A Quantum Information Theoretic Approach to Tractable Probabilistic Models

Pedro Zuidberg Dos Martires [1]

Abstract

By recursively nesting sums and products, probabilistic circuits have emerged in recent years as an attractive class of generative models as they enjoy, for instance, polytime marginalization of random variables. In this note we study these machine learning models using the framework of quantum information theory, leading to the introduction of *positive unital circuits* (PUnCs), which generalize circuit evaluations over positive real-valued probabilities to circuit evaluation over positive semidefinite matrices. As a consequence, PUnCs strictly generalize probabilistic circuits as well as the recently introduced class of PSD circuits.

1. Introduction

Probabilistic circuits (PCs) (Darwiche, 2003; Poon & Domingos, 2011) belong to an unusual class of probabilistic models: they are highly expressive but at the same time also tractable. For instance, so-called decomposable probabilistic circuits (Darwiche, 2001a) allow for the computation of marginals in time polynomial in the size of the circuit. Zhang et al. (2020) noted that it is exactly this restriction to positive values that limits the expressive efficiency (or succinctness) of PCs (Martens & Medabalimi, 2014; de Colnet & Mengel, 2021). In particular, the positivity constraint on the set of elements that PCs operate on prevents them from modelling negative correlations between variables. Circuits that are incapable of modelling negative correlations, i.e. circuits that can only combine probabilities in an additive fashion, are also called monotone circuits (Shpilka & Yehudayoff, 2010). This restricted expressiveness can be combatted by the use of so-called *non-monotone* circuits, where subtractions are allowed as a third operation (besides sums and products). Interestingly, Valiant (1979) showed that a mere single subtraction can render non-monotone circuits exponentially more expressive than monotone circuits a result that has recently been refined for decomposable circuits (Loconte et al., 2024b).

As shown in (Harviainen et al., 2023; Agarwal & Bläser, 2024), non-monotone circuits do, however, introduce an important complication: if non-monotone circuits are not designed carefully, verifying whether a circuit encodes a valid probability distribution or not is an NP-hard problem. This does also render learning the parameters of a circuit practically infeasible.

Using the concept of *positive operator valued measures* from quantum information theory, which encode random event as positive semidefinite matrices we are able to construct non-monotone circuits that nonetheless encode proper (normalized) probability distributions by construction. Our work falls into a line of recent works presented in the circuit literature (Sladek et al., 2023; Loconte et al., 2024c; Wang & Van den Broeck, 2024; Loconte et al., 2024b). However, our work is the first that establishes this deep connection between concepts in quantum information theory and tractable probabilistic models. Furthemorer, the circuits class of probabilistic unital circuits that we introduce generalizes both the probabilistic circuits as well as PSD circuits (Sladek et al., 2023; Loconte et al., 2024c;b)(Sladek et al., 2023)[1].

2. A Primer on Quantum Information Theory

A widely used and elegant framework to describe measurements of quantum systems is the so-called *positive operator-valued measure* (POVM) formalism. While POVMs have physical interpretations in terms of quantum information and quantum statistics, we will only be interested in their mathematical prperties as we use them to show that circuits (defined in Section 3) form valid probability distributions. We refer the reader to (Nielsen & Chuang, 2001) for an in-depth exposition on the topic, as well as quantum computing and quantum information theory in general.

Definition 2.1 (Positive Semidefinite)**.** An $B \times B$ Hermitian matrix H is called positive semidefinite (PSD) if and only if $\forall \mathbf{x} \in \mathbb{C}^B : \mathbf{x}^* H \mathbf{x} \geq 0$, where $\mathbf{x}^*$ denotes the conjugate transpose and $\mathbb{C}^B$ the B-dimensional space of complex numbers.

Definition 2.2 (POVM (Nielsen & Chuang, 2001, Page 90))**.**

[1] Department of Computer Science, Örebro University, Sweden. Correspondence to: Pedro Zuidberg Dos Martires <pedro.zuidberg-dos-martires@oru.se>.

[1] PSD circuits were later on rebranded as sum of squares circuits (Loconte et al., 2024b)

A positive operator-valued measure is a set of PSD matrices $\{E(i)\}_{i=0}^{I-1}$ (I being the number of possible measurement outcomes) that sum to the identity:

$$\sum_{i=0}^{I-1} E(i) = \mathbb{1}, \tag{1}$$

Before defining the probability of a specific i occurring, we need the notion of a density matrix (von Neumann, 1927):

Definition 2.3 (Density Matrix (Nielsen & Chuang, 2001, Page 102))**.** A density matrix ρ is a PSD matrix of trace one, i.e. $\mathrm{Tr}[\rho] = 1$.

