

# Recoverability from direct quantum correlations

S Di Giorgio<sup>1,2,3</sup>  P Mateus<sup>1,2</sup> and B Mera<sup>1,2</sup>

<sup>1</sup> Departamento de Matemática, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal

<sup>2</sup> Instituto de Telecomunicações, 1049-001 Lisboa, Portugal

E-mail: [serena.giorgio@tecnico.ulisboa.pt](mailto:serena.giorgio@tecnico.ulisboa.pt)

Received 19 August 2019

Accepted for publication 26 February 2020

Published 15 April 2020



CrossMark

## Abstract

We address the problem of compressing density operators defined on a finite dimensional Hilbert space which assumes a tensor product decomposition. In particular, we look for an *efficient procedure* for learning the *most likely* density operator, according to ‘Jaynes’ principle, given a *chosen set* of partial information obtained from the unknown quantum system we wish to describe. For complexity reasons, we restrict our analysis to *tree-structured* sets of bipartite marginals. We focus on the tripartite scenario, where we solve the problem for the couples of measured marginals which are compatible with a quantum Markov chain, providing then an algebraic necessary and sufficient condition for the compatibility to be verified. We introduce the generalization of the procedure to the n-partite scenario, giving some preliminary results. In particular, we prove that if the pairwise Markov condition holds between the subparts then the choice of the *best* set of tree-structured bipartite marginals can be performed efficiently. Moreover, we provide a new characterization of quantum Markov chains in terms of *quantum Bayesian updating processes*.

Keywords: quantum Markov chains, maximum von Neumann entropy, bipartite correlations, quantum trees

## 1. Introduction

The problem of efficiently compressing density operators can be related to the one addressed by Jaynes [1] for probability distributions. We are interested in determining an *efficient procedure* for inferring the *most likely density operator* from partial information about the system we wish to describe, with the freedom of choosing the *set of partial information* to be collected. With a complete set of measurements, quantum tomography techniques are able to infer with

<sup>3</sup> Author to whom any correspondence should be addressed.

maximum accuracy the density operator that most likely describes the given quantum system, but the needed resources to perform the former increases exponentially with the number of degrees of freedom of the system. A clever choice of a partial set of measurements should optimize the data collection, lead up to an efficient learning procedure and keep a good accuracy. The necessity of dealing with a statistically relevant number of degrees of freedom motivates machine learning [2] and quantum machine learning techniques [3–5].

### 1.1. The maximum entropy estimator

In his seminal paper [1], Jaynes wrote:

‘Information theory provides a constructive criterion for setting up probability distribution on the basis of partial knowledge and leads to a type of statistical inference which is called the maximum-entropy estimate. It is the least biased estimate possible on the given information, i.e. it is maximally noncommittal with regard to missing information.’

Since quantum information theory provides a well-defined generalization the Shannon entropy, the von Neumann entropy, it can be used to state a quantum ‘Jaynes’ principle and, therefore, to obtain a maximally noncommittal estimator for density operators with regard to the partial information collected. Moreover, due to the concavity of both Shannon and von Neumann entropy, the maximization problem has a unique solution and the desired estimator is uniquely determined. We choose therefore to infer from the given measurements, the density operators that maximizes the von Neumann entropy.

### 1.2. Learning from direct correlations

A common approximation for multipartite physical system description consists in cutting the correlations after the first neighbour. Direct dependencies between random variables are usually less struggling to be measured and it results in a theoretical exponential gain in data collection. Consider a multipartite quantum system, described by a Hilbert space  $\mathcal{H}_{X_1, \dots, X_n} = \mathcal{H}_{X_1} \otimes \dots \otimes \mathcal{H}_{X_n}$ , with  $\dim \mathcal{H}_{X_i} = O(d)$ , for all  $i = 1, \dots, n$ . To infer the state of the whole system, represented by a density matrix  $\rho_{X_1, \dots, X_n}$ , one needs an exponential amount of resources, more concretely,  $O(d^{2n})$ . However, if one restricts to bipartite correlations only, the amount of resources needed to approximately reproduce the state scales polynomially with  $n$ ,  $O(n^2 d^4)$ . This approximation restricts the set of learnable states: for pure-multipartite correlated systems, such as the GHZ state, we expect not to be accurately recoverable.

### 1.3. Restriction to trees

Since density operators generalize classical probability distributions, finding the density operator that maximizes the von Neumann entropy given the complete set of bipartite correlations, would solve the analogous problem for classical probability distributions. To study an efficient learning procedure, the hardness results of the classical problem need to be taken in account.

Given a finite set of classical random variables, the optimization problem of finding the probability distribution that maximizes the Shannon entropy given a general collection of bipartite marginals, can be stated as a graph inference problem. With the provided information, the classical system can be represented by a graph where the vertices represent the random variables and the edges represent the direct dependencies between them. Then, graphical models, such as Bayesian Networks and Markov random fields [2, 6], provide the learning techniques for the maximum entropy estimator. Inferring a general graphical structure is NP-hard [7], and so is finding an approximate solution [8]. The only structures for which a general efficient learning solution is known are trees, as even learning 2-polytrees is NP-hard [9]. Nevertheless, there

exists an efficient algorithm to obtain the *optimal tree*—the Chow–Liu algorithm [10]. Then, any set of random variables can be efficiently approximated by the probability distribution describing its most likely tree. It has been an open problem to find richer structures than tree Markov random fields that can be learned efficiently.

Motivated by these classical results, we look for an efficient learning procedure for the maximum von Neumann entropy estimator given a subset of bipartite marginals which is tree-structured. This means that, representing the joint quantum system by a graph where the vertices label the subsystems and where the edges represent the measured bipartite marginals, the resulting graph is a tree. Observe that the problem we are analyzing can be stated also as searching for a subclass of density operators for which the classical results of learning via graphical models can be extended.

The efficiency of the aforementioned learning techniques is mainly due to the factorization of the joint probability distribution that occurs when the conditional independence condition between the interested subparts holds. As we are going to see in detail, also when the system involves quantum correlations, the quantum generalization of conditional independence to quantum states results in an algebraic recovery of the joint in terms of the interested subparts. However, whereas the graphical structure of a classical system naturally encodes the conditional independence properties between them, the one involving quantum correlations does not. Further conditions, not explicit from the graphical structure, need in general to be verified. Many attempts have been done for developing appropriate generalizations of graphical models for density operators [11, 12], but none of them naturally encodes the required properties which result in a learning simplification without further conditions [13].

#### 1.4. Problem

We wish to find an *efficient procedure* for learning the density operator that maximizes the von Neumann entropy from a subset of (compatible, cf definition 1) tree-structured bipartite marginals.

**1.4.1. Results.** We consider the simplest nontrivial tree, i.e. a tripartite quantum system where two marginals are known (section 2.1). This analysis provides insight for the multipartite scenario (section 6).

*Tripartite case:* we provide an algebraic recovery procedure of the tripartite density operator given two bipartite marginals when they are compatible with a quantum Markov chain—definition 2. We find that there exists an algebraic recovery procedure, namely the Petz recovery map, for the maximum entropy estimator whenever there exists a quantum Markov chain in the compatibility set of the provided marginals—theorem 1. Indeed, quantum Markov chains represent the subset of tripartite density operators that strictly contains classical probability distributions, i.e., such that between the two non adjacent quantum states, the (quantum) conditional independence condition holds—corollary 1. Then, we provide a necessary and sufficient condition for efficiently verifying if the given couple of marginals is compatible with a quantum Markov chain—theorem 2. Moreover, we give a criterion for determining which couple of possible bipartite marginals, out of the three possible, provides *the best* maximum entropy estimator, meaning the one that minimizes the relative entropy distance with the *unknown* density operator. We prove the best estimator to be the one with minimum von Neumann entropy—theorem 5, which, if all the three couple of measured marginals are compatible with a quantum Markov chain, can be obtained by discarding the marginal with minimum quantum mutual information—theorem 6. Both the conditions of theorems 2 and 6 are algebraic, allowing the efficiency of the entire learning procedure and, possibly, an easier generalization to the multipartite case.

Furthermore, observing that the problem of entropy maximization can be stated as multiple-step minimum entropy updating (section 4), we give a further characterization of quantum Markov chains as the commutativity of a diagram of quantum Bayesian updating processes—theorem 4.

*Multipartite case:* We provide some preliminary results about a possible generalization of the notion of the procedure to the multipartite scenario. In particular, we prove that the global Markov condition, properly extended to density operators, is sufficient to have an efficient recovery of the maximum entropy estimator given a tree structured set of bipartite marginals and, additionally, it is sufficient to efficiently choose an optimal tree for the estimator—proposition 3.

## 2. Maximum entropy estimator

Let  $\mathcal{X} = \{X_1, \dots, X_n\}$ , with  $0 < n < \infty$ , be a set labelling the parts of a multipartite quantum system  $\mathcal{X}$ . The physical system  $\mathcal{X}$  can be described by an Hermitian, positive-semidefinite and trace one operator  $\rho_{\mathcal{X}}$ , namely a density operator in the Liouville space  $\mathcal{L}(\mathcal{H}_{\mathcal{X}})$ , where  $\mathcal{H}_{\mathcal{X}}$  is a separable Hilbert space on its subparts  $\mathcal{H}_{\mathcal{X}} := \otimes_{i=1}^n \mathcal{H}_{X_i}$ . We denote by  $d_{\mathcal{X}}$  the dimension of the Hilbert space  $\mathcal{H}_{\mathcal{X}}$ , which is always assumed to be finite in the whole manuscript. Running several times the same experiment, we can collect many copies of the unknown system; on each of them, we can perform a measurement, obtaining the set of expectation values  $\{\langle \Theta_i \rangle\}_{i \in I}$ , where  $\Theta_i$  are positive Hermitian operators acting on the full joint Hilbert space  $\mathcal{H}_{\mathcal{X}}$ .

**Proposition 1.** *The density operator  $\tilde{\rho} \in \mathcal{L}(\mathcal{H}_{\mathcal{X}})$  that maximizes the von Neumann entropy, denoted  $S(\rho)$ , is given by*

$$\tilde{\rho}_{\mathcal{X}} = \frac{1}{Z} \exp \left( \sum_{i \in I} \lambda_i \Theta_i \right), \quad \text{with } Z = \text{Tr} \left[ \exp \left( \sum_{i \in I} \lambda_i \Theta_i \right) \right] \quad (1)$$

and in which  $\{\lambda_i\}_{i \in I}$  are Lagrange multipliers which are obtained by solving the equations  $\text{Tr}[\rho \Theta_i] = \langle \Theta_i \rangle$ .

