

A probabilistic graphical model of quantum systems

Chen-Hsiang Yeang
*Institute of Statistical Science
Academia Sinica
Taipei, Taiwan
chyeang@stat.sinica.edu.tw*

Abstract—Quantum systems are promising candidates of future computing and information processing devices. In a large system, information about the quantum states and processes may be incomplete and scattered. To integrate the distributed information we propose a quantum version of probabilistic graphical models. Variables in the model (quantum states and measurement outcomes) are linked by several types of operators (unitary, measurement, and merge/split operators). We propose algorithms for three machine learning tasks in quantum probabilistic graphical models: a belief propagation algorithm for inference of unknown states, an iterative algorithm for simultaneous estimation of parameter values and hidden states, and an active learning algorithm to select measurement operators based on observed evidence. We validate these algorithms on simulated data and point out future extensions toward a more comprehensive theory of quantum probabilistic graphical models.

Keywords-quantum states; probabilistic graphical models; belief propagation; density matrix;

I. INTRODUCTION

Quantum computation and information processing have become active areas in computation and information theory. Unique features of quantum systems – such as superposition of states, measurement interference and entanglement – already provide valuable tools solving prime factorization [1], search [2], encryption [3], teleportation [4], and other problems. Various efforts of bridging machine learning and quantum mechanics have also been proposed. In the implementation of quantum computers or quantum optical devices, one often needs to estimate hidden quantum states or to characterize unknown devices. Estimation of quantum states and processes is tackled by a variety of methods such as quantum state tomography [5], quantum process tomography [6], maximum likelihood estimation methods [7], Bayesian methods [8], [9], and maximum entropy methods [10]. Despite their effectiveness, these methods discard the structure of a quantum system by lumping all the components into a black box. This may not be an adequate approach for a large system that is partially measured (only some quantum states are measured) and partially identified (only some quantum processes are known). In such setting integration of the distributed information is essential to characterize a large quantum system.

Integration of distributed information in a large system is well handled by probabilistic graphical models [11]. Distributed information in the model can be mapped to local structures of the graph such as nodes, edges or cliques. Efficient algorithms and heuristic methods have been invented to infer the states of the variables (e.g., [12], [13]), estimate unknown parameters (e.g., [14], [15]), or identify the structure of the model (e.g., [16]). Therefore, it is desirable to map the probabilistic graphical models and their efficient algorithms into the quantum systems.

Despite quantum computation and information processing are still at the incipient stage, the benefits of characterizing a large-scale quantum system with probabilistic graphical models can be anticipated. For instance, a practical quantum computer will consist of many quantum gates. Probabilistic graphical models can assist engineers to simulate the global and local operations of the circuit, reconstruct the states from partial measurements, design the test vectors to probe the system, and reverse-engineer an unknown circuit. In addition, in the future a large-scale, distributed network may convey a large amount of encrypted quantum information from many users. To reconstruct the faithful quantum information a router may need to collect, integrate and pass information from its neighbors. For both applications probabilistic graphical models will be a relevant and important component.

The goal of this paper is to set up the initial steps for this mapping. We introduce the fundamental concepts of a quantum probabilistic graphical model – including the basic ingredients, the underlying assumptions, the likelihood function formulation, and several examples of the model. Following the introduction we propose algorithms for three machine learning tasks in quantum probabilistic graphical models. For inference, we derive a quantum belief propagation algorithm from the sum-product algorithm of factor graphs. For parameter estimation, we develop an iterative algorithm combining quantum state estimation and gradient ascent. For active learning, we propose an incremental algorithm to select measurement operators according to an information theoretic criterion. Description of each algorithm is accompanied with empirical justifications on simulated data. Finally we state the limitations of the current versions of models and algorithms and discuss their future extensions.

II. A GRAPHICAL MODEL OF QUANTUM STATES

In this section we will first provide a background introduction about quantum information processing extracted from [17]. We then define probabilistic graphical models in a quantum system and elicit several special cases of quantum graphical models.

A. Overview of quantum information processing

A *quantum state* denotes an unobservable distribution which gives rise to various observable physical quantities. Mathematically it is a vector in a complex space. The Dirac notation designates $|\psi\rangle$ as a column vector, and the row vector $\langle\psi| \equiv |\psi\rangle^\dagger$ as the Hermitian adjoint (conjugate transpose) of $|\psi\rangle$. $|\langle\psi|\psi\rangle|^2 = 1$.

In classical computation, a *bit* denotes the binary state of a physical entity (e.g., the voltage level of a transistor). In quantum computation, a *quantum bit (qubit)* denotes a quantum state which gives rise to binary observable outcomes (e.g., the spin of an

electron). Unlike classical bits, a qubit consists of continuum of quantum states. Each quantum state is a linear combination of two basis vectors:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle. \quad (1)$$

where $|\alpha|^2 + |\beta|^2 = 1$. $|0\rangle$ and $|1\rangle$ are conventionally defined as column vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ but can be replaced by other orthonormal basis vectors.

$|\psi\rangle$ is not directly observable but can be probed through measurements. The probabilities of observing $|0\rangle$ and $|1\rangle$ are $|\langle 0|\psi\rangle|^2 = |\alpha|^2$ and $|\langle 1|\psi\rangle|^2 = |\beta|^2$.

A state in equation (1) is called a *pure state*. The underlying state is a fixed vector despite its measurement outcomes possess uncertainties. In contrast, a *mixed state* denotes a probabilistic mixture of pure states. Mathematically, it is convenient to express both pure and mixed states as *density matrices* [17]. Suppose $\{|\phi_i\rangle\}$ are an ensemble of pure states $|\phi_i\rangle$ and p_i their mixing probabilities respectively. The density matrix of the state is defined by

$$\rho = \sum_i p_i |\phi_i\rangle \langle \phi_i|. \quad (2)$$

Quantum states can be extended to multiple qubits. If there is no coupling (or *entanglement* in the terminology of quantum computation) of two quantum states ρ_1 and ρ_2 , then the joint state is their *tensor product*. Suppose A and B are $m \times n$ and $p \times q$ matrices respectively, then their tensor product is an $mp \times nq$ matrix:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}. \quad (3)$$

where a_{ij} 's are coefficients of A . Conversely, the density matrix of each uncoupled quantum state is a *partial trace* of the joint density matrix over each subspace. Suppose $C = A \otimes B$, then the partial traces of C over A and B are:

$$\begin{aligned} B_{ij} &= \text{tr}_A(C)_{ij} = \sum_k C_{ik,jk}. \\ A_{ij} &= \text{tr}_B(C)_{ij} = \sum_k C_{ki,kj}. \end{aligned} \quad (4)$$

A *quantum operation* $\mathcal{Q} : \rho \rightarrow \rho'$ transforms a density matrix ρ into ρ' . In this work we only consider unitary transformations U where $UU^\dagger = U^\dagger U = I$. The output density matrix has the following form:

$$\rho' = \mathcal{Q}(\rho) = U\rho U^\dagger. \quad (5)$$

Another important class of quantum operations are measurement operators that convert the quantum states into discrete outcomes with probability. In this work we consider a special case of measurement operators $\{\psi_m\}$ consisting of orthonormal basis vectors ψ_m indexed by m . The probability of observing each possible outcome is $\langle \psi_m | \rho | \psi_m \rangle$.

B. Motivation and definition of a quantum probabilistic graphical model

A quantum system constitutes a collection of quantum states, quantum operations and measurements. From a machine learning perspective, there are three types of problems pertaining to the characterization of a quantum system:

- 1) *Inference*. Given the quantum operations of the system and the measurement outcomes of some part of the system, retrieve the underlying quantum states or predict the unobserved measurement outcomes of the system.
- 2) *Parameter estimation*. Given the measurement outcomes, reconstruct the quantum operations and hidden quantum states of the system.
- 3) *Active learning*. Select the measurement operators that optimize the efficiency of inference and parameter estimation of the system.

Since these problems have been intensively studied in the statistical models of classical systems, it is desirable to find their counterparts in a quantum system.

Probabilistic graphical models are a class of multivariate statistical models where the joint probability mass (density) functions are decomposed into the products of terms with smaller number of variables [11]. A great number of statistical models (e.g., Bayesian networks, hidden Markov models, factor graphs) belong to this family. Many algorithms and heuristic methods on inference (e.g., [12], [13]), parameter estimation (e.g., [14], [15]) and active learning (e.g., [18], [19]) have been developed for probabilistic graphical models. Therefore, we propose the formulation of a *Quantum Probabilistic Graphical Model* (abbreviated as a Quantum Graphical Model or QGM) and the quantum version of algorithms on inference, parameter estimation, and active learning.

We construct a quantum graphical model based on the following simplifying assumptions. First, we are able to prepare a large sample with an identical quantum state. Second, there is no noise or coupling with the unknown external environment when quantum states undergo quantum operations. Third, there is no feedback in the system. Thus the inputs and outputs of each operation are clearly defined, and the graph contains no loops. Fourth, two 1-qubit quantum states are merged into a joint state by a tensor product. Similarly, a joint 2-qubits quantum state is split into two marginal 1-qubit quantum states by taking the partial trace over each dimension. In other words, entanglement of quantum states is destructed if they undergo merge or split operations.

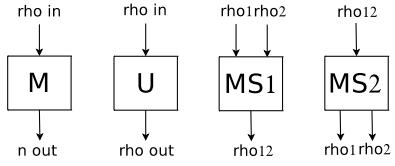
We define a quantum graphical model as a bipartite, directed graph $G = (V_S \cup V_O, E)$ consisting two types of nodes. $V_S = V_\rho \cup V_{\vec{n}}$ constitutes unobservable, continuous quantum states V_ρ and observable, discrete measurement outcomes $V_{\vec{n}}$. Each node $v_\rho \in V_\rho$ corresponds to a density matrix ρ , and each node $v_{\vec{n}} \in V_{\vec{n}}$ corresponds to a vector \vec{n} for discrete measurement outcomes over a sample. For instance, $\vec{n} = (n_0, n_1)$ denotes the numbers of observing 0 and 1 over $n_0 + n_1$ measurements. $V_O = V_U \cup V_M \cup V_{MS}$ constitutes unitary operators V_U , measurement operators V_M , and merge/split operators V_{MS} . Each node $v_U \in V_U$, $v_M \in V_M$ or $v_{MS} \in V_{MS}$ corresponds to an operator U , M or MS .

State and operation nodes are connected in the fashion of block diagrams. The input and output states of an operator are adjacent to the operator in G . The directions of the inputs/outputs depend on the direction of information flow in G . G is a Directed, Acyclic Graph (DAG) according to assumption 3 .

C. Likelihood functions of QGMs

The joint likelihood of a QGM is the probability of a joint configuration of all quantum states and measurement outcomes. According to assumption 2, uncertainties dwell on the quantum

Figure 1. Operators in the quantum graphical models. From left to right: measurement, unitary, merge and split operators



nature of the hidden states alone. The probability of the statistical outcome \vec{n} of measurement M conditioned on the hidden state ρ is (the left block in Figure 1):

$$P(\vec{n}|M, \rho) = \prod_{i=1}^d \langle \psi_i | \rho | \psi_i \rangle^{n_i}. \quad (6)$$

where $\{\psi_i\}_{i=1}^d$ constitutes a set of orthonormal basis vectors for M and n_i is the i^{th} component of \vec{n} .

The relation between the input and output density matrices of a unitary operator is deterministic. Thus the conditional probability of the output of a unitary matrix is an indicator function (the middle left block in Figure 1):

$$P(\rho_2|U, \rho_1) = \delta(\rho_2, U\rho_1 U^\dagger). \quad (7)$$

where $\delta(x_1, x_2) = 1$ if $x_1 = x_2$ and 0 otherwise.

