Quantum Tensor Networks, Stochastic Processes, and Weighted Automata

24th International Conference on Artificial Intelligence and Statistics, 2021

Sandesh Adhikary* — University of Washington
Siddarth Srinivasan* — University of Washington
Jacob Miller — MILA & DIRO, Université de Montréal
Guillaume Rabusseau — MILA & DIRO, Université de Montréal
Byron Boots — University of Washington

* denotes equal contribution

Link to the full paper

Abstract

We show equivalence between various probabilistic models developed independently in the physics, machine learning, and formal language communities.

▶ Modeling joint probability distributions over sequences has been studied from many perspectives.

▶ Specifically, *matrix product states (MPS)* [1] and associated quantum tensor networks in physics bear striking resemblance to stochastic processes and weighted automata.

▶ We show how exactly these model classes relate, focusing on the uniform (stationary) versions of quantum tensor networks.

The Non-Terminating Limit

▶ Within quantum tensor networks, we focus on *uniform* MPS where all tensor cores are identical.

▶ The probability of any sequence $y_1, \cdots, y_N$ in a non-terminating uMPS is computed by marginalizing over infinitely many future observations

$$P(y_1, \cdots, y_T) = \lim_{N\to\infty} \sum_{y_N} \cdots \sum_{y_{T+1}} P(y_1, \cdots, y_T, y_{T+1}, \cdots, y_N)$$

▶ We assume the tensor networks are semi-infinitely long, so marginal probabilities computed from the fixed boundary to any core are not affected by the distance from the boundary.

Quantum Tensor Networks

▶ **uBM = Uniform Born Machine**
▶ **uLPS = Uniform Locally Purified State**
▶ **uMPS = Uniform Matrix Product State**

Stochastic Processes

▶ **HMM = Hidden Markov Model**
▶ **NOOM = Norm Observable Operator Model**
▶ **HQMM = Hidden Quantum Markov Model**
▶ **PSR = Predictive State Representation**

Weighted Automata

▶ **PA = Probabilistic Automata**
▶ **QWA = Quadratic Weighted Automata**
▶ **SWA = Stochastic Weighted Automata**

uMPS = PSR = SWA

▶ A **uMPS** represents the joint distribution of a sequence $y_1 \cdots y_N$ as a tensor-train decomposition with matrices $\mathbf{A}_{y_i}$, an evaluation functional $\vec{\sigma}^\dagger$, and an initial state vector $\vec{\rho}_0$

$$P(y_1, \cdots, y_N) = \vec{\sigma}^\dagger \mathbf{A}_{y_N} \cdots \mathbf{A}_{y_1} \vec{\rho}_0$$

▶ The *effective* evaluation functional of a **non-terminating uMPS** converges to the fixed point $\vec{\sigma}^{*\dagger}$ of its transfer operator $\sum_y \mathbf{A}_y$

$$N \to \infty \implies \vec{\sigma}^{*\dagger} \textstyle\sum_y (\mathbf{A}_y)^{N-t} \to \vec{\sigma}^{*\dagger} \quad \text{where} \quad \vec{\sigma}^{*\dagger} \sum_y \mathbf{A}_y = \vec{\sigma}^{*\dagger}$$

▶ An (uncontrolled) **PSR** is also a tensor-train decomposition of the joint distribution over sequences, but its evaluation functional $\vec{\sigma}^\dagger$ is constrained to be the fixed point of its transfer operator. PSRs are thus equivalent to uMPS.

▶ An **SWA** is a weighted automaton that computes a probability distribution. These models are also equivalent to PSRs and uMPS.