**Proof.** This follows from taking the variation of the function

$$S(\rho) - \sum_{i \in I} \lambda_i (\text{Tr}[\rho \Theta_i] - \langle \Theta_i \rangle). \quad (2)$$

By concavity of  $S$  the solution is a maximum point and it is unique. □

In our case, we are given a set of bipartite marginals  $\{\rho_{X_i X_j}\}$ , which can be probed by a complete set of observables in the associated bipartite Hilbert spaces. If we are given a Hilbert space  $\mathcal{H}_{\mathcal{X}}$ , then there exist a set of Hermitian operators  $\{\Lambda_j^{(X)}: j = 0, \dots, d_{\mathcal{X}}^2 - 1\}$  which are complete in the sense that any linear operator and, in particular any observable, can be written as a linear combination of the latter. This basis of operators can be chosen to be orthonormal with respect to the Hilbert–Schmidt inner product and, additionally, to be traceless, so that  $\{i\Lambda_j^{(X)}: j = 0, \dots, d_{\mathcal{X}}^2 - 1\}$  span a Lie algebra  $\mathfrak{su}(d_{\mathcal{X}})$ , where  $\Lambda_0^{(X)} := id_{\mathcal{X}}$  is the identity on  $\mathcal{H}_{\mathcal{X}}$  and it corresponds to the remaining generator of  $u(d_{\mathcal{X}})$ . For each  $i \in I$ , we denote by  $\{\Lambda_k^{(X_i)}: k = 0, \dots, d_{X_i}^2 - 1\}$  a chosen complete set of observables for  $X_i$ , as before. For each  $X_i X_j$ , the set  $\{\Lambda_k^{(X_i)} \otimes \Lambda_l^{(X_j)}: k = 0, \dots, d_{X_i}^2 - 1, l = 0, \dots, d_{X_j}^2 - 1\}$  forms a complete set of observables for the bipartite system  $X_i X_j$ . We extend the operators  $\Lambda_k^{(X_i)} \otimes \Lambda_l^{(X_j)}$

acting on  $\mathcal{H}_{X_i} \otimes \mathcal{H}_{X_j}$  in a natural way to act on the joint Hilbert space  $\mathcal{H}_{\mathcal{X}}$  by taking the tensor product with the identity on the relevant factors. By abuse of notation we denote by  $\Lambda_k^{(X_i)} \Lambda_l^{(X_j)}$  the extended operators. As a consequence of proposition 1, the density operator  $\tilde{\rho}_{X_1 \dots X_n}$  maximizing the von Neumann entropy assumes the form

$$\tilde{\rho}_{\mathcal{X}} = \frac{1}{Z} \exp \left( \sum_{i,j} \sum_{\substack{k=0 \\ k^2+l^2 \neq 0}}^{d_{X_i}^2-1} \sum_{l=0}^{d_{X_j}^2-1} \lambda_{kl}^{(X_i X_j)} \Lambda_k^{(X_i)} \Lambda_l^{(X_j)} \right), \quad (3)$$

where, as before,  $\{\lambda_{jk}^{(X_i X_j)}\}$  are the Lagrange multipliers of the optimization problem, constrained by the partial traces on the given marginals. The state of equation (3) exists because the set of density operators satisfying the given constraints is convex and non-empty. Indeed, *a priori*, the set of bipartite marginals is the output of a set of measurements performed on an existing quantum system described by an  $n$ -partite density operator. We are then assuming the given set of provided marginals to be *compatible*.

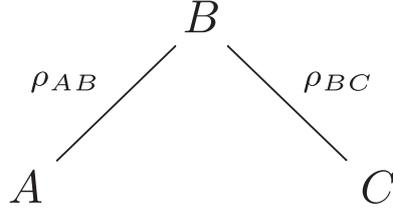
**Definition 1** (Compatibility and compatibility set). Consider a set of density operators  $\mathcal{C} = \{\rho_{\mathcal{Y}} \in \mathcal{L}(\mathcal{H}_{\mathcal{Y}})\}_{\mathcal{Y} \in \mathcal{K}}$ , where  $\mathcal{K}$  is a family of subsystems of  $\mathcal{X}$  which is a cover, i.e.,  $\bigcup_{\mathcal{Y} \in \mathcal{K}} \mathcal{Y} = \mathcal{X}$ . We say that  $\mathcal{C}$  is a compatible set of marginals if there exists at least one density operator  $\rho$  over the joint Hilbert space  $\mathcal{H}_{\mathcal{X}}$  such that  $\text{Tr}_{\bar{\mathcal{Y}}}(\rho) = \rho_{\mathcal{Y}}$  for all  $\rho_{\mathcal{Y}} \in \mathcal{C}$ . Here  $\text{Tr}_{\bar{\mathcal{Y}}}(\cdot)$  denotes the partial trace over the complementary factors of the Hilbert space  $\bar{\mathcal{Y}} = \mathcal{X} \setminus \mathcal{Y}$ .

Moreover, we denote by  $\text{Comp}(\mathcal{C})$  the set of density operators over  $\mathcal{H}_{\mathcal{X}}$  s.t.  $\text{Tr}_{\bar{\mathcal{Y}}}(\rho) = \rho_{\mathcal{Y}}$  for every  $\rho_{\mathcal{Y}} \in \mathcal{C}$ . We say for each  $\rho \in \text{Comp}(\mathcal{C})$  that  $\rho$  is compatible with  $\rho_{\mathcal{Y}}$ , for any  $\mathcal{Y} \in \mathcal{K}$ . Additionally, we also say that  $\rho$  is compatible with  $\mathcal{C}$ .

The problem of determining if a given set of density operators is compatible is known as *the quantum marginal problem* is QMA complete [14, 15]. This problem reduces to determining the maximum entropy estimator (compatible with the marginals), and therefore, the latter is QMA hard. In particular, a necessary condition, also sufficient for classical probability distributions, for a set of density operators defined on overlapping Hilbert spaces to be compatible is to coincide on their intersection. Formally, given  $\rho_{\mathcal{Y}_1}, \rho_{\mathcal{Y}_2} \in \mathcal{C}$  compatible, if  $\mathcal{Y}_1 \cap \mathcal{Y}_2 \neq \emptyset$  then  $\text{Tr}_{\bar{\mathcal{Y}_1 \cap \mathcal{Y}_2}}(\rho_{\mathcal{Y}_1}) = \text{Tr}_{\bar{\mathcal{Y}_1 \cap \mathcal{Y}_2}}(\rho_{\mathcal{Y}_2})$ .

Our goal, as stated in the introduction, is to have an efficient recovery of the maximum entropy estimator given a set of bipartite marginals. For that reason and taking into account the results known for the classical case (see the introduction), we restrict to the case where the set of bipartite marginals is tree structured. In the next subsection we consider the tripartite case in detail.

**Remark 1.** Also in the tripartite case, we are going to restrict the problem to trees where just two marginals out of the three possible are taken in account. This not only to gain some insight about the generalization to the multipartite scenario, but also since the recovery problem for a tripartite probability distribution given all the three possible bipartite marginals is open [16–18]. Moreover, moving to the quantum scenario, also the compatibility problem for just a couple of overlapping marginals is open [19, 20]. We are then going to assume the set of the two given marginal density operators compatible.



**Figure 1.** Graph associated to system of bipartite marginals  $\{\rho_{AB}, \rho_{BC}\}$ . Each vertex represents a quantum system to which is associated a density operator by partial tracing the marginals. The edges correspond to direct correlations between the vertices, associated to mixed states defined over the edge Hilbert space.

### 2.1. The tripartite case

Let us denote  $\mathcal{X} = \{A, B, C\}$ , and assume we are given access to marginals  $\{\rho_{AB}, \rho_{BC}\}$ . See the associated graph in figure 1.

In equation (3), we denote by  $\lambda_{kl} := \lambda_{kl}^{(AB)}$  and  $\eta_{kl} := \lambda_{kl}^{(BC)}$ . We then have,

$$\tilde{\rho}_{ABC} = \frac{1}{Z} \exp \left( \sum_{\substack{k=0 \\ k^2+l^2 \neq 0}}^{d_A^2-1} \sum_{l=0}^{d_B^2-1} \lambda_{kl} \Lambda_k^A \Lambda_l^B + \sum_{\substack{k=0 \\ k^2+l^2 \neq 0}}^{d_B^2-1} \sum_{l=0}^{d_C^2-1} \eta_{kl} \Lambda_k^B \Lambda_l^C \right). \quad (4)$$

Because for each  $X_i X_j$ , the set  $\{\Lambda_k^{(X_i)} \otimes \Lambda_l^{(X_j)} : k = 0, \dots, d_i^2 - 1, l = 0, \dots, d_j^2 - 1\}$  forms a complete set of observables for the bipartite system  $X_i X_j$ , we can write,

$$\begin{aligned} \rho_{AB} &= \frac{1}{d_A d_B} id_{AB} + \sum_{\substack{k=0 \\ k^2+l^2 \neq 0}}^{d_A^2-1} \sum_{l=0}^{d_B^2-1} \alpha_{kl} \Lambda_k^{(A)} \Lambda_l^{(B)}, \\ \rho_{BC} &= \frac{1}{d_B d_C} id_{BC} + \sum_{\substack{k=0 \\ k^2+l^2 \neq 0}}^{d_B^2-1} \sum_{l=0}^{d_C^2-1} \beta_{kl} \Lambda_k^{(B)} \Lambda_l^{(C)}, \end{aligned} \quad (5)$$

where  $\alpha_{kl}, \beta_{kl} \in \mathbb{R}$ . Additionally, the constraints imposed by the marginals can be cast in the form

$$\begin{aligned} \text{Tr} \left[ \Lambda_k^{(A)} \Lambda_l^{(B)} \left( \tilde{\rho}_{ABC} - \rho_{AB} \otimes \frac{id_C}{d_C} \right) \right] &= 0, \\ \text{Tr} \left[ \Lambda_k^{(B)} \Lambda_l^{(C)} \left( \tilde{\rho}_{ABC} - \frac{id_A}{d_A} \otimes \rho_{BC} \right) \right] &= 0, \end{aligned} \quad (6)$$

where  $k$  and  $l$  range in the dimensions of the space of linear operators of the appropriate subsystem.

From the partition function  $Z$ , one can obtain, by differentiation, a system of equations which allow to solve for the Lagrange multipliers in terms of the known parameters  $\{\alpha_{kl}\}$  and  $\{\beta_{kl}\}$ . This is a standard procedure in statistical mechanics. Determining if the optimization problem described above has an algebraic solution is not trivial: the Lie operators in the

exponent of equation (4) in general do not commute, which makes the analysis of the constraints imposed in equation (6) mathematically hard to manage. Many techniques for overcoming similar problems are object of study, see for example reference [21] for trace inequalities or references [22, 23] for operator inequalities. Instead, here we focus our attention on a well-behaved subset of density operators—quantum Markov chains [24]. In the next section, we recall the definition of a quantum Markov chain and prove a necessary and sufficient condition for a given pair of bipartite marginals to be compatible with a quantum Markov chain.

### 3. Marginals' compatibility with quantum Markov chains

**Definition 2.** (Quantum Markov chain [21]). A tripartite state  $\rho_{ABC}$  over  $\mathcal{H}_{ABC}$  is called a quantum Markov chain (QMC) in order  $A - B - C$  if there exists a recovery map  $\mathcal{R}_{B \rightarrow BC} : \mathcal{L}(\mathcal{H}_B) \rightarrow \mathcal{L}(\mathcal{H}_{BC})$  such that

$$\rho_{ABC} = (\mathcal{I}_A \otimes \mathcal{R}_{B \rightarrow BC})(\rho_{AB}), \tag{7}$$

where a recovery map is an arbitrary trace-preserving completely positive (CPTP) map, see reference [2], and  $\mathcal{I}_A$  denotes the identity map on  $\mathcal{L}(\mathcal{H}_A)$ .

A QMC can be characterized from an information-theoretical point of view due to the following result.