Definition 2.4 (Event Probability (Nielsen & Chuang, 2001, Page 102))**.** Let ρ be a density matrix and let i denote an event with $E(i)$ being the corresponding element from the POVM. The probability of the event i happening, i.e. measuring the outcome i, is given by

$$p(i) = \mathrm{Tr}[\rho E(i)] \tag{2}$$

Proposition 2.5. *The expression in Equation 2 defines a valid probability distribution.*

Proof. While this is a well-known result we were not able to identify a concise proof in the literature. We therefore provide one in Appendix A.1. □

Given that the $E(i)$'s completely describe the event i such that its event probability can be computed, they represent the quantum state of a system. This quantum state (represented by a matrix) lives in a certain Hilbert space. The changes that a quantum state can undergo are then described by so-called *quantum operations* acting on the Hilbert space. We can construct such operations using Kraus' theorem.

e theorm a
e to make it
different from
dia

Theorem 2.6 (Kraus' Theorem(Kraus et al., 1983))**.** *Let $\mathcal{H}$ and $\mathcal{G}$ be Hilbert spaces of dimension N and M respectively, and Φ be a quantum operation between $\mathcal{H}$ and $\mathcal{G}$. Then, there are matrices $\{K_j\}_{j=1}^{D}$ (with $D \leq NM$) mapping $\mathcal{H}$ to $\mathcal{G}$ such that for any state $E(i)$*

$$\Phi(E(i)) = \sum_{j=1}^{D} K_j E(i) K_j^* \tag{3}$$

provided that $\sum_j K_j^* K_j \leq \mathbb{1}$ (in the Loewner order sense).

Proof. See (Nielsen & Chuang, 2001, Chapter 8) □

The K_j matrices are usually referred to as Kraus operators.

3. Positive Unital Circuits

A popular subclass of probabilistic circuits are so-called structured decomposable probabilistic circuits (Darwiche, 2011) that are also smooth (Darwiche, 2001b). The advantage of this circuit subclass is that they can be implemented in a rather straightforward fashion on modern AI accelerators, as demonstrated by Peharz et al. (2019; 2020). For the sake of exposition, we will limit ourselves in first instance to such circuits that adhere to structured decomposability and will generalize to (non-structured) decomposable circuits in Section 5. For a detailed account on these different circuit properties we refer the reader to (Vergari et al., 2021).

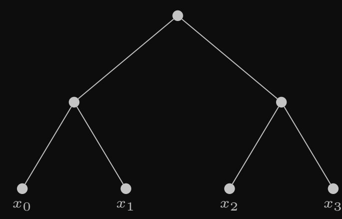

Figure 1: Partition circuit over four binary variables x_i with $i \in \{0, 1, 2, 3\}$, which are given as inputs to the circuit. The internal nodes of the partition circuit correspond the computation units.

Zuidberg Dos Martires (2024) introduced an abstraction for these smooth structured decomposable circuits in the form of partition trees. We further refine this introducing the concept of *partition circuits*. We give such a circuit in Figure 1. [2]

Definition 3.1 (Partition Circuit)**.** A partition circuit over a set of variables is a parametrized computation graph taking the form of a binary tree. The partition circuit consists of two kinds of computation units: *leaf* and *internal* units, as well as a single *root*. Units at the same distance from the root form a layer. Furthermore, let p_k denote the root unit or an internal unit. The unit p_k then receives its inputs from two units in the previous layer, which we denote by p_{k_l} and p_{k_r}. Each computation unit is input to exactly one other unit, except the root unit, which is the input to no other unit.

3.1. Positive Operator Circuits

Using the concept of partition circuits we construct positive operator circuits. Positive operator circuits can be thought of as generalizing circuit evaluations with probabilities to circuit evaluations with PSD matrices. Note, in the definition below we use O_k instead of p_k to make this generalization explicit.

Definition 3.2 (Positive Operator Circuit)**.** Let $\mathbf{x}=\{x_0, \ldots, x_{N-1}\}$ be a set of M categorical variables with domains of size S. We define an operator circuit

[2]The concept of a partition tree, and hence a partition circuit, is related to the concept of a variable tree (Pipatsrisawat & Darwiche, 2008) and a pseudo-tree (Dechter & Mateescu, 2007). However, partition circuits emphasize an interpretation as computation graphs, which the others do not.

as a partition circuit whose computation units take the following functional form:

$$O_k(\mathbf{x}_k) = \begin{cases} E_{x_k}, & \text{if } k \text{ is leaf} \\ \Phi_k\Big(O_{k_l}(\mathbf{x}_{k_l}) \otimes O_{k_l}(\mathbf{x}_{k_l})\Big) & \text{else,} \end{cases} \tag{4}$$

where the E_{x_k}'s are quantum state matrices, and where the Φ_k quantum operations.

Note that using the Kronecker product between $O_{k_l}(\mathbf{x}_{k_l})$ and $O_{k_l}(\mathbf{x}_{k_l})$ is a sensible choice as it describes the joint state of both subsystems.

Proposition 3.3. *Positive operator circuits are PSD.*

Proof. We know that all the leaves carry PSD matrices as they describe quantum states. Passing these on recursively to the quantum operations in the internal units retains the positive semi-defiteness as the Kronecker product between two PSD matrices is again PSD. □

3.2. Constructing a Probability Distribution

In Section 2 we saw that we can construct a probability distribution using a density matrix ρ and a positive operator-valued measure, with the latter being a set of PSD matrices (cf. Definition 2.2) that sum to the unit matrix. Using a positive operator circuit $O(\mathbf{x})$ we indeed have a set of PSD matrices. Namely, on for each instantiation of the $\mathbf{x}$ variables. We now introduce *positive unit preserving operator circuits* for which also the summation to the unit matrix holds.