A merge operator combines two quantum states of lower dimensions into a joint quantum state of a higher dimension. According to assumption 4, the uncoupled joint state is the tensor product of the two inputs. Hence the conditional probability is also an indicator function (the middle right block in Figure 1):

$$P(\rho_{12}|MS, \rho_1, \rho_2) = \delta(\rho_{12}, \rho_1 \otimes \rho_2). \quad (8)$$

Reciprocally, a split operator divides a joint quantum state into two marginal states. Following assumption 4 and the inverse operation of tensor products, the density matrix of each marginal quantum state is the partial trace of the joint density matrix over the subspace of another marginal state. This again yields an indicator function for the conditional probabilities (the right block in Figure 1):

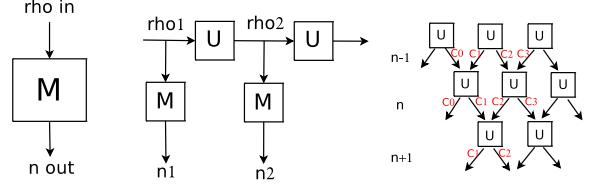
$$\begin{aligned} P(\rho_1|MS, \rho_{12}) &= \delta(\rho_1, \text{tr}_2(\rho_{12})). \\ P(\rho_2|MS, \rho_{12}) &= \delta(\rho_2, \text{tr}_1(\rho_{12})). \end{aligned} \quad (9)$$

From equations (6)-(9) the joint likelihood of a QGM is

$$\begin{aligned} L &= \prod_{U \in V_U} \delta(\rho_{U_{out}}, U\rho_{U_{in}} U^\dagger) \cdot \\ &\quad \prod_{MS \in V_{MS_1}} \delta(\rho_{MS_{12}}, \rho_{MS_1} \otimes \rho_{MS_2}) \cdot \\ &\quad \prod_{MS \in V_{MS_2}} \delta(\rho_{MS_{12}}, \text{tr}_2(\rho_{MS_{12}})) \delta(\rho_{MS_2}, \text{tr}_1(\rho_{MS_{12}})) \cdot \\ &\quad \prod_{M \in V_M} \prod_i \langle \psi_{M_i} | \rho_M | \psi_{M_i} \rangle^{n_{M_i}}. \end{aligned} \quad (10)$$

where $\rho_{U_{in}}$ and $\rho_{U_{out}}$ denote the input and output density matrices of unitary operator U , V_{MS_1} and V_{MS_2} the collections of merge and split operators, $\rho_{MS_{12}}$ the joint state, and ρ_{MS_1} , ρ_{MS_2} the corresponding marginal states, ρ_M the input quantum state for measurement M , n_{M_i} the count of the i^{th} outcome for measurement M , and ψ_{M_i} the i^{th} basis for measurement M .

Figure 2. Examples of quantum graphical models. Left: a binary measurement of a quantum state. The input is a density matrix ρ and the output is a vector of measurement outcomes \vec{n} . Middle: a quantum hidden Markov model. Density matrices ρ 's are hidden states and measurement outcomes \vec{n} 's are observed. Right: a quantum Markov random field. For simplicity the measurement and merge/split operators are not shown in the graph.



$\rho_{C_0 C_1}^n = \rho_{C_2 C_3}^n \equiv \rho^n$ and $\rho_{C_1 C_2}^{n+1} \equiv \rho^{n+1}$ as the joint states at layer n and $n+1$. Equation (12) then becomes:

$$\rho^{n+1} = U(\text{tr}_1(\rho^n) \otimes \text{tr}_2(\rho^n))U^\dagger. \quad (13)$$

III. INFERENCE ON THE PROBABILISTIC QUANTUM GRAPHICAL MODELS

Inference on QGMs denotes evaluation of the posterior probabilities of hidden quantum states or unobserved measurement outcomes conditioned on observed measurement outcomes. In classical graphical models, a variety of well-known message-passing algorithms have been proposed (e.g., [12], [13], [15]). In this section we describe a belief propagation algorithm on QGMs (abbreviated as QBP) and demonstrate the accuracy of the algorithm theoretically and experimentally.

A. Description of the Quantum Belief Propagation Algorithm

Figure 3. Quantum Belief Propagation Algorithm

Inputs: A QGM G , evidence \mathcal{E} of observed measurement outcomes.
Outputs: Posterior probabilities of each quantum state and unobserved measurement outcome of G conditioned on \mathcal{E} .

- 1) Initialize each message with a uniform function $m(x) = 1$.
- 2) Set the observed variable \rightarrow factor messages according to evidence \mathcal{E} .
- 3) Iteratively update messages until all messages converge.
 - a) Variable \rightarrow factor message update: equation (14).
 - b) Factor \rightarrow variable message update: equations (19)-(21).
- 4) Evaluate the belief function for each variable: $b(x) = \prod_{f \in N(x)} m_{f \rightarrow x}(x)$, where $N(x)$ denotes the first neighbors of variable x in G .

The steps of QBP are summarized in Figure 3. It is a variation of the sum-product algorithm on classical factor graphs [13]. A message is a function associated with a directed pair of adjacent variable ($v \in V_S$) and factor ($f \in V_O$) nodes in G . Information pertaining to the evidence of G is propagated via message functions throughout the network. Message passing stops when all message functions converge. The belief function (marginal probability) of each variable is the product of all incident message functions.

The key components of sum-product are evidence functions and message updates. An evidence function \mathcal{E} fixes a collection of variables to specific values. There are two types of messages: variable \rightarrow factor messages and factor \rightarrow variable messages. For a variable whose value is not fixed by \mathcal{E} , the variable \rightarrow factor message is updated by the product of factor \rightarrow variable messages incident to the variable:

$$m_{x \rightarrow f}(x) = \prod_{f' \in N(x) \setminus f} m_{f' \rightarrow x}(x). \quad (14)$$

A factor \rightarrow variable message is updated by marginalizing the product of the factor and incoming messages over the non-target variables:

$$m_{f \rightarrow x}(x) = \int f(x, N(f) \setminus x) \prod_{y \in N(f) \setminus x} m_{y \rightarrow f}(y) dy. \quad (15)$$

We translate sum-product into a QBP by constructing a factor graph from a QGM. Variables in the factor graph correspond to the density matrices ρ and measurement outcomes \vec{n} of the QGM. An evidence function $\mathcal{E}(\vec{n}; \vec{n}_O) = \delta(\vec{n}, \vec{n}_O)$ fixes the measurement outcome \vec{n} to the observed values \vec{n}_O . Factors in the factor graph

correspond to the constraints imposed by the operators in the QGM. There are three types of factors:

- 1) A unitary operator U incurs a deterministic relation of input and output density matrices $\rho_2 = U\rho_1U^\dagger$. The corresponding factor in a factor graph is an indicator function:

$$f(\rho_1, \rho_2) = \delta(\rho_2, U\rho_1U^\dagger). \quad (16)$$

- 2) For a measurement operator M , the probability of each outcome \vec{n} is given by a multinomial distribution parameterized by $\langle \psi_k | \rho | \psi_k \rangle$. The corresponding factor is:

$$f(\rho, \vec{n}) = \binom{N}{\vec{n}} \prod_k \langle \psi_k | \rho | \psi_k \rangle^{n_k}. \quad (17)$$

The multiplying factor $\binom{N}{\vec{n}}$ denotes the number of permutations consistent with \vec{n} . N is the total number of observations.