**Proposition 2.** [25] *A tripartite state  $\rho_{ABC}$  is a QMC in the order  $A - B - C$  if and only if  $I_\rho(A : C|B) = 0$ , where  $I_\rho(A : C|B) := S(\rho_{AB}) + S(\rho_{BC}) - S(\rho_B) - S(\rho_{ABC})$  is the quantum conditional mutual information.*

We recall that in case of classical random variables with finite domains, when we are given two bipartite marginals  $p(A, B)$  and  $p(B, C)$ , the compatibility condition  $\sum_a p(A = a, B = b) = \sum_c p(B = b, C = c)$  is necessary and sufficient for  $A$  and  $C$  to be independent conditioned on  $B$ , i.e.  $I(A : C|B) = 0$ . Then, the set of quantum Markov chains includes the set of classical tripartite probability distributions.

Proposition 2 is equivalent to the statement: a QMC  $A - B - C$  is a tripartite quantum state for which the strong subadditivity of the von Neumann entropy,

$$S(\rho_{AB}) + S(\rho_{BC}) \geq S(\rho_B) + S(\rho_{ABC}), \tag{8}$$

holds with equality [26], which trivially results in the fact that a QMC  $A - B - C$  maximizes the von Neumann entropy given its two bipartite marginals  $AB$  and  $BC$ . Furthermore, the quantum systems  $A$  and  $C$  are said to be *quantum conditionally independent* given the quantum system  $B$ .

A QMC always admits as a recovery channel the *rotated Petz recovery map* [27, 28]:

$$\mathcal{P}_{B \rightarrow BC}^t(X) := \rho_{BC}^{\frac{1+i}{2}} \left( \rho_B^{-\frac{1+i}{2}} X \rho_B^{-\frac{1-i}{2}} \right) \rho_{BC}^{\frac{1-i}{2}}, \quad \text{for any } X \in \mathcal{L}(\mathcal{H}_B), t \in \mathbb{R}. \tag{9}$$

In particular, for  $t = 0$ , the map is known as the *Petz recovery map or transpose map*:

$$\mathcal{P}_{B \rightarrow BC}(X) := \rho_{BC}^{\frac{1}{2}} \rho_B^{-\frac{1}{2}} X \rho_B^{-\frac{1}{2}} \rho_{BC}^{\frac{1}{2}}, \quad \text{for any } X \in \mathcal{L}(\mathcal{H}_B). \quad (10)$$

Note that in the previous formulas,  $\rho_B$  and  $X$  are understood as elements of  $\mathcal{L}(\mathcal{H}_{BC})$  by extending it in the natural way, i.e.,  $\rho_B \otimes id_C$ . In reference [21] it is shown that the Petz recovery map is indeed a recovery map, i.e., a CPTP map.

Our first result comes as a natural corollary of the previous stated results.

**Theorem 1.** *Given bipartite marginals  $\{\rho_{AB}, \rho_{BC}\}$  compatible with a QMC in the order  $A-B-C$ , say  $\rho_{ABC}$ , then the solution of the maximum entropy estimator  $\tilde{\rho}_{ABC}$  is precisely equal to  $\rho_{ABC}$ . Moreover,  $\tilde{\rho}_{ABC}$  can be algebraically recovered via the Petz map  $\mathcal{P}_{B \rightarrow BC}(\cdot)$ , concretely:*

$$\tilde{\rho}_{ABC} = \rho_{BC}^{\frac{1}{2}} \rho_B^{-\frac{1}{2}} \rho_{AB} \rho_B^{-\frac{1}{2}} \rho_{BC}^{\frac{1}{2}}. \quad (11)$$

**Remark 2.** We can also recover a tripartite density operator from  $\rho_{BC}$  through  $\mathcal{P}_{B \rightarrow AB}(\cdot)$ :

$$\rho_{AB}^{\frac{1}{2}} \rho_B^{-\frac{1}{2}} \rho_{BC} \rho_B^{-\frac{1}{2}} \rho_{AB}^{\frac{1}{2}}, \quad (12)$$

and by uniqueness, because the von Neumann entropy is concave, they are the same.

In alternative to the Petz recovery map, in presence of marginals compatible with a QMC, we can solve efficiently the associated optimization problem, i.e., determine the Lagrange multipliers in equation (4) due to the following result by Petz [29]:

**Lemma 1.** [29] *Assume that  $\rho_{ABC}$  is invertible. Then the equality holds in the strong subadditivity inequality (SSA) if and only if  $\log \rho_{ABC} - \log \rho_{AB} = \log \rho_{BC} - \log \rho_B$ .*

Our next result, provides a necessary and sufficient condition for the given marginals  $\{\rho_{AB}, \rho_{BC}\}$  to be compatible with a QMC. The condition is moreover algebraic, resulting in being easily verifiable.

**Theorem 2.** *Two bipartite marginals  $\{\rho_{AB}, \rho_{BC}\}$  are compatible with a QMC in  $\mathcal{L}(\mathcal{H}_{ABC})$  in the order  $A - B - C$  if and only if  $\text{Tr}_A(\rho_{AB}) = \text{Tr}_C(\rho_{BC})$  and the operator  $\Theta_{ABC} = \rho_{BC}^{\frac{1}{2}} \rho_B^{-\frac{1}{2}} \rho_{AB}^{\frac{1}{2}}$  is normal.*

Before giving the proof of theorem 2, we will need one further result.

**Theorem 3.** ([25]). *A tripartite density operator  $\rho_{ABC} \in \mathcal{L}(\mathcal{H}_{ABC})$  satisfies the SSA with equality, i.e.,  $I_\rho(A : C|B) = 0$ , if and only if there exists a decomposition of the Hilbert space  $\mathcal{H}_B$  of the form*

$$\mathcal{H}_B = \bigoplus_j \mathcal{H}_{B_j^L} \otimes \mathcal{H}_{B_j^R}, \quad \text{such that } \rho_{ABC} = \bigoplus_j p_j \rho_{AB_j^L} \otimes \rho_{B_j^R C}, \quad (13)$$

with  $p_j \geq 0$ , for all  $j$ ,  $\sum_j p_j = 1$  and the states  $\rho_{AB_j^L} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_{B_j^L})$  and  $\rho_{B_j^R C} \in \mathcal{L}(\mathcal{H}_{B_j^R} \otimes \mathcal{H}_C)$ .

**Proof.** ( $\Rightarrow$ ) Using compatibility, the following decompositions are a direct consequence of theorem 3

$$\rho_{AB} = \bigoplus_j p_j \rho_{AB_j^L} \otimes \rho_{B_j^R}, \quad \text{with } \rho_{B_j^R} = \text{Tr}_C(\rho_{B_j^R C}), \quad (14)$$

$$\rho_{BC} = \bigoplus_j p_j \rho_{B_j^L} \otimes \rho_{B_j^R C}, \quad \text{with } \rho_{B_j^L} = \text{Tr}_A(\rho_{AB_j^L}), \quad (15)$$

$$\rho_B = \bigoplus_j p_j \rho_{B_j^L} \otimes \rho_{B_j^R}. \quad (16)$$

It is now trivial to see that  $\text{Tr}_A(\rho_{AB}) = \text{Tr}_C(\rho_{BC})$ , as required. It remains to check that  $\Theta_{ABC}$  is normal. Notice that, from the above decompositions, we have

$$\begin{aligned} \Theta_{ABC} &= \rho_{BC}^{\frac{1}{2}} \rho_B^{-\frac{1}{2}} \rho_{AB}^{\frac{1}{2}} \\ &= \left( \bigoplus_j p_j \rho_{B_j^L} \otimes \rho_{B_j^R C} \right)^{\frac{1}{2}} \left( \bigoplus_k p_k \rho_{B_k^L} \otimes \rho_{B_k^R} \right)^{-\frac{1}{2}} \left( \bigoplus_l p_l \rho_{AB_l^L} \otimes \rho_{B_l^R} \right)^{\frac{1}{2}} \\ &= \bigoplus_j p_j \rho_{AB_j^L}^{\frac{1}{2}} \otimes \rho_{B_j^R}^{\frac{1}{2}}. \end{aligned} \quad (17)$$

and also

$$\Theta_{ABC}^\dagger = \left( \rho_{BC}^{\frac{1}{2}} \rho_B^{-\frac{1}{2}} \rho_{AB}^{\frac{1}{2}} \right)^\dagger = \rho_{AB}^{\frac{1}{2}} \rho_B^{-\frac{1}{2}} \rho_{BC}^{\frac{1}{2}} = \bigoplus_j p_j \rho_{AB_j^L}^{\frac{1}{2}} \otimes \rho_{B_j^R}^{\frac{1}{2}} = \Theta_{ABC}, \quad (18)$$

hence we conclude that  $\Theta_{ABC}$  is self-adjoint, therefore normal.

( $\Leftarrow$ ) Consider the operator

$$\varrho_{ABC} = \Theta_{ABC} \Theta_{ABC}^\dagger = \rho_{BC}^{\frac{1}{2}} \rho_B^{-\frac{1}{2}} \rho_{AB} \rho_B^{-\frac{1}{2}} \rho_{BC}^{\frac{1}{2}}. \quad (19)$$

We have:

- (a)  $\varrho_{ABC}$  is a density operator over  $\mathcal{H}_{ABC}$ . This can be seen from the fact that the map

$$\begin{aligned} \mathcal{P}_{B \rightarrow BC} : \mathcal{L}(\mathcal{H}_B) &\rightarrow \mathcal{L}(\mathcal{H}_{BC}) \\ \sigma &\mapsto \rho_{BC}^{\frac{1}{2}} \rho_B^{-\frac{1}{2}} \sigma \rho_B^{-\frac{1}{2}} \rho_{BC}^{\frac{1}{2}} \end{aligned} \quad (20)$$

is a CPTP map, as mentioned above, and it extends to a CPTP map  $\mathcal{I}_A \otimes \mathcal{P}_{B \rightarrow BC}$  which yields, when applied to  $\rho_{AB}$ ,  $\varrho_{ABC}$ .

- (b)  $\varrho_{ABC}$  is compatible with  $\{\rho_{AB}, \rho_{BC}\}$ . Indeed, we have,

$$\text{Tr}_A(\varrho_{ABC}) = \rho_{BC}^{\frac{1}{2}} \rho_B^{-\frac{1}{2}} \text{Tr}_A(\rho_{AB}) \rho_B^{-\frac{1}{2}} \rho_{BC}^{\frac{1}{2}} = \rho_{BC}^{\frac{1}{2}} \rho_B^{-\frac{1}{2}} \rho_B \rho_B^{-\frac{1}{2}} \rho_{BC}^{\frac{1}{2}} = \rho_{BC}, \quad (21)$$

where we have used  $\text{Tr}_A(\rho_{AB}) = \rho_B$ . Moreover, since  $\Theta_{ABC}$  is normal and  $\text{Tr}_C(\rho_{BC}) = \rho_B$ , we have,

$$\begin{aligned} \text{Tr}_C(\varrho_{ABC}) &= \text{Tr}_C(\Theta_{ABC} \Theta_{ABC}^\dagger) = \text{Tr}_C(\Theta_{ABC}^\dagger \Theta_{ABC}) \\ &= \text{Tr}_C \left( \rho_{AB}^{\frac{1}{2}} \rho_B^{-\frac{1}{2}} \rho_{BC} \rho_B^{-\frac{1}{2}} \rho_{AB}^{\frac{1}{2}} \right) = \rho_{AB}. \end{aligned} \quad (22)$$

- (c) By (a) and (b) and definition 2,  $\varrho_{ABC}$  is a QMC. □

From theorem 2, we have the following corollary:

**Corollary 1.** *The space of pairs of bipartite marginals  $\{\rho_{AB}, \rho_{BC}\}$  which is compatible with a QMC is strictly included in the space of pairs of bipartite marginals  $\{\rho_{AB}, \rho_{BC}\}$  which are compatible.*

The last statement seems rather intuitive, but cannot be seen immediately due to the fact that the quantum marginal problem, also for two overlapping marginals, is open. To provide a counter-example, we numerically generated a random 3-qubit density matrix, from which we obtained two compatible marginals via partial trace. Then, we check the compatibility condition with a quantum Markov chain in theorem 2, given the two marginals. We repeat the process until the latter conditions fails.

#### 4. Quantum Bayesian updating

The *principle of minimum discrimination information* [30, 31], is intrinsically related to the maximum entropy principle and it is at the base of inferential updating of probability distributions. Given a *prior* joint probability distribution  $q_{\mathbf{X}}$  describing the random variables  $\mathbf{X} := \{X_1, \dots, X_n\}$  and an additional set of new information about the system, the most unbiased *posterior*  $p_{\mathbf{X}}$  corresponds to the one that minimizes the Kullback–Leibler divergence with the prior distribution under the constraints given by the additional information. If the prior is the most un-informative one, i.e. the uniform distribution, then the problem is equivalent to finding the probability distribution that maximizes the Shannon entropy with the given constraints. In reference [32], the choice of the Kullback–Leibler divergence as a functional for inferential updating is explained in light of the maximum entropy principle and some designed criteria.

The learning problem for multipartite quantum states we are proposing here can be stated a multiple-step inferential updating procedure, where starting from the uniform distribution, i.e. the maximally mixed state, the marginals are the additional information set available at each step.

In reference [33], a generalization to the quantum realm of the Bayesian updating procedure is proposed. The Kullback–Leibler divergence is generalized by the von Neumann relative entropy, not by mere replacement, but deriving it from the same designed criteria and the maximum (von Neumann) entropy principle.

**Definition 3** (Quantum Bayesian updating). Given a quantum system  $\mathcal{X} = \{X_1, \dots, X_n\}$  described by the joint Hilbert space  $\mathcal{H}_{\mathcal{X}} = \bigotimes_{i=1}^n \mathcal{H}_{X_i}$ , let  $\varrho_{\mathcal{X}} \in \mathcal{L}(\mathcal{H}_{\mathcal{X}})$  be our a prior knowledge about it, and  $\{\langle \Theta_j \rangle\}_{j \in J}$  a set of expectation values corresponding to a set of observables  $\{\Theta_j\}_{j \in J}$  be the additional set of information. Then, via the principle of minimum updating, the posterior density operator assumes the form

$$\rho_{\mathcal{X}} = \exp \left( \lambda id_{\mathcal{X}} + \sum_{j \in J} \alpha_j \Theta_j + \log \varrho_{\mathcal{X}} \right), \tag{23}$$

where the Lagrange multipliers  $\lambda$  and  $\{\alpha_j\}_{j \in J}$  are determined by the constraints  $\text{Tr}(\rho) = 1$  and  $\text{Tr}(\rho \Theta_j) = \langle \Theta_j \rangle$ ,  $j \in J$ . We indicate the updating process of equation (23) by the diagram

$$\varrho_{\mathcal{X}} \xrightarrow{\{\Theta_j, \langle \Theta_j \rangle\}_{j \in J}} \rho_{\mathcal{X}}. \tag{24}$$

Observing that  $S(\rho_X) = -S(\rho_X \| \frac{id_X}{d_X})$ , it follows that the maximum entropy density operator of equation (4) is equivalent to the output of the process:

$$\frac{id_{ABC}}{d_{ABC}} \xrightarrow{\{\rho_{AB}, \rho_{BC}\}} \tilde{\rho}_{ABC}. \tag{25}$$

We could split the updating process into two processes, i.e:

$$\frac{id_{ABC}}{d_{ABC}} \xrightarrow{\rho_{AB}} \tilde{\sigma}'_{ABC} \xrightarrow{\rho_{BC}} \tilde{\sigma}_{ABC}, \tag{26}$$

or

$$\frac{id_{ABC}}{d_{ABC}} \xrightarrow{\rho_{BC}} \tilde{\varrho}'_{ABC} \xrightarrow{\rho_{AB}} \tilde{\varrho}_{ABC}, \tag{27}$$

and, classically, one can show, all these updating processes, equations (25)–(27), are equivalent. However, in the quantum scenario, this is not the case due to the non-commutative nature of the observables involved. In reference [34], it is shown that, in contrast to the classical case, where the conditional mutual information is always a measure of the distance of a tripartite quantum state to a general Markov chain, in the quantum case it is not. In particular,

$$I_\rho(A : C|B) \geq S(\rho_{ABC} \| \mathcal{R}_{B \rightarrow BC}(\rho_{AB})), \tag{28}$$

where  $\mathcal{R}_{B \rightarrow BC}(\cdot)$  is an arbitrary recovery map. Recently, tighter versions of the above inequality have been the object of study by several authors [35]. The equality holds when  $I_\rho(A : C|B) = 0$ , i.e., for a QMC. The fact that the three above processes, equations (25)–(27), lead to different density operators corresponds to the non-commutativity of a diagram. Moreover, in the next theorem we provide necessary and sufficient conditions for the diagram to commute, which provides a characterization of QMCs in terms of QBU.

**Theorem 4.** *Given a set of marginals  $\{\rho_{AB}, \rho_{BC}\}$ , we then have a commutative diagram of quantum Bayesian updating processes:*

$$\begin{array}{ccc} \frac{id_{ABC}}{d_{ABC}} & \xrightarrow{\rho_{AB}} & \tilde{\sigma}'_{ABC} \\ \rho_{BC} \downarrow & & \downarrow \rho_{BC} \\ \tilde{\varrho}'_{ABC} & \xrightarrow{\rho_{AB}} & \tilde{\rho}_{ABC} \end{array}$$

*if and only if they are compatible with a QMC in the order  $A - B - C$ .*

To provide a proof of this theorem we will need the following lemma.

**Lemma 2.** [22, 23] *Given a tripartite quantum state  $\rho_{ABC} \in \mathcal{L}(\mathcal{H}_{ABC})$  compatible with  $\{\rho_{AB}, \rho_{BC}\}$ , the following operator-inequalities hold:*

- (a)  $\text{Tr}_{AB}[\rho_{ABC}(\log \rho_{ABC} - \log \rho_{AB} + \log \rho_B - \log \rho_{BC})] \succeq 0$ ,
- (b)  $\text{Tr}_{AB}[\rho_{AB}(\log \rho_{AB} - \log \rho_{ABC} - \log \rho_B + \log \rho_{BC})] \succeq 0$ ,

*where the equality holds if and only if  $\rho_{ABC}$  is a QMC in the order  $A - B - C$ . Above,  $\succeq 0$  stands for positive semidefinite.*

**Proof.** ( $\Leftarrow$ ) We will now show that if  $\{\rho_{AB}, \rho_{BC}\}$  is compatible with a QMC  $\tilde{\rho}_{ABC}$  in the order  $A - B - C$ , then  $\tilde{\sigma}_{ABC} = \tilde{\varrho}_{ABC} = \tilde{\rho}_{ABC}$ .

Using the fact that  $S(\rho) = -S(\rho \| \frac{id_{ABC}}{d_{ABC}})$ , the first step of the updating process is obtained by maximizing the von Neumann entropy, yielding

$$\tilde{\sigma}'_{ABC} = \rho_{AB} \otimes \frac{id_C}{d_C} \quad \text{and} \quad \tilde{\varrho}'_{ABC} = \frac{id_A}{d_A} \otimes \rho_{BC}. \quad (29)$$

Recalling that  $\text{Comp}(\rho_{BC}) := \{\rho_{ABC} \in \mathcal{L}(\mathcal{H}_{ABC}) : \text{Tr}_A(\rho_{ABC}) = \rho_{BC}\}$ . Then

$$\tilde{\sigma}_{ABC} = \arg \min_{\rho_{ABC} \in \text{Comp}(\rho_{BC})} S(\rho_{ABC} \| \tilde{\sigma}'_{ABC}). \quad (30)$$

It then follows that, since  $\tilde{\rho}_{ABC} \in \text{Comp}(\rho_{BC})$ ,

$$S\left(\tilde{\rho}_{ABC} \| \rho_{AB} \otimes \frac{id_C}{d_C}\right) \geq S\left(\tilde{\sigma}_{ABC} \| \rho_{AB} \otimes \frac{id_C}{d_C}\right). \quad (31)$$

Analogously,  $\text{Comp}(\rho_{AB}) = \{\rho_{ABC} \in \mathcal{L}(\mathcal{H}_{ABC}) : \text{Tr}_C(\rho_{ABC}) = \rho_{AB}\}$ . Then

$$\tilde{\varrho}_{ABC} = \arg \min_{\rho_{ABC} \in \text{Comp}(\rho_{AB})} S(\rho_{ABC} \| \tilde{\varrho}'_{ABC}). \quad (32)$$

It then follows that, since  $\tilde{\rho}_{ABC} \in \text{Comp}(\rho_{AB})$ ,

$$S\left(\tilde{\rho}_{ABC} \| \frac{id_A}{d_A} \otimes \rho_{BC}\right) \geq S\left(\tilde{\varrho}_{ABC} \| \frac{id_A}{d_A} \otimes \rho_{BC}\right). \quad (33)$$

The SSA inequality can be obtained as a special case as the contractibility property of the quantum relative entropy under CPTP maps [36, 37]:

$$S(\rho \| \sigma) \geq S(\Phi(\rho) \| \Phi(\sigma)), \quad \text{for all } \rho, \sigma \text{ and } \Phi \text{ a CPT map.} \quad (34)$$

Indeed, set  $\rho = \rho_{ABC} \in \text{Comp}(\rho_{AB}) \cap \text{Comp}(\rho_{BC})$ , then:

(a) For  $\sigma = \rho_{AB} \otimes \frac{id_C}{d_C}$  and  $\Phi(\cdot) = \text{Tr}_A(\cdot)$ , we get

$$S\left(\rho_{ABC} \| \rho_{AB} \otimes \frac{id_C}{d_C}\right) \geq S\left(\rho_{BC} \| \rho_B \otimes \frac{id_C}{d_C}\right). \quad (35)$$

(b) For  $\sigma = \frac{id_A}{d_A} \otimes \rho_{BC}$  and  $\Phi(\cdot) = \text{Tr}_C(\cdot)$ , we get

$$S\left(\rho_{ABC} \| \frac{id_A}{d_A} \otimes \rho_{BC}\right) \geq S\left(\rho_{AB} \| \frac{id_A}{d_A} \otimes \rho_B\right). \quad (36)$$

In both equations (35) and (36) the equality holds for  $\rho_{ABC} = \tilde{\rho}_{ABC}$ . Using the inequalities (31) and (35) with  $\rho_{ABC} = \tilde{\sigma}_{ABC}$  and the equation (35) with  $\rho_{ABC} = \tilde{\rho}_{ABC}$ ,

$$\begin{aligned}
 S\left(\tilde{\rho}_{ABC}\|\rho_{AB}\otimes\frac{id_C}{d_C}\right) &\geq S\left(\tilde{\sigma}_{ABC}\|\rho_{AB}\otimes\frac{id_C}{d_C}\right) \\
 &\geq S\left(\rho_{BC}\|\rho_B\otimes\frac{id_C}{d_C}\right) = S\left(\tilde{\rho}_{ABC}\|\rho_{AB}\otimes\frac{id_C}{d_C}\right). \quad (37)
 \end{aligned}$$

It then follows that  $\tilde{\rho}_{ABC} = \tilde{\sigma}_{ABC}$ .