Definition 3.4. We call a quantum operation *unital* if we have that

$$\Phi_k(\mathbb{1}_{k_l} \otimes \mathbb{1}_{k_r}) = \Phi_k(\mathbb{1}_{k_l k_r}) = \mathbb{1}_k, \tag{5}$$

where $\mathbb{1}_k$, $\mathbb{1}_{k_l}$, $\mathbb{1}_{k_l k_r}$, and $\mathbb{1}_{k_r}$ denote unit matrices of appropriate size,

Proposition 3.5. *Unital quantum operations are valid in the sense that the inequality $\sum_j K_j^* K_j \leq \mathbb{1}$ holds for all unital quantum operations.*

Proof. See Appendix A.2 □

Definition 3.6. We call a positive operator circuit *unital* if we have that the quantum operation Φ_k are unital, and if the sets $\{E_{x_k}\}_{x_k \in \Omega(X_k)}$ form a POVM for each random variables X_k.

We will also refer to positive unital operator circuits as positive unital circuits, or PUnCs.

Proposition 3.7. *Let $\mathbf{X}$ denote a set of random variables with sample space $\Omega(\mathbf{X})$. Then the set $\{\widehat{O}(\mathbf{x})\}_{\mathbf{x} \in \Omega(\mathbf{X})}$ of positive unital circuits forms a POVM.*

Proof. See Appendix A.3 □

Theorem 3.8. *Let ρ be a density matrix and $\widehat{O}(\mathbf{x})$ a positive unital circuit. The function*

$$p_{\mathbf{X}}(\mathbf{x}) = \mathrm{Tr}\Big[\widehat{O}(\mathbf{x})\rho\Big] \tag{6}$$

is a proper probability distribution over the random variables $\mathbf{X}$ with sample space $\Omega(\mathbf{X})$.

Proof. This follows from Proposition 2.5 and Proposition 3.7 □

One of the outstanding properties of probabilisitc circuits is that they are tractable – in the sense that they allow for polytime marginalization of random variables. Positive unital circuits retain this property.

Proposition 3.9. *Positive unital circuits allow for tractable marginalization.*

Proof. (Sketch) The proof is rather straightforward and hinges on the fact that the quantum operations in the internal coputation units are computable in poytime and on the fact that the marginalization of a random variables is performed by pushing the sum to the corresponding leaf in the partition circuit. Anologous to the proof of Proposition 3.7. □

4. Special Cases

We will now make certain structural assumption on the matrices representing the quantum states and the functional form of the quantum operations Φ. By doing so, we obtain the PSD circuits introduced by Sladek et al. (2023) and the (strucutred-decomposable) probabilistic circuits as described by Peharz et al. (2020) as special cases of positive unital circuits (Section 4.1 and Section 4.2 respectively).

First, however, we note our formulation of PUnCs already encompasses canonical polyadic tensor decomposition (Carroll & Chang, 1970) – a popular choice in the circuit literature (Shih et al., 2021; Loconte et al., 2024a) to merge partitions that uses the Hadamard product instead of the Kronecker product.

Specifically, we observe that the Hadamard product between two matrices A and B can be rewritten using Kronecker product;

$$A \circ B = P(A \otimes B)P^*, \tag{7}$$

where P is the semi-unitary partial permutation matrix selecting a principal submatrix (Visick, 2000, Corollary 2). This also means that a quantum operation involving q Hadamard product can be rewritten using a Kronecker product:

$$\begin{aligned}\Phi(A \circ B) &= \sum_i K_i (A \circ B) K_i^* \\ &= \sum_i K_i P (A \otimes B) P^* K_i^* \\ &= \sum_i K_i' (A \otimes B) K_i'^* = \Phi'(A \otimes B)\end{aligned} \tag{8}$$

Note that, for Φ' to be unital it suffices that $\sum_i K_i K_i^* = \mathbb{1}$ as P is semi-unitary ($PP^* = \mathbb{1}$).

From the discussion above we conclude that we can safely limit the discussion to circuits with Kronecker products as circuits with Hadamard products follow as a special case.

4.1. Pure Quantum States

As the matrices that represent quantum states are PSD we can decompose them as follows using the spectral theorem:

$$q = v_j \otimes v_j^*, \tag{9}$$

with the $vcircuit_j$'s denoteing the eigenvector. A special case are then so-called pure states for which we have

$$q = \lambda v \otimes v^*, \tag{10}$$

We will show now that by restricting PUnCs to performing operations on pure quantum states gives us the special case of PSD circuits as introduced by Sladek et al. (2023), which we define first using a partition circuit.