▶ **The Negative Probability Problem:** It is undecidable whether a given set of PSR, uMPS, or SWA parameters will assign negative probability to some arbitrary length-sequence [2].

uLPS = HQMM

▶ A **uLPS** is a type of uMPS [3] where model parameters are expressed in a special Schmidt decomposition, which ensures that the model always assigns non-negative probabilities $P(y_1, \cdots, y_N)$ computed as follows

$$\underbrace{\left(\sum_{\beta_L} \overline{\mathbf{K}}^T_{\beta_L, L} \otimes \mathbf{K}^T_{\beta_L, L}\right)}_{\vec{\sigma}} \left(\sum_{\beta} \overline{\mathbf{K}}_{\beta, y_N} \otimes \mathbf{K}_{\beta, y_N}\right) \cdots \left(\sum_{\beta} \overline{\mathbf{K}}_{\beta, y_1} \otimes \mathbf{K}_{\beta, y_1}\right) \underbrace{\left(\sum_{\beta_R} \overline{\mathbf{K}}_{\beta_R, R} \otimes \mathbf{K}_{\beta_R, R}\right)}_{\vec{\rho}_0}$$

▶ An **HQMM** is a type of PSR [4] where model parameters are constructed similarly to uLPS, but the evaluation functional is fixed to be the vectorized identity $\vec{\mathbb{I}}^T$. As with a uLPS, HQMMs always return non-negative probabilities $P(y_1, \cdots, y_N)$ computed as

$$\vec{\mathbb{I}}^T \left(\sum_{\beta} \overline{\mathbf{K}}_{\beta, y_N} \otimes \mathbf{K}_{\beta, y_N}\right) \cdots \left(\sum_{\beta} \overline{\mathbf{K}}_{\beta, y_1} \otimes \mathbf{K}_{\beta, y_1}\right) \vec{\rho}_0$$

▶ We demonstrate that non-terminating uLPS and HQMMs are equivalent model classes.

▶ We also show that NOOMs are a *strict* subset of HQMMs or uLPS.