Analogously, using the inequalities (33) and (36) with  $\rho_{ABC} = \tilde{\varrho}_{ABC}$  and the equation (36) with  $\rho_{ABC} = \tilde{\rho}_{ABC}$ , we get  $\tilde{\rho}_{ABC} = \tilde{\varrho}_{ABC}$ .

( $\Rightarrow$ ) We will now show that if  $\tilde{\sigma}_{ABC} = \tilde{\varrho}_{ABC}$  then  $\{\rho_{AB}, \rho_{BC}\}$  is compatible with a QMC  $\tilde{\rho}_{ABC}$  in the order  $A - B - C$  and  $\tilde{\rho}_{ABC} = \tilde{\sigma}_{ABC} = \tilde{\varrho}_{ABC}$ .

To provide explicit representations of  $\tilde{\sigma}_{ABC}$  and  $\tilde{\varrho}_{ABC}$  as exponentials of some operators we write, using invertibility,

$$\begin{aligned}
 \rho_{AB} &= \exp\left(\sum_{i=0}^{d_A^2-1}\sum_{j=0}^{d_B^2-1}A_{ij}\Lambda_i^{(A)}\otimes\Lambda_j^{(B)}\right), \\
 \rho_{BC} &= \exp\left(\sum_{i=0}^{d_B^2-1}\sum_{j=0}^{d_C^2-1}C_{ij}\Lambda_i^{(B)}\otimes\Lambda_j^{(C)}\right), \quad (38)
 \end{aligned}$$

for  $\{A_{ij}\}$  and  $\{B_{ij}\}$  real coefficients. For convenience, we also define the partition functions

$$Z_{AB} = \exp(-A_{00}) \quad \text{and} \quad Z_{BC} = \exp(-C_{00}). \quad (39)$$

We then have, using definition 3,

$$\tilde{\sigma}_{ABC} = \exp\left(\sum_{i=0}^{d_B^2-1}\sum_{j=0}^{d_C^2-1}\eta_{ij}\Lambda_i^{(B)}\Lambda_j^{(C)} + \log \tilde{\sigma}'_{ABC}\right). \quad (40)$$

where  $\{\eta_{ij}\}$  are the Lagrange multipliers in the definition. After some algebra, we can write

$$\begin{aligned}
 \log \tilde{\sigma}_{ABC} &= (\eta_{00} - \log Z_{BC} - \log d_C) id_{ABC} + \sum_{i=0}^{d_B^2-1}\sum_{j=1}^{d_C^2-1}\eta_{ij}\Lambda_i^{(B)}\Lambda_j^{(C)} \\
 &\quad + \sum_{i=1}^{d_A^2-1}\sum_{j=0}^{d_B^2-1}A_{ij}\Lambda_i^{(A)}\Lambda_j^{(B)} + \sum_{i=1}^{d_B^2-1}(A_{0i} + \eta_{i0})\Lambda_i^{(B)}. \quad (41)
 \end{aligned}$$

Similarly,

$$\tilde{\varrho}_{ABC} = \exp\left(\sum_{i=0}^{d_A^2-1}\sum_{j=0}^{d_B^2-1}\lambda_{ij}\Lambda_i^{(A)}\Lambda_j^{(B)} + \log \tilde{\varrho}'_{ABC}\right). \quad (42)$$

where  $\{\lambda_{ij}\}$  are the Lagrange multipliers in the definition. After some algebra, we can write

$$\begin{aligned} \log \tilde{\varrho}_{ABC} &= (\lambda_{00} - \log Z_{AB} - \log d_A) id_{ABC} + \sum_{i=1}^{d_A^2-1} \sum_{j=0}^{d_B^2-1} \lambda_{ij} \Lambda_i^{(A)} \Lambda_j^{(B)} \\ &+ \sum_{i=0}^{d_B^2-1} \sum_{j=1}^{d_C^2-1} C_{ij} \Lambda_i^{(B)} \Lambda_j^{(C)} + \sum_{i=1}^{d_B^2-1} (C_{i0} + \lambda_{0i}) \Lambda_i^{(B)}. \end{aligned} \quad (43)$$

Since  $\tilde{\sigma}_{ABC} = \tilde{\varrho}_{ABC}$ , by hypothesis, it follows that  $\log \tilde{\sigma}_{ABC} = \log \tilde{\varrho}_{ABC}$  and, hence,

$$\eta_{00} - \log Z_{BC} - \log d_C = \lambda_{00} - \log Z_{AB} - \log d_A, \quad (44)$$

$$A_{ij} = \lambda_{ij}, \quad \text{for } i = 1, \dots, d_A^2 - 1 \quad \text{and} \quad j = 0, \dots, d_B^2 - 1, \quad (45)$$

$$C_{ij} = \eta_{ij}, \quad \text{for } i = 0, \dots, d_B^2 - 1 \quad \text{and} \quad j = 1, \dots, d_C^2 - 1, \quad (46)$$

$$A_{0i} + \eta_{i0} = C_{i0} + \lambda_{0i}, \quad \text{for } i = 1, \dots, d_B^2 - 1. \quad (47)$$

Observing that,

$$\sum_{i=1}^{d_A^2-1} \sum_{j=0}^{d_B^2-1} \lambda_{ij} \Lambda_i^{(A)} \Lambda_j^{(B)} = \log \rho_{AB} - \log Z_{AB} id_{AB} - \sum_{i=1}^{d_B^2-1} A_{0i} \Lambda_i^{(B)}, \quad (48)$$

$$\sum_{i=0}^{d_B^2-1} \sum_{j=1}^{d_C^2-1} \eta_{ij} \Lambda_i^{(B)} \Lambda_j^{(C)} = \log \rho_{BC} - \log Z_{BC} id_{BC} - \sum_{i=1}^{d_B^2-1} C_{i0} \Lambda_i^{(B)}, \quad (49)$$

it follows that the two reconstructed states can be written as

$$\log \tilde{\sigma}_{ABC} = \log \rho_{AB} + \log \rho_{BC} + (\eta_{00} - \log Z_{BC} - \log d_C) id_{ABC} + \sum_{i=1}^{d_B^2-1} (\eta_{i0} - C_{i0}) \Lambda_i^{(B)}, \quad (50)$$

$$\log \tilde{\varrho}_{ABC} = \log \rho_{AB} + \log \rho_{BC} + (\lambda_{00} - \log Z_{AB} - \log d_A) id_{ABC} + \sum_{i=1}^{d_B^2-1} (\lambda_{0i} - A_{0i}) \Lambda_i^{(B)}. \quad (51)$$

We now define the Hermitian operator  $\theta_B$  by the following two equivalent formulas (consequence of  $\log \tilde{\sigma}_{ABC} = \log \tilde{\varrho}_{ABC}$ ):

$$\begin{aligned} \log \theta_B &:= (\eta_{00} - \log Z_{BC} - \log d_C) id_B + \sum_{i=1}^{d_B^2-1} (\eta_{i0} - C_{i0}) \Lambda_i^{(B)} \\ &= (\lambda_{00} - \log Z_{AB} - \log d_A) id_B + \sum_{i=1}^{d_B^2-1} (\lambda_{0i} - A_{0i}) \Lambda_i^{(B)}. \end{aligned} \quad (52)$$

Observe that  $\theta_B$  can, as usual, be extended by tensoring with the relevant identity maps to the whole Hilbert space  $\mathcal{H}_{ABC}$ . We conclude that  $\tilde{\sigma}_{ABC} = \tilde{\varrho}_{ABC}$  implies the existence of an Hermitian  $\theta_B$  such that

$$\log \tilde{\sigma}_{ABC} = \log \rho_{AB} + \log \rho_{BC} + \log \theta_B = \log \tilde{\varrho}_{ABC}. \quad (53)$$

Plugging in  $\rho_{ABC} = \tilde{\sigma}_{ABC}$  and its logarithm as in equation (53), in lemma 2, we get:

- (a)  $\text{Tr}_B[\rho_{BC}(\log \rho_B + \log \theta_B)] \succeq 0$ ,
- (b)  $-\text{Tr}_B[\rho_B(\log \rho_B + \log \theta_B)]id_C \succeq 0$ .

The second inequality is equivalent to  $\text{Tr}_B[\rho_B(\log \rho_B + \log \theta_B)] \leq 0$ . By tracing over  $C$  the first inequality we get  $\text{Tr}_B[\rho_B(\log \rho_B + \log \theta_B)] \geq 0$ . Therefore,

$$\text{Tr}_B[\rho_B(\log \rho_B + \log \theta_B)] = 0. \quad (54)$$

From this equation, observing that  $\rho_B = \text{Tr}_{AC}(\tilde{\sigma}_{ABC})$ , we can write

$$\text{Tr}_B[\text{Tr}_{AC}(\tilde{\sigma}_{ABC})(\log \rho_B + \log \theta_B)] = \text{Tr}_{ABC}[\tilde{\sigma}_{ABC} id_A \otimes (\log \rho_B + \log \theta_B) \otimes id_C] = 0. \quad (55)$$

Adding and subtracting both  $\log(\rho_{AB}) \otimes id_C$  and  $id_A \otimes \log(\rho_{BC})$ , which for simplicity we write without the identity factors, we get

$$\text{Tr}_{ABC}[\tilde{\sigma}_{ABC} (\log \tilde{\sigma}_{ABC} + \log \rho_B - \log \rho_{AB} - \log \rho_{BC})] = 0, \quad (56)$$

and this equation is equivalent to  $I_\rho(A : C|B) = 0$ , i.e., equivalent to the statement that the SSA inequality holds with equality.  $\square$

### 5. Best two out of three

In section 3, we provided a way of efficiently reconstructing a tripartite quantum state given two of its bipartite marginals subject to the compatibility condition—theorems 1 and 2. In general, if we are given the three possible bipartite density operators, we still have a residual degree of freedom, namely, the choice of the pair of bipartite density operators from which one will recover the tripartite estimator. We are going to show that the *best* tripartite maximum von Neumann entropy estimator is the one out of three with *minimum* von Neumann entropy—theorem 5.