Definition 4.1. Let $\mathbf{x}$=$\{x_0, \ldots, x_{N-1}\}$ be a set of N categorical variables with domains of size S. A PSD circuit is a partition circuit whose computation units take the following functional form:

$$v_k(\mathbf{x}_k) = \begin{cases} U_k \times e_{x_k}, & \text{if } k \text{ leaf} \\ U_k \times \left(v_{k_l}(\mathbf{x}_k) \otimes v_{k_r}(\mathbf{x}_k)\right), & \text{else} \end{cases} \tag{11}$$

where the U_k's are semi-unitary matrices. The probability $p(\mathbf{x})$ is computed via

$$p(\mathbf{x}) = v_{root}^*(\mathbf{x}) \times \rho \times v_{root}(\mathbf{x}), \tag{12}$$

where ρ is a density matrix.

Note that in the original formulation Sladek et al. (2023) used non-semi-unitary matrices. However, Loconte & Vergari (2024) have recently shown that there is no loss in expressiveness when using semi-unitary matrices.

To show that PSD circuits are a special case of PUnCs we now impose the following restriction on the quantum operations Φ_k:

$$\Phi_k(O_{k_l} \otimes O_{k_l}) = K_k \left(O_{k_l} \otimes O_{k_r}\right) K_k^* \tag{13}$$

That is, we limit the quantum operation to having only a single pair of Kraus operators. For the quantum operation to be unital we need to have $K_{k1} K_{k1}^* = \mathbb{1}$. That is, K_{k1} has to be semi-unitary,

Furthermore, we make the following choice in the leaves:

$$E_{x_k} = K_k \left(e_{x_k} \otimes e_{x_k}^*\right) K_k^*, \tag{14}$$

where the set $\{e_{x_k}\}_{x_k \in \Omega(X_k)}$ is a complete set of orthonormal basis vectors, and K_k is again semi-unitary.

We can show that this choice for E_{x_k} forms a POVM. Firstly, by observing that each E_{x_k} is PSD. Secondly, by verifying the completeness of the set of operators:

$$\begin{aligned}\sum_{x_k \in \Omega(X_k)} E_{x_k} &= \sum_{x_k \in \Omega(X_k)} K_k \left(e_{x_k} \otimes e_{x_k}^*\right) K_k^* \\ &= K_k \left(\sum_{x_k \in \Omega(X_k)} e_{x_k} \otimes e_{x_k}^* \right) K_k^* \\ &= K_k \mathbb{1} K_k^* \\ &= \mathbb{1}\end{aligned} \tag{15}$$

Definition 4.2. We call a positive unital circuit pure if Equation 13 and Equation 14 hold.

Proposition 4.3. *For computation units of a pure positive unital circuit and a PSD circuit it holds that*

$$\forall k : O_k(\mathbf{x}_k) = v_k(\mathbf{x}_k) \otimes v_k^*(\mathbf{x}_k). \tag{16}$$

given that $U_k = K_k$

Proof. See Appendix A.4 □

Corollary 4.4. *Pure positive unital circuits perform operations on pure quantum states exclusively.*

Proof. This follows immediately from Proposition 4.3. □

Proposition 4.5. *A PSD circuit and a pure PUnC encode the same probability distribution if* $U_k = K_k$ *for each unit* k.

Proof. See Section A.5 □

In this section we have shown that by making specific choices in the functional form of the leaves and the internal units of a positive unital circuit that we recover the the special case PSD circuits and its variants (Loconte et al., 2024c;b). Furthermore, our analysis also provides the rather satisfying interpretation of PSD circuits as quantum circuits acting on pure states exclusively.

4.2. Statistical Mixtures of Eigenstates

5. Dropping Structured Decomposability

6. Related Work

6.1. Squared Curcuits

The work closest related to ours are the sum of compatible square circuits by Loconte et al. (2024b). As shown in Section **??** we recover sum of square circuits (introduced earlier as PSD circuits by Sladek et al. (2023)) if we use the functional form in Equation **??**. A non-polyadic decomposition would have resulted in a computation graph closer to that of *inception circuits* introduced by Wang & Van den Broeck (2024). Such a choice, however, increases the computation cost per unit in the partition circuit from quadratic (using the vector representation) to cubic (using the operator representation).

Our positive unit preserving operator circuits provide also a different perspective on constructing non-monotone circuits. While the methods described in (Sladek et al., 2023; Loconte et al., 2024c;b; Wang & Van den Broeck, 2024) regard such circuits as sum of squares, we interpret them as probabilistic events described by positive semidefinite matrices that are combined within a circuit using unit preserving bilinear forms. As such, we also establish a strong link between the circuit literature and quantum information theory.

6.2. Tensor Networks

As already pointed out by Loconte et al. (2024c;b) squared circuits and monotone circuits share many similarities with *tensor networks* (Orús, 2014; White, 1992) developed in the condensed matter physics community and have in recent years also been applied to tackle problems in supervised as well as unsupervised machine learning (Cheng et al., 2019; Han et al., 2018; Stoudenmire & Schwab, 2016). In this regard and given that tensor networks originate in the physics community, it is rather surprising that tensor networks have so far, and to the best of our knowledge, not been formulated using POVMs.