- 3) A merge/split operator MS gives a deterministic relation $\rho_{12} = \rho_1 \otimes \rho_2$ between the joint state ρ_{12} and marginal states ρ_1, ρ_2 . The corresponding factor is:

$$f(\rho_{12}, \rho_1, \rho_2) = \delta(\rho_{12}, \rho_1 \otimes \rho_2). \quad (18)$$

The QBP in Figure 3 is the sum-product on the factor graph described above. The variable \rightarrow factor messages are updated by equation (14). For factor \rightarrow variable messages updates, we substitute equations (16)-(18) into equation (15) and obtain the following rules.

- 1) A unitary operator U . Suppose ρ_1 and ρ_2 are the input and output density matrices respectively. Then the following updates hold:

$$\begin{aligned} m_{U \rightarrow \rho_2}(\rho_2) &= m_{\rho_1 \rightarrow U}(U^\dagger \rho_2 U). \\ m_{U \rightarrow \rho_1}(\rho_1) &= m_{\rho_2 \rightarrow U}(U \rho_1 U^\dagger). \end{aligned} \quad (19)$$

- 2) A measurement operator M . Suppose ρ and \vec{n} are the density matrix of the hidden state and the measurement outcome respectively. Then the following updates hold:

$$\begin{aligned} m_{M \rightarrow \vec{n}}(\vec{n}) &= \int \binom{N}{\vec{n}} \prod_k \langle \psi_k | \rho | \psi_k \rangle^{n_k} m_{\rho \rightarrow}(\rho) d\rho. \\ m_{M \rightarrow \rho}(\rho) &= \sum_{\vec{n}} \binom{N}{\vec{n}} \prod_k \langle \psi_k | \rho | \psi_k \rangle^{n_k} m_{\vec{n} \rightarrow M}(\vec{n}). \end{aligned} \quad (20)$$

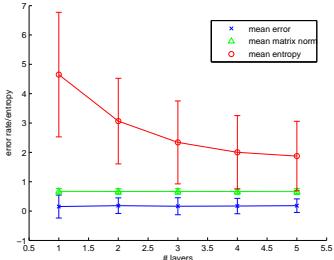
- 3) A merge/split operator MS . Suppose ρ_{12} is the density matrix of the joint state and ρ_1 and ρ_2 the two marginal states. Then the following updates hold:

$$\begin{aligned} m_{MS \rightarrow \rho_{12}}(\rho_{12}) &= \int m_{\rho_1 \rightarrow MS}(\rho_1) m_{\rho_2 \rightarrow MS}(\rho_2) \delta(\rho_{12}, \rho_1 \otimes \rho_2) d\rho_1 d\rho_2 \\ &= m_{\rho_1 \rightarrow MS}(tr_2(\rho_{12})) m_{\rho_2 \rightarrow MS}(tr_1(\rho_{12})). \\ m_{MS \rightarrow \rho_1}(\rho_1) &= \int m_{\rho_{12} \rightarrow MS}(\rho_{12}) m_{\rho_2 \rightarrow MS}(\rho_2) \delta(\rho_{12}, \rho_1 \otimes \rho_2) d\rho_2 d\rho_{12} \\ &= \int m_{\rho_{12} \rightarrow MS}(\rho_{12} \otimes \rho_2) m_{\rho_2 \rightarrow MS}(\rho_2) d\rho_2. \\ m_{MS \rightarrow \rho_2}(\rho_2) &= \int m_{\rho_{12} \rightarrow MS}(\rho_{12}) m_{\rho_1 \rightarrow MS}(\rho_1) \delta(\rho_{12}, \rho_1 \otimes \rho_2) d\rho_1 d\rho_{12} \\ &= \int m_{\rho_{12} \rightarrow MS}(\rho_1 \otimes \rho_2) m_{\rho_1 \rightarrow MS}(\rho_1) d\rho_1. \end{aligned} \quad (21)$$

B. Empirical results

We built Quantum Markov Random Fields on 2-qubit quantum states by varying the number of layers from 1 to 5 and fixing the number of unitary operators in each layer to be 5. For each grid structure we randomly generated 100 initial density matrices in

Figure 4. Error and uncertainty of inference results on 2D QMRFs with 1-5 layers. Blue crosses: average Frobenius norm distances between the true and MAP density matrices. Green triangles: average norms of the true density matrices. Red circles: average entropies of posterior probabilities of density matrices.



layer 1 and unitary operators of the following forms:

$$\begin{aligned}
 \rho &= \rho_1 \otimes \rho_2, \\
 \rho_i &= \sum_{j=0}^1 p_{ij} |\phi_{ij}\rangle\langle\phi_{ij}|, i = 1, 2, \\
 |\phi_{i0}| &= \begin{pmatrix} \cos\zeta_i & \sin\zeta_i \end{pmatrix}, \\
 |\phi_{i1}| &= \begin{pmatrix} -\sin\zeta_i & \cos\zeta_i \end{pmatrix}, \\
 U &= U_1 \otimes U_2, \\
 U_i &= \begin{pmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{pmatrix}, i = 1, 2.
 \end{aligned} \quad (22)$$

Simulated measurement outcomes were fed into QBP to evaluate the posterior probability of each quantum state.

We gauged the quality of inference results with two quantities. Errors were measured by the average Frobenius norm distance between the true density matrix and the maximum a posteriori density matrix:

$$\epsilon = \frac{1}{K} \sum_s \sqrt{\sum_i \sum_j |\rho_{sij} - \hat{\rho}_{sij}|^2}. \quad (23)$$

where K denotes the number of quantum state variables and s the index of quantum state variables in the QMRF, ρ_s its simulated state, and $\hat{\rho}_s$ the MAP state inferred from QBP. Uncertainty was measured by the average differential entropy of the inferred posterior probability of each ρ :

$$h = -\frac{1}{K} \sum_s \int P(\rho_s | \mathcal{E}) \log(P(\rho_s | \mathcal{E})) d\rho_s. \quad (24)$$

where \mathcal{E} denotes observed evidence.

Figure 4 shows the inference errors (blue crosses), matrix norms of the quantum states (green triangles), and uncertainties (red circles) with varying numbers of layers in the model. As expected, the mean errors are substantially smaller than the matrix norms of the quantum states. Errors arise from quantization of density matrices in integration instead of the inference procedures. Therefore, the means and standard deviations of the errors are insensitive to the number of layers in the model. In contrast, uncertainties decrease with the number of layers in the model, since posterior probabilities are more concentrated as information from more layers is included.

IV. PARAMETER ESTIMATION ON QUANTUM HIDDEN MARKOV MODELS

Parameter estimation in QGMs denotes simultaneous determination of unitary operators and hidden quantum states in the system

from the discrete measurement outcomes. In this work we consider a special case of parameter estimation. The QGM is a quantum hidden Markov model where the unitary operator U is identical at each step. We describe an iterative algorithm to reconstruct U and ρ and validate the algorithm on simulated data.

A. Description of the iterative algorithm for parameter/state estimation on QHMMs

The log likelihood function of a QHMM is derived from equation (11):

$$\begin{aligned}
 L(\rho, U) &= \sum_{i=0}^T \sum_k n_{ik} \log \langle \psi_k | U^i \rho U^{\dagger i} | \psi_k \rangle, \\
 &\equiv \sum_{i=0}^T \sum_k n_{ik} \log \langle y_{i,k} | \rho | y_{i,k} \rangle.
 \end{aligned} \quad (25)$$

where T denotes the number of steps, ρ the density matrix of the initial step, U the unitary matrix at each step, and $\{\psi_k\}$ the orthonormal measurement basis vectors. Define $\langle y_{i,k} | \equiv \langle \psi_k | U^i$. The goal is to find the maximizer of $L(\rho, U)$ subjected to the constraints that (1) ρ is a valid density matrix, (2) U is unitary.

Simultaneous optimization of ρ and U is difficult as it requires solving complicated equations with Lagrange multipliers. We adopt an EM-like approach by iteratively fixing one set of variables (ρ or U) and estimating the others. Iteration terminates when both ρ and U converge to fixed points. Figure 5 summarizes the procedures of the iterative algorithm.

Figure 5. An algorithm for parameter and hidden state estimation of QHMMs

Inputs: Measurement outcomes for each step of the QHMM.

Outputs: Maximum likelihood unitary matrix \hat{U} and initial density matrix $\hat{\rho}$.

- 1) Initialize U as a random unitary matrix and ρ as a density matrix with random orthonormal bases and mixture coefficients.
- 2) Iterate the following two subroutines until both U and ρ converge.
- 3) E-step, fix U and optimize ρ :
 - a) Initialize ρ as a random density matrix (equation (26)).
 - b) Iterate the following routines until both r_l and $|\phi_l\rangle\langle\phi_l|$ converge.
 - i) r_l optimization: fix $|\phi_l\rangle\langle\phi_l|$ and apply equation (32) to find r_l .
 - ii) $|\phi_l\rangle\langle\phi_l|$ optimization: fix r_l and apply equations (33), (27), and (28) to find $|\phi_l\rangle\langle\phi_l|$.
- 4) M-step, fix ρ and optimize U :
 - a) Choose K random unitary matrices as the initial U 's.
 - b) For each initial U , iteratively apply equation (39) to update U^t until it converges.
 - c) Find the supremum over the local optimum matrices of U .

Estimation of ρ with a fixed U is achieved by a quantum state reconstruction algorithm described in [7]. A valid density matrix can be expressed as

$$\rho = \sum_l r_l |\phi_l\rangle\langle\phi_l|. \quad (26)$$

where $|\phi_l\rangle\langle\phi_l|$'s are the pure states bases (phases) and r_l 's their mixture coefficients (magnitudes). To estimate ρ we consider a small perturbation of ρ where $r'_l = r_l + \delta r_l$ and $|\phi'_l\rangle\langle\phi'_l| = S|\phi_l\rangle\langle\phi_l|S^\dagger$. Each $|\phi_l\rangle\langle\phi_l|$ is rotated by an infinitesimal unitary transformation:

$$S = e^{i\varepsilon G} \approx 1 + i\varepsilon G. \quad (27)$$

The rotated phases are

$$\begin{aligned}
 |\phi'_l\rangle\langle\phi'_l| &= S|\phi_l\rangle\langle\phi_l|S^\dagger \\
 &\approx |\phi_l\rangle\langle\phi_l| + i\varepsilon \text{tr}([G, |\phi_l\rangle\langle\phi_l|]). \quad (28)
 \end{aligned}$$

where $[A, B] \equiv AB - B^\dagger A$. The perturbation of log likelihood function is

$$\delta L(r_l, |\phi_l\rangle\langle\phi_l|) = \sum_l \left(\frac{\partial L(r_l, |\phi_l\rangle\langle\phi_l|)}{\partial \rho} - 1 \right) \delta r_l + \sum_l (L(r_l, |\phi'_l\rangle\langle\phi'_l|) - L(r_l, |\phi_l\rangle\langle\phi_l|)). \quad (29)$$

The $-\delta r_l$ terms arise from the constraint $\sum_l r_l = 1$. Equation (29) is reduced to

$$\delta L(\rho) = \sum_l (\langle\phi_l|R|\phi_l\rangle - 1) \delta r_l + i\varepsilon \text{tr}(G[\rho, R]). \quad (30)$$

where R denotes the operator

$$R = \sum_{i,k} \frac{n_{ik}}{\langle y_{i,k} | \rho | y_{i,k} \rangle} |y_{i,k}\rangle\langle y_{i,k}|. \quad (31)$$

Estimation of ρ is achieved by iteratively optimizing r_l and $|\phi_l\rangle\langle\phi_l|$. For fixed $|\phi_l\rangle\langle\phi_l|$, find the r_l that set the first term in equation (29) to 0. This is carried out by iterative scaling: updating r_l 's according to equation (32) until they converge.

$$r_l^t = r_l^{t-1} \sum_{i,k} \frac{n_{ik}}{\sum_j r_j^{t-1} |\langle y_{i,k} | \phi_l \rangle|^2}. \quad (32)$$

For fixed r_l , find the $|\phi_l\rangle\langle\phi_l|$ that forces the second term in equation (29) to be positive. This is carried out by setting

$$G = i[\rho, R] = i[\rho R - R^\dagger \rho]. \quad (33)$$

and substituting G into equation (28). r_l and $|\phi_l\rangle\langle\phi_l|$ are alternatively optimized until they converge to fixed points.

