Neural Network Quantum State Tomography

Mo Abedi
Supervised by Chae-Yeun Park

Department of Physics

June 14, 2021

Outline

1. Mathematical Preliminaries
2. RBM Quantum State Tomography
3. RNN Quantum State Tomography
4. Summary

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries

- Basic idea: can we reconstructing quantum states from measurements? How to do so?

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries

- Basic idea: can we reconstructing quantum states from measurements? How to do so?

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries

- Basic idea: can we reconstructing quantum states from measurements? How to do so?

Definition

An informationally complete positive-operator valued measure [1] (POVM), Π_i is the set of operators on $\mathcal{H}$ such that:

$$\Pi_i \geq 0 \quad \text{Semi-Positivity}$$

$$\sum_i \Pi_i = 1$$

$$p_i = \mathrm{Tr}\left(\rho \Pi_i\right) \quad \text{Born rule}$$

$$\{\Pi_i\} = \mathrm{span}(\mathcal{B}(\mathcal{H})) \quad \text{Informational Completeness}$$

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries

- Given informational completeness, we can invert this relationship

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries

- Given informational completeness, we can invert this relationship

Preposition

The density operator can be written as:

$$\rho = \sum_{ij} p_i T_{ij}^{-1} \Pi_j$$

Where:

$$T_{ij} = \mathrm{Tr}\left(\Pi_i \Pi_j\right)$$

Is the called the overlap matrix.

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries

- Given informational completeness, we can invert this relationship

Preposition

The density operator can be written as:

$$\rho = \sum_{ij} p_i T_{ij}^{-1} \Pi_j$$

Where:

$$T_{ij} = \mathrm{Tr}\left(\Pi_i \Pi_j\right)$$

Is the called the overlap matrix.

- Note that the overlap matrix may not be invertable

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries

Example

A set of POVMs are given by Pauli-4 Tetrahedral operators:

$$\Pi_{Tetra} = \left\{ \Pi_i = \frac{1}{4}(1 + s_i \cdot \sigma) \right\}$$

$$s_0 = (0,0,1) \quad s_1 = \left(\frac{2\sqrt{2}}{3}, 0, -\frac{1}{3} \right)$$

$$s_2 = \left(-\frac{2\sqrt{2}}{2}, \sqrt{\frac{2}{3}}, \frac{1}{3} \right) \quad s_3 = \left(-\frac{2\sqrt{2}}{3}, -\sqrt{\frac{2}{3}}, -\frac{1}{3} \right)$$

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries

Example

A set of POVMs are given by Pauli-4 Tetrahedral operators:

$$\Pi_0 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \Pi_2 = \frac{1}{12}\begin{pmatrix} 2 & -\sqrt{2} - \sqrt{6}i \\ -\sqrt{2} + \sqrt{6}i & 4 \end{pmatrix}$$

$$\Pi_1 = \frac{1}{6}\begin{pmatrix} 1 & \sqrt{2} \\ \sqrt{2} & 2 \end{pmatrix} \quad \Pi_3 = \frac{1}{4}\begin{pmatrix} 2 & -\sqrt{2} + \sqrt{6}i \\ -\sqrt{2} - \sqrt{6}i & 4 \end{pmatrix}$$

$$T^{-1} = \begin{pmatrix} 5 & -1 & -1 & -1 \\ -1 & 5 & -1 & -1 \\ -1 & -1 & 5 & -1 \\ -1 & -1 & -1 & 5 \end{pmatrix}$$

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries

- Well then what is the problem? Experimentally, you cannot know probabilities p_i, you can only measure frequencies f_i.

Mathematical Preliminaries

- Well then what is the problem? Experimentally, you cannot know probabilities p_i, you can only measure frequencies f_i.
- Number of operators scales exponentially as 4^N different operators for N qubits.

Mathematical Preliminaries

- Well then what is the problem? Experimentally, you cannot know probabilities p_i, you can only measure frequencies f_i.
- Number of operators scales exponentially as 4^N different operators for N qubits.

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries

- Well then what is the problem? Experimentally, you cannot know probabilities p_i, you can only measure frequencies f_i.
- Number of operators scales exponentially as 4^N different operators for N qubits.