Using our POVM formulation for PUnCs is, however, not only theoretically elegant but has also practical benefits: using this formalism leads to circuits that are normalized by construction, and we can perform learning simply by optimizing the likelihood. We contrast this to the sweeping algorithms that are usually deployed in the tensor network literature, where blocks of variables are optimized one at the time while the remaining variables are held constant. It appears that this approach is inspired by the variational ansatz taken in the density-matrix renormalization group algorithm (White, 1992), which is the original algorithm for tensor networks. Alternatives to this sweeping approach have also been proposed, such as an intricate Riemannian gradient descent optimization scheme that conserves unitarity of matrices, but are rather costly to run.

We would also like to point out a theoretical result from the tensor network literature stating that picking a complex-valued parametrization instead of a real-valued one can lead to an arbitrarily large reduction in the number of parameters (Glasser et al., 2019). While this result is formulated with respect to (exact) low rank tensor decomposition and does not apply directly to the problem of learning parameters via gradient descent, we consider this observation to be a strong theoretical indicator for the superiority of complex numbers over real numbers when parametrizing probabilistic circuits. A similar argument has also been made by Gao et al. (2022) in the context of hidden Markov models.

6.3. Theoretical Studies

First theoretical results on expressive power of polynomial functions, i.e. circuits and tensor networks, were presented in the tensor network literature in the context of tensor decomposition (Glasser et al., 2019) and complex-valued hidden Markov models (Gao et al., 2022). Recently, the works of Loconte et al. (2024c;b) and Wang & Van den Broeck (2024) have studied the relationship of different circuits classes more carefully, as well. Additionally, Loconte et al. (2024b) pointed out links between tensor networks and the circuit literature and were able to generalize earlier results from the tensor network literature by Glasser et al. (2019).

Finally, we would also like to point out theoretical results n the statistical relational AI literature. Specifically, Buchman & Poole (2017b), and Kuželka (2020) noted that using only real-valued parametrizations (including negatives (Buchman & Poole, 2017a)), i.e. monotone functions exclusively, does not allow for fully expressive models.

7. Conclusions & Future Work

Based on first principles from quantum information theory, we constructed positive operator circuits – a novel class of probabilistic tractable models. A main hurdle to voercome to make PUnCs practical machine learning models is to find expressive yet efficient functional forms of the bilinear forms and their parametrizations. A key challenge to overcome in this respect is the efficient parametrization of (expressive) unitary matrices as this is a notoriously difficult feat (Kiani et al., 2022; Jing et al., 2017; Lezcano-Casado & Martınez-Rubio, 2019; Mhammedi et al., 2017; Wisdom et al., 2016). Furthermore, we have only presented a specific functional form for the bilinear forms.

cite riguzzi quantum stuff here

future work, are there tractable circuits for which a exponential quantum advantage exists?

References

Agarwal, S. and Bläser, M. Probabilistic generating circuits-demystified. In *Forty-first International Conference on Machine Learning*, 2024.

Buchman, D. and Poole, D. Negative probabilities in probabilistic logic programs. *International Journal of Approximate Reasoning*, 83:43–59, 2017a.

Buchman, D. and Poole, D. Why rules are complex: Real-valued probabilistic logic programs are not fully expressive. In *UAI*, 2017b.

Carroll, J. D. and Chang, J.-J. Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young decomposition. *Psychometrika*, 35(3):283–319, 1970.

Cheng, S., Wang, L., Xiang, T., and Zhang, P. Tree tensor networks for generative modeling. *Physical Review B*, 99 (15):155131, 2019.

Darwiche, A. Decomposable negation normal form. *Journal of the ACM (JACM)*, 48(4):608–647, 2001a.

Darwiche, A. On the tractable counting of theory models and its application to truth maintenance and belief revision. *Journal of Applied Non-Classical Logics*, 11(1-2): 11–34, 2001b.

Darwiche, A. A differential approach to inference in bayesian networks. *Journal of the ACM*, 50(3):280–305, 2003.

Darwiche, A. Sdd: A new canonical representation of propositional knowledge bases. In *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.

de Colnet, A. and Mengel, S. A compilation of succinctness results for arithmetic circuits. In *International Conference on Principles of Knowledge Representation and Reasoning, KR 2021*, 2021.

Dechter, R. and Mateescu, R. And/or search spaces for graphical models. *Artificial intelligence*, 171(2-3): 73–106, 2007.

Gao, X., Anschuetz, E. R., Wang, S.-T., Cirac, J. I., and Lukin, M. D. Enhancing generative models via quantum correlations. *Physical Review X*, 12(2):021037, 2022.

Glasser, I., Sweke, R., Pancotti, N., Eisert, J., and Cirac, I. Expressive power of tensor-network factorizations for probabilistic modeling. In *Advances in neural information processing systems*, 2019.

Han, Z.-Y., Wang, J., Fan, H., Wang, L., and Zhang, P. Unsupervised generative modeling using matrix product states. *Physical Review X*, 8(3):031012, 2018.