**Theorem 5.** *Let  $\rho \in \mathcal{L}(\mathcal{H}_{ABC})$  be an unknown quantum state describing a tripartite quantum system  $ABC$ . Given the bipartite marginals  $\{\rho_{AB}, \rho_{BC}, \rho_{AC}\}$ , we define  $\mathcal{C} = \{\rho_{XY}, \rho_{YZ}\}$  with  $X, Y, Z \in A, B, C$  to be one of the three possible pairs of marginals, and  $\tilde{\rho}_C \in \text{Comp}(\mathcal{C})$  to be the maximum von Neumann entropy estimator. Then  $\tilde{\rho}$  minimizing the relative entropy with respect to the unknown state  $\rho_{ABC}$  is the one with minimum von Neumann entropy*

$$\tilde{\rho} = \arg \min_{\mathcal{C}} S(\tilde{\rho}_C). \quad (57)$$

**Proof.** By linearity of the trace, the von Neumann relative entropy between  $\rho_{ABC}$  and  $\tilde{\rho}_C$ :

$$S(\rho_{ABC}||\tilde{\rho}_C) = -S(\rho_{ABC}) - \text{Tr}[\rho_{ABC} \log \tilde{\rho}_C]. \quad (58)$$

$\tilde{\rho}_C$  has the form derived in equation (4). Then, there exist Hermitian operators  $\theta_{XY} \in \mathcal{L}(\mathcal{H}_{XY})$ ,  $\theta_{YZ} \in \mathcal{L}(\mathcal{H}_{YZ})$  and  $\theta_Y \in \mathcal{L}(\mathcal{H}_Y)$  (naturally extended to act on the joint Hilbert space) such that:

$$\log \tilde{\rho}_{\mathcal{C}} = \theta_{XY} + \theta_{YZ} + \theta_Y. \tag{59}$$

Plugging in (59) in (58) and using the fact that both  $\rho_{ABC}$  and  $\tilde{\rho}_{XYZ}$  are in the compatibility set of  $\mathcal{C}$ , it immediately follows that

$$S(\rho_{ABC}||\tilde{\rho}_{\mathcal{C}}) = S(\tilde{\rho}_{\mathcal{C}}) - S(\rho_{ABC}). \tag{60}$$

Since the term  $S(\rho_{ABC})$  is independent on the choice of  $\mathcal{C}$ :

$$\tilde{\rho} = \arg \min_{\mathcal{C}} S(\rho_{ABC}||\tilde{\rho}_{\mathcal{C}}) = \arg \min_{\mathcal{C}} S(\tilde{\rho}_{\mathcal{C}}). \tag{61}$$

□

The estimator we are proposing here is then:

$$\tilde{\rho} = \arg \min_{\mathcal{C}} \max_{\rho \in \text{Comp}(\mathcal{C})} S(\rho). \tag{62}$$

At this level, the efficiency of the choice of the optimal set of marginals is strictly related to the number of possibilities, which makes the direct generalization to the multipartite scenario inefficient. As we are going to see in detail in section 6, the number of possible choices increases exponentially with the number of variables. The Chow–Liu learning algorithm [10], solves the problem in the case of probability distributions. The next corollary generalizes the Chou–Liu main argument to QMCs., which will give an hint for the possible generalization of the Chow–Liu algorithm to *Markov quantum trees* (cf definition 6). Moreover, we are going to see that is sufficient that the compatibility condition with a QMC holds for the estimator obtained via Chow–Liu algorithm to be the optimal one.

**Theorem 6.** *Having a tripartite quantum system ABC described by an unknown  $\rho \in \mathcal{L}(\mathcal{H}_{ABC})$  and given the bipartite marginals  $\{\rho_{AB}, \rho_{BC}, \rho_{AC}\}$ . If for every pair  $\mathcal{C} = \{\rho_{XY}, \rho_{YZ}\}$  with  $X, Y, Z \in \{A, B, C\}$  there exists a QCM in the order  $X - Y - Z$  (cf theorem 2), then the QMC  $\tilde{\rho}_{XYZ}$  minimizing the relative entropy with respect to the unknown state  $\rho_{ABC}$  is the one recovered from the pair*

$$\tilde{\mathcal{C}} = \arg \max_{\mathcal{C}} \{I_{\rho}(X : Y) + I_{\rho}(Y : Z)\}. \tag{63}$$

**Proof.** Observe that:

$$\begin{aligned} S(\rho_{ABC}||\tilde{\rho}_{XYZ}) &= \text{Tr} [\rho_{ABC} (\log \rho_{ABC} - \log \tilde{\rho}_{XYZ})] \\ &= -S(\rho_{ABC}) - \text{Tr} [\rho_{ABC} (\log \rho_{XY} + \log \rho_{YZ} - \log \rho_Y)] \\ &= -S(\rho_{ABC}) + S(\rho_{XY}) + S(\rho_{YZ}) - S(\rho_Y), \end{aligned} \tag{64}$$

where in the second line we used lemma 1. Now, adding and subtracting  $S(\rho_X)$ ,  $S(\rho_Y)$  and  $S(\rho_Z)$ , we immediately get

$$S(\rho_{ABC}||\tilde{\rho}_{XYZ}) = - [I_{\rho}(X : Y) + I_{\rho}(Y : Z)] + \sum_{W \in \{A, B, C\}} S(\rho_W) - S(\rho_{ABC}). \tag{65}$$

Since the two last terms are independent on the choice of pair of bipartites, we get the desired result. □

Then, in case all the subsets  $C$  are compatible with a QMC, the min–max estimator is the one obtained excluding from the set of marginals the one with lowest quantum mutual information. The compatibility conditions with QMCs are necessary for this result to hold. Indeed, relaxing the compatibility conditions, now we are going to determine an analogous form to equation (65) for a general maximum entropy estimator equation (57).

Take equation (59) and add and subtract  $S(\rho_{XY}), S(\rho_{YZ}), S(\rho_Y)$ . We then observe that

$$(i) S(\rho_{XY}) + S(\rho_{YZ}) - S(\rho_Y) = - [I_\rho(X : Y) + I_\rho(Y : Z)] + \sum_{W \in \{A,B,C\}} S(\rho_W), \quad (66)$$

$$\begin{aligned} (ii) \text{Tr} [\rho_{ABC} \log \tilde{\rho}_{XYZ}] &= \text{Tr} [\rho_{ABC} (\theta_{XY} + \theta_{YZ} + \theta_Y)] \\ &= \text{Tr}_{XY} [\text{Tr}_Z(\rho_{ABC})\theta_{XY}] + \text{Tr}_{YZ} [\text{Tr}_X(\rho_{ABC})\theta_{YZ}] + \text{Tr}_Y [\text{Tr}_{XZ}(\rho_{ABC})\theta_Y] \\ &= \text{Tr}_{XY} [\text{Tr}_Z(\tilde{\rho}_{XYZ})\theta_{XY}] + \text{Tr}_{YZ} [\text{Tr}_X(\tilde{\rho}_{XYZ})\theta_{YZ}] + \text{Tr}_Y [\text{Tr}_{XZ}(\tilde{\rho}_{XYZ})\theta_Y] \\ &= -S(\tilde{\rho}_{XYZ}); \end{aligned} \quad (67)$$

$$(iii) S(\tilde{\rho}_{XYZ}) - S(\rho_{XY}) - S(\rho_{YZ}) + S(\rho_Y) = -I_\rho(X : Z|Y) \quad (68)$$

Therefore equation (57) can be rewritten as it follows

$$S(\rho_{ABC} || \tilde{\rho}_{XYZ}) = - [I_\rho(X : Y) + I_\rho(Y : Z)] - I_\rho(X : Z|Y) + \sum_{W \in \{A,B,C\}} S(\rho_W) - S(\rho_{ABC}). \quad (69)$$

Comparing equation (69) with equation (65), we can see that, in general, the choice of the two marginals with maximum mutual informations between the subparts, is not the optimal one for the maximum entropy estimator. Relaxing the compatibility conditions between just one pair with a QMC, the additional term in equation (69), namely the quantum conditional mutual information of the constructed maximum entropy estimator,  $I_\rho(X : Z|Y)$ , is different from zero. The result of equation (69) does not lead, at first sight, to a simplification of equation (61).

## 6. The multipartite case

In the previous sections, while considering the tripartite case, we learned that in order to have an algebraic recovery procedure, we need to have a tree structure and an additional constraint regarding the conditional mutual information. It is then natural to suppose that in the general multipartite case one would need a set of additional constraints, which generalize the one obtained previously, and this motivates the following definitions.

**Definition 4** (Quantum graph and quantum tree). A quantum graph is a triple  $(\mathcal{X}, \{\mathcal{H}_X\}_{X \in \mathcal{X}}, \rho, \mathcal{G})$ , where  $\mathcal{X} = \{X_1, \dots, X_n\}$  labels quantum systems described by the associated Hilbert spaces  $\mathcal{H}_X, X \in \mathcal{X}$ , with the  $n$ -partite composite system described by  $\mathcal{H}_{\mathcal{X}} = \mathcal{H}_{X_1} \otimes \dots \otimes \mathcal{H}_{X_n}, \rho \in \mathcal{L}(\mathcal{H}_{\mathcal{X}})$  is an  $n$ -partite density operator and  $\mathcal{G} = (\mathcal{X}, E)$  is an undirected graph. Whenever the underlying graph of a quantum graph is a tree, we call the structure a quantum tree.

**Definition 5.** Let  $\mathcal{G} = (V, E)$  be an undirected graph and let  $U$  and  $W$  be non overlapping subsets of  $V$ . We say that a subset  $Z$ , disjoint from  $U$  and  $W$ , separates  $U$  and  $W$  if every path connecting a vertex in  $U$  and a vertex in  $W$  necessarily overlaps with  $Z$ . We say that  $Z$  is a separator for  $U$  and  $W$ .

**Definition 6** (Markov quantum field and Markov quantum tree). A Markov quantum field is a quantum graph  $(\mathcal{X}, \{\mathcal{H}_X\}_{X \in \mathcal{X}}, \rho, \mathcal{G})$ , where the density operator  $\rho$  satisfies the global Markov property: for all  $U, W \subset \mathcal{X}$  such that there exists a separator  $Z$  for  $U$  and  $W$  then

$$I_\rho(U : W | Z) = 0. \tag{70}$$

Whenever the underlying graph of a Markov quantum field is a tree, we call the structure a Markov quantum tree.

Definition 6 includes the most general Markov property, i.e. the global property. Observe that given a Markov quantum field, then any quantum subtree is a quantum Markov tree.

Let  $\rho_{\mathcal{X}} \in \mathcal{L}(\mathcal{H}_{\mathcal{X}})$  be an unknown density operator that describes an  $n$ -partite physical system labelled by  $\mathcal{X}$ . Let  $\mathcal{C}_{\mathcal{T}} = \{\rho_{X_i X_j}, \{X_i, X_j\} \in E(\mathcal{T})\}$  a subset of bipartite marginals, graphically representable by one of its spanning trees  $\mathcal{T}$ . The quantum state  $\tilde{\rho}_{\mathcal{C}_{\mathcal{T}}}$  that maximizes the von Neumann entropy under the constraints of compatibility with the marginals in  $\mathcal{C}_{\mathcal{T}}$  has the form derived in equation (3). Then, there exists a set of Hermitian operators  $\{\theta_{X_i} \in \mathcal{L}(\mathcal{H}_{X_i}), X_i \in V(\mathcal{T})\}$  and  $\{\theta_{X_i X_j} \in \mathcal{L}(\mathcal{H}_{X_i X_j}), \{X_i, X_j\} \in E(\mathcal{T})\}$ , naturally extended to act on the joint Hilbert space, such that:

$$\log \tilde{\rho}_{\mathcal{C}_{\mathcal{T}}} = \sum_{\{X_i, X_j\} \in E(\mathcal{T})} \theta_{X_i X_j} + \sum_{i=1}^n \theta_{X_i}, \tag{71}$$

where  $\deg X_i$  is the *degree of the node*  $X_i$ , i.e. the number of edges linked to the node. If the quantum tree is a Markov quantum tree, then we have,

$$\log \tilde{\rho}_{\mathcal{C}_{\mathcal{T}}} = \sum_{\{X_i, X_j\} \in E(\mathcal{T})} \log \rho_{X_i X_j} - \sum_{i=1}^n (\deg X_i - 1) \log \rho_{X_i}, \tag{72}$$

The combinatorial factor is obtained by considering the spanning Markov quantum tree  $\mathcal{T}$  over  $\mathcal{X}$  as a tripartite one on  $ABC$  with  $A = X_i \in \mathcal{X}$ ,  $B = X_j \in \mathcal{X}$ ,  $C = \mathcal{X} \setminus \{X_i, X_j\}$ , where  $X_i$  and  $X_j$  are chosen such that there are no edges between  $X_i$  and any vertices in  $\mathcal{X} \setminus \{X_i, X_j\}$ , i.e.,  $A - B - C$  is a Markov quantum tree, for which we can use lemma 1. Then, consider the Markov quantum subtree  $\mathcal{X} \setminus \{X_i, X_j\}$  as a tripartite quantum tree  $A' - B' - C'$  with  $A = X_k \in \mathcal{X} \setminus X_i$ ,  $B = X_l \in \mathcal{X} \setminus X_i$ ,  $C = \mathcal{X} \setminus \{X_i, X_k, X_l\}$ , apply lemma 1, and iterate the procedure until the remaining subgraph is bipartite. It is then clear that each vertex comes with a factor of its degree minus one.