Definition

The likelihood is a measure of the degree of belief in the hypothesis that for a particular data set $\mathcal{D}$, the system was prepared in the quantum state ρ [2]. For QST, we have the multinomial distribution:

$$\mathcal{L}(\mathcal{D}|\rho) = \mathcal{N} \prod_{i}^{k} p_i^{f_i} = \frac{n!}{\prod_i^k f_i!} \prod_{i}^{k} \mathrm{Tr}(\rho \Pi_i)^{f_i}$$

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries

- Consider the negative log instead which we call the cost function

$$\mathcal{C} = -\sum_i f_i \ln p_i = -\sum_i f_i \ln \left[\mathrm{Tr}\left(\rho \Pi_i\right)\right] = D_{KL}\left(\mathcal{D}|\rho\right) + \mathbb{H}_{\mathcal{D}}$$

Where D_{KL} is the KL divergence, a measure of how close the actual probability distribution is to our measured data. We want to minimise this.

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries

- Before talking about NN, I will give an example of an MLE algorithm by (Hradil et al., 2004). Idea from calculus of variations

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries

- Before talking about NN, I will give an example of an MLE algorithm by (Hradil et al., 2004). Idea from calculus of variations

Preposition

The variation of $\mathcal{C}$ w.r.t. ρ is given by:

$$\delta\mathcal{C}(\rho) = \mathcal{C}(\rho - \delta\rho) - \mathcal{C}(\rho) = \text{Tr}\left((R-1)\rho(R-1)\right)$$
$$\delta\rho = (R-1)\,\rho + \rho\,(R-1)$$

And is 0 when:

$$R\rho = \rho R = \rho$$

Where:

$$R = -\sum_i \frac{f_i}{p_i}\Pi_i$$

Mo Abedi
Department of Physics
University of Cologne

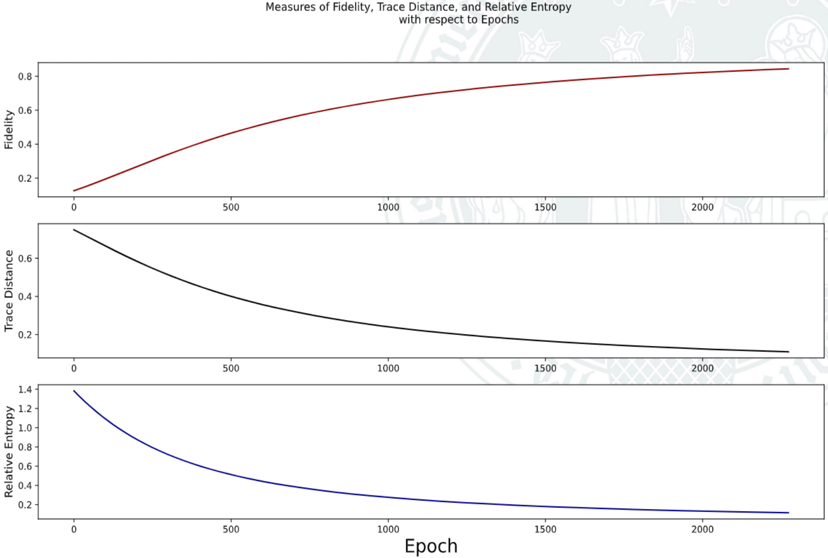
Mathematical Preliminaries

Algorithm ($R\rho R$)

- Start with maximally mixed state $\rho = \frac{1}{\dim(\mathcal{H})}\mathbb{1}$ and some precision ϵ and set the trace distance $TD > \epsilon$
- While $TD > \epsilon$:
  Calculate $R_{(k)}$
  Compute trace distance $\frac{1}{2}\mathrm{Tr}\left(|R_{(k)}\rho_{(k)} - \rho_{(k)}|\right) = TD$
  Compute $\delta\rho_{(k)} = \left(\left(R_{(k)} - 1\right)\right)\rho_{(k)} + \rho_{(k)}\left(R_{(k)} - 1\right)\right)$
  Update $\rho_{(k+1)} = \rho_{(k)} + \alpha\delta\rho_{(k)}$

Mo Abedi
Department of Physics
University of Cologne

Mathematical Preliminaries



Mo Abedi
Department of Physics
University of Cologne

RBM QST

- Generative models are suitable for this problem. Our implementation based on (Carrasquilla et al., 2019).