Harviainen, J., Ramaswamy, V. P., and Koivisto, M. On inference and learning with probabilistic generating circuits. In *Uncertainty in Artificial Intelligence*, pp. 829–838. PMLR, 2023.

Jing, L., Shen, Y., Dubcek, T., Peurifoy, J., Skirlo, S., LeCun, Y., Tegmark, M., and Soljačić, M. Tunable efficient unitary neural networks (eunn) and their application to rnns. In *International Conference on Machine Learning*, pp. 1733–1741. PMLR, 2017.

Kiani, B., Balestriero, R., LeCun, Y., and Lloyd, S. projunn: efficient method for training deep networks with unitary matrices. *Advances in Neural Information Processing Systems*, 35:14448–14463, 2022.

Kraus, K., Böhm, A., Dollard, J. D., and Wootters, W. *States, Effects, and Operations Fundamental Notions of Quantum Theory: Lectures in Mathematical Physics at the University of Texas at Austin*. Springer, 1983.

Kuzelka, O. Complex markov logic networks: Expressivity and liftability. In *Conference on Uncertainty in Artificial Intelligence*, pp. 729–738. PMLR, 2020.

Lezcano-Casado, M. and Martınez-Rubio, D. Cheap orthogonal constraints in neural networks: A simple parametrization of the orthogonal and unitary group. In *International Conference on Machine Learning*, pp. 3794–3803. PMLR, 2019.

Loconte, L. and Vergari, A. On faster marginalization with squared circuits via orthonormalization. *arXiv preprint arXiv:2412.07883*, 2024.

Loconte, L., Mari, A., Gala, G., Peharz, R., de Campos, C., Quaeghebeur, E., Vessio, G., and Vergari, A. What is the relationship between tensor factorizations and circuits (and how can we exploit it)? *arXiv preprint arXiv:2409.07953*, 2024a.

Loconte, L., Mengel, S., and Vergari, A. Sum of squares circuits. *arXiv preprint arXiv:2408.11778*, 2024b.

Loconte, L., Sladek, A. M., Mengel, S., Trapp, M., Solin, A., Gillis, N., and Vergari, A. Subtractive mixture models via squaring: Representation and learning. In *Proceedings of the International Conference on Learning Representations*, 2024c.

Martens, J. and Medabalimi, V. On the expressive efficiency of sum product networks. *arXiv preprint arXiv:1411.7717*, 2014.

Mhammedi, Z., Hellicar, A., Rahman, A., and Bailey, J. Efficient orthogonal parametrisation of recurrent neural networks using householder reflections. In *International Conference on Machine Learning*, pp. 2401–2409. PMLR, 2017.

Nielsen, M. A. and Chuang, I. L. Quantum computation and quantum information. *Phys. Today*, 54(2):60, 2001.

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

Peharz, R., Vergari, A., Stelzner, K., Molina, A., Shao, X., Trapp, M., Kersting, K., and Ghahramani, Z. Random sum-product networks: A simple and effective approach to probabilistic deep learning. In *Uncertainty in Artificial Intelligence*, 2019.

Peharz, R., Lang, S., Vergari, A., Stelzner, K., Molina, A., Trapp, M., Van den Broeck, G., Kersting, K., and Ghahramani, Z. Einsum networks: Fast and scalable learning of tractable probabilistic circuits. In *International Conference on Machine Learning*, 2020.

Pipatsrisawat, K. and Darwiche, A. New compilation languages based on structured decomposability. In *AAAI*, volume 8, pp. 517–522, 2008.

Poon, H. and Domingos, P. Sum-product networks: A new deep architecture. In *2011 IEEE International Conference on Computer Vision Workshops*. IEEE, 2011.

Shih, A., Sadigh, D., and Ermon, S. Hyperspns: Compact and expressive probabilistic circuits. *Advances in Neural Information Processing Systems*, 2021.

Shpilka, A. and Yehudayoff, A. Arithmetic circuits: A survey of recent results and open questions. *Foundations and Trends® in Theoretical Computer Science*, 5(3–4): 207–388, 2010.

Sladek, A. M., Trapp, M., and Solin, A. Encoding negative dependencies in probabilistic circuits. In *The 6th Workshop on Tractable Probabilistic Modeling*, 2023.

Stoudenmire, E. and Schwab, D. J. Supervised learning with tensor networks. *Advances in neural information processing systems*, 2016.

Valiant, L. G. Negation can be exponentially powerful. In *Proceedings of the eleventh annual ACM symposium on theory of computing*, pp. 189–196, 1979.

Vergari, A., Choi, Y., Liu, A., Teso, S., and Van den Broeck, G. A compositional atlas of tractable circuit operations for probabilistic inference. In *Advances in Neural Information Processing Systems*, 2021.

Visick, G. A quantitative version of the observation that the hadamard product is a principal submatrix of the kronecker product. *Linear Algebra and Its Applications*, 304(1-3):45–68, 2000.

von Neumann, J. Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1927:245–272, 1927.