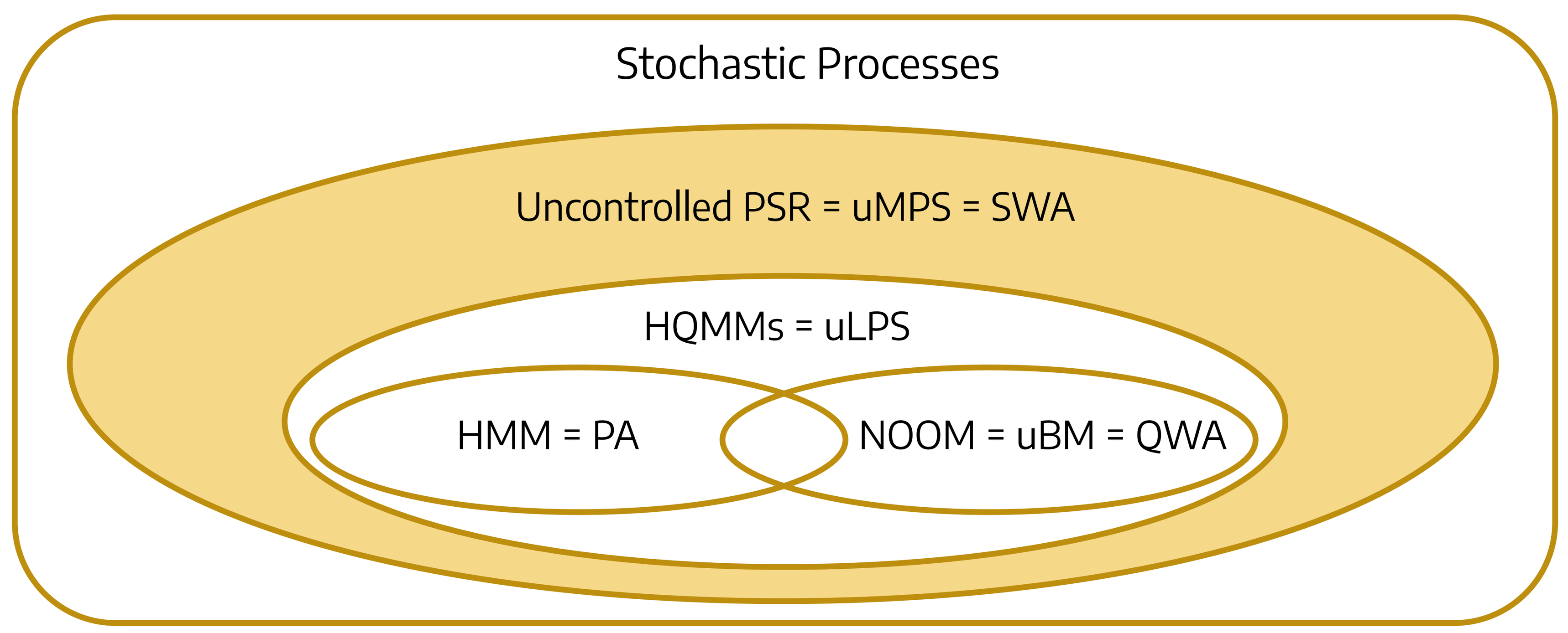

Figure 1 : *All quantum tensor networks assumed to be non-terminating. The shaded area is potentially empty. Two model classes are equivalent if any joint distribution over sequences that can be represented by a model in one class (with finite parameters) can be represented exactly by a model in the other class (also with finite parameters).*

uBM = NOOM = QWA

▶ **uBMs** are a subset of uMPS where probabilities of sequences are computed with an additional the absolute-square to ensure non-negative probabilities.

$$P(y_1, \cdots, y_N) = \left|\vec{\alpha}^{\,\dagger} \mathbf{A}^{y_N} \ldots \mathbf{A}^{y_1} \vec{\omega}_0\right|^2$$

▶ **NOOMs** are a subset of PSRs that enforces non-negative probabilities through the 2-norm.

$$P(y_1, \cdots, y_N) = \left|\left|\boldsymbol{\phi}_{y_N} \cdots \boldsymbol{\phi}_{y_1} \vec{\psi}_0\right|\right|_2^2$$

▶ **QWAs** are a subset of SWAs designed to ensure non-negative probabilities through the 2-norm, similar to NOOMs.

▶ We show that non-terminating uBMS, NOOMs and QWA are equivalent model classes.

Non-negative uMPS = HMM = PA

▶ It is well known that an **HMM** is equivalent to a PSR/uMPS where all model parameters are non-negative and $\vec{\sigma}$ is fixed to be the ones vector $\vec{\mathbb{1}}$.

▶ A **PA** is an SWA with non-negative parameters, and is equivalent to an HMM.

▶ HMMs avoid the negative probability problem at the cost of reduced expressiveness.

▶ We show that NOOMs/uBMs are also restrictive model classes as they do not encompass all finite dimensional HMMs i.e. HMM ⊄ NOOM/uBM.

Remarks

▶ Whether the gap between HQMMs/uLPS and PSRs/uMPS is empty is unknown; this gap **is not** empty in the non-uniform case.

▶ Connections raise the possibility of adopting learning algorithms from one framework to another; e.g., learning algorithm for HQMMs builds in physical (TP/CP/HP) constraints, unlike tensor networks