The construction above allows for an algebraic recovery of the state by iteratively applying the Petz recovery map. In the following paragraph, we show that the global Markov condition results in an efficient choice of the best tree.

**Choosing the best tree:** given an unknown  $n$ -partite quantum system, its bipartite correlations can be represented by a complete graph. Its number of possible spanning trees, i.e., tree subgraphs which include all vertices in the graph, is given by Cayley’s formula [38],  $n^{n-2}$ , which grows exponentially with the number of vertices. Because of this, we can not choose the best tree efficiently. In the classical scenario, one possible solution is to use the Chow–Liu algorithm [10].

Recall that the Chow–Liu algorithm, cf [appendix](#), provides an efficient way to find the optimal tree minimizing the Kullback–Leibler divergence between the actual probability distribution,  $p(X_1 = x_1, \dots, X_n = x_n)$ , and a probability distribution associated to a spanning tree,

$p^{\mathcal{T}}(X_1 = x_1, \dots, X_n = x_n)$ . In [10], it is shown that the Kullback–Leibler divergence can be written as:

$$D_K(p, p^{\mathcal{T}}) = - \sum_{\{X_i, X_j\} \in E(\mathcal{T})} I(X_i, X_j) + \sum_{i=1}^n H(X_i) - H(X_1, \dots, X_n), \quad (73)$$

where  $H(\cdot)$  is the Shannon entropy and  $I(X_i, X_j)$  is the classical mutual information between  $X_i$  and  $X_j$ . The only term that depends on the choice of tree  $\mathcal{T} = (V(\mathcal{T}) = \mathcal{X}, E(\mathcal{T}))$  is the first one, therefore, minimizing the Kullback–Leibler divergence is equivalent to maximizing

$$\sum_{\{X_i, X_j\} \in E(\mathcal{T})} I(X_i, X_j), \quad (74)$$

to which problem the Chow–Liu algorithm provides an efficient solution.

The relative entropy between the unknown density operator and one of its quantum tree maximum entropy estimators can be written as

$$S(\rho_{\mathcal{X}} \| \tilde{\rho}_{\mathcal{C}_{\mathcal{T}}}) = -S(\rho_{\mathcal{X}}) + \text{Tr}(\rho_{\mathcal{X}} \log \tilde{\rho}_{\mathcal{C}_{\mathcal{T}}}) = S(\tilde{\rho}_{\mathcal{C}_{\mathcal{T}}}) - S(\rho_{\mathcal{X}}). \quad (75)$$

where we have used the compatibility conditions to perform a calculation similar to that of equation (67). It then follows that equation (62) still holds in the  $n$ -partite scenario. An alternative form for equation (75), which generalizes the one obtained by Chow and Liu for probability distributions equation (73), is now derived. Adding and subtracting to equation (75) the terms  $\sum_{\{X_i, X_j\} \in E(\mathcal{T})} S(\rho_{X_i X_j})$  and  $\sum_{i=1}^n (\text{deg } X_i - 1) S(\rho_{X_i})$ , observing that

$$\sum_{\{X_i, X_j\} \in E(\mathcal{T})} S(\rho_{X_i X_j}) - \sum_{i=1}^n (\text{deg } X_i - 1) S(\rho_{X_i}) = - \sum_{\{X_i, X_j\} \in E(\mathcal{T})} I_{\rho}(X_i, X_j) + \sum_{i=1}^n S(\rho_{X_i}), \quad (76)$$

and setting

$$\Delta S(\tilde{\rho}_{\mathcal{C}_{\mathcal{T}}}) := \sum_{\{X_i, X_j\} \in E(\mathcal{T})} S(\rho_{X_i X_j}) - \sum_{i=1}^n (\text{deg } X_i - 1) S(\rho_{X_i}) - S(\tilde{\rho}_{\mathcal{C}_{\mathcal{T}}}), \quad (77)$$

Equation (75) assumes the form

$$S(\rho_{\mathcal{X}} \| \tilde{\rho}_{\mathcal{C}_{\mathcal{T}}}) = - \sum_{\{X_i, X_j\} \in E(\mathcal{T})} I_{\rho}(X_i, X_j) - \Delta S(\tilde{\rho}_{\mathcal{C}_{\mathcal{T}}}) + \sum_{i=1}^n S(\rho_{X_i}) - S(\rho_{\mathcal{X}}). \quad (78)$$

When  $\Delta S(\tilde{\rho}_{\mathcal{C}}) = 0$ , again the best tree is the one that maximizes the term

$$\sum_{\{X_i, X_j\} \in E(\mathcal{T})} I_{\rho}(X_i, X_j). \quad (79)$$

This last problem can be efficiently solved using the Chow–Liu algorithm where the mutual information on the bipartite subparts is replaced by its quantum generalization.

Next, we are going to study the conditions the correlations inside the given system have to satisfy in order to have  $\Delta S(\tilde{\rho}_{\mathcal{C}_{\mathcal{T}}}) = 0$ . Consider the following iterative construction. Since  $\mathcal{T}$  is a

tree, there exists a leaf, i.e., a vertex with degree 1 call it  $X_{l_1} \in \mathcal{X}$  and denote by  $\text{ad}(X_{l_1})$  its adjacent vertex. Set  $\mathcal{T}_1 = (V_1, E_1) := \mathcal{T}$ . Now consider the chain  $X_{l_1} - \text{ad}(X_{l_1}) - V_1 \setminus \{X_{l_1}, \text{ad}(X_{l_1})\}$ . The associated quantum conditional mutual information reads

$$I_\rho(X_{l_1} : V_1 \setminus \{X_{l_1}, \text{ad}(X_{l_1})\} | \text{ad}(X_{l_1})) = S(\rho_{X_{l_1} \text{ad}(X_{l_1})}) + S(\rho_{V_1 \setminus \{X_{l_1}\}}) - S(\rho_{\text{ad}(X_{l_1})}) - S(\rho_{V_1}). \quad (80)$$

Observe that  $\rho_{V_1} = \tilde{\rho}_{\mathcal{T}}$ . Set  $\mathcal{T}_2 := (V_2, E_2)$ , where  $V_2 = V_1 \setminus \{X_{l_1}\}$  and  $E_2$  is obtained naturally from  $E_1$  by dropping  $\{X_{l_1}, \text{ad}(X_{l_1})\}$ . It is trivial to see that  $\mathcal{T}_2$  is a tree and, thus, we can find a leaf  $X_{l_2} \in V_2$ . Consider now the chain  $X_{l_2} - \text{ad}(X_{l_2}) - V_2 \setminus \{X_{l_2}, \text{ad}(X_{l_2})\}$ . The associated quantum conditional mutual information reads

$$I_\rho(X_{l_2} : V_2 \setminus \{X_{l_2}, \text{ad}(X_{l_2})\} | \text{ad}(X_{l_2})) = S(\rho_{X_{l_2} \text{ad}(X_{l_2})}) + S(\rho_{V_2 \setminus \{X_{l_2}\}}) - S(\rho_{\text{ad}(X_{l_2})}) - S(\rho_{V_2}). \quad (81)$$

It is now clear that  $S(\rho_{V_1 \setminus \{X_{l_1}\}})$  in equation (80) cancels with  $S(\rho_{V_2})$  upon summing the two equations. Iteratively, we can build a tree  $\mathcal{T}_{i+1} = (V_{i+1}, E_{i+1})$  from a tree  $\mathcal{T}_i = (V_i, E_i)$  with a chosen leaf  $X_{l_i}$ , by setting  $V_{i+1} = V_i \setminus \{X_{l_i}\}$  and dropping the edge  $\{X_{l_i}, \text{ad}(X_{l_i})\}$  from  $E_i$  to obtain  $E_{i+1}$ . This construction can be performed until one obtains the last chain  $X_{l_{n-2}}$

$$\begin{aligned} & \sum_{i=1}^{n-2} I_\rho(X_{l_i} : V_i \setminus \{X_{l_i}, \text{ad}(X_{l_i})\} | \text{ad}(X_{l_i})) \\ &= \sum_{i=1}^{n-2} \left[ S(\rho_{X_{l_i} \text{ad}(X_{l_i})}) - S(\rho_{\text{ad}(X_{l_i})}) \right] + S(\rho_{V_{n-2} \setminus \{X_{l_{n-2}}\}}) - S(\rho_{V_1}) \\ &= \sum_{\{X_i, X_j\} \in E(\mathcal{T})} S(\rho_{X_i X_j}) - \sum_{i=1}^n (\text{deg}(X_i) - 1) S(\rho_{X_i}) - S(\tilde{\rho}_{\mathcal{T}}), \end{aligned} \quad (82)$$

where we noticed that  $V_{n-2} \setminus \{X_{l_{n-2}}\}$  is the last edge missing in the sum on the second line and also that  $\bigcup_{i=1}^{n-2} \{\text{ad}(X_{l_i})\}$  is  $\mathcal{X}$  except two vertices that have degree one. Therefore, we have

$$\Delta S(\tilde{\rho}_{\mathcal{T}}) = \sum_{i=1}^{n-2} I_\rho(X_{l_i} : V_i \setminus \{X_{l_i}, \text{ad}(X_{l_i})\} | \text{ad}(X_{l_i})). \quad (83)$$

This equation shows that  $\Delta S(\tilde{\rho}_{\mathcal{T}}) \geq 0$ , because each term in the sum is non-negative. To proceed, we will apply the chain rule for quantum conditional mutual information:

$$I_\rho(A : C_1 \dots C_n | B) = \sum_{j=1}^n I_\rho(A : C_j | B C_1 \dots C_{j-1}), \quad (84)$$

where the first term in the sum is defined to be  $I_\rho(A : C_1 | B)$ . Motivated by the order appearing in the Chain rule, we introduce an order in  $V_i \setminus \{X_{l_i}, \text{ad}(X_{l_i})\} = \{X_{w_i(j)}\}_{j=1}^{n-i}$ , so that

$$\begin{aligned} \Delta S(\tilde{\rho}_{\mathcal{C}_T}) &= \sum_{i=1}^{n-2} I_\rho(X_{l_i} : X_{w_i(1)} \dots X_{w_i(n-i)} | \text{ad}(X_{l_i})) \\ &= \sum_{i=1}^{n-2} \sum_{j=1}^{n-i} I_\rho(X_{l_i} : X_{w_i(j)} | \text{ad}(X_{l_i}) X_{w_i(1)} \dots X_{w_i(j-1)}). \end{aligned} \tag{85}$$

To have  $\Delta S(\tilde{\rho}_{\mathcal{C}_T}) = 0$  is equivalent to having each term in the sum zero.