Estimation of U with a fixed ρ is achieved by gradient ascent. We derive the gradient of log likelihood by differentiating equation (25) with respect to U ,

$$\frac{\partial L}{\partial U} = \sum_{i,k} \frac{n_{ik}}{\langle \psi_k | U^i \rho U^{\dagger i} | \psi_k \rangle} \frac{\partial}{\partial U} \langle \psi_k | U^i \rho U^{\dagger i} | \psi_k \rangle. \quad (34)$$

We derive the following equalities from matrix calculus [21]:

$$\begin{aligned} d(f^\dagger(U)Cf(U)) &= (I \otimes f^\dagger(U)C)df(U) + (f^\dagger(U)C^\dagger \otimes I)df^\dagger(U). \\ f(U) &= U^{\dagger i} |\psi_k\rangle. \\ df(U) &= [(\langle \psi_k | \otimes I) (\sum_{j=1}^i U^{i-j} \otimes (U^\dagger)^{j-1})]dU^\dagger. \end{aligned} \quad (35)$$

where $C \equiv \rho$ is a constant matrix for a fixed ρ , dU is a flattened column vector of differentials of components du_{pq} , and dU^\dagger is the Hermitian adjoint of dU . From these equalities,

$$\begin{aligned} d(f^\dagger(U)Cf(U)) &= (I \otimes \langle \psi_k | U^i C) \cdot \\ &\quad [(\langle \psi_k | \otimes I) (\sum_{j=1}^i U^{i-j} \otimes (U^\dagger)^{j-1})]dU^\dagger + \\ &\quad ((\langle \psi_k | U^i C^\dagger \otimes I) \cdot \\ &\quad [(\sum_{j=1}^i (U^\dagger)^{i-j} \otimes U^{j-1}) (I \otimes \langle \psi_k |)]dU. \end{aligned} \quad (36)$$

By substituting equation (36) into equation (34), the gradient of log likelihood on the (p, q) component becomes

$$\begin{aligned} (\nabla L(U))_{pq} &= \sum_{i,k} \frac{n_{ik}}{\langle \psi_k | U^i \rho U^{\dagger i} | \psi_k \rangle} \cdot \{ (\langle \psi_k | U^i \rho [(\langle \psi_k | \otimes I) \\ &\quad (\sum_{j=1}^i U^{i-j} \otimes (U^\dagger)^{j-1})])_{qp} + \\ &\quad \{ [(\sum_{j=1}^i (U^\dagger)^{i-j} \otimes U^{j-1}) (\langle \psi_k | \otimes I)] \rho (U^\dagger)^i |\psi_k\rangle \}_{pq} \}. \end{aligned} \quad (37)$$

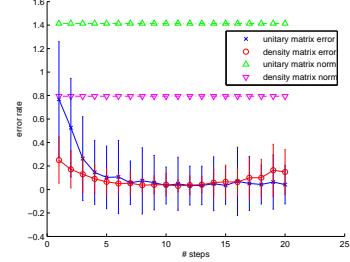
To restrict the gradient on the manifold of unitary matrices, we introduce a Riemannian gradient proposed by [22]:

$$\tilde{\nabla} L(U) = U(\nabla L(U))^\dagger U. \quad (38)$$

The gradient ascent update of U becomes

$$U^t = U^{t-1} + \tilde{\nabla} L(U^{t-1}) du^t. \quad (39)$$

Figure 6. Estimation errors for density and unitary matrices on 1D QHMMs with 1-20 layers. Blue crosses: average Frobenius norm distances between the true and estimated unitary matrices. Red circles: average Frobenius norm distances between the true and estimated density matrices. Green triangles: average norms of the true unitary matrices. Magenta triangles: average norms of the true density matrices.



where U^t denotes the unitary matrix in the t^{th} iteration and du^t is a scalar differential. Following the update rule of equation (39) U^t converges to a local optimum.

Equation (39) does not guarantee to find the global optimum of L . When U is embedded in a low dimensional space (e.g., a 2×2 real matrix), we observe $L(U)$ is nearly periodic. The number of peaks equals to the number of steps in the QHMM. Thus we can shift the initial U of gradient ascent multiple times and find the supremum of the results over multiple trials.

B. Empirical results

We applied the parameter estimation algorithm to the 1-qubit QHMMs. 100 random QHMMs with a varying number of steps (from 1-20) were generated. Estimation errors were measured by the average Frobenius distances between the true (density or unitary) matrices and the estimated matrices over the 100 random trials. Figure 6 shows the average estimation errors of the unitary matrices (blue crosses) and density matrices (red circles) with varying numbers of steps in QHMMs. Several observations validate the accuracy of the iterative algorithm. First, both types of errors are substantially smaller than the matrix norms of the underlying unitary (green triangles) and density (magenta triangles) matrices. Second, both types of errors decrease with the number of steps in the QHMMs in the first 10 data points. Third, estimation errors of unitary matrices are larger than those of the density matrices when the number of steps ≤ 6 because the data contains more information of ρ than U . For an n -step QHMM the initial density matrix appears in all the n terms of measurement outcomes, but the unitary matrix only appears in the last $n-1$ terms (equation (11)). Fourth, both types of errors converge to similar values as more steps are included in the model. This is also sensible as the effect of the additional information on density matrices is diluted when more steps are included.

V. DETERMINATION OF THE MEASUREMENT OPERATORS

A measurement of a discrete outcome projects a quantum state onto a one-dimensional space. The number of projections carried by a fixed set of orthonormal basis vectors is smaller than the number of parameters in a density or unitary matrix. Therefore, multiple orthonormal basis sets are required in order to completely determine the underlying density or unitary matrix.

One critical issue is to design the measurement operators that maximize the expected information between the estimated quantity (quantum state or operator) and measurement outcomes. This problem is intensively studied in experimental design [18] and active learning [19]. In this section we describe a method of determining the measurement operators and demonstrate its utility in empirical analysis.

A. An information theoretic criteria for choosing measurement operators

In this work we consider the simplest case of a quantum system: determining a hidden quantum state ρ via measurements. Suppose multiple sets of orthonormal basis $\vec{\psi}_1, \dots, \vec{\psi}_m$ are applied, then the joint likelihood of ρ and measurement outcomes $\vec{n}_1, \dots, \vec{n}_m$ is

$$L(\rho, \vec{n}_1, \dots, \vec{n}_m; \vec{\psi}_1, \dots, \vec{\psi}_m) = P(\rho) \prod_{i=1}^m \prod_k \langle \psi_{ik} | \rho | \psi_{ik} \rangle^{n_{ik}}. \quad (40)$$

For notation simplicity denote $(\vec{\psi}_1, \dots, \vec{\psi}_m) \equiv \vec{\psi}$ and $(\vec{n}_1, \dots, \vec{n}_m) \equiv \vec{n}$.

One sensible criterion for designing the measurement operators $\vec{\psi}'$ is to maximizes the mutual information between the underlying density matrix ρ and the expected measurement outcomes \vec{n}' , conditioned on all observed data:

$$\begin{aligned} I(\rho; \vec{n}' | \vec{\psi}, \vec{n}, \vec{\psi}') &= \sum_{\vec{n}'} \int d\rho P(\rho | \vec{\psi}, \vec{n}) P(\vec{n}' | \rho, \vec{\psi}') \cdot \\ &\quad \log\left(\frac{P(\rho; \vec{n}' | \vec{\psi}, \vec{n}, \vec{\psi}')} {P(\rho | \vec{\psi}, \vec{n}, \vec{\psi}') P(\vec{n}' | \vec{\psi}, \vec{n}, \vec{\psi}')} \right) \\ &= \sum_{\vec{n}'} \int d\rho P(\rho | \vec{\psi}, \vec{n}) P(\vec{n}' | \rho, \vec{\psi}') \cdot \\ &\quad \log\left(\frac{P(\vec{n}' | \rho, \vec{\psi}')}{P(\vec{n}' | \vec{\psi}, \vec{n}, \vec{\psi}')} \right). \end{aligned} \quad (41)$$

$P(\rho | \vec{\psi}, \vec{n})$ is the posterior probability of ρ conditioned on the measurement outcomes of m orthonormal basis sets:

$$\begin{aligned} P(\rho | \vec{\psi}, \vec{n}) &= \frac{1}{Z} \prod_{i=1}^m \prod_k \langle \psi_{ik} | \rho | \psi_{ik} \rangle^{n_{ik}}. \\ Z &= \int d\rho \prod_{i=1}^m \prod_k \langle \psi_{ik} | \rho | \psi_{ik} \rangle^{n_{ik}}. \end{aligned} \quad (42)$$

Multiplying factors $\binom{N}{n_i}$ are absent in equation (42) because the observed data $\vec{n}_1, \dots, \vec{n}_m$ follow fixed permutation orders. $P(\vec{n}' | \rho, \vec{\psi}')$ denotes the probability of observing a hypothetical measurement outcome \vec{n}' conditioned on the quantum state ρ and the measurement basis set $\vec{\psi}'$. A multiplying factor $\binom{N}{n'}$ is required since the permutation order of the hypothetical measurement outcome is not yet fixed:

$$P(\vec{n}' | \rho, \vec{\psi}') = \binom{N}{n'} \prod_k \langle \psi'_k | \rho | \psi'_k \rangle^{n'_k}. \quad (43)$$

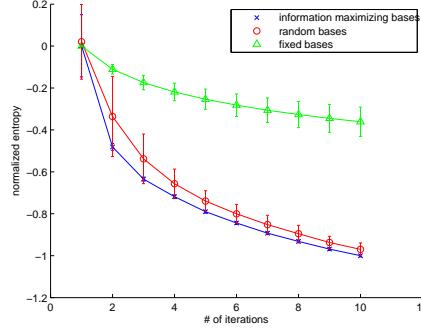
$P(\vec{n}' | \vec{\psi}, \vec{n}, \vec{\psi}')$ is obtained by integrating $P(\rho | \vec{\psi}, \vec{n}) P(\vec{n}' | \rho, \vec{\psi}')$ over ρ :

$$P(\vec{n}' | \vec{\psi}, \vec{n}, \vec{\psi}') = \int d\rho \frac{1}{Z} \prod_{i=1}^m \prod_k \langle \psi_{il} | \rho | \psi_{il} \rangle^{n_{il}}. \quad (44)$$

By substituting equations (42), (43) and (44) into equation (41),

$$\begin{aligned} I(\rho; \vec{n}' | \vec{\psi}, \vec{n}, \vec{\psi}') &= \sum_{\vec{n}'} \int d\rho \frac{1}{Z} \prod_{i=1}^m \prod_k \langle \psi_{ik} | \rho | \psi_{ik} \rangle^{n_{ik}} \cdot \\ &\quad \left(\log\left(\binom{N}{n'}\right) + \sum_k n'_k \log(\langle \psi'_k | \rho | \psi'_k \rangle^{n'_k}) - \right. \\ &\quad \left. \log\left(\int d\rho \frac{1}{Z} \prod_{i=1}^m \prod_k \langle \psi_{il} | \rho | \psi_{il} \rangle^{n_{il}}\right) \right. \\ &\quad \left. - \log\left(\binom{N}{n'} \prod_k \langle \psi'_k | \rho | \psi'_k \rangle^{n'_k}\right) \right]. \end{aligned} \quad (45)$$

Figure 7. Active learning of measurement bases. Posterior distribution uncertainty (entropy) of density matrices with the measurement outcomes from multiple basis sets. Blue crosses: incrementally select basis sets by mutual information. Red circles: basis sets with random phases. Green triangles: canonical bases using columns of the identity matrix.



We propose an incremental algorithm for choosing the measurement operators based on equation (45). Initially set a random basis set. In each iteration, apply the current basis set to acquire the measurement outcomes. Conditioned on the basis sets and measurement outcomes in all the preceding steps, find the new basis set that maximizes equation (45). Continue the iteration until a fixed number of basis sets are selected.

B. Empirical results

We applied the active learning algorithm to infer 1-qubit quantum states from multiple sets of measurement bases. 10 distinct 1-qubit density matrices were randomly generated. For each density matrix, 10 trials of inference processes were undertaken with random initial orthonormal bases. We then iteratively generated the measurement outcomes and altered the measurement bases. Three methods were applied to select the orthonormal bases in each iteration: (1) incrementally chose the basis sets to maximize the mutual information score (equation (45)), (2) select random orthonormal basis sets, (3) a canonical orthonormal basis set $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ are repetitively chosen. To gauge the performance of each method we evaluated the differential entropies of the density matrix posterior probabilities in 10 iterations. Figure 7 shows the means and standard deviations of the density matrix entropies in each iteration. Several observations confirm the utility of the active learning scheme. First, entropy decreases with the number of basis sets for all three methods. This is sensible since more measurement outcomes lower the uncertainty of the density matrix. Second, both the mutual information criteria and random bases yield substantially lower entropy values compared to the canonical bases. Canonical bases have a poor performance since the increasing amount of data can only determine the diagonal entries of the density matrix but are completely uninformative about the off-diagonal entries. Third, the mean entropy values from the mutual information criteria (blue crosses) are consistently lower than the random bases (red circles) across 10 iterations. Although the difference margin is much smaller than the margins from the

canonical bases, it indicates the gain of including the information from the observed data in determining measurement bases. Fourth, after the first iteration the standard deviations of entropy values from the mutual information criteria are much smaller than those from the random bases. The mutual information criteria adjust the subsequent bases so that the entropies quickly converge to the fixed values insensitive of the initial choice. In contrast, random bases yield a much slower convergence of entropy values, as the standard deviations remain high in the first 5 iterations.

VI. DISCUSSION AND CONCLUSION

In this study we build probabilistic graphical models on quantum systems and propose algorithms for inference, parameter estimation and active learning on quantum probabilistic graphical models. To our knowledge this is one of the first attempts to build probabilistic graphical models on quantum systems.

Our modeling framework treats density matrices as hidden variables and applies operations of classical probabilities to them. This is probably the most obvious but not the only approach to model a quantum system. Other possible frameworks exist (for instance, [17] and [23]).

Our work is an initial step toward a comprehensive understanding of quantum probabilistic graphical models. Many open problems remain in this direction. Here we list several important issues to be addressed in the future work.

Parameter estimation of generic QGMs yields complicated likelihood functions and challenging optimization problems. In this work we only consider the simplest settings (QHMMs for parameter estimation). Better approximation methods to handle the complicated tasks need to be developed.

Similarly, for mathematical convenience we impose several simplifying assumptions in constructing QGMs, such as no coupling with external environment and unitary transformations between density matrices and orthonormal measurement bases. For more useful and practical applications these assumptions need to be relaxed or removed.

Structure learning of a QGM is not addressed in this work. It is NP-hard in classical graphical models [20] and is likely to be as hard in quantum systems. Various heuristic methods on structure learning exist in classical systems (e.g., [16]). It will be of interest to find the quantum version of these methods.

ACKNOWLEDGMENT

The author would like to thank Yuan-Chin Chang for the feedback on the manuscript and Manfred Warmuth for the discussion about density matrices.

REFERENCES

- [1] Shor, P.W.: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comp.* 26(5):1484-1509. 1997.
- [2] Grover, L.K.: Quantum mechanics helps in search for a needle in a haystack. *Phys. Rev. Lett.* 79(2):325. 1997.
- [3] Ekert, A.K.: Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.* 67(6):661-663. 1991.
- [4] Bennett, C.H., Brassard, G., Crepeau, C., Jozsa, R., Peres, A., Wootters, W.: Teleporting an unknown quantum state via dual classical and EPR channels. *Phys. Rev. Lett.* 70:1895-1899. 1993.
- [5] Leonhardt, U.: Measuring the quantum state of light. Cambridge University Press, New York, 1997.
- [6] Chuang, I.L., Nielsen, M.A.: Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt.* 44(11-12):2455-2467, 1997.
- [7] Hradil, Z., Rehacek, J., Fiurasek, J., Jezek, M.: Maximum-likelihood methods in quantum mechanics. *Quantum State Estimation*, vol. 649. 59-112. Lecture Notes in Physics. Springer Berlin/Heidelberg, 2004.
- [8] Schack, R., Brun, T.A., Caves, C.M.: Quantum Bayes rule. *Phys. Rev. A.* 64:014305-1-014305-4. 2001.
- [9] Fuchs, C.A., Schack, R.: Unknown quantum states and operations, a Bayesian view. *Quantum State Estimation*, vol. 649. 147-187. Lecture Notes in Physics. Springer Berlin/Heidelberg, 2004.
- [10] Buzek, V.: Quantum tomography from incomplete data via *MaxEnt* principle. *Quantum State Estimation*, vol. 649. 190-234. Lecture Notes in Physics. Springer Berlin/Heidelberg, 2004.
- [11] Jordan, M.: Graphical models. *Stat. Sci.* 19:140-155. 2004.
- [12] Pearl, J.: Probabilistic inference in intelligent systems. Morgan Kaufmann, San Mateo. 1988.
- [13] Kschischang, F.R., Frey, B.J.: Factor graphs and the sum-product algorithm. *IEEE Trans. Inform. Theo.* 47(2):498-519, 2001.
- [14] Heckerman, D.: A tutorial on learning with Bayesian networks. *Learning in graphical models*. 301-354. The MIT Press, Cambridge, MA, USA. 1999.
- [15] Rabiner, L.R., Juang, B.H.: An introduction to hidden Markov models. *IEEE ASSP Magazine* 3:4-16. 1986.
- [16] Friedman, N.: Learning Bayesian networks in the presence of missing values and hidden variables. *Proc. of the Fourteenth International Conference on Machine Learning*, 125-133, 1997.
- [17] Nielsen, M.A., Chuang, I.L.: Quantum computation and quantum information. Cambridge University Press, New York. 2000.
- [18] Fedorov, V.V.: Theory of optimal experiments. Academis Press, New York. 1972.
- [19] Tong, S., Koller, D.: Active learning for parameter estimation in Bayesian networks. *NIPS Proceedings*. 2000.
- [20] Chickering, D.M., Geiger, D., Heckerman, D.: Learning Bayesian networks is NP-hard. Microsoft Research Technical Report. MSR-TR-94-17. 1994.
- [21] Brookes, M.: The matrix reference manual. <http://www.ee.ic.ac.uk/hpstaff/dm/matrix/intro.html>.
- [22] Abrudan, T., Eriksson, J., Koivunen, V.: Conjugate gradient algorithm for optimization under unitary matrix constraint. *Signal Processing* 89:1704-1714. 2009.
- [23] Warmuth, M.K., Kuzmin, D.: A Bayesian probability calculus for density matrices. *NIPS Proceedings*, 2006.