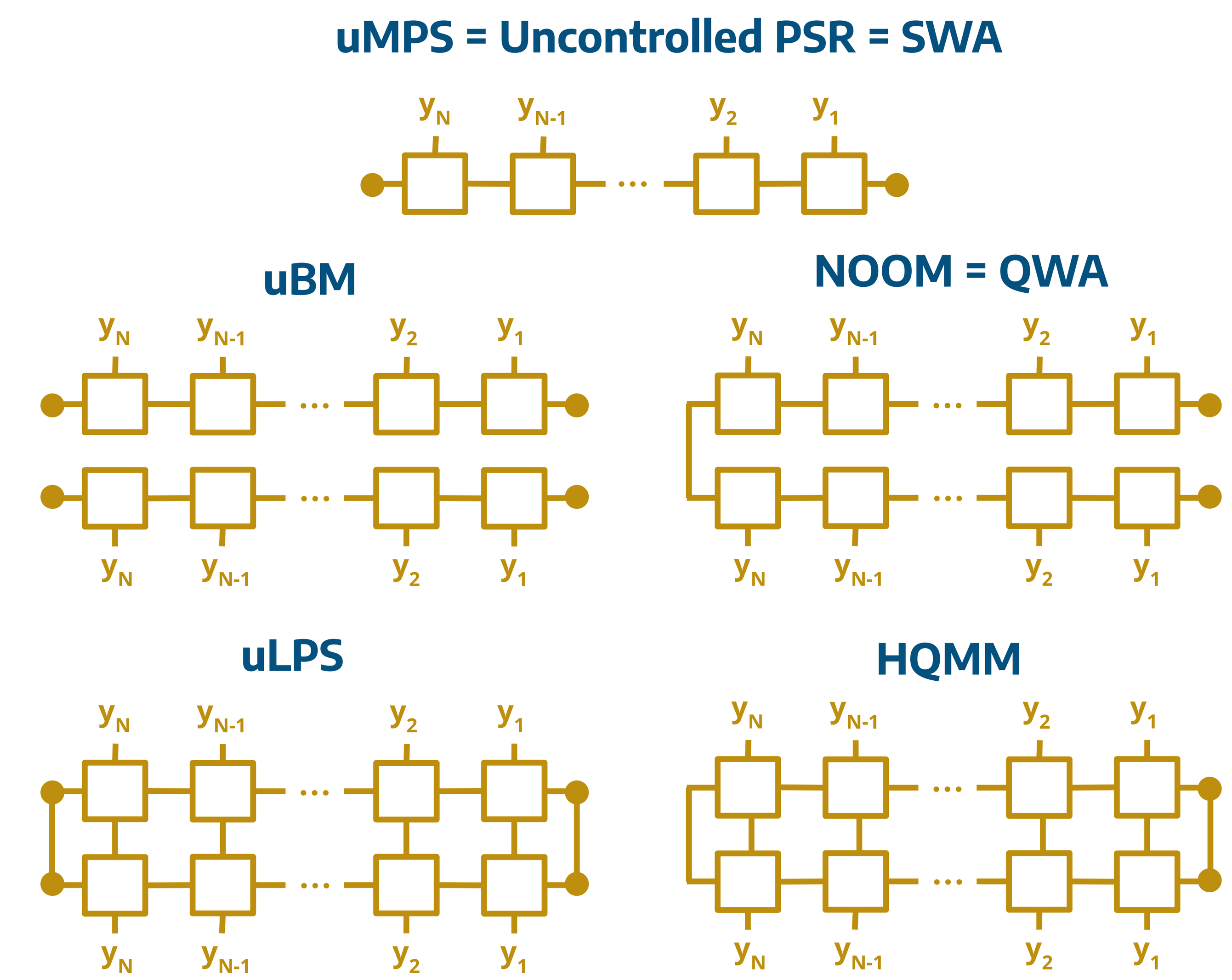

Figure 2 : *Tensor network diagrams of all models considered. At the boundaries, filled circles are vectors and connecting lines (without circles) represent the vectorized identity.*

Summary of Results

▶ **HQMM = uLPS**
▶ **NOOM ⊂ HQMM**
▶ **HMM ⊄ NOOM**
▶ **NOOM = uBM**
▶ **uBM ⊄ HMM**

References

[1] Perez-Garcia, D., Verstraete, F., Wolf, M. M., and Cirac,J. I. (2006). Matrix product state representations. arXiv preprint quant-ph/0608197

[2] Wiewiora, E. W. (2008). Modeling probability distributions with predictive state representations. PhD thesis, UC San Diego.

[3] Glasser, I., Sweke, R., Pancotti, N., Eisert, J., and Cirac,I. (2019). Expressive power of tensor-network factorizations for probabilistic modeling. In Advances in Neural Information Processing Systems.

[4] Adhikary, S., Srinivasan, S., Gordon, G., and Boots., B. (2020). Expressiveness and learning of hidden quantum Markov models. In Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics.