, c-MPS, LPS, and c-LPS for learning HMM, c-HMM, HQMM, and c-HQMM from the data on different data sets: a) biofam data set of family life states from the Swiss Household Panel biographical survey [Müller et al., 2007]; data sets from the UCI Machine Learning Repository [Dua and Graff, 2017]: b) Lymphography, c) SPECT Heart, d) Congressional Voting Records, e) Primary Tumor, and f) Solar Flare.

8 Conclusion

This paper proposes a new class of hidden Markov models, that we called circular Hidden Quantum Markov Models (c-HQMMs). c-HQMMs can be used to model temporal data in quantum datasets (with classical datasets as a special case). We proved that c-HQMMs are equivalent to circular LPS models with positive-semidefinite constraints on certain matrix structure in the LPS decomposition. Leveraging this result, we proposed an MLE based algorithm for learning c-HQMMs from data via c-LPS. We evaluated the proposed learning approach on six real datasets, demonstrating the advantage of c-HQMMs on multiple datasets as compared to HQMMs, circular HMMs, and HMMs.

Acknowledgments and Disclosure of Funding

This research was supported by the Defense Advanced Research Projects Agency (DARPA) Quantum Causality [Grant No. HR00112010008].

References

Sandesh Adhikary, Siddarth Srinivasan, and Byron Boots. Learning quantum graphical models using constrained gradient descent on the stiefel manifold. *arXiv preprint arXiv:1903.03730*, 2019.

Sandesh Adhikary, Siddarth Srinivasan, Geoff Gordon, and Byron Boots. Expressiveness and learning of hidden quantum markov models. In *International Conference on Artificial Intelligence and Statistics*, pages 4151–4161. PMLR, 2020.

Sandesh Adhikary, Siddarth Srinivasan, Jacob Miller, Guillaume Rabusseau, and Byron Boots. Quantum tensor networks, stochastic processes, and weighted automata. In *International Conference on Artificial Intelligence and Statistics*, pages 2080–2088. PMLR, 2021.

Nafiz Arica and FT Yarman Vural. A shape descriptor based on circular hidden markov model. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 1, pages 924–927. IEEE, 2000.

Jinhai Cai, Ming Ee, and Robert Smith. Image retrieval using circular hidden markov models with a garbage state. In *Proceedings of the Image and Vision Computing Conference New Zealand 2007*, pages 115–120. University of Waikato, 2007.

Lewis A Clark, Wei Huang, Thomas M Barlow, and Almut Beige. Hidden quantum markov models and open quantum systems with instantaneous feedback. In *ISCS 2014: Interdisciplinary Symposium on Complex Systems*, pages 143–151. Springer, 2015.

Douglas A Coast, Richard M Stern, Gerald G Cano, and Stanley A Briller. An approach to cardiac arrhythmia analysis using hidden markov models. *IEEE Transactions on biomedical Engineering*, 37(9):826–836, 1990.

Jon Dattorro. *Convex optimization & Euclidean distance geometry*. Lulu. com, 2010.

Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. URL `http://archive.ics.uci.edu/ml`.

Ivan Glasser, Ryan Sweke, Nicola Pancotti, Jens Eisert, and Ignacio Cirac. Expressive power of tensor-network factorizations for probabilistic modeling. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL `https://proceedings.neurips.cc/paper/2019/file/b86e8d03fe992d1b0e19656875ee557c-Paper.pdf`.

Hajo Holzmann, Axel Munk, Max Suster, and Walter Zucchini. Hidden markov models for circular and linear-circular time series. *Environmental and Ecological Statistics*, 13(3): 325–347, 2006.

Mohammad Ali Javidian, Vaneet Aggarwal, and Zubin Jacob. Learning circular hidden quantum markov models: A tensor network approach. *arXiv preprint arXiv:2111.01536*, 2021.

Biing Hwang Juang and Laurence R Rabiner. Hidden markov models for speech recognition. *Technometrics*, 33(3):251–272, 1991.

Timo Koski. *Hidden Markov models for bioinformatics*, volume 2. Springer Science & Business Media, 2001.

Karl Kraus, Arno Böhm, John D Dollard, and WH Wootters. States, effects, and operations: fundamental notions of quantum theory. lectures in mathematical physics at the university of texas at austin. *Lecture notes in physics*, 190, 1983.

Anders Krogh, Michael Brown, I Saira Mian, Kimmen Sjölander, and David Haussler. Hidden markov models in computational biology: Applications to protein modeling. *Journal of molecular biology*, 235(5):1501–1531, 1994.

Rogemar S Mamon and Robert James Elliott. *Hidden Markov models in finance*, volume 4. Springer, 2007.

Oscar Mickelin and Sertac Karaman. On algorithms for and computing with the tensor ring decomposition. *Numerical Linear Algebra with Applications*, 27(3):e2289, 2020.

Alex Monras, Almut Beige, and Karoline Wiesner. Hidden quantum markov models and non-adaptive read-out of many-body states. *arXiv: Quantum Physics*, 2010.

Simone Montangero, Montangero, and Evenson. *Introduction to Tensor Network Methods*. Springer, 2018.

Nicolas Séverin Müller, Matthias Studer, and Gilbert Ritschard. Classification de parcours de vie à l'aide de l'optimal matching. *XIVe Rencontre de la Société francophone de classification (SFC 2007)*, pages 157–160, 2007.

Ara V Nefian and Monson H Hayes. Hidden markov models for face recognition. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, volume 5, pages 2721–2724. IEEE, 1998.

Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117–158, 2014.

Román Orús. Tensor networks for complex quantum systems. *Nature Reviews Physics*, 1(9): 538–550, 2019.

Ivan V Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33 (5):2295–2317, 2011.

Lawrence Rabiner and Biinghwang Juang. An introduction to hidden markov models. *ieee assp magazine*, 3(1):4–16, 1986.

Ismail Shahin. Enhancing speaker identification performance under the shouted talking condition using second-order circular hidden markov models. *Speech Communication*, 48 (8):1047–1055, 2006.

Adam Siepel and David Haussler. Combining phylogenetic and hidden markov models in biosequence analysis. *Journal of Computational Biology*, 11(2-3):413–428, 2004.

Siddarth Srinivasan, Geoff Gordon, and Byron Boots. Learning hidden quantum markov models. In *International Conference on Artificial Intelligence and Statistics*, pages 1979–1987. PMLR, 2018.

E Miles Stoudenmire and David J Schwab. Supervised learning with quantum-inspired tensor networks. *arXiv preprint arXiv:1605.05775*, 2016.

Wenqi Wang, Vaneet Aggarwal, and Shuchin Aeron. Efficient low rank tensor ring completion. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5697–5705, 2017.

Wenqi Wang, Yifan Sun, Brian Eriksson, Wenlin Wang, and Vaneet Aggarwal. Wide compression: Tensor ring nets. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9329–9338, 2018.

Andrew D Wilson and Aaron F Bobick. Parametric hidden markov models for gesture recognition. *IEEE transactions on pattern analysis and machine intelligence*, 21(9): 884–900, 1999.

Qibin Zhao, Guoxu Zhou, Shengli Xie, Liqing Zhang, and Andrzej Cichocki. Tensor ring decomposition. *arXiv preprint arXiv:1606.05535*, 2016.

Qibin Zhao, Masashi Sugiyama, Longhao Yuan, and Andrzej Cichocki. Learning efficient tensor representations with ring-structured networks. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 8608–8612. IEEE, 2019.

Y-C Zheng and B-Z Yuan. Text-dependent speaker identification using circular hidden markov models. In *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*, pages 580–581. IEEE Computer Society, 1988.

Walter Zucchini and Iain L MacDonald. *Hidden Markov models for time series: an introduction using R*. Chapman and Hall/CRC, 2009.