Mo Abedi
Department of Physics
University of Cologne

RBM QST

- Generative models are suitable for this problem. Our implementation based on (Carrasquilla et al., 2019).
- The categorical input data is one-hot encoded:

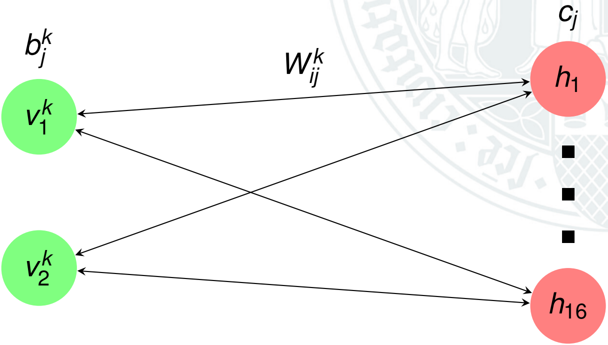
$$\begin{pmatrix} 0 \\ 3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Mo Abedi
Department of Physics
University of Cologne

RBM QST

- Generative models are suitable for this problem. Our implementation based on (Carrasquilla et al., 2019).
- The categorical input data is one-hot encoded:

$$\begin{pmatrix} 0 \\ 3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$



-

Mo Abedi
Department of Physics
University of Cologne

RBM QST

- Want to minimise the cost function (KL) divergence. Generally hard so instead take cost function to be (CD) (Salakhutdinov et al., 2007) (the difference between KL divergences at 0 and k Gibbs sample steps)

Mo Abedi
Department of Physics
University of Cologne

RBM QST

- Want to minimise the cost function (KL) divergence. Generally hard so instead take cost function to be (CD) (Salakhutdinov et al., 2007) (the difference between KL divergences at 0 and k Gibbs sample steps)
- The activation probabilities are calculated as follows:

$$p\left(v_i^k|h,\lambda\right) = \frac{\exp\left(b_i^k + \sum_j h_j W_{ij}^k\right)}{\sum_l \exp\left(b_i^l + \sum_j h_j W_{ij}^l\right)} = \text{sm}\left(b_i^k + \sum_j h_j W_{ij}^k\right)$$

$$p\left(h_j|v,\lambda\right) = \sigma\left(c_j + \sum_i \sum_k v_i^k W_{ij}^k\right)$$

Mo Abedi
Department of Physics
University of Cologne

RBM QST

- Want to minimise the cost function (KL) divergence. Generally hard so instead take cost function to be (CD) (Salakhutdinov et al., 2007) (the difference between KL divergences at 0 and k Gibbs sample steps)
- The activation probabilities are calculated as follows:

$$p\left(v_i^k|h,\lambda\right) = \frac{\exp\left(b_i^k + \sum_j h_j W_{ij}^k\right)}{\sum_l \exp\left(b_i^l + \sum_j h_j W_{ij}^l\right)} = \text{sm}\left(b_i^k + \sum_j h_j W_{ij}^k\right)$$

$$p\left(h_j|v,\lambda\right) = \sigma\left(c_j + \sum_i \sum_k v_i^k W_{ij}^k\right)$$

- After calculating probabilities, we sample using Binomial/Bernoulli or Multinomial/Categorical distributions