**Proposition 3.** *The quantity  $\Delta S(\tilde{\rho}_{\mathcal{C}_T})$  is always non-negative. It is zero iff*

$$I_\rho(X_{l_i} : X_{w_i(j)} | \text{ad}(X_{l_i}) X_{w_i(1)} \dots X_{w_i(j-1)}) = 0, \quad \text{for all } j = 1, \dots, n - i \text{ and } i = 1, \dots, n - 2. \tag{86}$$

□

We conclude that if we are given a Markov quantum field, then the choice of the best tree is efficient and the recovery procedure is algebraic. Notice that the set of conditions obtained in proposition 3 is polynomial, namely  $O(n^2)$ , and these are in general weaker than the global Markov property. This can perhaps be used as a hint towards relaxing the computationally demanding verification of the global Markov property on the provided marginals.

### 7. Conclusions and outlook

In this manuscript, we proposed a way to compress a subset of density operators according to a generalization to the quantum realm of the ‘Jaynes’ max entropy principle, given a chosen set of partial information. Focusing on the tripartite case, with access to two bipartite marginals, we provided a necessary and sufficient algebraic condition for compatibility with a QMC. The recovery procedure through the Petz map is algebraic and efficient. The recovery procedure goes as follows:

- (a) Measure the three bipartite marginals  $\rho_{AB}, \rho_{BC}, \rho_{AC}$ ;
- (b) Check for every couple  $\rho_{XY}, \rho_{YZ}$  with  $X, Y, Z \in \{A, B, C\}$  the compatibility condition with a quantum Markov chain  $X - Y - Z$  (theorem 2);
- (c) If for all the three couples of marginals the compatibility holds, compute the quantum mutual information  $I(X : Y)$  and discard the  $\rho_{XY}$  with minimum  $I(X : Y)$ ;
- (d) From the two remaining marginals, via Petz recovery map, construct the min–max tripartite estimator (theorem 1).

In theorem 6, we provided a new characterization of QMCs in terms of a commutative diagram of quantum Bayesian updating processes. This hints on a possible category-theoretical characterization of QMCs, which requires further investigation.

Through the notion of a Markov quantum tree, we were able to generalize the Chow–Liu algorithm to density operators. In fact, the results of this manuscript indicate that the classical theory of learning probability distributions via maximum entropy estimation can be extended naturally to Markov quantum trees.

We speculate that, due to the additional term in equation (69) (see the paragraph after), it might be possible to extend our results to approximate quantum Markov chains (see reference [21]), however, one would have to understand if it is possible to have an efficient compatibility condition, i.e., the analogue of theorem 2 for the case of approximate quantum Markov chains.

We would like to be able to relax the global Markov condition in definition 6 and the result of proposition 3 provides a hint that such relaxation might be possible. This is desirable because, in general, the global Markov property seems computationally demanding, for a classical computer, to be verified. Another approach, would be to understand if a quantum computer can learn an even wider class of density operators efficiently.

Another possible direction towards extending the space of learnable quantum states is to enlarge the Hilbert space by an ancilla, which would allow to have a QMC, thus, efficiently learnable, however, this ancilla would have to be subject to certain conditions in order not to end up with a state which would be far from the unknown state in the relative entropy sense.

## Acknowledgments

This work was supported by European funds namely via H2020 project SPARTA, national funds through FCT, Fundação para a Ciência e a Tecnologia, under contract IT (UID/EEA/50008/2019), PREDICT (PTDC/CCI-CIF/29877/2017), project QuantMining (POCI-01-0145-FEDER-031826) and internal IT projects QBigData and RAPID. Serena Di Giorgio also acknowledges the FCT PhD Grant PD/BD/114332/2016 and the DP-PMI FCT programme.

## Appendix. The Chow–Liu algorithm

Given a set of random variables  $V = \{X_1, \dots, X_n\}$  for which we have access to the bipartite correlations described by the joint probability distributions  $\{p_{X_i X_j}\}$  we can build a weighted complete graph  $\mathcal{G}$  whose vertices label the random variables and the edges correspond to the bipartite probability distributions,  $p_{X_i X_j}$ , weighted by the classical mutual informations  $I(X_i, X_j)$ , given by

$$I(X_i, X_j) = \sum_{x_i, x_j} p_{X_i X_j}(x_i, x_j) \log \frac{p_{X_i}(x_i) p_{X_j}(x_j)}{p_{X_i X_j}(x_i, x_j)}.$$

The Chow–Liu algorithm allows us to efficiently construct the maximum weighted spanning tree. Explicitly, sort the values  $\{I(X_i, X_j) = I_\alpha\}_{\alpha=1}^{N=\frac{1}{2}n(n-1)}$  in descending order,  $I_1 \geq I_2 \geq \dots \geq I_N$ , then the algorithm proceeds to build a tree  $\mathcal{T}$  iteratively as follows:

[Initialization]  $\mathcal{G}_0 = (V, E_0)$  where  $V = \{X_1, \dots, X_n\}$  and  $E_0 = \emptyset$ .

[Iterative Step] Let  $\{X_i, X_j\}$  be the pair associated to  $\alpha \in \{1, \dots, N\}$ . Build a graph  $\mathcal{G}_\alpha = (V, E_{\alpha-1})$ , where  $E_\alpha$  is obtained as follows

$$E_\alpha = \begin{cases} E_{\alpha-1} \cup \{\{X_i, X_j\}\}, & \text{if } \mathcal{G}_\alpha = (V, E_\alpha) \text{ is a tree,} \\ E_{\alpha-1}, & \text{otherwise.} \end{cases}$$

The graph  $\mathcal{G}_N = \mathcal{T}$  is the desired maximum weighted tree.

## ORCID iDs

S Di Giorgio  <https://orcid.org/0000-0002-4719-1663>

## References

- [1] Jaynes E T 1957 Information theory and statistical mechanics *Phys. Rev.* **106** 620
- [2] Bishop C M 2006 *Pattern Recognition and Machine Learning* (Berlin: Springer)
- [3] Jacob B, Wittek P, Pancotti N, Reberntrost P, Nathan W and Lloyd S 2017 Quantum machine learning *Nature* **549** 195
- [4] Schuld M, Sinayskiy I and Petruccione F 2015 An introduction to quantum machine learning *Contemp. Phys.* **56** 172–85
- [5] Carleo G and Troyer M 2017 Solving the quantum many-body problem with artificial neural networks *Science* **355** 602–6
- [6] Pearl J 2001 Bayesian networks, causal inference and knowledge discovery *Technical Report* (Los Angeles, CA: UCLA Cognitive Systems Laboratory)
- [7] Chickering D M 1996 Learning Bayesian networks is NP-complete *Learning from Data* (New York: Springer) pp 121–30
- [8] Paul D and Luby M 1993 Approximating probabilistic inference in Bayesian belief networks is np-hard *Artif. Intell.* **60** 141–53
- [9] Dasgupta S 1999 Learning polytrees *Proc. 15th Conf. on Uncertainty in Artificial Intelligence UAI'99* pp 134–41
- [10] Chow C and Liu C 1968 Approximating discrete probability distributions with dependence trees *IEEE Trans. Inf. Theory* **14** 462–7
- [11] Leifer M S and Spekkens R W 2013 Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference *Phys. Rev. A* **88** 052130
- [12] Fitzsimons J F, Jones J A and Vedral V 2015 Quantum correlations which imply causation *Sci. Rep.* **5** 18281
- [13] Horsman D, Heunen C, Pusey M F, Barrett J and Spekkens R W 2017 Can a quantum state over time resemble a quantum state at a single time? *Proc. R. Soc. A* **473** 20170395
- [14] Liu Y-K 2006 Consistency of local density matrices is QMA-complete *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques* (Berlin: Springer) pp 438–49
- [15] Liu Y-K, Christandl M and Frank V 2007 Quantum computational complexity of the n-representability problem: Qma complete *Phys. Rev. Lett.* **98** 110503
- [16] Hall W 2007 Compatibility of subsystem states and convex geometry *Phys. Rev. A* **75** 032102
- [17] Chen L, Gittsovich O, Modi K and Piani M 2014 Role of correlations in the two-body-marginal problem *Phys. Rev. A* **90** 042314
- [18] Sontag D and Jaakkola T S 2008 New outer bounds on the marginal polytope *Advances in Neural Information Processing Systems* pp 1393–400
- [19] Tyc T and Jan V 2015 Quantum marginal problems *Eur. Phys. J. D* **69** 209
- [20] Huber F M Quantum states and their marginals: from multipartite entanglement to quantum error-correcting codes *PhD Thesis* University of Siegen
- [21] Sutter D 2018 *Approximate Quantum Markov Chains* (Cham: Springer) pp 75–100
- [22] Kim I H 2012 Operator extension of strong subadditivity of entropy *J. Math. Phys.* **53** 122204
- [23] Ruskai M B 2013 Remarks on Kim's strong subadditivity matrix inequality: extensions and equality conditions *J. Math. Phys.* **54** 102202
- [24] Omar F and Renner R 2015 Quantum conditional mutual information and approximate Markov chains *Commun. Math. Phys.* **340** 575–611
- [25] Hayden P, Jozsa R, Petz D and Winter A 2004 Structure of states which satisfy strong subadditivity of quantum entropy with equality *Commun. Math. Phys.* **246** 359–74
- [26] Lieb E H and Ruskai M B 1973 Proof of the strong subadditivity of quantum-mechanical entropy *J. Math. Phys.* **14** 1938–41
- [27] Petz D 1986 Sufficient subalgebras and the relative entropy of states of a von neumann algebra *Commun. Math. Phys.* **105** 123–31
- [28] Wilde M M 2015 Recoverability in quantum information theory *Proc. R. Soc. A* **471** 20150338
- [29] Petz D 2003 Monotonicity of quantum relative entropy revisited *Rev. Math. Phys.* **15** 79–91
- [30] Kullback S 1997 *Information Theory and Statistics* (Courier Corporation)
- [31] Kullback S and Khairat M A et al 1966 A note on minimum discrimination information *Ann. Math. Stat.* **37** 279–80
- [32] Vanslette K 2017 Entropic updating of probabilities and density matrices *Entropy* **19** 664

- [33] Vanslette K 2018 The quantum Bayes rule and generalizations from the quantum maximum entropy method *J. Phys. Commun.* **2** 025017
- [34] Sutter D, Omar F and Renner R 2016 Universal recovery map for approximate Markov chains *Proc. R. Soc. A* **472** 20150623
- [35] Brandao F G S L, Harrow A W, Oppenheim J and Strelchuk S 2015 Quantum conditional mutual information, reconstructed states, and state redistribution *Phys. Rev. Lett.* **115** 050501
- [36] Lindblad G 1973 Entropy, information and quantum measurements *Commun. Math. Phys.* **33** 305–22
- [37] Lesniewski A and Ruskai M B 1999 Monotone Riemannian metrics and relative entropy on noncommutative probability spaces *Commun. Math. Phys.* **40** 5702–24
- [38] Cayley C 1889 A theorem on trees *Quart. J. Math.* **23** 376–8