Efficient estimation and verification of quantum many-body systems

Dissertation zur Erlangung des Doktorgrades Dr. rer. nat.
der Fakultät für Naturwissenschaften der Universität Ulm

vorgelegt von

Milan Holzäpfel

aus Sersheim

2019

Universität Ulm
Fakultät für Naturwissenschaften
Institut für Theoretische Physik

Dekan:	Prof. Dr. Peter Dürre
Erstgutachter:	Prof. Dr. Martin B. Plenio
Zweitgutacher:	Prof. Dr. Joachim Ankerhold
Tag der Prüfung:	29. Juli 2019

Summary

Tensor reconstruction as well as quantum state and process estimation, verification and certification are the main subjects of this thesis. Quantum systems and tensors share the property that the complexity of their description grows exponentially with the number n of subsystems or indices, respectively. We discuss methods which, in many instances, are *efficient* or *scalable* in the sense that the necessary resources grow only polynomially with n.

The restriction to polynomial resources requires that states, processes and tensors are represented in an efficient way. We focus on matrix product state (MPS)/tensor train (TT) representations and also employ projected entangled pair state (PEPS) and hierarchical Tucker representations in some places. In Chapter 4, the resources required by a state estimation method based on MPSs and maximum likelihood estimation (MLE) are discussed and it is shown that polynomial resources are sufficient to achieve a constant estimation error for a particular family of states. In addition, we show that the MPS-MLE method can be adapted to estimate states of infinite-dimensional systems with continuous variables from measurement data of quadrature amplitudes. We further improve the MPS-MLE method by combining it with so-called locally purified MPSs.

The Choi-Jamiołkowski isomorphism provides a way to represent unitary and non-unitary quantum processes as pure and mixed quantum states, respectively. In Chapter 5, we show how this correspondence can be used to construct efficient representations of unitary and general, completely positive quantum processes. Determining a quantum process by estimating a corresponding quantum state is known as ancilla-assisted process tomography (AAPT), where the terms *tomography* and *estimation* are used synonymously. We show that this approach enables efficient quantum process tomography by using existing efficient estimators for quantum states. The feasibility of this approach is demonstrated for unitary processes by means of numerical simulations including unitary circuits and local Hamiltonians.

The interaction range of a Hamiltonian is the maximal distance of subsystems coupled by the Hamiltonian and we call a family of Hamiltonians *local* if its interaction range is bounded by a constant which is independent of n. It was shown that the time evolution induced by a local Hamiltonian on a one-dimensional linear chain can be computed with resources exponential in evolution time but polynomial in n. We extend this result to lattices of arbitrary spatial dimension in Chapter 6, replacing the polynomial scaling with a pseudo-polynomial one. Furthermore, we introduce a verification or certification method for a pure product state which evolves under a local Hamiltonian and show that the necessary resources obey the same scaling. This method relies on a so-called parent Hamiltonian, a concept from the theory of matrix product states.

In Chapter 7, we discuss results from an ion trap experiment performed by Ben Lanyon, Christine Maier and others. The experiment implements time evolution under an approximately local Hamiltonian on 8 and 14 qubits. The resulting state was determined with efficient estimation methods and the estimate was verified with an efficient certification method. The certificate is again derived from a parent Hamiltonian. The results confirm both the capability of the ion trap experiment to act as a high-fidelity quantum simulator and the power of MPS-based methods for state estimation and verification in real-world physical systems.

Tensors are central to representing quantum states, discrete multivariate probability distributions and many other data sets. In Chapter 8, we explore similarities between an efficient quantum state estimation method and several other methods for tensor reconstruction. In this way, an efficient method for reconstructing low-rank tensors is obtained which is applicable to different tensor representations and which combines advantages of several previous proposals.

Publications

This thesis includes material from the following publications and preprints:

- M. Holzäpfel, T. Baumgratz, M. Cramer, and M. B. Plenio (2015). "Scalable reconstruction of unitary processes and Hamiltonians". Phys. Rev. A 91(4), 042129. DOI: `10.1103/PhysRevA.91.042129`. arXiv: `1411.6379 [quant-ph]`. © 2015 American Physical Society. See Chapter 5.
- M. Holzäpfel and M. B. Plenio (2017). "Efficient certification and simulation of local quantum many-body Hamiltonians". arXiv: `1712.04396 [quant-ph]`. See Chapter 6.
- B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos (2017). "Efficient tomography of a quantum many-body system". Nat. Phys. 13, 1158–1162. DOI: `10.1038/nphys4244`. arXiv: `1612.08000 [quant-ph]`. Published by Macmillan Publishers Limited, part of Springer Nature. See Chapter 7 and Section 4.1.
- M. Holzäpfel, M. Cramer, N. Datta, and M. B. Plenio (2018). "Petz recovery versus matrix reconstruction". J. Math. Phys. 59(4), 042201. DOI: `10.1063/1.5009658`. arXiv: `1709.04538 [quant-ph]`. Published by AIP Publishing. See Sections 2.2.1 and 8.1.

The following publications are not covered:

- N. Friis, O. Marty, C. Maier, C. Hempel, M. Holzäpfel, P. Jurcevic, M. B. Plenio, M. Huber, C. F. Roos, R. Blatt, and B. P. Lanyon (2018). "Observation of entangled states of a fully-controlled 20 qubit system". Phys. Rev. X 8 (2), 021012. DOI: `10.1103/PhysRevX.8.021012`. arXiv: `1711.11092 [quant-ph]`.
- D. Suess and M. Holzäpfel (2017). "mpnum: A matrix product representation library for Python". J. Open Source Softw. 2(20), 465. DOI: `10.21105/joss.00465`.

Contents

Summary iii
Publications v
Contents vii

I. Introduction 1

1. Basic concepts 3
  1.1. Elements of statistical estimation theory 3
  1.2. Notation 7
  1.3. Quantum states 11
  1.4. Quantum measurements 13

2. Tensor representations 17
  2.1. Tensor rank 19
  2.2. Matrix product states and tensor trains (MPS/TT) 20
    2.2.1. Definitions and basic properties 20
    2.2.2. Efficient methods 24
    2.2.3. Examples 30
  2.3. Projected entangled pair states (PEPS) 31

3. Quantum state tomography 37
  3.1. Basic concepts 38
    3.1.1. Complexity 38
    3.1.2. Informationally complete measurements 41
    3.1.3. Pauli matrices 43
  3.2. Standard estimators 45
    3.2.1. Linear inversion 45
    3.2.2. Bayesian mean estimation 46
    3.2.3. Maximum likelihood estimation 47

3.3. Efficient estimation 49
  3.3.1. Maximum likelihood estimation with MPSs 49
  3.3.2. Singular value thresholding with MPSs 52
  3.3.3. Reconstruction of matrix product operators 54
3.4. Efficient verification 55
  3.4.1. Parent Hamiltonian certificate 55
  3.4.2. Markov entropy decomposition certificate 59

II. Scalable estimation and verification of quantum systems 61

4. Efficient maximum likelihood estimation (MLE) 63
4.1. Resources for constant estimation error 63
4.2. Continuous variables 65

5. Scalable ancilla-assisted process tomography (AAPT) 73
5.1. Local ancilla-assisted process tomography 73
5.2. Hamiltonian reconstruction 78
5.3. Numerical simulations 79
  5.3.1. Unitary quantum circuits 81
  5.3.2. Hamiltonian reconstruction 83

6. Efficient verification and simulation of local time evolution 89
6.1. Summary 89
6.2. Lieb–Robinson bounds 92
6.3. Efficient verification 97
6.4. Efficient simulation 104
  6.4.1. Trotter decomposition 105
  6.4.2. Arbitrary lattice 107
  6.4.3. Hypercubic lattice 112
6.5. Discussion 119

7. Tomography of an ion trap quantum simulator 121
7.1. The ion trap quantum simulator experiment 122
7.2. Measurements 123
7.3. State estimation 125
7.4. Certificate 125
  7.4.1. Empirical parent Hamiltonian 126

7.4.2. Measurement uncertainty 127
7.5. Results 129
7.6. Conclusion 133

8. Tensor reconstruction 135
8.1. Reconstruction of low-rank matrices 137
8.1.1. Reconstruction 137
8.1.2. Stability 139
8.2. Tensor reconstruction 141
8.2.1. Notation and basic facts 142
8.2.2. Reconstruction 147
8.2.3. Recursive measurements 151
8.2.4. Comparison 155
8.2.5. Recursive measurements II 160
8.2.6. Related work and conclusion 161

9. Conclusions and outlook 167

A. Appendix 173
A.1. Interaction picture 173
A.2. Selected inequalities 175
A.3. Metric spaces 177

Bibliography 181

Lists and indices 193
List of figures 193
List of tables 194
Acronyms 194

Part I.

Introduction

Chapter 1.

Basic concepts

Chance plays a fundamental role in quantum mechanics because measurement outcomes are inherently random. Measurement outcomes are random samples from a probability distribution and this distribution is determined by the measurement configuration and the quantum state of the system. In deriving a quantum state or any other information from the results of a quantum measurement, one estimates one or more parameters of a probability distribution from a random sample. For this reason, this chapter starts by reviewing selected concepts from statistical estimation theory. Afterwards, basic properties of quantum states and measurements as well as relevant notation are introduced.

1.1. Elements of statistical estimation theory

Setting. The central object in statistical estimation theory[1] is a joint probability density function (PDF) $p(x) = p(x_1, \dots, x_m)$ of m random variables $X_1, \dots, X_m$. We will refer to the random variables mainly by their value $x = (x_1, \dots, x_m)$. We consider $x_i \in \mathbb{R}^n$, i.e. $x \in \mathbb{R}^{mn}$. At this point, the division of the mn values into m vectors of n values is arbitrary and depends on the application: The PDF may describe e.g. m repetitions of an experiment where the i-th run of the experiment produces the observation $x_i \in \mathbb{R}^n$. A PDF has the following properties:

$$p(x) \geq 0 \qquad \int_{\mathbb{R}^{mn}} p(x) \, \mathrm{d}x = 1 \qquad x = (x_1, \dots, x_m) \tag{1.1}$$

---

[1] See e.g. Mood et al. 1974; Lehmann and Casella 2003; Jaynes 2003.

Other basic objects of our interest are functions $f, g\colon \mathbb{R}^{mn} \longrightarrow \mathbb{R}$ of observations $x \in \mathbb{R}^{mn}$. Their mean, covariance and variance are given by:

$$\mathbb{E}_p(f) := \int f(x)p(x)\,\mathrm{d}x \tag{1.2a}$$

$$\mathbb{V}_p(f,g) := \mathbb{E}_p(fg) - \mathbb{E}_p(f)\mathbb{E}_p(g) \tag{1.2b}$$

$$\mathbb{V}_p(f) := \mathbb{V}_p(f,f) \tag{1.2c}$$

A quick inspection of the definitions shows that the mean is linear and that the covariance is bilinear and symmetric in its arguments. For functions $f, g, h\colon \mathbb{R}^{mn} \longrightarrow \mathbb{R}$ and a scalar $a \in \mathbb{R}$, we have:

$$\mathbb{E}(a(f+g)) = a[\mathbb{E}(f) + \mathbb{E}(g)] \tag{1.3a}$$

$$\mathbb{V}(f,g) = \mathbb{V}(g,f) \tag{1.3b}$$

$$\mathbb{V}(af,g) = a\mathbb{V}(f,g) \tag{1.3c}$$

$$\mathbb{V}(f+g,h) = \mathbb{V}(f,h) + \mathbb{V}(g,h) \tag{1.3d}$$

Related to the division of $x \in \mathbb{R}^{mn}$ into m vectors of n entries, we also use functions $f, g\colon \mathbb{R}^{n} \longrightarrow \mathbb{R}$. The sample mean of such a function f is a function $\mathbb{R}^{mn} \longrightarrow \mathbb{R}$, $x \longmapsto \mathbb{E}_x(f)$ whose value is given by

$$\mathbb{E}_x(f) := \frac{1}{m}\sum_{i=1}^{m} f(x_i), \qquad x = (x_1, \ldots, x_m) \in \mathbb{R}^{mn}. \tag{1.4a}$$

We define the sample covariance and variance[2] of functions $f, g\colon \mathbb{R}^{n} \longrightarrow \mathbb{R}$ in the same way as the covariance and variance (Equation (1.2)):

$$\mathbb{V}_x(f,g) := \mathbb{E}_x(fg) - \mathbb{E}_x(f)\mathbb{E}_x(g) \tag{1.4b}$$

$$\mathbb{V}_x(f) := \mathbb{V}_x(f,f) \tag{1.4c}$$

The estimation problem. In estimation problems, it is assumed that the PDF $p(x)$ depends on a parameter θ which cannot be observed directly. The dependence is indicated by a subscript, $p_\theta(x) = p(x)$ (sometimes, we omit this subscript). The goal is to estimate a function value $g(\theta)$ using a random sample x from $p_\theta(x)$ and prior information on the structure of $p_\theta(x)$. We use the assumption that $X_1, \ldots, X_m$

---

[2] The sample variance is often defined to be $\frac{m}{m-1}\mathbb{V}_x(f)$ because that is an unbiased estimator of $\mathbb{V}_p(f)$. We discuss this matter below.

are random variables which are independent and indentically distributed (i.i.d.). If m corresponds to the number of independent repetitions of a given experiment, assuming i.i.d. random variables $X_1, \dots, X_m$ is often justified. In the i.i.d. case, the probability density factorizes into $p_\theta(x) = p_\theta^{(1)}(x_1) \dots p_\theta^{(1)}(x_m)$ (cf. (1.8) below).

An estimator is a function $\epsilon \colon \mathbb{R}^{mn} \to \mathbb{R}$ which takes a sample x from p_θ and returns an estimate $\epsilon(x)$ of $g(\theta)$. A commonly used measure for an estimator's accuracy is given by the mean squared error (MSE):

$$\mathrm{MSE}(\epsilon) := \mathbb{E}_{p_\theta}[(\epsilon - g(\theta))^2] \tag{1.5}$$

The MSE is given by the squared difference between the estimated value $\epsilon(x)$ and the true value $g(\theta)$, averaged over p_θ. A short computation shows that the MSE is the sum of two contributions:

$$\mathrm{MSE}(\epsilon) = [b(\epsilon)]^2 + v(\epsilon) \qquad b(\epsilon) := \mathbb{E}_{p_\theta}(\epsilon) - g(\theta) \qquad v(\epsilon) := \mathbb{V}_{p_\theta}(\epsilon) \tag{1.6}$$

The two terms $b(\epsilon)$ and $v(\epsilon)$ are called the *bias* and *variance* of the estimator ϵ. The bias is sometimes also called the estimator's *systematic error*.[3] An estimator is called *unbiased* if its bias is equal to zero. In general, it is desirable to use an estimator which has minimal MSE or which minimizes another reasonable measure of estimation accuracy. It is well possible that a biased estimator achieves a smaller MSE than a given unbiased estimator and it is also possible that no unbiased estimator exists for a particular problem.[4] Despite the negative connotations of the word *biased*, a biased estimator may be the best or even the only choice.

Suppose that for $m \in \{1, 2, 3, \dots\}$, there is a probability density $p_{\theta,m}$ of m random variables and a corresponding estimator ϵ_m. Such a sequence $(\epsilon_m)_m$ of estimators is called *consistent* if

$$\lim_{m \to \infty} \mathrm{MSE}(\epsilon_m) = 0. \tag{1.7}$$

If a sequence of estimators is consistent, their MSE vanishes as m increases. If m corresponds to the number of repetitions of an experiment, an estimator should be consistent because we want to obtain estimates with smaller expected error for larger numbers of repetitions. In the following, we discuss two examples in the i.i.d. setting, which is a natural assumption for m repetitions of an experiment.

---

[3] Lehmann and Casella 2003; Schwemmer et al. 2015.

[4] E.g. Shang et al. 2014. See also Section 3.2.1.

Simple examples in the i.i.d. setting. Here, we introduce two simple and well-known estimators which are used in Chapter 7. In this section, we denote the common PDF by $p_m(x) := p_\theta(x)$, $x = (x_1, \dots, x_m) \in \mathbb{R}^{mn}$. Since the random variables $X_1, \dots, X_m$ are assumed to be i.i.d., the distribution $p_m(x)$ factorizes into

$$p_m(x) = p(x_1)p(x_2)\dots p(x_m). \tag{1.8}$$

We restrict our attention to estimands of the following form: Given a function $g\colon \mathbb{R}^n \longrightarrow \mathbb{R}$, we want to estimate the mean

$$g_0 := \mathbb{E}_p(g). \tag{1.9}$$

An estimator of g_0 is given by the sample mean,

$$\epsilon(x) := \mathbb{E}_x(g). \tag{1.10}$$

The estimator from the last equation, obtained by inserting the function g into the sample mean $\mathbb{E}_x(\,\cdot\,)$, is called a plug-in estimator.[5] The estimator ϵ has the following properties:

$$\mathbb{E}_{p_m}(\epsilon) = \mathbb{E}_p(g), \tag{1.11a}$$

$$\mathrm{MSE}(\epsilon) = v(\epsilon) = \mathbb{V}_{p_m}(\epsilon) = \mathbb{V}_{p_m}(\mathbb{E}_x(g)) = \frac{\mathbb{V}_p(g)}{m}. \tag{1.11b}$$

The properties can be obtained using Lemma 1.1 below. Equation (1.11) shows that the estimator ϵ is unbiased. In addition, it is consistent if $\mathbb{V}_p(g)$ is finite.