Mo Abedi
Department of Physics
University of Cologne

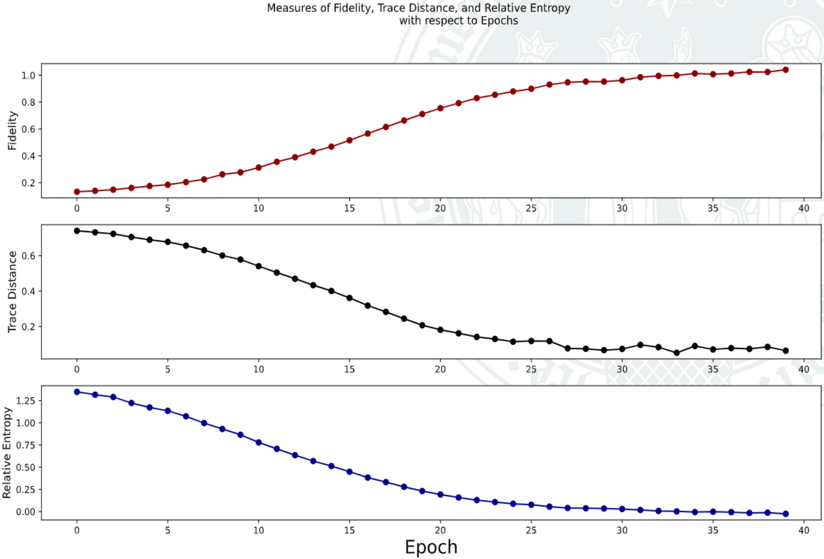
RBM QST

Algorithm (CD_k algorithm for Quantum State Tomography)

- For epoch in total epochs:
- For mini-batch in training data:
- From input data v, sample h after calculating $p(h|v)$
- For k steps:
- Sample v' after calculating $p(v'|h')$ (h for $k = 1$)
- Sample h' after calculating $p(h'|v')$
- Calculate $\left\langle \frac{\partial E(v,h)}{\partial \lambda} \right\rangle - \left\langle \frac{\partial E(v',h')}{\partial \lambda} \right\rangle$
- Update parameters λ

Mo Abedi
Department of Physics
University of Cologne

RBM QST



Mo Abedi
Department of Physics
University of Cologne

RBM QST

- Dont always need the density matrix. Can estimate expectations values directly:

$$\langle \mathcal{O} \rangle = \sum_i \mathcal{Q}_i^{\mathcal{O}} p_i$$

$$\mathcal{Q}_i^{\mathcal{O}} = \sum_j \mathrm{Tr}\left(\mathcal{O}\Pi_j\right) T_{ji}^{-1}$$

Mo Abedi
Department of Physics
University of Cologne

RBM QST

- Dont always need the density matrix. Can estimate expectations values directly:

$$\langle \mathcal{O} \rangle = \sum_i \mathcal{Q}_i^{\mathcal{O}} p_i$$

$$\mathcal{Q}_i^{\mathcal{O}} = \sum_j \mathrm{Tr}\left(\mathcal{O}\Pi_j\right) T_{ji}^{-1}$$

- Have to take into account local depolarising noise.

Mo Abedi
Department of Physics
University of Cologne

RBM QST

- Dont always need the density matrix. Can estimate expectations values directly:

$$\langle \mathcal{O} \rangle = \sum_i \mathcal{Q}_i^{\mathcal{O}} p_i$$

$$\mathcal{Q}_i^{\mathcal{O}} = \sum_j \mathrm{Tr}\left(\mathcal{O}\Pi_j\right) T_{ji}^{-1}$$

- Have to take into account local depolarising noise.
- In general, RBMs take long time to train for small noise [3].

Mo Abedi
Department of Physics
University of Cologne

RNN QST

- Sequential feed forward networks with notion of "memory" [5]. Used often in word prediction and machine translation [6].

Mo Abedi
Department of Physics
University of Cologne

RNN QST

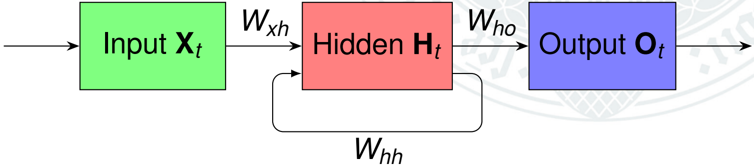
- Sequential feed forward networks with notion of "memory" [5]. Used often in word prediction and machine translation [6].
- Diagrammatically a FNN is :

Mo Abedi
Department of Physics
University of Cologne

RNN QST

- Sequential feed forward networks with notion of "memory" [5]. Used often in word prediction and machine translation [6].
- Diagrammatically a FNN is :

- And RNN:



Mo Abedi
Department of Physics
University of Cologne

RNN QST

- Mathematically:

$$\mathbf{H}_t = \mathcal{F}_1\left(\mathbf{X}_t W_{xh} + \mathbf{H}_{t-1} W_{hh} + b_h\right)$$
$$\mathbf{O}_t = \mathcal{F}_2\left(\mathbf{H}_t W_{ho} + b_o\right)$$

Mo Abedi
Department of Physics
University of Cologne

RNN QST

- Mathematically:

$$\mathbf{H}_t = \mathcal{F}_1\left(\mathbf{X}_t W_{xh} + \mathbf{H}_{t-1} W_{hh} + b_h\right)$$
$$\mathbf{O}_t = \mathcal{F}_2\left(\mathbf{H}_t W_{ho} + b_o\right)$$

- Simple RNNs fail to capture long-term dependencies and cause vanishing or exploding gradients [5].

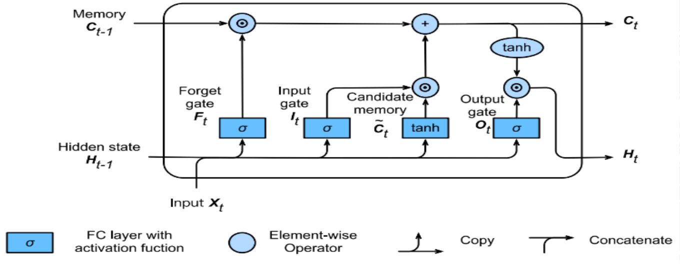
$$\mathcal{C} = \sum_{t=1}^{T} \mathcal{C}_t$$

$$\frac{\partial \mathcal{C}}{\partial W_{hh}} \sim \sum_{k}^{t} \frac{\partial \mathbf{H}_t}{\partial \mathbf{H}_k} \frac{\partial \mathbf{H}_k}{\partial W_{hh}}$$
$$\sim \sum_{k}^{t} (W_{hh}^{T})^{t-k} \mathbf{H}_k$$

Mo Abedi
Department of Physics
University of Cologne

RNN QST

- Have to introduce better long term dependencies. Examples are LSTM and GRU. LSTM have the following structure [5]:



$$\mathbf{O}_t = \sigma\left(\mathbf{X}_t W_{xo} + \mathbf{H}_{t-1} W_{ho} + b_o\right) \quad \tilde{\mathbf{C}}_t = \tanh\left(\mathbf{X}_t W_{xc} + \mathbf{H}_{t-1} W_{hc} + b_c\right)$$

$$\mathbf{I}_t = \sigma\left(\mathbf{X}_t W_{xi} + \mathbf{H}_{t-1} W_{hi} + b_i\right) \quad \mathbf{C}_t = \mathbf{F}_t \circ \mathbf{C}_{t-1} + \mathbf{I}_t \circ \tilde{\mathbf{C}}_t$$

$$\mathbf{F}_t = \sigma\left(\mathbf{X}_t W_{xf} + \mathbf{H}_{t-1} W_{xf} + b_f\right) \quad \mathbf{H}_t = \mathbf{O}_t \circ \tanh(\mathbf{C}_t)$$

Mo Abedi
Department of Physics
University of Cologne

RNN QST

- The cell state represents the long term memory of our network

Mo Abedi
Department of Physics

RNN QST

- The cell state represents the long term memory of our network
- The first "forgetful layer" tells the the cell state what to keep and get rid of

Mo Abedi
Department of Physics
University of Cologne

RNN QST

- The cell state represents the long term memory of our network
- The first "forgetful layer" tells the the cell state what to keep and get rid of
- The next two layers represent the what new information to include in the cell state

Mo Abedi
Department of Physics
University of Cologne

RNN QST

- The cell state represents the long term memory of our network
- The first "forgetful layer" tells the the cell state what to keep and get rid of
- The next two layers represent the what new information to include in the cell state
- The final layer represents the cumulative output of the cell (multiplied with the cell "memory")

Mo Abedi
Department of Physics
University of Cologne

RNN QST

- The cell state represents the long term memory of our network
- The first "forgetful layer" tells the the cell state what to keep and get rid of
- The next two layers represent the what new information to include in the cell state
- The final layer represents the cumulative output of the cell (multiplied with the cell "memory")
- Solves the vanishing gradient as you have extra terms:

$$\frac{\partial \mathcal{C}}{\partial W_{hh}} \sim \frac{\partial \mathbf{C}_t}{\partial \mathbf{C}_k}$$

Mo Abedi
Department of Physics
University of Cologne

RNN QST

- In RNN-QST, the input is a string of N-qubit measurements $(m_1, ..m_n)$ ie. for four qubits, we have $(0, 2, 1, 3)$.