Wang, B. and Van den Broeck, G. On the relationship between monotone and squared probabilistic circuits. In *The 7th Workshop on Tractable Probabilistic Modeling*, 2024.

White, S. R. Density matrix formulation for quantum renormalization groups. *Physical review letters*, 69(19):2863, 1992.

Wisdom, S., Powers, T., Hershey, J., Le Roux, J., and Atlas, L. Full-capacity unitary recurrent neural networks. *Advances in neural information processing systems*, 29, 2016.

Zhang, H., Holtzen, S., and Broeck, G. On the relationship between probabilistic circuits and determinantal point processes. In *Conference on Uncertainty in Artificial Intelligence*, pp. 1188–1197. PMLR, 2020.

Zuidberg Dos Martires, P. Probabilistic neural circuits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

Appendix

A. Proofs

A.1. Proof of Proposition 2.5

Proposition 2.5. *The expression in Equation 2 defines a valid probability distribution.*

Proof. While this is a well-known result we were not able to identify a concise proof in the literature and provide therefore, for the sake of completeness here. To this end, we first show that $p(i) \geq 1$, for each i.

$$\begin{aligned} p(i) = \mathrm{Tr}[E(i)\rho] &= \mathrm{Tr}[D(i)D^*(i)\rho] \\ &= \mathrm{Tr}[D(i)^*\rho D(i)]. \end{aligned} \tag{17}$$

Here we used the fact that $E(i)$ is PSD and factorized it into the product $D(i)D^*(i)$. Then we used the fact that the trace is invariant under cyclical shifts. Clearly, the matrix $D(i)^*\rho D(i)$ is PSD as we have for every vector x:

$$x^* D(i)^* \rho D(i) x = y^* \rho y \geq 0 \quad \text{with } y = Dx. \tag{18}$$

As the trace of a PSD matrix is positive we have that $p(i)$ is a positive real number, i.e. $p(i){\geq}0$ for every i.

Secondly, we show that $p(i)$ is normalized:

$$\sum_{i=1}^{N} p(i) = \sum_{i=1}^{N} \mathrm{Tr}[E(i)\rho] = \mathrm{Tr}\left[\sum_{i=1}^{N} E(i)\rho\right] = \mathrm{Tr}[\rho], \tag{19}$$

where we used Equation 1. Exploiting the fact that the trace of a density matrix is one gives us indeed $\sum_{i=1}^{N} p(i) = 1$, and we can conclude that $p(i)$ is a valid probability distribution. □

A.2. Proof of Proposition 3.5

Proposition 3.5. *Unital quantum operations are valid in the sense that the inequality $\sum_j K_j^* K_j \leq \mathbb{1}$ holds for all unital quantum operations.*

Proof. The (sufficient and necessary) condition that $\sum_j K_j^* K_j \leq \mathbb{1}$ stems from fact that we wish to bound the probability of a state $\Phi(E(i))$ is less than or equal to 1, cf. (Nielsen & Chuang, 2001, Proof of Theorem 8.1). That is, we wish to have:

$$\mathrm{Tr}[\Phi(E(i))\rho] \leq 1. \tag{20}$$

We will now show that for unital quantum operations this holds by construction and that the condition $\sum_j K_j^* K_j$ is implied. We start with the probability for the state $\sum_i E(i) = \mathbb{1}$:

$$\mathrm{Tr}[\Phi(\mathbb{1})\rho] = \mathrm{Tr}\left[\sum_j K_j \mathbb{1} K_j^* \rho\right] = \mathrm{Tr}\left[\sum_j K_j K_j^* \rho\right] = \mathrm{Tr}[\rho] = 1 \tag{21}$$

Alternatively, we write this also as:

$$\mathrm{Tr}[\Phi(\mathbb{1})\rho] = \mathrm{Tr}\left[\sum_j K_j \mathbb{1} K_j^* \rho\right] = \sum_j \mathrm{Tr}\big[K_j \mathbb{1} K_j^* \rho\big] = \sum_j \mathrm{Tr}\big[\gamma^* K_j \mathbb{1} K_j^* \gamma\big] \tag{22}$$

with $\rho = \gamma\gamma^*$.

Similarly we also write for the probability of the arbitrary state σ denoting the sum over any subset of the POVM $\{E(i)\}_{i=1}^{N}$:

$$\mathrm{Tr}[\Phi(\sigma)\rho] = \sum_j \mathrm{Tr}\big[\gamma^* K_j \sigma K_j^* \gamma\big]. \tag{23}$$

We now need that $\mathrm{Tr}[\Phi(\sigma)\rho] \leq 1$, which is equivalent to:

$$1 - \mathrm{Tr}[\Phi(\sigma)\rho] \geq 0 \Leftrightarrow \mathrm{Tr}[\Phi(\mathbb{1})\rho] - \mathrm{Tr}[\Phi(\sigma)\rho] \geq 0 \tag{24}$$

$$\Leftrightarrow \sum_j \mathrm{Tr}\left[\gamma^* K_j \mathbb{1} K_j^* \gamma\right] - \sum_j \mathrm{Tr}\left[\gamma^* K_j \sigma K_j^* \gamma\right] \geq 0 \tag{25}$$

$$\Leftrightarrow \sum_j \mathrm{Tr}\left[\gamma^* K_j \left(\mathbb{1} - \sigma\right) K_j^* \gamma\right] \geq 0 \tag{26}$$