If an estimator is given by a sample mean of a function as in Equation (1.10), we call $v(\epsilon)$ and $\sqrt{v(\epsilon)}$ the *variance of the mean* and the *standard deviation of the mean*. As an estimator for $v(\epsilon)$, we propose

$$v_\epsilon(x) := \frac{1}{m-1}\mathbb{V}_x(g). \tag{1.12}$$

The following Lemma 1.1[6] shows that $v_\epsilon(x)$ is an unbiased estimator of the variance $v(\epsilon)$ as follows:

$$\mathbb{E}_{p_m}(v_\epsilon) = \frac{1}{m-1}\mathbb{E}_{p_m}(\mathbb{V}_x(g)) = \mathbb{V}_{p_m}(\mathbb{E}_x(g)) = \mathbb{V}_{p_m}(\epsilon) = v(\epsilon) \tag{1.13}$$

Note that $\sqrt{v_\epsilon(x)}$ is a *biased* estimator of the standard deviation of the mean $\sqrt{v(\epsilon)}$.

---

[5] See Section 4, in particular Eqs. (4.13) and (4.17) in Section 4.3, of Efron and Tibshirani (1993).

[6] See e.g. Lehmann and Casella (2003). The lemma is reproduced from the author's contributions to Lanyon, Maier, et al. (2017).

Lemma 1.1 *For two functions* $f, g\colon \mathbb{R}^n \to \mathbb{R}$, *we have*

$$\mathbb{E}_{p_m}[\mathbb{E}_x(f)] = \mathbb{E}_p(f), \tag{1.14a}$$

$$\mathbb{V}_{p_m}[\mathbb{E}_x(f), \mathbb{E}_x(g)] = \frac{1}{m}\mathbb{V}_p(f,g), \tag{1.14b}$$

$$\frac{1}{m}\mathbb{V}_p(f,g) = \frac{1}{m-1}\mathbb{E}_{p_m}[\mathbb{V}_x(f,g)]. \tag{1.14c}$$

Proof First equation:

$$\mathbb{E}_{p_m}[\mathbb{E}_x(f)] = \int \frac{1}{m}\sum_{i=1}^{m} f(x_i)p(x_1)\dots p(x_m)\mathrm{d}x_1\dots\mathrm{d}x_m = \frac{m}{m}\mathbb{E}_p(f). \tag{1.15}$$

For the second and third equation, we first compute

$$\begin{aligned}\mathbb{E}_{p_m}[\mathbb{E}_x(f)\mathbb{E}_x(g)] &= \int \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} f(x_i)g(x_j)p(x_1)\dots p(x_m)\mathrm{d}x_1\dots\mathrm{d}x_m \\ &= \frac{m}{m^2}\mathbb{E}_p(fg) + \frac{m^2-m}{m^2}\mathbb{E}_p(f)\mathbb{E}_p(g) \\ &= \frac{1}{m}\mathbb{V}_p(f,g) + \mathbb{E}_p(f)\mathbb{E}_p(g). \end{aligned} \tag{1.16}$$

This provides

$$\begin{aligned}\mathbb{V}_{p_m}[\mathbb{E}_x(f), \mathbb{E}_x(g)] &= \mathbb{E}_{p_m}[\mathbb{E}_x(f)\mathbb{E}_x(g)] - \mathbb{E}_{p_m}[\mathbb{E}_x(f)]\,\mathbb{E}_{p_m}[\mathbb{E}_x(g)] \\ &= \frac{1}{m}\mathbb{V}_p(f,g) \end{aligned} \tag{1.17}$$

and

$$\mathbb{E}_{p_m}[\mathbb{V}_x(f,g)] = \mathbb{E}_p(fg) - \mathbb{E}_{p_m}[\mathbb{E}_x(f)\mathbb{E}_x(g)] = (1-\frac{1}{m})\mathbb{V}_p(f,g), \tag{1.18}$$

which completes the proof. ■

1.2. Notation

Definitions are indicated by the notation $a := b^2$ or $b^2 =: a$, which are both taken to mean that a is defined as b^2. Given three sets A, B and C, the Cartesian product is defined as $A \times B \times C := \{(a,b,c)\colon a \in A, b \in B, c \in C\}$. Assuming suitable equivalence relations, the Cartesian product becomes associative, i.e. $A \times (B \times C) = (A \times B) \times C = A \times B \times C$. Powers of sets are given by $A^1 := A$ and $A^n := A \times A^{n-1}$. The notation $A \subset B$ and $A \subseteq B$ both denote $x \in A \Rightarrow x \in B$

whereas strict subsets are indicated by $A \subsetneq B$ which means that both $A \subset B$ and $A \neq B$ hold. The expression $C = A \dot{\cup} B$ implies that $C = A \cup B$ and $A \cap B = \emptyset$. The sets $\{B_i \colon i\}$ are called a partition of the set A if $A = \dot{\bigcup}_i B_i$. Let A, B be sets and $f \colon A \to B$ a map. For $C \subset A$, the restriction $f|_C$ is defined as $f|_C \colon C \to B$, $(f|_C)(x) := f(x)$. We use the notation $[m : n] := \{m, m+1, \dots, n\}$ and $[n] := [1 : n]$. The complex conjugate of a complex number $a \in \mathbb{C}$ is denoted by $\overline{a}$.

Given two functions f and g, we write $f = \mathcal{O}(g)$ if there are positive constants c_1, c_2, c_3 such that $f(n) \leq c_1 g(c_2 n)$ holds for all $n \geq c_3$. We say that f is *at most polynomial* in n and write $f = \mathcal{O}(\mathrm{poly}(n))$ if there is a polynomial g such that $f = \mathcal{O}(g)$. A function f is called *at most exponential* in n, denoted by $f = \mathcal{O}(\exp(n))$, if there are constants $c_1, c_2 > 0$ such that $f(n) \leq c_1 \exp(c_2 n)$ holds for all n. A function $f(n)$ is called *quasi-polynomial* in n if $f(n) = \mathcal{O}(\exp(c_1 (\log n)^{c_2})) = \mathcal{O}(n^{c_1 (\log n)^{c_2 - 1}})$ with constants $c_1, c_2 > 0$. A function is called *poly-logarithmic* in n if $f(n) = \mathcal{O}((\log n)^{c_1})$ with a positive constant c_1.

The dimension of a vector space V is denoted by $\dim(V)$ or d_V. Most vector spaces are finite-dimensional and infinite-dimensional vectors spaces are not treated rigorously. Most vector spaces are complex vector spaces but real vector spaces occur as well. Vectors from a vector space V are denoted by $x \in V$ or, in bra–ket notation, by $|x\rangle := x \in V$. Sometimes, the differing notation $x = |x\rangle\langle x|$ is used, but this is indicated. An inner product on V is denoted by $\langle x, y \rangle$ or $\langle x|y \rangle := \langle x, y \rangle$ and satisfies $\langle x, y \rangle = \overline{\langle y, x \rangle}$ and $\langle ax + by, z \rangle = \overline{a} \langle x, z \rangle + \overline{b} \langle y, z \rangle$, where $x, y, z \in V$ and $a, b \in \mathbb{C}$. An inner product induces the norm $\|x\| := \sqrt{\langle x, x \rangle}$.

The set of bounded, linear maps from a vector space V to another vector space W is denoted by $\mathcal{B}(V; W)$ and $\mathcal{B}(V) := \mathcal{B}(V; V)$ is the set of bounded, linear operators on V. The set $\mathcal{B}(V; W)$ is a vector space and a basis of $\mathcal{B}(V; W)$ is called a matrix basis or operator basis. The sets of linear maps and operators on tensor products of vector spaces V_1, V_2, W_1, W_2 may be denoted as $\mathcal{B}(V_1, V_2; W_1, W_2) := \mathcal{B}(V_1 \otimes V_2; W_1 \otimes W_2)$ and $\mathcal{B}(V_1, V_2) := \mathcal{B}(V_1 \otimes V_2)$. In the bra–ket notation, the notations $|uv\rangle := |uv\rangle_{1,2} := |u\rangle_1 \otimes |v\rangle_2 := |u\rangle \otimes |v\rangle$ all denote the tensor product of $|u\rangle \in V_1$ and $|v\rangle \in V_2$.

The (algebraic) dual space is $V^* := \mathcal{B}(V; \mathbb{C})$, the dual or Hermitian adjoint vector of x is denoted by $x^* \in V^*$ and it is defined in terms of the inner product as $x^*(y) := \langle x, y \rangle$. The bra–ket notation for dual vectors is $\langle x| := x^*$. Given inner product spaces V and W, the (Hermitian) adjoint of a linear map $A \in \mathcal{B}(V; W)$ is denoted by $A^* \in \mathcal{B}(W; V)$ and it is defined by the property $\langle A^* y, x \rangle = \langle y, Ax \rangle$ for all $x \in V$ and $y \in W$. The notation $A^\dagger$ for the adjoint is not used because it may be confused with the pseudoinverse A^+ (see below). A linear operator A

is called self-adjoint or Hermitian if it satisfies $A = A^*$. The sets of (trace-free) Hermitian linear operators on V are denoted by $\mathcal{B}_H(V) := \{A \in \mathcal{B}(V) \colon A = A^*\}$ and $\mathcal{B}_0(V) := \{A \in \mathcal{B}_H(V) \colon \operatorname{Tr}(A) = 0\}$. The sets $\mathcal{B}_H(V)$ and $\mathcal{B}_0(V)$ are a real vector spaces. A linear operator $A \in \mathcal{B}(V)$ on a complex vector space V is called positive semidefinite (positive definite) if $\langle x, Ax\rangle$ is real and $\langle x, Ax\rangle \geq 0$ ($\langle x, Ax\rangle > 0$) holds for all $x \in V$. A positive semidefinite linear operator on a complex vector space is also Hermitian.[7] The image (range) and kernel of a linear map $A \in \mathcal{B}(V; W)$ are $\operatorname{im}(A) := \{Ax \colon x \in V\}$ and $\ker(A) := \{x \in V \colon Ax = 0\}$. The orthogonal complement of a vector subspace $Z \subset V$ is given by $Z^\perp := \{x \in V \colon \forall y \in Z \colon \langle x, y\rangle = 0\}$. A linear map $A \in \mathcal{B}(V; W)$ is a partial isometry if it satisfies $\|Ax\| = \|x\|$ for all $x \in [\ker(A)]^\perp$.[8] A linear operator $A \in \mathcal{B}(V)$ is the orthogonal projection onto a vector subspace $Z \subset V$ if $[\ker(A)]^\perp = Z$ and $Ax = x$ for all $x \in Z$. A linear operator $U \in \mathcal{B}(V)$ is unitary if $\|Ux\| = \|x\|$ for all $x \in V$. It satisfies $UU^* = U^*U = \mathrm{id}_V$, where id_V is the identity map on V.

Vectors from $\mathbb{C}^n$ are treated as column vectors (or matrices from $\mathbb{C}^{n\times 1}$). On $V = \mathbb{C}^n$, the standard inner product $\langle x, y\rangle = \sum_{i=1}^n \overline{x_i} y_i$ is usually used. Here, the induced norm is the vector 2-norm $\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$. In this case, adjoint vectors and adjoint linear maps are given by the conjugate transpose vectors and maps. The n-by-n identity matrix is denoted by $\mathbb{1}_n$.

Any matrix $A \in \mathbb{C}^{m\times n}$ admits a *singular value decomposition (SVD)*[9]

$$A = USV^*, \qquad S \in \mathbb{C}^{r\times r}, \qquad U \in \mathbb{C}^{m\times r}, \qquad V \in \mathbb{C}^{n\times r} \tag{1.19}$$

where $r = \operatorname{rk}(A)$ is A's rank. The matrix S is positive definite and diagonal. Its entries $S = \operatorname{diag}(\sigma_1, \dots, \sigma_r)$ are called *singular values* of A and we take them to be ordered non-increasingly. If necessary, we set $\sigma_i := 0$ for the further singular values with $i > r$. The matrices U and V are partial isometries satisfying $U^*U = V^*V = \mathbb{1}_r$. The matrices U and S can be obtained from an eigendecomposition of the Hermitian matrix $AA^* = US^2U^*$. The third matrix V could then be obtained as $V^* = S^{-1}U^*A$.[10] Equation (1.19) is called a *compact* SVD. In contrast, a non-compact SVD is given by

$$A = USV^*, \qquad S \in \mathbb{C}^{m\times n}, \qquad U \in \mathbb{C}^{m\times m}, \qquad V \in \mathbb{C}^{n\times n}. \tag{1.20}$$

---

[7] Horn and Johnson 1991a, Section 7.1.

[8] E.g. Horn and Johnson 1991b, Definition 3.1.7.

[9] Eckart and Young 1936. See also Horn and Johnson 1991a, Section 7.3, Horn and Johnson 1991b, Chapter 3 or Bhatia 1997, Section I.2.

[10] Due to the matrix inverse, this approach can be numerical unstable if small non-zero singular values occur.

Here, the only non-zero entries of S are $S_{ii} = \sigma_i$ for $i \in \{1 \dots r\}$ and the matrices U and V are unitary.

The set of matrices $\mathbb{C}^{m\times n}$ is a vector space and an inner product on $\mathbb{C}^{m\times n}$ is the Hilbert–Schmidt inner product $\langle A, B\rangle := \mathrm{Tr}(A^*B)$, $A, B \in \mathbb{C}^{m\times n}$. Given a matrix $A \in \mathbb{C}^{m\times n}$, the trace norm, Frobenius norm and operator norm are defined as[11]

$$\|A\|_{(1)} := \sum_{i=1}^{r} \sigma_i, \tag{1.21a}$$

$$\|A\|_{(2)} := \|A\|_{\mathrm{Fr}} := \sqrt{\sum_{i=1}^{r} \sigma_i^2}, \tag{1.21b}$$

$$\|A\|_{(\infty)} := \sigma_1. \tag{1.21c}$$

These three norms are the so-called Schatten-p-norms for $p \in \{1, 2, \infty\}$. They are unitarily invariant, which means that $\|UAV\| = \|A\|$ holds for all unitary $U \in \mathcal{B}(\mathbb{C}^m)$ and $V \in \mathcal{B}(\mathbb{C}^n)$. Unitarily invariant matrix norms are submultiplicative because they satisfy $\|ABC\| \le \|A\|_{(\infty)}\|B\|\|C\|_{(\infty)} \le \|A\|\|B\|\|C\|$.[12] Note that the Hilbert–Schmidt inner product induces the Frobenius norm, i.e. $\|A\|_{(2)} = \mathrm{Tr}(A^*A)$. The trace norm is elsewhere also called nuclear norm and denoted by $\|\cdot\|_* := \|\cdot\|_{(1)}$. The trace distance is $D(M, N) := \frac{1}{2}\|M - N\|_{(1)}$ where $M, N \in \mathbb{C}^{m\times n}$.

The Moore–Penrose pseudoinverse $A^+ \in \mathbb{C}^{n\times m}$ of $A \in \mathbb{C}^{m\times n}$ is the only matrix which satisfies the following four equations:[13]

$$AA^+A = A \qquad A^+AA^+ = A^+ \qquad (A^+A)^* = A^+A \qquad (AA^+)^* = AA^+ \tag{1.22}$$

We refer to A^+ as *the pseudoinverse* of A. If $A = USV^*$ is a compact SVD, the pseudoinverse is $A^+ = VS^{-1}U^*$. The pseudoinverse has the properties $AA^+ = P_{\mathrm{im}(A)}$ and $A^+A = P_{(\ker(A))^\perp}$, where $P_{\mathrm{im}(A)}$ and $P_{(\ker(A))^\perp}$ denote the orthogonal projections onto A's image $\mathrm{im}(A)$ and onto the orthogonal complement of its null space $\ker(A)$. The adjoint and the pseudoinverse satisfy $(A^*)^* = A$, $(A^+)^+ = A$ and $(A^*)^+ = (A^+)^*$. The equivalent relations $\mathrm{im}(A) = [\ker(B)]^\perp$ and $\ker(A) = [\mathrm{im}(B)]^\perp$ hold for the adjoint $B = A^*$ and the pseudoinverse $B = A^+$. Given a matrix A, we define $A_\tau^+ := (A_\tau)^+$ and A_τ is given by A with singular values smaller than or equal to τ replaced by zero.

---

[11] E.g. Horn and Johnson 1991b, Section 3.5 or Bhatia 1997, Section IV.2.

[12] Bhatia 1997, Proposition IV.2.4.

[13] Moore 1935, 1939; Penrose 1955; Rado 1956. See also Greville 1966.

1.3. Quantum states

The central object of the present work is the state of a quantum system which consists of a finite number of n distinguishable subsystems. For the most part, we consider finite-dimensional systems such that the Hilbert space of system $i \in \{1 \dots n\}$ is isomorphic to $\mathcal{H}_i := \mathbb{C}^{d_i}$ where $d_i < \infty$ is subsystem i's dimension. The set $\Lambda := [n] = \{1 \dots n\}$ refers to the complete system and a subset $t \subset \Lambda$ refers to a subsystem composed of the given basic subsystems. A subsystem $k \in \Lambda$ is also called a *site* and a subsystem $t \subset \Lambda$ may be called a *supersite*. The Hilbert space of $t \subset \Lambda$ and its dimension are denoted by

$$\mathcal{H}_t := \bigotimes_{i \in t} \mathcal{H}_i \qquad\qquad d_t := \prod_{i \in t} d_i \tag{1.23}$$

The state of a quantum system with Hilbert space $\mathcal{H}$ is described by a density matrix and the set of all such density matrices is

$$\mathcal{D}(\mathcal{H}) := \left\{\rho \in \mathcal{B}(\mathcal{H})\colon \rho \text{ is positive semidefinite and } \operatorname{Tr}(\rho) = 1\right\} \tag{1.24}$$

If $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$ is a density matrix of the complete system, the reduced or marginal state ρ_t of $t \subset \Lambda$ is

$$\rho_t := \operatorname{Tr}_{t^c}(\rho) \tag{1.25}$$

where $t^c := \Lambda \setminus t$. The so-called partial trace over $t \subset \Lambda$ is defined as

$$\operatorname{Tr}_t := T_1^{(t)} \otimes \dots \otimes T_n^{(t)}, \qquad T_i^{(t)} = \begin{cases} \mathcal{B}(\mathcal{H}_i) \to \mathcal{B}(\mathcal{H}_i), & \sigma \mapsto \sigma \text{ if } i \notin t, \\ \mathcal{B}(\mathcal{H}_i) \to \mathbb{C}, & \sigma \mapsto \operatorname{Tr}(\sigma) \text{ if } i \in t. \end{cases} \tag{1.26}$$

A quantum state is called pure or mixed if the rank of the density matrix equals or exceeds one, respectively. We set $\mathcal{H}^{(1)} := \{|\psi\rangle \in \mathcal{H}\colon \langle\psi|\psi\rangle = 1\}$. For any density matrix $\rho \in \mathcal{D}(\mathcal{H})$ of unit rank, there is a vector $|\psi\rangle \in \mathcal{H}^{(1)}$ such that $\rho = |\psi\rangle\langle\psi|$. Accordingly, a pure state is described by a unit rank density matrix or by a state vector.

In the following, we discuss bipartite correlations in quantum states. For this purpose, we restrict to $n = 2$. There are no correlations between subsystems 1 and 2 in a given state if the density matrix is a tensor product $\rho = \sigma_1 \otimes \sigma_2$, $\sigma_i \in \mathcal{D}(\mathcal{H}_i)$, $i \in \{1, 2\}$. For pure states, this is equivalent to $|\psi\rangle = |\phi_1\rangle \otimes |\phi_2\rangle$, $|\phi_i\rangle \in \mathcal{H}_i^{(1)}$. Correlations between subsystems 1 and 2 exist in all density matrices and state vectors which are not of this form.

The Schmidt rank[14] of a general vector $|\psi\rangle \in \mathcal{H}_{12}$ or state vector $|\psi\rangle \in \mathcal{H}_{12}^{(1)}$ is

$$\mathrm{SR}(1:2)_{|\psi\rangle} := \min\Big\{r: \quad |\psi\rangle = \sum_{j=1}^{r} \big|x_j^{(1)}\big\rangle \otimes \big|x_j^{(2)}\big\rangle, \quad \big|x_j^{(i)}\big\rangle \in \mathcal{H}_i,\ i \in \{1,2\}\Big\}. \tag{1.27}$$

The operator Schmidt rank (OSR)[15] of a linear operator $\rho \in \mathcal{B}(\mathcal{H}_{12})$ or a density matrix $\rho \in \mathcal{D}(\mathcal{H}_{12})$ is

$$\mathrm{OSR}(1:2)_\rho := \min\Big\{r: \quad \rho = \sum_{j=1}^{r} X_j^{(1)} \otimes X_j^{(2)}, \quad X_j^{(i)} \in \mathcal{B}(\mathcal{H}_i),\ i \in \{1,2\}\Big\}. \tag{1.28}$$

Note that the operator Schmidt rank is not the same as the so-called Schmidt number.[16] Dividing a multipartite system into two parts, bipartite OSRs such as $\mathrm{OSR}(23:145)_\rho$ are already covered by the given definition. Note that $\mathrm{OSR}(1:2)_{|\psi\rangle\langle\psi|} = [\mathrm{SR}(1:2)_{|\psi\rangle}]^2$. For $|\psi\rangle \neq 0$ and $\rho \neq 0$, we have

$$\mathrm{SR}(1:2)_{|\psi\rangle} \in \{1,\dots,\min\{d_1,d_2\}\}, \tag{1.29}$$

$$\mathrm{OSR}(1:2)_\rho \in \{1,\dots,\min\{d_1^2,d_2^2\}\}. \tag{1.30}$$

Both Schmidt ranks share the property that they assume their minimal value if and only if there are no correlations between subsystems 1 and 2. If the Schmidt ranks' logarithms are taken, the minimal value zero is assumed if and only if there are no correlations in a given state. It can be shown that the Schmidt ranks are non-increasing under local operations.[17] This property justifies using the Schmidt ranks or their logarithms as correlations measures.[18] For any density matrix or state vector, there is an arbitrarily close vector or matrix of maximal (operator) Schmidt rank. A stable measure of correlations can nevertheless be obtained by considering minimal ranks over small neighbourhoods in some distance measure.[19]

---

[14] See e.g. Nielsen and Chuang 2007, where the Schmidt rank of a pure state is called Schmidt number.

[15] Nielsen et al. 2003; Datta and Vidal 2007.

[16] The Schmidt number of a density matrix $\rho \in \mathcal{D}(\mathcal{H}_{12})$ is defined in terms of decompositions $\rho = \sum_i p_i |\phi_i\rangle\langle\phi_i|$ where $p_i \geq 0$, $\sum_i p_i = 1$ and $|\phi_i\rangle \in \mathcal{H}_{12}^{(1)}$ (not necessarily orthogonal). The Schmidt number of ρ is equal to r if there is at least one $|\phi_i\rangle$ which has Schmidt rank r in any decomposition and if there is a decomposition such that all $|\phi_i\rangle$ have Schmidt rank at most r (Terhal and Horodecki 2000).

[17] A local operation is given by a completely positive, trace non-increasing map from $\mathcal{B}(\mathcal{B}(\mathcal{H}_i))$ and such maps are discussed in some detail in Section 5.1. For the OSR, see Holzäpfel et al. 2018, Corollary 11. The Schmidt rank of a pure state is non-increasing even if local operations and classical communication (LOCC) are allowed (Nielsen 1999, Theorem 1).

[18] Cf. e.g. Henderson and Vedral 2001.

[19] Cf. e.g. Renner 2005.

At this point, a comment on different types of correlations is in order. Generally, correlations can be divided into classical correlations and quantum correlations but there are different possibilities where to draw the line between the two categories. The Schmidt rank quantifies *quantum entanglement* because this is the only type of correlations in pure states; entanglement is a (or even *the*) form of quantum correlations. The operator Schmidt rank, on the other hand, is called a measure of total (classical and quantum) correlations. For mixed states, the distinction between classical and quantum correlations is achieved by defining a set of all states containing only classical correlations and no quantum correlations at all. All remaining states then contain at least some quantum correlations. The first choice for the set of purely classically correlated states is the set of separable states; a separable state is a state which can be written as $\rho = \sum_j p_j \sigma_j^{(1)} \otimes \sigma_j^{(2)}$ where $\{p_j\}$ is a probability distribution and where $\sigma_j^{(i)} \in \mathcal{D}(\mathcal{H}_i)$.[20] All non-separable states contain entanglement and with this choice, entanglement is the only type of quantum correlations. An alternative choice for the set of purely classically correlated states is the set of classical-classical states which contains all states which can be written as $\rho = \sum_{jk} p_{jk} |\phi_j^{(1)}\rangle\langle\phi_j^{(1)}| \otimes |\phi_k^{(2)}\rangle\langle\phi_k^{(2)}|$ where $|\phi_j^{(i)}\rangle \in \mathcal{H}_i$ are orthonormal and where $\{p_{jk}\}$ is a bivariate probability distribution.[21] With this choice, all separable states which are not in the set of classical-classical states do not contain entanglement but some other form of quantum correlations referred to as quantum discord.[22] The operator Schmidt rank assumes its minimal value only for product states which do not contain any correlations at all. Therefore, the operator Schmidt rank quantifies total (classical and quantum) correlations independently of the choice for the division between quantum and classical correlations.[23]

1.4. Quantum measurements

There are two important differences between measurements in a classical system and measurements in quantum mechanics: First, the outcome of a quantum mechanical measurement is described by a probability distribution over all possible outcomes of the measurement and this holds true even for a perfect measurement on a perfectly

---

[20] Werner 1989.

[21] E.g. Piani et al. 2008, and references therein.

[22] Henderson and Vedral 2001; Ollivier and Zurek 2002.

[23] This is a property which the operator Schmidt rank shares with the quantum mutual information, which is also a measure of total (classical and quantum) correlations (Cerf and Adami 1997; Groisman et al. 2005).

Table 1.1: Properties of projective measurements $\{P_i\}$, generalized measurements $\{M_i\}$ and positive operator-valued measures $\{N_i\}$. The table provides the probability p_i for obtaining outcome $i \in \{1, \ldots, \mu\}$ and the state ρ_i after outcome i has been obtained.

Properties			p_i	ρ_i	Relation
$P_i \in \mathcal{B}(A)$	$P_i P_j = \delta_{ij} P_i$	$\sum_{i=1}^{\mu} P_i = \mathbb{1}$	$\mathrm{Tr}(\rho P_i)$	$P_i \rho P_i / p_i$	–
$M_i \in \mathcal{B}(A; B)$	–	$\sum_{i=1}^{\mu} M_i^* M_i = \mathbb{1}$	$\mathrm{Tr}(M_i \rho M_i^*)$	$M_i \rho M_i^* / p_i$	$M_i = P_i$
$N_i \in \mathcal{B}(A)$	$N_i \geq 0$	$\sum_{i=1}^{\mu} N_i = \mathbb{1}$	$\mathrm{Tr}(\rho N_i)$	–	$N_i = M_i^* M_i$

controlled system. As a consequence, any measurement needs to be repeated many times to obtain meaningful statistics and statistical estimation plays a central role in inferring properties or the complete state of a quantum system from measurements. Second, quantum mechanical measurements either have the potential to strongly perturb or even destroy the state of the system or they are limited to delivering very restricted amounts of information on the system. This property implies that repeated measurements are only possible if the system in question is prepared anew for each measurement, providing many identically prepared copies of the system in a given state, one for each repetition of the measurement.

In the following, we discuss the formal description of quantum measurements carried out on a state $\rho \in \mathcal{D}(\mathcal{H}_A)$ of a system with Hilbert space $\mathcal{H}_A$ of finite dimension $d_A < \infty$. For simplicity, we assume a finite number μ of distinguishable outcomes of the measurement. The discussion can be generalized to an uncountably infinite number of distinguishable outcomes and infinite-dimensional Hilbert spaces; we discuss such measurements non-rigorously in Section 4.2 below. The μ distinguishable outcomes of a given measurement are denoted by $i \in \{1 \ldots \mu\}$. By p_i, we denote the probability that outcome i will be obtained and the quantum state of the system after outcome i has been obtained is denoted by ρ_i. We will discuss projective measurements, generalized measurements and measurements described by a positive operator-valued measure (POVM). Table 1.1 provides an overview of their properties.

The most basic quantum mechanical measurement is a *projective measurement*. It is specified by a sequence of mutually orthogonal projections P_i which determine outcome probabilities via $p_i = \mathrm{Tr}(\rho P_i)$. The projections must sum to the identity, i.e. $\sum_{i=1}^{\mu} P_i = \mathbb{1}$, which ensures that $\sum_{i=1}^{\mu} p_i = 1$ holds. The state after outcome i has been obtained is $\rho_i = P_i \rho P_i / p_i \in \mathcal{D}(\mathcal{H}_A)$. The maximal number of distinguishable outcomes of a projective measurement is $\mu \leq d_A$.

A Hermitian linear operator $H \in \mathcal{B}(\mathcal{H}_A)$ is called an *observable* and its eigendecomposition specifies a projective measurement: The eigendecomposition can be written as $H = \sum_{i=1}^{\mu} \lambda_i P_i$ where $\lambda_1 > \lambda_2 > \dots$ are the eigenvalues of H without repetitions of degenerate eigenvalues. The eigenprojections $\{P_i\}$ form a projective measurement and measurement outcomes are specified by $i \in \{1, \dots, \mu\}$ or, equivalently, by an eigenvalue λ_i of H. The expectation value of H in the state ρ is given by the mean eigenvalue, $\langle H \rangle_\rho := \sum_{i=1}^{\mu} p_i \lambda_i$, and we observe that $\langle H \rangle_\rho = \operatorname{Tr}(\rho H)$ holds. Conversely, each projection P_i is an observable whose expectation value is the outcome probability of outcome i since $p_i = \operatorname{Tr}(\rho P_i) = \langle P_i \rangle_\rho$ holds. (The same holds true for the elements of a positive operator-valued measure (POVM).)

As mentioned already, the maximal number of outcomes of a projective measurement is $\mu \leq d_A$ and the state after measurement is $\rho_i \in \mathcal{B}(\mathcal{H}_A)$. A *generalized measurement* is more general in that it allows for an arbitrary, even infinite number of outcomes and it also allows the state after measurement to belong to a different system with Hilbert space $\mathcal{H}_B$, i.e. $\rho_i \in \mathcal{B}(\mathcal{H}_B)$. A *generalized measurement* is specified by a sequence of linear maps $M_i \in \mathcal{B}(\mathcal{H}_A; \mathcal{H}_B)$. Outcome probabilities and states after measurement are given by $p_i = \operatorname{Tr}(M_i \rho M_i^*)$ and $\rho_i = M_i \rho M_i^* / p_i$, respectively. The property $\sum_i M_i^* M_i = \mathbb{1}$ ensures $\sum_i p_i = 1$.

Measurements described by a positive operator-valued measure (POVM) are as general as generalized measurements but they omit the description of the system state after measurement. This is convenient if the state after measurement is discarded because the system is prepared again for the next repetition of the same measurement. A POVM is given by a sequence of positive-semidefinite linear operators $N_i \in \mathcal{B}(A)$ which satisfy $\sum_i N_i = \mathbb{1}$ and specify outcome probabilities $p_i = \operatorname{Tr}(\rho N_i)$.

Table 1.1 summarizes the similarities and differences between the mentioned types of measurements. Note that any projective measurement is also a valid generalized measurement (set $M_i := P_i$) and that a POVM can in turn be constructed from a generalized measurement via $N_i := M_i^* M_i$.

Chapter 2.

Tensor representations

State vectors and density matrices of n distinguishable quantum systems with dimensions $(d_1, \dots, d_n)$ correspond to tensors with n indices and shape $d_1 \times \dots \times d_n$ or $d_1^2 \times \dots \times d_n^2$, respectively. As the number of entries grows exponentially with n, efficient tensor representations are urgently needed. For this reason, we introduce the matrix product state and operator tensor representations in this chapter.[1] They are also known as tensor train representation and the last section of this chapter briefly touches upon so-called projected entangled pair states, a generalization to higher-dimensional lattices.

A tensor is a collection of complex numbers organized by n indices:

$$M \in \mathbb{C}^{d_1 \times \dots \times d_n}, \qquad M_{i_1,\dots,i_n} \in \mathbb{C}, \qquad i_k \in \{1, \dots, d_k\}, \qquad k \in \{1, \dots, n\}. \tag{2.1}$$

The number n is also called the number of *modes* of the tensor. The tensor M could e.g. contain the joint probability distribution of n discrete random variables or the components of a finite-dimensional quantum state vector or density matrix in some vector or matrix basis. In the case of uniform dimensions $d := d_1 = \dots = d_n$, the tensor M has d^n components and this exponential growth of the storage cost with the number of modes is called the *curse of dimensionality*.[2]

The practical limitations imposed by the curse of dimensionality are illustrated by the following example. A contemporary personal computer with random access memory (RAM) size of 8 GiB (= 2^{33} bytes) has the capability to store 2^{30} real numbers in double precision (8 bytes per number), which equals the number of entries of a tensor with $d = 4$ and $n = 15$. Even a supercomputer with 2 PiB (= 2^{51} bytes) distributed RAM[3] could only store a tensor of size $d = 4$ and $n = 24$.