Mo Abedi
Department of Physics
University of Cologne

RNN QST

- In RNN-QST, the input is a string of N-qubit measurements $(m_1, ..m_n)$ ie. for four qubits, we have $(0, 2, 1, 3)$.
- The RNN can predict the n-th measurement given the previous elements, ie. $p(m_4|021)$

Mo Abedi
Department of Physics
University of Cologne

RNN QST

- In RNN-QST, the input is a string of N-qubit measurements $(m_1, ..m_n)$ ie. for four qubits, we have $(0, 2, 1, 3)$.
- The RNN can predict the n-th measurement given the previous elements, ie. $p(m_4|021)$
- Can then recover the full distribution:

$$p(m_1...m_n) = p(m_1)p(m_2|m_1)....p(m_n|m_1...m_{n-1})$$

Mo Abedi
Department of Physics
University of Cologne

RNN QST

- In RNN-QST, the input is a string of N-qubit measurements $(m_1, ..m_n)$ ie. for four qubits, we have $(0, 2, 1, 3)$.
- The RNN can predict the n-th measurement given the previous elements, ie. $p(m_4|021)$
- Can then recover the full distribution:

$$p(m_1...m_n) = p(m_1)p(m_2|m_1)....p(m_n|m_1...m_{n-1})$$

- RNN train faster than RBMs [3]. The required training data for an RNN linearly with N, which is remarkable.

Mo Abedi
Department of Physics
University of Cologne

Summary

- QST is important to find unknown states from measurement and/or certify known states (ie. might be affected by depolarising noise)

Mo Abedi
Department of Physics
University of Cologne

Summary

- QST is important to find unknown states from measurement and/or certify known states (ie. might be affected by depolarising noise)
- Traditional methods like MLE are too slow for large number of qubits

Mo Abedi
Department of Physics
University of Cologne

Summary

- QST is important to find unknown states from measurement and/or certify known states (ie. might be affected by depolarising noise)
- Traditional methods like MLE are too slow for large number of qubits
- Generative modelling from provide ways to earn probability distributions of measurements and are a excellent choice for the problem at hand

Mo Abedi
Department of Physics
University of Cologne

Summary

- QST is important to find unknown states from measurement and/or certify known states (ie. might be affected by depolarising noise)
- Traditional methods like MLE are too slow for large number of qubits
- Generative modelling from provide ways to earn probability distributions of measurements and are a excellent choice for the problem at hand
- RBMs and RNNs can be used for QST, the latter scales linearly with number of qubits

Mo Abedi
Department of Physics
University of Cologne

Bibliography I

[1] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

[2] Zdeněk Hradil, Jaroslav Řeháček, Jaromír Fiurášek, and Miroslav Ježek. *3 Maximum-Likelihood Methodsin Quantum Mechanics*, pages 59–112. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. ISBN 978-3-540-44481-7. doi: 10.1007/978-3-540-44481-7_3. URL https://doi.org/10.1007/978-3-540-44481-7_3.

[3] Juan Carrasquilla, Giacomo Torlai, Roger G. Melko, and Leandro Aolita. Reconstructing quantum states with generative models. *Nature Machine Intelligence*, 1(3): 155–161, Mar 2019. ISSN 2522-5839. doi: 10.1038/s42256-019-0028-1. URL http://dx.doi.org/10.1038/s42256-019-0028-1.

Mo Abedi
Department of Physics
University of Cologne

Bibliography II

[4] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. Restricted boltzmann machines for collaborative filtering. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, page 791–798, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595937933. doi: 10.1145/1273496.1273596. URL https://doi.org/10.1145/1273496.1273596.

[5] Robin M. Schmidt. Recurrent neural networks (rnns): A gentle introduction and overview. *CoRR*, abs/1912.05911, 2019. URL http://arxiv.org/abs/1912.05911.

[6] Shuoheng Yang, Yuxin Wang, and Xiaowen Chu. A survey of deep learning techniques for neural machine translation. *CoRR*, abs/2002.07526, 2020. URL https://arxiv.org/abs/2002.07526.

The End

All questions welcome!

Mo Abedi
Department of Physics
University of Cologne