The last line only holds if $\mathbb{1} - \sigma$ is PSD, which is indeed the case as σ only sums over a subset of the POVM. Substracting the sum of this subset from $\mathbb{1}$ leaves us with a sum over the remaining elements of the POVM. As this is a a sum over PSD matrices the sum over the remaining elements is again PSD. This concludes the proof as we have shown that Equation 20 is satisfied by construction. □

A.3. Proof of Proposition 3.7

Proposition 3.7. *Let $\mathbf{X}$ denote a set of random variables with sample space $\Omega(\mathbf{X})$. Then the set $\{\widehat{O}(\mathbf{x})\}_{\mathbf{x}\in\Omega(\mathbf{X})}$ of positive unital circuits forms a POVM.*

Proof. Given that positive unital circuits are by definition positive operator circuits we already have that:

$$\forall \mathbf{x} \in \Omega(\mathbf{X}) : \widehat{O}(\mathbf{x}) \text{ is PSD.} \tag{27}$$

Next we show that $\sum_{\mathbf{x}\in\Omega(\mathbf{X})} \widehat{O}(\mathbf{x}) = \mathbb{1}$. Here we observe that in the computation units we have:

$$\begin{aligned}\sum_{\mathbf{x}_k\in\Omega(\mathbf{X}_k)} \widehat{O}_k(\mathbf{x}_k) &= \sum_{\mathbf{x}_{k_l}}\sum_{\mathbf{x}_{k_r}} \Phi(\widehat{O}_k(\mathbf{x}_{k_l}) \otimes \widehat{O}_k(\mathbf{x}_{k_r})) \qquad (28)\\ &= \Phi\left(\sum_{\mathbf{x}_{k_l}} \widehat{O}_k(\mathbf{x}_{k_l}) \otimes \sum_{\mathbf{x}_{k_r}} \widehat{O}_k(\mathbf{x}_{k_r})\right)\end{aligned}$$

This lets us push down the summation of a specific variable to the corresponding leaf where the variable is given as input, where we then have:

$$\sum_{\mathbf{x}_k\in\Omega(\mathbf{X}_k)} \widehat{O}_k(\mathbf{x}_k) = \sum_{x_k\in\Omega(X_k)} E_{x_k} = \mathbb{1}_{\mathbb{k}} \tag{29}$$

We now exploit that the completely positive maps in a positive unital circuit are unital, which gives us indeed $\sum_{\mathbf{x}\in\Omega(\mathbf{X})} \widehat{O}(\mathbf{x}) = \mathbb{1}$. □

A.4. Proof of Proposition 4.3

Proposition 4.3. *For computation units of a pure positive unital circuit and a PSD circuit it holds that*

$$\forall k : O_k(\mathbf{x}_k) = v_k(\mathbf{x}_k) \otimes v_k^*(\mathbf{x}_k). \tag{16}$$

given that $U_k = K_k$

Proof. We start the proof at the leaf where we have

$$O_k(x_k) = U_k\left(e_{x_k} \otimes e_{x_k}^*\right) U_k^* = (U_k e_{x_k}) \otimes \left(e_{x_k}^* U_k^*\right) \tag{30}$$

and Equation 16 holds almost by definition. In the computation units into which the leaves feed, we then have

$$\begin{aligned} O_k &= U_k\left(O_{k_l} \otimes O_{k_r}\right) U_k^* \\ &= U_k\left(v_{k_l} \otimes v_{k_l}^* \otimes (v_{k_r} \otimes v_{k_r}^*)\right) U_k^* \\ &= \left(U_k\left(v_{k_l} \otimes v_{k_r}\right)\right) \otimes \left(\left(v_{k_l}^* \otimes v_{k_r}^*\right) U_k^*\right) \\ &= v_k \otimes v_k^*. \end{aligned} \tag{31}$$

where we omitted the explicit dependencies on the variables x_k, x_{k_l}, and x_{k_r}. Repeating this argument recursively until the root of the circuit concludes the proof. □

A.5. Proof of Proposition 4.5

Proposition 4.5. *A PSD circuit and a pure PUnC encode the same probability distribution if* $U_k = K_k$ *for each unit* k.

Proof. The proof starts by simply plugging in the vector representation of O_{root} (obtained in the proof of Proposition 4.3) into the expression $\mathrm{Tr}[O_{\text{root}}\rho]$ and rather straightforwardöy get:

$$\mathrm{Tr}[O_{\text{root}}\rho] = \mathrm{Tr}\left[\left(v_{\text{root}} \otimes v^*_{\text{root}}\right) \times \rho\right] \tag{32}$$

$$= v^*_{\text{root}} \times \rho \times v_{\text{root}}, \tag{33}$$

which is indeed the same probability as defined in Defintion 4.1. □