---

[1] Parts of Section 2.2.1 reproduce parts of the original publication Holzäpfel et al. 2018 with the permission of AIP Publishing. See Footnote 15 on Page 21.

[2] E.g. Grasedyck et al. 2013.

[3] *Top500 supercomputer list* 2018.

Table 2.1: Tensor notation and terminology. Maximal local and bond dimensions are denoted by $d = \max_k d_k$ and $D = \max_k D_k$. In the present work, the physics notation is used.

Names	Physics	Mathematics
Tensor Multidimensional array	$T \in \mathbb{C}^{d_1 \times \cdots \times d_n}$	$T \in \mathbb{C}^{n_1 \times \cdots \times n_d}$
Shape Size	$d_1 \times \cdots \times d_n$	$n_1 \times \cdots \times n_d$
Mode	$k \in \{1 \ldots n\}$	$k \in \{1 \ldots d\}$
Index	$i_k \in \{1 \ldots d_k\}$	$i_k \in \{1 \ldots n_k\}$
Component Entry	$T_{i_1 i_2 \ldots i_n} \in \mathbb{C}$	$T_{i_1 i_2 \ldots i_d} \in \mathbb{C}$
Number of modes/indices Dimension of the tensor Order/degree/rank of the tensor	n	d
Size of mode/index k Dimension of mode/index k Local dimension k	d_k	n_k
Maximal mode/index size Maximal local dimension	$d = \max_{k \in \{1 \ldots n\}} d_k$	$n = \max_{k \in \{1 \ldots d\}} n_k$
Bond dimension Representation rank	D_k	r_k

The real part of a quantum density matrix of only 24 qubits corresponds to such a tensor. Storing all components of a tensor becomes infeasible quickly as the number n of modes grows. Manifold applications for n-mode tensors call for efficient methods for storing and processing tensors. We call a given method *efficient* if its cost increases at most polynomially with n.

The basis for any efficient tensor method is an efficient tensor representation, i.e. a scheme which specifies all components of a tensor with storage cost polynomial in n. The storage cost of a tensor representation is usually determined by a suitably defined tensor rank and several concepts of tensor rank are reviewed in Section 2.1. Section 2.2 then continues by presenting the MPS/TT representation and it is divided into three subsections covering basic properties, examples and efficient methods. The MPS/TT representation is limited in that it assumes a one-dimensional structure. Consequently, matrix product states were very successful in applications to one-dimensional quantum systems but are not well-suited for quantum states of higher-dimensional quantum lattice systems. An appropriate extension of the MPS/TT representation is given by the PEPS representation which is briefly discussed in Section 2.3.

The non-uniform terminology of tensor properties can be summarized as follows. The tensor T from Equation (2.1) is said to have *size* or *shape* $d_1 \times \cdots \times d_n$. The complex (or real) numbers $T_{i_1 \dots i_n}$ are the *components* or *entries* of the tensor. The number n has many names: It is called the *number of modes*, the *number of indices* or the *dimension* of the tensor. It is also called the *rank*, *degree* or *order* of the tensor (note that both n and $\mathrm{rk}_{\mathrm{CP}}(T)$ defined in (2.3) may be called *tensor rank*). The *size* or *dimension* of the k-th mode (the k-th index) is given by d_k. The shape of a tensor is often taken as $d_1 \times \cdots \times d_n$ (as above) or as $n_1 \times \cdots \times n_d$ (with n and d swapped). The former convention is frequently used in quantum physics literature[4] while the latter convention is more prevalent in the literature on numerical mathematics.[5] The discussed terminology and the two conventions are summarized in Table 2.1. (The bond dimension or representation rank is introduced in Section 2.2 below.)

2.1. Tensor rank

To prepare our discussion of tensor representations, this section introduces different notions of tensor rank, which are related the storage cost of corresponding representations. We shall explain the basic idea of a tensor representation at the example of a low-rank matrix. Consider a matrix $A \in \mathbb{C}^{d_1 \times d_2}$ of rank $r := \mathrm{rk}(A)$. The matrix rank can be defined as follows:[6]

$$\mathrm{rk}(A) = \min \left\{ r : \quad A = \sum_{j=1}^{r} x_j^{(1)} \left(x_j^{(2)} \right)^*, \quad x_j^{(k)} \in \mathbb{C}^{d_k}, k \in \{1, 2\} \right\}. \tag{2.2}$$

This definition implies that there are matrices $B \in \mathbb{C}^{d_1 \times r}$, $C \in \mathbb{C}^{r \times d_2}$ which represent the matrix as $A = BC$.[7] The matrix A can be considered to have low rank if its rank is much smaller than the number of rows and columns in the matrix; in this case, the storage cost for the representation, $(d_1 + d_2)r$, is much smaller than the storage cost $d_1 d_2$ for the full matrix, i.e. $r(d_1 + d_2) \ll d_1 d_2$.

---

[4] E.g. Vidal 2004.

[5] E.g. Oseledets 2011.

[6] Use e.g. an SVD (Section 1.2) and $\mathrm{rk}(BC) \leq \min\{\mathrm{rk}(B), \mathrm{rk}(C)\}$ (e.g. Horn and Johnson 1991a, 0.4.5(c)).

[7] The matrices B and C can be obtained e.g. from an SVD of the matrix. Cf. Section 1.2.

Direct generalization of (2.2) to tensors results in the following definition:

$$\mathrm{rk}_{\mathrm{CP}}(M) := \min\left\{r: \quad M = \sum_{j=1}^{r} x_j^{(1)} \otimes \cdots \otimes x_j^{(n)}, \quad x_j^{(k)} \in \mathbb{C}^{d_k}, k \in \{1 \ldots n\}\right\}. \tag{2.3}$$

The quantity $\mathrm{rk}_{\mathrm{CP}}(M)$ is called the *CP rank* or just *tensor rank* of M. CP is an abbreviation for *CANDECOMP/PARAFAC* or *canonical polyadic*. The tensor representation defined in (2.3) is called the CP decomposition and its storage cost is only drn, where $d = d_1 = \cdots = d_n$ and $r = \mathrm{rk}_{\mathrm{CP}}(M)$. There is an extensive amount of work concerning the CP decomposition but it suffers from the fact that computing the CP rank is NP-hard.[8] In this work, we will not use the CP decomposition.

An alternative definition of tensor rank is obtained by considering a subset $t \subset \Lambda$ of all modes $\Lambda := \{1 \ldots n\}$. The *$t$-matricization* or *$t$-unfolding* of a tensor M is denoted by $\mathcal{M}_t(M) \in \mathbb{C}^{d_t \times d'_t}$. It is the matrix with $d_t := \prod_{k \in t} d_k$ rows and $d'_t := \prod_{k \in \Lambda \setminus t} d_k$ columns which has the same entries as M.[9] The *t-unfolding rank* or *t-rank* of M is now defined as the matrix rank of its t-matricization:[10]

$$\mathrm{rk}_t(M) := \mathrm{rk}(\mathcal{M}_t(M)) \tag{2.4}$$

Different selections of subsets $t \subset \Lambda$ now are related to different tensor representations such as the Tucker representation, the MPS/TT representation and the hierarchical Tucker representation.[11] The MPS/TT representation is introduced in the next section and plays central role in this work. The other mentioned representations occur in Chapter 8.

2.2. Matrix product states and tensor trains (MPS/TT)

2.2.1. Definitions and basic properties

This section summarizes the main properties of the matrix product state (MPS) or tensor train (TT) tensor representations. Note that Section 2.2.2 discusses the representation in greater detail, introduces a graphical notation for the representations discussed here and also discusses related numerical algorithms.

---

[8] Håstad 1990; Kolda and Bader 2009.

[9] E.g. Grasedyck et al. 2013. A more formal definition is provided in Definition 8.10.

[10] This terminology is used at least for $t = \{k\}, k \in \Lambda$ by Kolda and Bader (2009) as well as Grasedyck et al. (2013).

[11] Grasedyck et al. 2013.

An MPS/TT representation[12] of a tensor $T \in \mathbb{C}^{d_1 \times \dots \times d_n}$ is given by

$$T_{i_1 \dots i_n} = G_1(i_1) G_2(i_2) \dots G_n(i_n), \qquad G_k(i_k) \in \mathbb{C}^{D_{k-1} \times D_k} \tag{2.5}$$

where $i_k \in \{1 \dots d_k\}$, $k \in \Lambda := \{1 \dots n\}$ and $D_0 = D_n = 1$.[13] The integers D_k are called *bond dimensions* or *representation ranks*. The representation's storage cost is proportional to ndD^2, where $d = \max_k d_k$ and $D_k = \max_k D_k$, which is linear in n if D is independent of n.

The tensor T from (2.5) has unfolding ranks $\mathrm{rk}_{[k]}(T) \leq D_k$, $k \in \{0 \dots n\}$ where $[k] = \{1 \dots k\}$. Conversely, any tensor T admits an MPS/TT representation with bond dimensions $D_k = \mathrm{rk}_{[k]}(T)$.

The MPS/TT tensor representation can be used to represent quantum state vectors, quantum density matrices and to purifications[14] of density matrices, which we set out to do in the following.[15] Using the MPS/TT representation to represent quantum state vectors constitutes the MPS representation in a strict sense. Let $\{|\phi^{(k)}_{i_k}\rangle\}_{i_k=1}^{d_k} \subset \mathcal{H}_k$ be an orthonormal basis of the k-th quantum system. An MPS representation of a pure state $|\psi\rangle \in \mathcal{H}^{(1)}_\Lambda$ is given by

$$\left\langle \phi^{(1)}_{i_1} \dots \phi^{(n)}_{i_n} \middle| \psi \right\rangle = G_1(i_1) G_2(i_2) \dots G_n(i_n) \tag{2.6}$$

where $D_0 = D_n = 1$, $G_k(i_k) \in \mathbb{C}^{D_{k-1} \times D_k}$ and $i_k \in \{1, \dots, d_k\}$.

The matrix product operator (MPO) representation[16] is obtained by applying the MPS/TT representation to linear operators $\rho \in \mathcal{B}(\mathcal{H}_\Lambda)$ or density matrices $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$ in the following way:

$$\left\langle \phi^{(1)}_{i_1} \dots \phi^{(n)}_{i_n} \right| \rho \left| \phi^{(1)}_{j_1} \dots \phi^{(n)}_{j_n} \right\rangle = G_1(i_1, j_1) G_2(i_2, j_2) \dots G_n(i_n, j_n) \tag{2.7}$$

where $D_0 = D_n = 1$, $G_k(i_k, j_k) \in \mathbb{C}^{D_{k-1} \times D_k}$ and $i_k, j_k \in \{1, \dots, d_k\}$. Alternatively, an MPO representation may be given in terms of operator bases: Let

---

[12] Fannes et al. 1992; Perez-Garcia et al. 2007; Schollwöck 2011; Oseledets 2011.

[13] A related representation is obtained by letting $D_0 = D_n$ be an arbitrary integer and by setting $T_{i_1 \dots i_n} = \mathrm{Tr}(G_1(i_1) G_2(i_2) \dots G_n(i_n))$. This representation has been called matrix product state with *periodic boundary conditions (PBC)* (in contrast to so-called *open boundary conditions (OBC)* in (2.5), cf. Perez-Garcia et al. 2007). Another name is *tensor chain* (e.g. Espig et al. 2012).

[14] A purification of $\rho \in \mathcal{D}(\mathcal{H}_A)$ is any $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ with $\langle\psi|\psi\rangle = 1$ and $\rho = \mathrm{Tr}_B(|\psi\rangle\langle\psi|)$. System B needs to have dimension $d_B \geq \mathrm{rk}(\rho)$. If $\rho = \sum_{j=1}^r \lambda_j |x_j\rangle\langle x_j|$ is an eigendecomposition and $\{|y_j\rangle : j \in [r]\}$ is an orthonormal sequence from $\mathcal{H}_B$, then $|\psi\rangle = \sum_{j=1}^r \sqrt{\lambda_j} |x_j\rangle \otimes |y_j\rangle$ is a purification of ρ.

[15] Parts of the following remainder of Section 2.2.1 were reproduced from Holzäpfel et al. (2018, Section II.B and Appendix A.5) with the permission of AIP Publishing.

[16] Zwolak and Vidal 2004; Verstraete et al. 2004. Note that the similar term matrix product density operator (MPDO) can refer to the PMPS representation (Verstraete et al. 2004) or to the MPO representation (De las Cuevas et al. 2013).

$\{F^{(k)}_{i_k} \colon i_k \in \{1 \dots d_k^2\}\}$ be a Hilbert–Schmidt orthonormal basis of $\mathcal{B}(\mathbb{C}^{d_k})$. An MPO representation of ρ is also given by

$$\left\langle F^{(1)}_{i_1} \otimes \dots \otimes F^{(n)}_{i_n}, \rho \right\rangle = G_1(i_1) G_2(i_2) \dots G_n(i_n) \tag{2.8}$$

where $D_0 = D_n = 1$, $G_k(i_k) \in \mathbb{C}^{D_{k-1} \times D_k}$ and $i_k \in \{1, \dots, d_k^2\}$. If the operator basis $F^{(k)}_{(i_k, j_k)} = \left|\phi^{(k)}_{i_k}\right\rangle\!\left\langle\phi^{(k)}_{j_k}\right|$ is used, Equation (2.8) turns into Equation (2.7).

The locally purified matrix product state (PMPS) representation[17] provides an alternative to the MPO representation for positive semidefinite linear operators such as density matrices. The purification[14] is given in terms of n ancilla systems of dimensions d'_k with orthonormal bases $\left\{\left|\epsilon^{(k)}_{i'_k}\right\rangle\right\}_{k=1}^{d'_k}$. A PMPS representation of a positive semidefinite $\rho \in \mathcal{B}(\mathcal{H}_\Lambda)$ is given by

$$\begin{aligned} \rho &= \operatorname{Tr}_{1' \dots n'}(|\Psi\rangle\langle\Psi|) & D_0 = D_n &= 1 \\ \left\langle \phi^{(1)}_{i_1} \epsilon^{(1)}_{i'_1} \dots \phi^{(n)}_{i_n} \epsilon^{(n)}_{i'_n} \middle| \Psi \right\rangle &= G_1(i_1, i'_1) G_2(i_2, i'_2) \dots G_n(i_n, i'_n) & i_k &\in \{1, \dots, d_k\} \\ G_k(i_k, i'_k) &\in \mathbb{C}^{D_{k-1} \times D_k} & i'_k &\in \{1, \dots, d'_k\} \end{aligned} \tag{2.9}$$

The linear operator ρ from the last equation is positive semidefinite by definition and it is a density matrix if $|\Psi\rangle$ has unit norm. The quantum state ρ can also be written as

$$\rho = MM^* \tag{2.10}$$

where the linear map M is defined by the MPO representation which uses the matrices $G_k(i_k, i'_k)$ from the PMPS representation (2.9), i.e.

$$\langle \phi^{(1)}_{i_1} \dots \phi^{(n)}_{i_n} | M | \epsilon^{(1)}_{i'_1} \dots \epsilon^{(n)}_{i'_n} \rangle = G_1(i_1, i'_1) G_2(i_2, i'_2) \dots G_n(i_n, i'_n). \tag{2.11}$$

Given the tensors G_k of a PMPS representation, the tensors $\tilde{G}_k$ of an MPO representation are given by

$$\tilde{G}_k(i_k, j_k) = \sum_{i'_k = 1}^{d'_k} G_k(i_k, i'_k) \otimes \overline{G_k(j_k, i'_k)} \tag{2.12}$$

---

[17] Verstraete et al. 2004; De las Cuevas et al. 2013. The structure of MPO and PMPS representations was used by Fannes et al. (1992) in the analysis of translationally invariant (TI) states of infinite spin chains. MPOs and PMPSs correspond to finitely correlated state (FCS) and C^*-FCS, respectively, where the infinite chain has been replaced by a finite chain and where the requirement of translational invariance has been dropped (Fannes et al. 1992, Definitions 2.2 and 2.4). In classical probability theory, a similar structure, called a hidden Markov model (HMM), has been used already in 1957 (Vidyasagar 2011, and references therein). For more information about other similar structures, refer to Kliesch et al. (2014a).

where the overline denotes the complex conjugate. Equation (2.12) shows that given a PMPS representation of bond dimension D, we can directly construct an MPO representation of bond dimension D^2.

As mentioned at the beginning of the section, the smallest bond dimensions D_k which allow for an exact representation of a tensor are determined by the tensor's unfolding ranks via $D_k = \mathrm{rk}_{[k]}(T)$. For the MPS, MPO and PMPS representations, these unfolding ranks correspond to the following Schmidt or operator Schmidt ranks, which is shown by comparing the definition of the matrix rank as per (2.2) with the definitions of the Schmidt ranks in Section 1.3. If $T_{|\psi\rangle}$ is the n-mode tensor from the MPS representation (2.6), then

$$\mathrm{rk}_{[k]}(T_{|\psi\rangle}) = \mathrm{SR}(1, \dots, k : k+1, \dots, n)_{|\psi\rangle}. \tag{2.13}$$

If T_ρ is the n-mode tensor from the MPO representation (2.8), then

$$\mathrm{rk}_{[k]}(T_\rho) = \mathrm{OSR}(1, \dots, k : k+1, \dots, n)_\rho. \tag{2.14}$$

If $T_{|\Psi\rangle}$ is the $2n$-mode tensor containing the entries of $|\Psi\rangle$ from the PMPS representation (2.9), then

$$\mathrm{rk}_{[2k]}(T_{|\Psi\rangle}) = \mathrm{SR}(1, a_1, \dots, k, a_k : k+1, a_{k+1}, \dots, n, a_n)_{|\Psi\rangle} \tag{2.15}$$

where a_k refers to the k-th ancilla system. Furthermore, if ρ is constructed according to (2.9), then

$$\begin{aligned} & \mathrm{OSR}(1, \dots, k : k+1, \dots, n)_\rho \\ \leq\ & \mathrm{OSR}(1, a_1, \dots, k, a_k : k+1, a_{k+1}, \dots, n, a_n)_{|\Psi\rangle\langle\Psi|} && \text{(2.16a)} \\ =\ & \left[\mathrm{SR}(1, a_1, \dots, k, a_k : k+1, a_{k+1}, \dots, n, a_n)_{|\Psi\rangle}\right]^2. && \text{(2.16b)} \end{aligned}$$

The preceding equations show that the amount of correlations as measured by the Schmidt or operator Schmidt rank determines the bond dimension of an exact MPS or MPO representation of a state vector or density matrix. Furthermore, the OSR of a density matrix implies only a lower bound on the bond dimension of an exact PMPS representation because a given density matrix can be represented by a family of purifications.[18]

We conclude this section with a short comparison of the MPO and PMPS representations, which can both be used to represent positive semidefinite operators such as density matrices. The MPO representation is conceptually simple and can represent not only positive semidefinite operators but arbitrary operators. Most numerical algorithms involve approximating an MPO by an MPO of smaller bond dimension and this process can replace a positive semidefinite operator with a linear operator which is no longer positive semidefinite. This can be a serious problem because verifying whether a given MPO is positive semidefinite has shown to be NP-hard in the number of sites n, i.e. a solution in polynomial (in n) time is unlikely.[19] Positivity, once destroyed, cannot be recovered efficiently in the general case. In a particular instance of a particular numerical scheme, however, it may well be possible to restore positivity efficiently.

---

[18] Unitary operators on the ancilla systems do not affect the represented density matrix but can change the value of $|\Psi\rangle$'s Schmidt rank.

Above, it was mentioned that a PMPS representation of bond dimension D implies that an MPO representation of bond dimension D^2 exists as well. However, an MPO representation with bond dimensions smaller than D^2 can exist. It has been shown that there is a family of quantum states on n systems which can be represented as an MPO with bond dimension independent of n but the bond dimension of any PMPS representation of those states increases with n.[20] The PMPS representation guarantees that the represented linear operator is positive semidefinite while the MPO representation can provide a bond dimension which scales more favorably with n in some cases. In conclusion, the relative merits of the MPO and PMPS representations of a mixed quantum state depend on the application.

2.2.2. Efficient methods

The MPS/TT representation can be efficient in the sense that the storage cost of the representation increases only linearly or polynomially with the number of modes n (if the maximal bond dimension D is constant or polynomial in n, respectively). A tensor representation is of limited use, however, without methods for operations involving tensors which are also efficient, i.e. whose processing time increases at most polynomially with the number of modes n. In this section, we discuss the MPS/TT representation in more detail and sketch the outlines of such methods. We discuss methods for computing inner products, operator-vector products as well as operator-operator products. Furthermore, we discuss the approximation of tensors by tensors of lower bond dimension and the problem of computing extremal eigenvalues of Hermitian, linear operators. A full description of the named methods can be found in the literature.[21]

---

[19] Kliesch et al. 2014a.

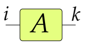
[20] De las Cuevas et al. 2013.

In the following, we assume that the maximal bond dimension D does not increase with the number of modes n. As in Section 2.2.1, vectors from $\mathcal{H}_\Lambda = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ are mapped isomorphically to tensors from $\mathbb{C}^{d_1 \times \cdots \times d_n}$ by means of orthonormal bases of each $\mathcal{H}_k$. In the same way, linear operators from $\mathcal{B}(\mathcal{H}_\Lambda)$ are mapped to tensors with $2n$ modes from $\mathbb{C}^{d_1 \times d_1 \times \cdots \times d_n \times d_n}$.

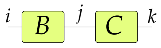
The MPS/TT representation. Any matrix A can be written as a product of two other matrices,

$$A = BC, \tag{2.17}$$

where B has m columns and C has m rows if m is not smaller than the rank of the matrix, i.e. $m \geq \operatorname{rk}(A)$ (cf. Equation (2.2)). A graphical representation of the matrix factorization from Equation (2.17) is given by

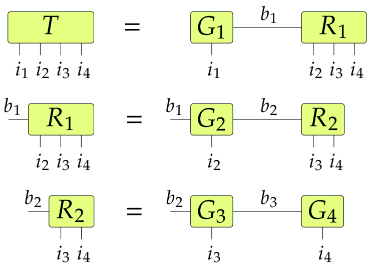


$$= A_{ik} = (BC)_{ik} = \sum_{j=1}^{m} B_{ij} C_{jk} =$$



(2.18)

A possible generalization for a four-mode tensor T is given by the following graphical equations:

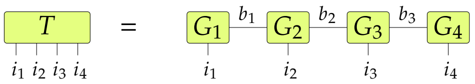


$b_1 \in \{1, \dots, D_1\}$

$b_2 \in \{1, \dots, D_2\}$

$b_3 \in \{1, \dots, D_3\}$ (2.19)

In each row, the tensors G_k and R_k are obtained from R_{k-1} by means of a (e.g. rank-revealing) matrix decomposition, introducing the index b_k (at the beginning and end, we have $R_0 = T$ and $R_3 = G_4$). In the end, the tensors G_k suffice to represent T:



(2.20)

---

[21] Schollwöck 2011; Oseledets 2011.

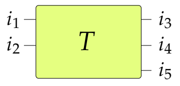
The last equation as a formula reads

$$\begin{aligned} T_{i_1 \dots i_n} &= \sum_{b_0=1}^{D_0} \cdots \sum_{b_n=1}^{D_n} [G_1]_{b_0 i_1 b_1} [G_2]_{b_1 i_2 b_2} \dots [G_n]_{b_{n-1} i_n b_n} && (2.21) \\ &= G_1(i_1) \dots G_n(i_n) && (2.22) \end{aligned}$$

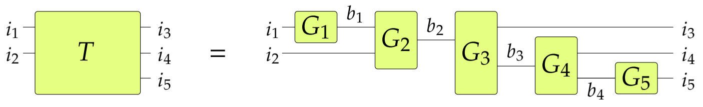
where $[G_k(i_k)]_{b_{k-1}, b_k} = [G_k]_{b_{k-1} i_k b_k}$. The tensors G_k are called *core tensors*, the indices b_k and their dimensions D_k are called *bond indices* and *bond dimensions* and we have introduced the dummy bond indices b_0 and b_n of dimensions $D_0 = D_n = 1$. The last equation is an MPS/TT representation of T as defined in (2.5). The procedure outlined in (2.19) therefore shows how to construct an MPS/TT representation of an unknown tensor T. If a rank-revealing matrix decomposition is used in each step, the obtained bond dimensions equal T's corresponding unfolding ranks.

The bond dimension/representation rank. Next, we show that the bond dimensions are related to the so-called unfolding matrices. The $\{1, 2\}$-matricization or unfolding of T is the following matrix of shape $d_1 d_2 \times d_3 d_4 d_5$:

$$[\mathcal{M}_{12}(T)]_{i_1 i_2, i_3 i_4 i_5} = \qquad (2.23)$$



We can write unfoldings of the type $\mathcal{M}_{[k]}(T)$ ($[k] = \{1 \dots k\}$) as a composition of linear maps constructed from the core tensors G_k:

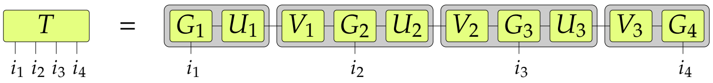


(2.24)

Because e.g. the bond index b_2 runs over $b_2 \in \{1, \dots, D_2\}$, decompositions of this kind imply that the following unfolding ranks cannot exceed the bond dimensions:

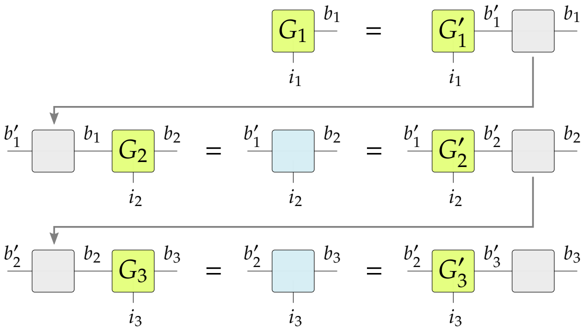
$$\operatorname{rk}\left(\mathcal{M}_{[k]}(T)\right) \le D_k \qquad (2.25)$$

We discuss how a given MPS/TT representation can be converted to a representation with $D_k = \operatorname{rk}(\mathcal{M}_{[k]}(T))$. A tensor T represented by cores G_k as shown in (2.20) is also represented by the following modified core tensors, under the condition that the rectangular matrices U_i and V_i satisfy $U_i V_i = \mathbb{1} \in \mathbb{C}^{D_k \times D_k}$:



(2.26)

For example, the matrices U_i and V_i could be used to increase the bond dimension of the representation beyond the previous value. The bond dimension of a given representation can be reduced to the unfolding rank by applying rank-revealing matrix decompositions starting as follows:

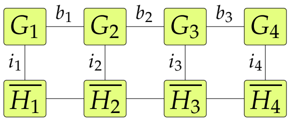


(2.27)

A rank-revealing matrix decomposition is used repeatedly to introduce the new indices b'_k whose dimension equals the rank of the corresponding matrix. Transformations which proceed through the chain step by step as shown in the last Equation are often called *sweeps*. A representation with $D_k = \operatorname{rk}(\mathcal{M}_{[k]}(T))$ for a single value of k is obtained by sweeping from both ends of the chain up to the desired point (D_k is changed twice). Alternatively, one can obtain a representation with $D_k = \operatorname{rk}(\mathcal{M}_{[k]}(T))$ for all k with a complete left-to-right sweep followed by a complete right-to-left sweep. (or vice-versa).[22] The mentioned operations are efficient because they involve at most one core tensor at a time (assuming bond dimensions which do not increase with n).

Norm and inner product. The standard inner product of two tensors S and T is given by $\langle S, T\rangle = \sum_{i_1\dots i_n} \overline{S_{i_1\dots S_{i_n}}} T_{i_1\dots i_n}$ and it induces a norm $\|T\| = \sqrt{\langle T, T\rangle}$ for tensors. Computing the standard inner product is another efficient operation:

$$\langle S, T\rangle \quad = \quad \sum_{i_1=1}^{d_1} \cdots \sum_{i_n=1}^{d_n} \overline{S_{i_1\dots i_n}} T_{i_1\dots i_n} \quad = \qquad (2.28)$$



Here, the contraction order is important: If we contract $G_1, \dots, G_d$ first, a tensor with d^n entries is created as intermediate result and the procedure is not efficient (in the figure, $n = 4$). However, if we contract in the order $G_1, H_1, G_2, H_2, \dots$, the size of any intermediate result is at most dD^2 and the procedure is efficient.

---

[22] Schollwöck 2011; Oseledets 2011.

Sums. As mentioned above, the MPS/TT representation can be expressed as follows:

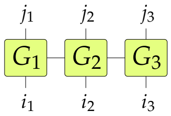
$$T_{i_1\dots i_n} = G_1(i_1)G_2(i_2)\dots G_n(i_n) \qquad G_k(i_k) \in \mathbb{C}^{D_{k-1}\times D_k} \qquad D_0 = D_n = 1 \tag{2.29}$$

The sum of two tensors T and S represented by core tensors G_k and H_k has a representation with bond dimension equal to the sum of the bond dimensions:

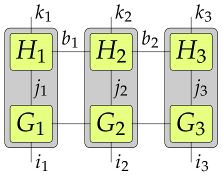
$$T_{i_1\dots i_n} + S_{i_1\dots i_n} = \begin{pmatrix} G_1(i_1) & H_1(i_1) \end{pmatrix} \begin{pmatrix} G_2(i_2) & 0 \\ 0 & H_2(i_2) \end{pmatrix} \dots \begin{pmatrix} G_n(i_n) \\ H_n(i_n) \end{pmatrix} \tag{2.30}$$

Operator representations and operator products. Consider two linear operators $A, B \in \mathcal{B}(\mathcal{H}_\Lambda)$. The following representation of A allows for the efficient computation of the product AB:

$$[A]_{i_1i_2i_3,j_1j_2j_3} =$$



$$[AB]_{i_1i_2i_3,k_1k_2k_3} =$$



(2.31)

The representation of A from the last equation is the matrix product operator (MPO) representation from (2.7). The product can be computed in the same way if A and B are suitable linear maps (instead of operators) and, in particular, if B is replaced by a vector $|\psi\rangle \in \mathcal{H}_\Lambda$. The last equation already shows that the bond dimension of the representation of the product equals the product of the bond dimensions of the individual representations.

Compression to/approximation by a tensor with smaller bond dimension. As the examples of the sum and product of two tensors have shown, operations on tensors often increase the bond dimension of the representation. This calls for methods for finding approximate representations of reduced bond dimension. If the SVD is used in the scheme to find the minimal exact bond dimension from Equation (2.27), approximate representations can be found by truncating small singular values, reducing the bond dimension.[23] Based on this approach, a quasi-optimal approximation can be found, meaning that the obtained approximation error is at most the minimal possible error (for a given bond dimension) times

---

[23] For reliable results, cores should be in a suitable *canonical* or *orthogonal* form, which is not discussed here. See e.g. Schollwöck 2011; Oseledets 2011.

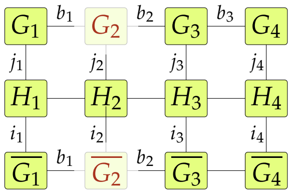
$\sqrt{n-1}$ (*SVD compression*).[24] As an alternative, one can try to solve the maximization problem $\arg\max_S |\langle T, S\rangle| / \|T\| \|S\|$ with a restriction on the bond dimension on S (*variational compression*).[25] It has the same solution as $\arg\max_S \langle S, T\rangle\langle T, S\rangle / \langle S, S\rangle$, whose efficient, heuristic solution is discussed in some detail in the next paragraph.

Computing eigenvectors/DMRG. Consider a Hermitian linear operator $A \in \mathcal{B}(\mathcal{H}_\Lambda)$ and a vector $v \in \mathcal{H}_\Lambda$, represented as an MPO and MPS, respectively. The result of the maximization

$$\arg\max_v \langle v, Av\rangle / \langle v, v\rangle \tag{2.32}$$

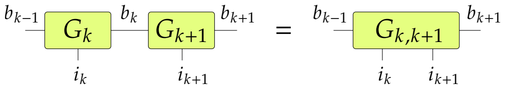
is an eigenvector of the largest eigenvalue of A. If A and v are represented as MPO and MPS with cores G_k and H_k, one can attempt to find such an eigenvector efficiently as follows. One can hope to increase the value of $\langle v, Av\rangle / \langle v, v\rangle$ by replacing G_k with a solution of[26]

$$\arg\max_{G_k} \frac{\langle v, Av\rangle}{\langle v, v\rangle}, \qquad \langle v, Av\rangle(\overline{G_2}, G_2) \quad =$$



. (2.33)

The right hand side of the Equation shows the tensor network which provides the value of $\langle v, Av\rangle$ as a function of the core tensors G_k and $\overline{G_k}$ (for $k = 2$). The tensor network can be contracted efficiently by contracting from both ends to mode k and the resulting tensor allows direct solution of the maximization on the left hand side. One attempts to find a global maximum by repeated local optimizations. Instead of reaching the global maximum, such an algorithm can get stuck in a local maximum. A partial remedy is provided by a modified optimization where two neighbouring cores are contracted into one supercore $G_{k,k+1}$:

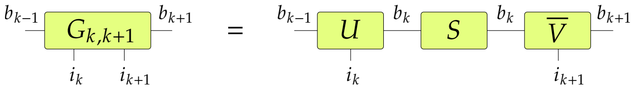


(2.34)

One step optimizes over a complete supercore, i.e. over a greater number of parameters, reducing but not eliminating the risk of getting stuck in a local maximum. After the optimization the supercore must be split into two normal cores, e.g. using an SVD:

---

[24] Vidal 2003, 2004; Oseledets 2011; Schollwöck 2011.
[25] Schollwöck 2011; Holtz et al. 2012.
[26] Schollwöck 2011; Holtz et al. 2012.



(2.35)

Because the splitting can increase the bond dimension, a truncation of singular values can be necessary to keep the bond dimension constant. The approximation error can be used as a means to gauge whether the chosen bond dimension is sufficiently large.[27] Further improvements and applications of the sketched variational algorithm have been demonstrated recently, with applications including the solution of a linear system specified by an MPO as well as the computation of inverses, extreme singular values and pseudoinverses of MPOs.[28] For the relation of these variational MPS/TT algorithms to earlier algorithms based on the density matrix renormalization group (DMRG) and for a more complete introduction to MPS/TT representations, we refer the reader to the literature.[29]

2.2.3. Examples

A product state

$$|\psi\rangle = |\psi\rangle_1 \otimes \cdots \otimes |\psi\rangle_n \tag{2.36}$$

admits an MPS representation with bond dimension one. The Greenberger–Horne–Zeilinger (GHZ) state

$$|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}}\big(|00\ldots0\rangle + |11\ldots1\rangle\big) \tag{2.37}$$

admits an MPS representation with bond dimension two because it is the sum of two MPSs with bond dimension 1. The W state

$$|\mathrm{W}_n\rangle = \frac{1}{\sqrt{n}}\big(|100\ldots0\rangle + |010\ldots0\rangle + \ldots\big) \tag{2.38}$$

admits an MPS representation with bond dimension n. It also admits an MPS representation with bond dimension two because its $1\ldots k|k+1\ldots n$ Schmidt ranks are equal to two, which is shown by writing

---

[27] Schollwöck 2011; Oseledets and Dolgov 2012.

[28] Kressner et al. 2014; Oseledets and Dolgov 2012; Dolgov and Savostyanov 2014; Lee and Cichocki 2015, 2016.

[29] Schollwöck 2011; Oseledets 2011.

$$|\mathrm{W}_n\rangle = \frac{1}{\sqrt{2}}\Big(\sqrt{n-k}|00\dots0\rangle_{1\dots k} \otimes |\mathrm{W}_{n-k}\rangle_{k+1\dots n} + \sqrt{k}|\mathrm{W}_k\rangle_{1\dots k} \otimes |00\dots0\rangle_{k+1\dots n}\Big).$$

An arbitrary state vector admits an MPS representation with bond dimension $D = d^{\lfloor n/2 \rfloor}$ (this is an upper bound on the Schmidt ranks $\mathrm{SR}(1,\dots,k : k+1,\dots,n)$).

A product observable

$$A = A_1 \otimes \dots \otimes A_n \tag{2.39}$$

admits an MPO representation with bond dimension one. In the case of uniform local dimensions $d_i = d$, an observable which acts on r qubits admits an MPO representation of bond dimension $D = d^r$ (upper bound on $\mathrm{OSR}(1,\dots,k : k+1,\dots,n)$). A local Hamiltonian on a one-dimensional, linear chain whose terms act on r neighbouring sites,

$$H = \sum_{i=1}^{n-r+1} \mathbb{1}_{1\dots i-1} \otimes h_i \otimes \mathbb{1}_{i+r\dots n}, \tag{2.40}$$

where $h_i \in \mathcal{B}(\mathcal{H}_i,\dots,\mathcal{H}_{i+r-1})$, admits an MPO representation of bond dimension $(n-r+1)d^r$. However, with an argument similar to the one used for the W state, one can show that an MPO representation with bond dimension independent of n is possible.[30]

Consider the time-evolved state $|\psi(t)\rangle = \exp(-\mathrm{i}Ht)|\psi\rangle$ where $|\psi\rangle$ is a product state $|\psi\rangle$ (Equation (2.36)) and H is a one-dimensional, local Hamiltonian H (Equation (2.40)) with $\|h_i\|_{(\infty)} \le J$ and r as well as J independent of n. The time-evolved state $|\psi(t)\rangle$ admits an approximate MPS representation of bond dimension $D = O(\exp(vt) + \mathrm{poly}(n/\epsilon))$ where ϵ is the approximation error. At fixed time t and approximation error ϵ, the bond dimension scales polynomially in n and the representation is efficient. This has been shown by Osborne (2006). The result is generalized to local Hamiltonians on arbitrary-dimensional lattices in Theorem 6.33 below.[31]

2.3. Projected entangled pair states (PEPS)

The MPS/TT representation as well as the related MPO and PMPS representations have a one-dimensional structure and they were observed to be useful for one-dimensional quantum systems in the examples from the last section as well as many other applications. One possible generalization of MPS and MPO representations to higher dimensions is given by projected entangled pair states (PEPSs) and projected entangled pair operators (PEPOs),[32] which are introduced in this section. We also prove a bound on the PEPO bond dimension of a unitary circuit composed of a limited number of unitary gates which act on a limited number of neighbouring sites.[33]

---

[30] See e.g. Section 6.1 of Schollwöck (2011). A similar scheme has been implemented in the function `mpnum.local_sum` of the open source library `mpnum` (Suess and Holzäpfel 2017).

[31] See Page 117 or Holzäpfel and Plenio 2017.

Let $\Lambda = \{1 \dots n\}$ be the set of all systems. The systems need not be in a linear chain but we assign the names $1, \dots, n$ to the sites of the system in an arbitrary order. Let $d(x)$ denote the dimension of system $x \in \Lambda$. Let $\left|\phi_i^{(x)}\right\rangle$ ($i \in \{1 \dots d(x)\}$) denote an orthonormal basis of system x. As above, the components of a pure state $|\psi\rangle$ on the n systems are given by

$$t_{i_1 \dots i_n} = \left\langle \phi_{i_1}^{(1)} \dots \phi_{i_n}^{(n)} \middle| \psi \right\rangle, \qquad i_x \in \{1 \dots d(x)\}. \tag{2.41}$$

The last equation shows that the pure state on n systems corresponds to a tensor t with n indices of shape $d(1) \times \cdots \times d(n)$. A PEPS representation of $|\psi\rangle$ or t is defined in terms of a graph (Λ, E) whose vertices correspond to sites $x \in \Lambda$ (Figure 2.1 left). The graph (Λ, E) is assumed to be simple and connected, i.e. each edge $e \in E$ connects exactly two distinct sites and each vertex is connected to any other vertex by a suitable sequence of edges. The set of neighbours of $x \in \Lambda$ is given by $N(x) = \{y \in \Lambda\colon \{x, y\} \in E\}$ and the number of neighbours (degree) is given by $z_x = |N(x)|$. We denote the edges involving $x \in \Lambda$ in some arbitrary, fixed order by $\{n_1^{(x)} \dots n_{z_x}^{(x)}\}$; i.e. $n_k^{(x)} = \{x, y\} \in E$ for one $y \in N(x)$. For each edge $e \in E$, choose a positive integer $D(e)$, called the bond dimension. The maximal local and bond dimension are denoted by $d = \max_{x \in \Lambda} d(x)$ and $D = \max_{e \in E} D(e)$. For $x \in \Lambda$, let G_x be a tensor of size $d(x) \times D(n_1^{(x)}) \times \cdots \times D(n_{z_x}^{(x)})$. Let $\{e_1 \dots e_{|E|}\} = E$ be an enumeration of all the edges. A PEPS representation[34] of the tensor t is given by (Figure 2.1 middle)

$$t_{i_1 \dots i_n} = \sum_{b(e_1)=1}^{D(e_1)} \cdots \sum_{b(e_{|E|})=1}^{D(e_{|E|})} \prod_{x=1}^{n} G_x\left[i_x, b\big(n_1^{(x)}\big), \dots, b\big(n_{z_x}^{(x)}\big)\right] \tag{2.42}$$

A PEPS representation of a pure quantum state is given by the combination of Equations (2.41) and (2.42). Any tensor or quantum state can be represented as

---

[32] Verstraete and Cirac 2004; Verstraete et al. 2008.

[33] The content of Section 2.3 is reproduced from Holzäpfel and Plenio 2017, Section 4.2.

[34] Verstraete and Cirac 2004; Schuch et al. 2007; Verstraete et al. 2008.

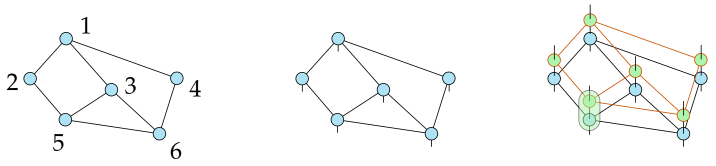


Figure 2.1: Left: Graph (Λ, E) whose vertices correspond to lattice sites $x \in \Lambda$. Lattice sites have been named $\{1 \dots 6\} = \Lambda$. Middle: Graphical representation of the tensor network which constitutes a PEPS representation of a quantum state on the lattice. Circles correspond to tensors, lines correspond to indices and lines which connect two circles indicate which indices are contracted (Equation (2.42); cf. Section 2.2.2). Right: The shaded box exemplifies how a PEPO representation of the operator product is obtained (Equation (2.44)).

PEPS if the bond dimensions $D(e)$ are chosen sufficiently large (cf. Lemma 2.2 below).

A projected entangled pair operator (PEPO) representation[35] of a linear operator G on n quantum systems is given by a PEPS representation of the following tensor:

$$t_{(i_1,j_1)\dots(i_n,j_n)} = \langle \phi^{(1)}_{i_1} \dots \phi^{(n)}_{i_n} | G | \phi^{(1)}_{j_1} \dots \phi^{(n)}_{j_n} \rangle \tag{2.43}$$

As for the MPO representation described above, the tensor t is treated as tensor with n indices and shape $[d(1)]^2 \times \dots \times [d(n)]^2$. Suppose that two linear operators G and H have PEPO representations given by tensors G_x and H_x with bond dimensions $D_G(e)$ and $D_H(e)$. The following formula provides the tensors of a PEPO representation of the operator product $F = GH$ (Figure 2.1 right):

$$\begin{aligned} & F_x\left[i_x, k_x, b\left(n^{(x)}_1\right), \dots, b\left(n^{(x)}_{z_x}\right)\right] \\ &= \sum_{j_x=1}^{d(x)} G_x\left[i_x, j_x, c\left(n^{(x)}_1\right), \dots, c\left(n^{(x)}_{z_x}\right)\right] H_x\left[j_x, k_x, d\left(n^{(x)}_1\right), \dots, d\left(n^{(x)}_{z_x}\right)\right] \end{aligned} \tag{2.44}$$

where $b\left(n^{(x)}_k\right) = \left(c\left(n^{(x)}_k\right), d\left(n^{(x)}_k\right)\right)$ $(k \in \{1 \dots z_x\})$. Equation (2.44) proves the following lemma:

Lemma 2.1 *Let G and H be operators with PEPS bond dimensions $D_G(e)$ and $D_H(e)$. The operator product $F = GH$ admits a PEPS representation with bond dimension $D_F(e) = D_G(e) D_H(e)$ $(e \in E)$.*

---

[35] E.g. Molnar et al. 2015.

The next lemma gives an explicit upper bound on the bond dimension of the PEPS representation of an arbitrary tensor:

Lemma 2.2 *Let t be a tensor with n indices and size $d(1) \times \cdots \times d(n)$. Then t admits a PEPS representation with maximal bond dimension $D \leq d^n$ where $d = \max_{x \in \Lambda} d(x)$.*

Remark 2.3 When applying Lemma 2.2 to an operator which acts non-trivially on a region $Y \subset \Lambda$ observe that this region Y must be connected in terms of the PEPS graph (Λ, E). □

Proof (of Lemma 2.2) Suppose that the connected, simple graph (Λ, E) is such that it admits a permutation $(x_1, \dots, x_n)$ of all the vertices such that $\{x_{k-1}, x_k\} \in E$ is a valid edge ($k \in \{2 \dots n\}$; such a permutation is called a *Hamiltonian path*). In this case, an MPS/TT representation of the suitably permuted tensor provides a valid PEPS representation with bond dimension[36] $D \leq d^{\lfloor n/2 \rfloor} < d^n$. However, the graph may not admit such a permutation.[37] In this case, we perform a depth-first search (DFS) on the graph to obtain a tree graph with the same vertices and a subset of the edges of the original graph (we can start the DFS on any vertex). Walking through the resulting tree graph in the DFS order visits each vertex at least once and each edge at most twice.[38] The tensor with indices permuted according to their first visit in the DFS order[39] can be represented as MPS/TT of bond dimension $d^{\lfloor n/2 \rfloor}$. Because each edge is visited at most twice, the resulting MPS can be converted to a PEPS with bond dimension $D \leq (d^{\lfloor n/2 \rfloor})^2 \leq d^n$. ■

The bond dimension of a unitary circuit can be bounded with Lemmata 2.1 and 2.2 as follows:

Lemma 2.4 *Let $U = U_1 \dots U_G$ a unitary circuit composed of G gates where each gate U_j acts non-trivially on at most K connected sites ($j \in \{1 \dots G\}$). Let at most L gates act on any site of the system. The unitary U admits an exact PEPO representation of bond dimension $D \leq d^{2KL}$ where $d = \max_{x \in \Lambda} d(x)$ is the maximum local dimension.*

Proof The statement is proven by repeating a simple counting argument which has been used before by Jozsa (2006) for one-dimensional MPS.[40] As the operator

---

[36] E.g. Schollwöck 2011.

[37] Example: A central vertex connected to three surrounding vertices.

[38] Tarry's algorithm returns a bidirectional double tracing, i.e. a walk over the graph which visits each edge exactly twice (J. L. Gross et al. 2014, Sec. 4.2.4). Omitting visits to already-visited vertices in this walk represents a depth-first search (J. L. Gross et al. 2014, Sec. 2.1.2).

[39] It would be equally permissible to use the second or a later visit in the DFS order.

[40] The argument could be improved by counting how often each edge is used instead of counting how often each site is acted upon; cf. Holzäpfel et al. (2015).

U_j acts non-trivially on at most K connected sites, it admits a PEPO representation with bond dimension $D' = (d^2)^K$ (Lemma 2.2). At each edge, the bond dimension of U is at most the product of the bond dimensions of the operators $U_1, \ldots, U_G$ (Lemma 2.1): $D_U(e) \le \prod_{j=1}^{G} D_{U_j}(e)$, $e \in E$. We have $D_{U_j}(e) \le D'$ for all edges and $D_{U_j}(e) = 1$ if the edge e involves a site x on which U_j acts as the identity. At most L of the G operators $U_1, \ldots, U_G$ act non-trivially on an arbitrary site x and this bounds U's bond dimension to $D_U(e) \le (D')^L = d^{2KL}$ for all edges. ∎

Chapter 3.

Quantum state tomography

Quantum state tomography (QST), which is also called quantum state estimation, refers to the task of determining the unknown state of a quantum system from measurement outcomes. The state of a quantum system is described by a density matrix ρ which is a linear operator on a Hilbert space $\mathcal{H}_\Lambda$ whose dimension d^n grows exponentially with the number n of subsystems.[1] Quantum measurements provide an estimate of $y = \mathcal{M}(\rho) \in \mathbb{R}^\mu$ where $\mathcal{M}$ is a linear map. Determining an arbitrary density matrix ρ requires $\mu \geq d^{2n} - 1$, corresponding to measurements whose complexity grows exponentially with n. (see Section 3.1.1 below). However, additional constraints can enable determining ρ uniquely from $y = \mathcal{M}(\rho)$ even if $\mu < d^{2n-1}$. A method for quantum state tomography which is *efficient* or *scalable* in the sense that the necessary measurement and post-processing effort increases at most polynomially with n can be given by the following steps:[2]

1. **Measure:** Choose a measurement $\mathcal{M} \in \mathcal{B}(\mathcal{B}_H(\mathcal{H}_\Lambda); \mathbb{R}^\mu)$ where $\mu = O(\text{poly}(n))$. Obtain an estimate y_{est} of $y = \mathcal{M}(\rho)$.
2. **Estimate:** Choose a subset $S \subsetneq \mathcal{D}(\mathcal{H}_\Lambda)$ or $S \subsetneq \mathcal{B}_H(\mathcal{H}_\Lambda)$ and determine a $\sigma_{\text{est}} \in S$ which satisfies $\mathcal{M}(\sigma_{\text{est}}) \approx y_{\text{est}}$. The set S contains all linear operators or density matrices which satisfy reasonable but unverified assumptions.
3. **Verify** that $\sigma_{\text{est}} \approx \rho$ using only $y_{\text{est}} \approx \mathcal{M}(\rho)$ or additional measurements of complexity polynomial in n. This step is also called *certification*.

The method is efficient if all steps are performed with $O(\text{poly}(n))$ storage space and processing time. Considering the tensor representations from the previous chapter, restricting the amount of correlations by bounding the Schmidt or operator Schmidt rank appears as an obvious constraint. In addition, it can be beneficial to restrict to positive semidefinite density matrices if the chosen estimation method admits that. If ρ is assumed to be pure or nearly pure, estimation is simplified by restricting to unit rank density matrices. The remainder of this chapter provides an introduction to standard ("inefficient") methods for quantum state tomography and also presents known methods for efficient quantum state estimation and verification.

---

[1] For simplicity, we set $d_k = d$, $k \in \Lambda$.

[2] Cramer et al. 2010.

3.1. Basic concepts

In this section, we discuss the complexity of quantum state tomography, explain when a measurement is called *informationally complete (IC)* and define frequently-used observables.

3.1.1. Complexity

At the beginning of the chapter, we already suggested that the complexity of quantum state tomography generally increases exponentially with the number n of subsystems. We proceed by discussing this question in more detail, in particular for the case where only an approximate description of a quantum state required; obtaining an approximate description of a quantum state could be easier than obtaining an exact description. Our main, unsurprising result is that an approximate description of a density matrix still requires storage space exponential in n (Lemma 3.1). Therefore, the effort for QST of an arbitrary state also grows exponentially with n.

The state of a quantum system is described by a state vector $|\psi\rangle \in \mathcal{H}_\Lambda^{(1)}$ or a density matrix $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$. In the case of n subsystems of uniform dimension $d \geq 2$, state vectors and density matrices are represented by vectors and matrices from $\mathbb{C}^N$ or $\mathbb{C}^{N\times N}$, respectively, where $N = d_\Lambda = d^n$. It appears inevitable that the difficulty of determining a quantum state increases exponentially with n since a vector or matrix with d^n or $(d^2)^n$ entries is to be determined. However, some states turn out to be easy to identify. Suppose that an observable $M = |\phi\rangle\langle\phi|$ is measured and yields the expectation value $\operatorname{Tr}(M\rho) \approx 1$ ($|\phi\rangle \in \mathcal{H}_\Lambda^{(1)}$). The only solution $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$ of the equation $\operatorname{Tr}(M\rho) = 1$ is $\rho = |\phi\rangle\langle\phi|$ (more on this example in the next subsection).

In a system comprised of n two-dimensional subsystems, effort polynomial in n is sufficient for the following problem which is related to QST and which was proposed by Aaronson (2007). Consider a probability distribution $\mathcal{P}$ over observables $M \in \mathcal{B}_H(\mathcal{H}_\Lambda)$ with $\|M\|_{(\infty)} \leq 1$. Instead of demanding, as in QST, that expectation values of all observables can be predicted with error at most γ, the task is restricted to predicting the expectation values of observables drawn from $\mathcal{P}$ with error at most γ and success probability $1-\epsilon$. The prediction is to be achieved using expectation values of accuracy $\sim \gamma\epsilon$ and the expectation values of m observables drawn from $\mathcal{P}$ are available. Remarkably, this data is already sufficient[3] with probability $1-\delta$ if m grows linearly with $(\gamma\epsilon)^{-4}\log^2(1/\gamma\epsilon)$, linearly with $\log(1/\delta)$ and, as a surprise, linearly with the number n of qubits. This approach does not provide a description of the density matrix ρ, but it shows that observables drawn from a probability distribution can in principle be predicted correctly with high probability and with effort linear in n.

Next, we show that specifying a general density matrix $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$ requires at least $\sim d^n - 1$ bits even if a (sufficiently small) finite error in trace distance is allowed ($d = d_i$, $i \in \Lambda$). Since describing the output of QST, which is a general density matrix, requires effort exponential in n, one can conclude that at least the same effort is required for QST itself in general.

Any method for specifying a density matrix approximately needs to distinguish between at least a certain finite number of regions in $\mathcal{D}(\mathcal{H}_\Lambda)$. To this end, consider a set C of subsets of $\mathcal{D}(\mathcal{H}_\Lambda)$ with the following properties:

$$\mathcal{D}(\mathcal{H}_\Lambda) = \bigcup_{A \in C} A \tag{3.1a}$$

where for each $A \in C$, there is an $M \in \mathcal{B}(\mathcal{H}_\Lambda)$ such that

$$A \subset B_\epsilon(M) := \left\{\rho \in \mathcal{D}(\mathcal{H}_\Lambda) \colon D(\rho, M) \leq \epsilon\right\} \tag{3.1b}$$

where $D(M,N) := \frac{1}{2}\|M-N\|_{(1)}$ is the trace distance ($M, N \in \mathcal{B}(\mathcal{H}_\Lambda)$). An arbitrary density matrix ρ can now be described approximately by choosing $A \in C$ such that $\rho \in A$. Any approximate description of a general density matrix gives rise to a set C with the given properties. The trace distance radius ϵ describes the accuracy of the description.

If there was a set C which satisfied (3.1) and contained only e.g. 2^n elements, then $\log_2(2^n) = n$ bits would be sufficient to enumerate C's elements and to represent general density matrices approximately. However, the following lemma shows that $|C| \geq c^{N-1}$, which implies that at least $\log_2(c^{N-1}) = (N-1)\log_2(c)$ bits are required for any approximate description of a density matrix. The constant is $c = \frac{1}{8\epsilon}$ and it satisfies $c > 1$ for $\epsilon < \frac{1}{8}$. In the case of uniform dimensions, $d_i = d$, $i \in \Lambda$, we can insert $N = d_\Lambda = d^n$. The approximate description of a density matrix $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$ accordingly requires at least $(d^n - 1)\log_2(c)$ bits.

---

[3] Aaronson 2007.

Lemma 3.1 *The set C from (3.1) has at least $(\frac{1}{8\epsilon})^{N-1}$ elements where $N := d_\Lambda$.*

In the following, we provide a proof sketch for the claimed scaling and a more involved rigorous proof of Lemma 3.1.

Proof (Sketch) Covering the set $\mathcal{H}_\Lambda^{(1)} = \{|\psi\rangle \in \mathcal{H}_\Lambda \colon \langle\psi|\psi\rangle = 1\}$ with ϵ-balls $B_\epsilon^{(v)}$ in vector-2 norm requires roughly $(1/\epsilon)^{2N-1}$ balls, where $N = d_\Lambda$.[4] This can be seen be relating $\mathcal{H}_\Lambda^{(1)}$ to the unit sphere in $\mathbb{R}^{2N}$, where a real vector contains real and imaginary parts of all components of $|\psi\rangle \in \mathcal{H}_\Lambda^{(1)}$ in some basis. We extend an ϵ-ball $B_\epsilon^{(v)}$ into $A_\epsilon := \{\mathrm{e}^{\mathrm{i}\alpha}|\psi\rangle \colon |\psi\rangle \in B_\epsilon^{(v)}, \alpha \in \mathbb{R}\}$. Heuristically, this is expected to reduce the number of sets required to cover $\mathcal{H}_\Lambda^{(1)}$ to about $(1/\epsilon)^{2N-2}$. The same scaling (with potentially different constants) applies to the number of elements in the set C since the norm in $\mathcal{H}_\Lambda$ and the trace distance in $\mathcal{D}(\mathcal{H}_\Lambda)$ are related by $\min_{\alpha\in\mathbb{R}} \||\psi\rangle - \mathrm{e}^{\mathrm{i}\alpha}|\phi\rangle\|/\sqrt{2} = \sqrt{1 - |\langle\psi|\phi\rangle|} \le \sqrt{1 - |\langle\psi|\phi\rangle|^2} = D(|\psi\rangle\langle\psi|, |\phi\rangle\langle\phi|)$.[5] ■

Proof Choose an orthonormal basis $\{|\phi_i\rangle \colon i \in [N]\}$ of $\mathcal{H}_\Lambda$ and set

$$\mathcal{E}(M) := \sum_{i=1}^{N} |\phi_i\rangle\langle\phi_i|M|\phi_i\rangle\langle\phi_i|, \quad x(M) := [\langle\phi_i|M|\phi_i\rangle]_{i=1}^N \quad (M \in \mathcal{B}(\mathcal{H}_\Lambda)). \tag{3.2}$$

Let $\|x\|_1 := \sum_{i=1}^N |x_i|$ denote the vector-1 norm on $\mathbb{R}^N$. For density matrices $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$, the relation $x(\rho) \in \Delta(N) := \{x \in \mathbb{R}^N \colon x_i \ge 0, \|x\|_1 = 1\}$ holds. For subsets $A \subset \mathcal{D}(\mathcal{H}_\Lambda)$, set $x(A) := \{x(\rho) \colon \rho \in A\}$. Let C satisfy (3.1). Let $A \in C$ and $\rho, \sigma \in A$. Then $D(\rho, \sigma) \le 2\epsilon$ and

$$\|x(\rho) - x(\sigma)\|_1 = \|\mathcal{E}(\rho) - \mathcal{E}(\sigma)\|_{(1)} \le \|\rho - \sigma\|_{(1)} \le 4\epsilon. \tag{3.3}$$

The first inequality holds because the map $\mathcal{E}$ is completely positive and trace-preserving.[6] This shows that $x(A) \subset B_{4\epsilon}^{(N)}(x(\rho))$ holds where

$$B_\epsilon^{(N)}(y) := \{z \in \Delta(N) \colon \|z - y\|_1 \le \epsilon\}, \quad y \in \mathbb{R}^N. \tag{3.4}$$

Furthermore, (3.1) implies that $\Delta(N) = \bigcup_{A\in C} x(A)$ holds. Applying a result by Knill (1995, Lemma 2.5 and Appendix) completes the proof. ■

---

[4] E.g. Nielsen and Chuang 2007, Section 4.5.4.

[5] Lemma A.7 and Nielsen and Chuang 2007, Eq. (9.99).

[6] E.g. Nielsen and Chuang 2007, Theorem 9.2.

3.1.2. Informationally complete measurements

In Section 1.4 above, we learned that outcomes of quantum mechanical measurements on a system in state $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$ provide statistical estimates of linear functionals $\operatorname{Tr}(M_i \rho)$ where $M_i \in \mathcal{B}_H(\mathcal{H}_\Lambda)$ are observables. A general measurement can thus be described with the following $\mathbb{R}$-linear map:

$$\mathcal{M}\colon \quad \mathcal{B}_H(\mathcal{H}_\Lambda) \to \mathbb{R}^\mu, \quad \rho \mapsto \left[\operatorname{Tr}(M_i\rho)\right]_{i=1}^{\mu} \quad (M_i \in \mathcal{B}_H(\mathcal{H}_\Lambda)). \tag{3.5}$$

Suppose that a statistical estimate y_{est} of $y = \mathcal{M}(\rho)$ is available. In order to estimate the unknown density matrix ρ, we can try to find a $\sigma \in \mathcal{D}(\mathcal{H}_\Lambda)$ which satisfies $\mathcal{M}(\sigma) \approx y_{\text{est}}$. Several estimators for the density matrix are discussed in the following Sections; in the remainder of this Section, we shall discuss conditions under which $y = \mathcal{M}(\rho)$ uniquely determines ρ.

The density matrix ρ is uniquely determined by $y = \mathcal{M}(\rho)$ if the map $\mathcal{M}$ is injective when restricted to the set of density matrices $\mathcal{D}(\mathcal{H}_\Lambda)$. In this case, we call $\mathcal{M}$ and the set of observables $\left\{M_1, \dots, M_\mu\right\}$ informationally complete (IC):[7]

Definition 3.2 The set of observables $\left\{M_1, \dots, M_\mu\right\} \subset \mathcal{B}_H(\mathcal{H}_\Lambda)$ and the induced, $\mathbb{R}$-linear map $\mathcal{M}\colon \mathcal{B}_H(\mathcal{H}_\Lambda) \to \mathbb{R}^\mu$, $\mathcal{M}(X) = [\operatorname{Tr}(M_i X)]_{i=1}^{\mu}$ are called informationally complete (IC) if the map $\mathcal{M}|_{\mathcal{D}(\mathcal{H}_\Lambda)}$ is injective. □

Recall that the set of trace-free observables is denoted by $\mathcal{B}_0(\mathcal{H}_\Lambda) = \{X \in \mathcal{B}_H(\mathcal{H}_\Lambda)\colon \operatorname{Tr}(X) = 0\}$. The map $\mathcal{M}$ has the following properties:

Lemma 3.3 *If $\mathcal{M}$ is injective, then $\mathcal{M}|_{\mathcal{B}_0(\mathcal{H}_\Lambda)}$ is injective.*

$\mathcal{M}|_{\mathcal{B}_0(\mathcal{H}_\Lambda)}$ is injective if and only if $\mathcal{M}|_{\mathcal{D}(\mathcal{H}_\Lambda)}$ is injective.

$\mathcal{M}|_{\mathcal{B}_0(\mathcal{H}_\Lambda)}$ and $\mathcal{M}$ are injective if and only if $\operatorname{rk}(\mathcal{M}|_{\mathcal{B}_0(\mathcal{H}_\Lambda)}) = (d_\Lambda)^2 - 1$ or $\operatorname{rk}(\mathcal{M}) = (d_\Lambda)^2$, respectively.

Let the set of observables $\left\{M_1, \dots, M_\mu\right\} \subset \mathcal{B}_H(\mathcal{H}_\Lambda)$ be a POVM or let it be such that there are coefficients $c_i \in \mathbb{R}$ such that $\mathbb{1} = \sum_{i=1}^{\mu} c_i M_i$. Under these conditions, $\mathcal{M}$ is injective if and only if $\mathcal{M}|_{\mathcal{B}_0(\mathcal{H}_\Lambda)}$ is injective where $\mathcal{M}$ is the induced map.

Proof The relations $\mathcal{D}(\mathcal{H}_\Lambda) \subset \mathcal{B}_0(\mathcal{H}_\Lambda) \subset \mathcal{B}_H(\mathcal{H}_\Lambda)$ imply the following implications: $\mathcal{M}$ injective $\Rightarrow \mathcal{M}|_{\mathcal{B}_0(\mathcal{H}_\Lambda)}$ injective $\Rightarrow \mathcal{M}|_{\mathcal{D}(\mathcal{H}_\Lambda)}$ injective.

Let $\mathcal{M}|_{\mathcal{B}_0(\mathcal{H}_\Lambda)}$ be non-injective and let $A \in \mathcal{B}_0(\mathcal{H}_\Lambda)$, $\mathcal{M}(A) = 0$ and $A \neq 0$. Set $\rho_\pm = (\mathbb{1} \pm A/\|A\|_{(\infty)})/d_\Lambda$. Then $\rho_\pm \in \mathcal{D}(\mathcal{H}_\Lambda)$ and $\mathcal{M}(\rho_+) = \mathcal{M}(\rho_-)$, i.e. $\mathcal{M}|_{\mathcal{D}(\mathcal{H}_\Lambda)}$ is not injective.

---

[7] Prugovečki 1977; Busch 1991; Renes et al. 2004.

The sets $\mathcal{B}_0(\mathcal{H}_\Lambda)$ and $\mathcal{B}_H(\mathcal{H}_\Lambda)$ are vector spaces over $\mathbb{R}$ of dimension $(d_\Lambda)^2 - 1$ and $(d_\Lambda)^2$, respectively.

If $\{M_1, \dots, M_\mu\}$ is a POVM, then $\mathbb{1} = \sum_i c_i M_i$ holds for $c_i = 1$. Let $\mathbb{1} = \sum_i c_i M_i$ hold and let $\mathcal{M}$ be non-injective. Let $A \in \mathcal{B}_H(\mathcal{H}_\Lambda)$, $\mathcal{M}(A) = 0$ and $A \neq 0$. Then $\operatorname{Tr}(A) = \operatorname{Tr}(\mathbb{1}A) = \sum_{i=1}^{\mu} c_i[\mathcal{M}(A)]_i = 0$. Accordingly, $A \in \mathcal{B}_0(\mathcal{H}_\Lambda)$, $\mathcal{M}(A) = 0$ and $A \neq 0$ hold and $\mathcal{M}|_{\mathcal{B}_0(\mathcal{H}_\Lambda)}$ is not injective. ■

The lemma states that injectivity of the maps $\mathcal{M}|_{\mathcal{B}_0(\mathcal{H}_\Lambda)}$ and $\mathcal{M}|_{\mathcal{D}(\mathcal{H}_\Lambda)}$ is equivalent. Therefore, the property that density matrices are positive semidefinite does not reduce the number of expectation values required to identify an arbitrary state. For uniform dimensions $d = d_i$, $i \in \Lambda$, at least $(d_\Lambda)^2 - 1 = d^n - 1$ expectation values are required to identify an arbitrary state.

While the fact that density matrices are positive semidefinite does not help with identifying an arbitrary state, it can be very helpful for identifying particular states such as pure states $|\phi\rangle \in \mathcal{H}_\Lambda^{(1)}$. As already suggested in the last subsection, suppose that we obtained an estimate of $\operatorname{Tr}(M\rho)$ where $M = |\phi\rangle\langle\phi|$. It is easy to see that the null space of $\mathcal{B}_0(\mathcal{H}_\Lambda) \to \mathbb{R}$, $\rho \mapsto \operatorname{Tr}(M\rho)$ has dimension $(d_\Lambda)^2 - 2$ but that the only solution $\sigma \in \mathcal{D}(\mathcal{H}_\Lambda)$ of $\operatorname{Tr}(M\sigma) = 1$ is $\sigma = |\phi\rangle\langle\phi|$. It is implied e.g. by[8]

$$D(\tau, |\phi\rangle\langle\phi|) \leq \sqrt{1 - \langle\phi|\tau|\phi\rangle} = \sqrt{1 - \operatorname{Tr}(M\tau)} \tag{3.6}$$

where $\tau \in \mathcal{D}(\mathcal{H}_\Lambda)$ and $D(M, N) = \frac{1}{2}\|M - N\|_{(1)}$ is the trace distance ($M, N \in \mathcal{B}(\mathcal{H}_\Lambda)$). An observed value of $\operatorname{Tr}(M\rho) \leq 1 - \epsilon$ thus implies that the trace distance of the unknown state ρ and $|\phi\rangle\langle\phi|$ is at most $\sqrt{\epsilon}$. Given orthonormal state vectors $|\phi_1^{(i)}\rangle, |\phi_2^{(i)}\rangle \in \mathcal{H}_i^{(1)}$ ($i \in \Lambda$), one can similarly show that the expectation values of the $2n$ observables $M_{ij} = \mathbb{1}_{i-1} \otimes |\phi_j^{(i)}\rangle\langle\phi_j^{(i)}| \otimes \mathbb{1}_{n-i}$ ($i \in \Lambda$, $j \in \{1, 2\}$) suffice to determine the system's state if it is in any of the 2^n orthogonal states $|\phi_{j_1}^{(1)}\rangle \otimes \cdots \otimes |\phi_{j_n}^{(n)}\rangle$, $j_i \in \{1, 2\}$, $i \in \Lambda$. Furthermore, it has been shown that a density matrix with small rank (e.g. an arbitrary pure state) can be distinguished from all other states with a POVM with fewer than $(d_\Lambda)^2$ elements and that even fewer elements are necessary if the state is to be distinguished only from states with at most the same rank.[9] These examples show that exploiting the positive-semidefiniteness of density matrices can greatly simplify determining the state of a quantum system.

---

[8] E.g. Nielsen and Chuang 2007, Eqs. (9.60), (9.110).

[9] Flammia et al. 2005; Finkelstein 2004; Heinosaari et al. 2013; Baldwin et al. 2014; Baldwin et al. 2016.

3.1.3. Pauli matrices

In the following, we discuss useful techniques for constructing POVMs and we define a number of POVMs constructed from eigenvectors of Pauli matrices, which constitute frequently used observables of quantum mechanical two-level systems.

We refer to a POVM on $\mathcal{H}$ by the set $\Pi \subset \mathcal{B}_H(\mathcal{H})$ containing the POVM's elements. For simplicity, index sets $\mathcal{I}_\Pi$, which enumerate elements via $\Pi = \{P_i \colon i \in \mathcal{I}_\Pi\}$, are not specified explicitly. Given POVMs Π_1 on $\mathcal{H}_1$ and Π_2 on $\mathcal{H}_2$, a POVM on $\mathcal{H}_1 \otimes \mathcal{H}_2$ is given by

$$\Pi_1 \otimes \Pi_2 := \big\{A \otimes B \colon A \in \Pi_1, B \in \Pi_2\big\}. \tag{3.7}$$

The POVM is IC if and only if both Π_1 and Π_2 are IC. A measurement of $\Pi_1 \otimes \Pi_2$ can be implemented by measuring Π_1 on system 1 and Π_2 on system 2 and the two measurements can be performed sequentially or simultaneously.

For quantum state estimation with maximum likelihood estimation (MLE), it is convenient to represent the measurement data of several POVMs as data of a single POVM. This is achieved by the uniformly weighted union[10]

$$\mathcal{U}_P(\{\Pi_1, \dots, \Pi_N\}) := \left\{\frac{1}{N} X \colon X \in \Pi_1 \cup \dots \cup \Pi_N\right\}. \tag{3.8}$$

The union enables us to consider an arbitrary number of distinct measurements with a single POVM because dividing Π_i's outcome probabilities by N provides corresponding outcome probabilities of $\mathcal{U}_P(\{\Pi_1, \dots, \Pi_N\})$. Choosing $i \in \{1, \dots, N\}$ uniformly at random and measuring Π_i implements a measurement of $\mathcal{U}_P(\{\Pi_1, \dots, \Pi_N\})$. If $\mathcal{U}_P(\{\Pi_1, \dots, \Pi_N\})$ is measured mN times, then m measurements of M_i ($i \in \{1, \dots, N\}$) are performed on average. This difference between mN measurements of $\mathcal{U}_P(\{\Pi_1, \dots, \Pi_N\})$ and m measurements of each Π_i ($i \in \{1, \dots, N\}$) vanishes for large M.

The Pauli matrices

$$\sigma_X := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_Y := \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \qquad \sigma_Z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{3.9}$$

are frequently-used observables for quantum-mechanical two-level systems (qubits). The matrices

$$P_0^{(\sigma)} := \frac{\mathbb{1}_2}{\sqrt{2}}, \qquad P_1^{(\sigma)} := \frac{\sigma_X}{\sqrt{2}}, \qquad P_2^{(\sigma)} := \frac{\sigma_Y}{\sqrt{2}}, \qquad P_3^{(\sigma)} := \frac{\sigma_Z}{\sqrt{2}} \tag{3.10}$$

---

[10] The notation accounts for the fact that taking unions in this way is not associative.

constitute a basis of $\mathbb{C}^{2\times 2}$ which is orthonormal in the Hilbert–Schmidt inner product. In multi-qubit systems, we use the notation

$$\sigma_a^{(k)} := \mathbb{1}_{k-1} \otimes \sigma_a \otimes \mathbb{1}_{n-k}, \quad a \in \{X, Y, Z\} \tag{3.11}$$

where σ_a acts on site $k \in \{1, \dots, n\}$ and where $\mathbb{1}_j$ denotes the identity operator on the first $k-1$ or last $n-k$ sites, respectively. We denote normalized eigenvectors and eigenprojectors by

$$\sigma_a |\Psi_{\sigma_a,s}\rangle = s|\Psi_{\sigma_a,s}\rangle, \qquad a \in \{X, Y, Z\}, \qquad s \in \{+1, -1\}, \tag{3.12a}$$

$$P_{\sigma_a,s} := |\Psi_{\sigma_a,s}\rangle\langle\Psi_{\sigma_a,s}|. \tag{3.12b}$$

The notation $P_{a,s} := P_{\sigma_a,s}$ is also used. The POVM which describes a projective measurement of σ_a is

$$\Pi^{(1)}_{\sigma_a} = \big\{P_{\sigma_a,s} \colon s \in \{+1, -1\}\big\}. \tag{3.13}$$

We define the tensor product observables

$$\sigma_a := \sigma_{a_1} \otimes \cdots \otimes \sigma_{a_r}, \quad a \in \{X, Y, Z\}^r \tag{3.14}$$

as well as the POVMs

$$\Pi^{(r)}_{\sigma_a} := \Pi^{(1)}_{\sigma_{a_1}} \otimes \cdots \otimes \Pi^{(1)}_{\sigma_{a_1}} = \left\{ P_{\sigma_{a_1},s_1} \otimes \cdots \otimes P_{\sigma_{a_r},s_r} \colon s \in \{+1, -1\}^r \right\}. \tag{3.15}$$

The POVM $\Pi^{(r)}_{\sigma_a}$ describes a measurement of σ_{a_k} on site k, where the r measurements can be carried out in any order. The POVMs

$$\Pi^{(1)}_{\sigma} := \mathcal{U}_P\left(\left\{\Pi^{(1)}_{\sigma_X}, \Pi^{(1)}_{\sigma_Z}, \Pi^{(1)}_{\sigma_Z}\right\}\right) = \left\{ \tfrac{1}{3} P_{\sigma_a,s} \colon \ a \in \{X, Y, Z\}, \ s \in \{+1, -1\} \right\} \tag{3.16}$$

and

$$\Pi^{(r)}_{\sigma} := \left(\Pi^{(1)}_{\sigma}\right)^{\otimes r} = \mathcal{U}_P\left(\left\{\Pi^{(r)}_{\sigma_a} \colon a \in \{X, Y, Z\}^r\right\}\right) \tag{3.17}$$

$$= \left\{ \tfrac{1}{3^r} P_{\sigma_{a_1},s_1} \otimes \cdots \otimes P_{\sigma_{a_r},s_r} \colon \quad a \in \{X, Y, Z\}^r, \ s \in \{+1, -1\}^r \right\} \tag{3.18}$$

are IC on one and r qubits, respectively. In a system with n sites, the last POVM applied to sites $k, \dots, k+r-1$ is denoted by

$$\Pi^{(n,r,k)}_{\text{Block}} := \left\{ \mathbb{1}_{k-1} \otimes A \otimes \mathbb{1}_{n-r-k+1} \colon \quad A \ \in \ \Pi^{(n,r,k)}_{\text{Local}} := \Pi^{(r)}_{\sigma} \right\}. \tag{3.19}$$

where $k \in \{1, \dots, n-r+1\}$. The POVM $\Pi^{(n,r,k)}_{\text{Local}}$ contains the same elements as $\Pi^{(r)}_{\sigma}$ but it is taken to act on sites $k, \dots, k+r-1$. The POVM

$$\Pi^{(n,r)}_{\text{Block}} := \mathcal{U}_P\left(\left\{ \Pi^{(n,r,k)}_{\text{Block}} \colon \quad k \in \{1, \dots, n-r-k+1\} \right\}\right) \tag{3.20}$$

provides complete information on all contiguous blocks of r neighbouring sites in a linear chain of n sites. This POVM will be useful in conjunction with the estimation of matrix product states.

3.2. Standard estimators

Three standard estimators for quantum state tomography are described in this section: Linear inversion, Bayesian mean estimation and maximum likelihood estimation (MLE). Their complexity increases exponentially with the number of subsystems unless they are suitably modified (see Section 3.3).

3.2.1. Linear inversion

A very basic approach to quantum state estimation consists in finding a solution $\chi \in \mathcal{B}_H(\mathcal{H}_\Lambda)$ of the linear system $\mathcal{M}(\chi) = y_{\text{est}}$, where y_{est} contains statistical estimates obtained from measurement data. The linear inversion estimator is based on this approach[11] and we define it as follows:[12]

Definition 3.4 Let $\mathcal{M} \colon \mathcal{B}_H(\mathcal{H}_\Lambda) \to \mathbb{R}^\mu$ be a linear map and let $y \in \mathbb{R}^\mu$ be given. Set $\rho_{\text{LI}} := \mathcal{M}^+(y)$ where $\mathcal{M}^+$ denotes the Moore–Penrose pseudoinverse.

ρ_{LI} minimizes $\|\mathcal{M}(\chi) - y\|$, $\chi \in \mathcal{B}_H(\mathcal{H}_\Lambda)$.[13] If $\text{rk}(\mathcal{M}) = (d_\Lambda)^2$ and $y = \mathcal{M}(\rho)$, then $\rho_{\text{LI}} = \rho$. □

If we set e.g. $\mathcal{M}(\chi) := [\text{Tr}(P^{(\sigma)}_i \chi)]^{\mu}_{i=1}$ with the normalized Pauli matrices $P^{(\sigma)}_i$ from (3.10), then the linear inversion estimator is particularly simple:

$$\rho_{\text{LI}} = \mathcal{M}^+(y) = \sum_{i=0}^{3} y_i P^{(\sigma)}_i \tag{3.21}$$

As is illustrated by the last equation, $\rho_{\text{LI}} = \mathcal{M}^+(y)$ provides an estimator if exact expectation values $y_i = [\mathcal{M}(\rho)]_i$ are replaced by their estimates. This estimator suffers from the fact that the estimate $\rho_{\mathrm{LI}} \in \mathcal{B}_H(A)$ may not be positive semidefinite, i.e. χ may not be a valid density matrix. If the original density matrix ρ which gave rise to the expectation values $y = \mathcal{M}(\rho)$ has zero as an eigenvalue, ρ_{LI} is likely to have small negative eigenvalues for estimated expectation values of any finite precision. Many practical applications of tomography can be adapted to work with an estimate which is not positive semidefinite.[14] Alternatively, one can use an estimator which returns a density matrix. Such an estimator necessarily has non-zero bias,[15] but, more importantly, it can bring the advantage of a smaller mean-squared error.[16] In the following subsections, we discuss two estimators which restrict estimates to the set of density matrices.

---

[11] E.g. Schwemmer et al. 2015, and references given therein.

[12] For the relevant properties of the pseudoinverse, see e.g. Horn and Johnson 1991a, Section 7.3, P8.

[13] On $\mathbb{R}^\mu$, the standard inner product and norm are assumed.

3.2.2. Bayesian mean estimation

The Bayesian mean estimator for quantum states can be constructed as follows. Denote the probability that one or more measurements on a state $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$ produce a particular result z by

$$p(z|\rho). \tag{3.22}$$

The probability distribution $p(z|\rho)$ can be constructed as function of ρ with knowledge of the performed measurements using the rules from Section 1.4 (an example is given in (3.27) below). Note that $p(z|\rho)$ is a probability distribution over measurement results z and that ρ is only a parameter. To construct the Bayesian mean estimate, we need to choose a measure for the set $\mathcal{D}(\mathcal{H}_\Lambda)$ of density matrices and a so-called *prior distribution* $p(\rho)$. The prior distribution $p(\rho)$ represents prior knowledge about the unknown quantum state. Together with the conditional probability $p(z|\rho)$, the prior distribution gives rise to the joint probability distribution

$$p(z, \rho) = p(z|\rho)p(\rho) = p(\rho|z)p(z) \tag{3.23}$$

of measurement results z and quantum states ρ (the last equation is known as *Bayes' theorem*). The marginal distribution $p(z)$ can be obtained from

$$p(z) = \int_{\mathcal{D}(\mathcal{H}_\Lambda)} p(z, \rho)\, \mathrm{d}\rho = \int_{\mathcal{D}(\mathcal{H}_\Lambda)} p(z|\rho)p(\rho)\, \mathrm{d}\rho. \tag{3.24}$$

---

[14] E.g. Schwemmer et al. 2015.

[15] Schwemmer et al. 2015.

[16] Hradil 1997; Fiurášek 2001; Shang et al. 2014.

The Bayesian mean estimate (BME) ρ_{BME} of ρ is then given by

$$\rho_{\mathrm{BME}} = \int_{\mathcal{D}(\mathcal{H}_\Lambda)} \rho p(\rho|z)\,\mathrm{d}\rho = \int_{\mathcal{D}(\mathcal{H}_\Lambda)} \rho \frac{p(z|\rho)p(\rho)}{p(z)}\,\mathrm{d}\rho \tag{3.25}$$

which is evaluated for the observed measurement result z.[17] Computing the Bayesian mean estimate can pose a challenge because it involves computing integrals over the set $\mathcal{D}(\mathcal{H}_\Lambda)$ of density matrices.

3.2.3. Maximum likelihood estimation

As before, let $p(z|\rho)$ denote the probability that measurements on ρ produce an outcome z. The maximum likelihood estimation (MLE) approach suggests the following estimate:[18]

$$\rho_{\mathrm{MLE}} = \arg\max_\rho \mathcal{L}(\rho) \qquad\qquad \mathcal{L}(\rho) = p(z|\rho) \tag{3.26}$$

where, again, the observed outcome is inserted for z. The likelihood function $\mathcal{L}$ assigns each state its so-called *likelihood* and the estimate ρ_{MLE} is a state with maximal likelihood. It should be noted that the likelihood function is *not* a probability distribution because no measure has been chosen for $\mathcal{D}(\mathcal{H}_\Lambda)$.

In the following, we specify the likelihood function for a POVM $\Pi = \{M_1, \ldots, M_\mu\} \subset \mathcal{B}_H(\mathcal{H}_\Lambda)$ and m measurements of the POVM. Measuring the POVM m times provides the sequence of outcomes $(x_1, \ldots, x_m)$, $x_k \in \{1, \ldots, \mu\}$, $k \in \{1, \ldots, m\}$. Denote by $z_i := |\{k \colon x_k = i\}|$ the number of occurrences of outcome i, i.e. $\sum_{i=1}^{\mu} z_i = m$. Denote by $f_i = z_i/m$ the relative frequencies of the different outcomes. The symbols $z = (z_1, \ldots, z_\mu)$ and $f = (f_1, \ldots, f_\mu)$ denote the vectors of counts and relative frequencies. Given a density matrix $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$, the probability for an outcome z of the m measurements is given[19] by the multinomial distribution:

$$p(z|\rho) = \binom{m}{z} \prod_{i=1}^{\mu} {p_i}^{z_i} \qquad p_i = \mathrm{Tr}(M_i\rho) \qquad \binom{m}{z} = \frac{m!}{z_1! \ldots z_\mu!} \tag{3.27}$$

We define the log-likelihood function L, which is maximized by the same states as the likelihood function $\mathcal{L}$, by

$$L(\rho) = \frac{1}{m} \log\left(\binom{m}{z}^{-1} \mathcal{L}(\rho)\right) = \sum_{i=1}^{\mu} f_i \log(\mathrm{Tr}(M_i\rho)). \tag{3.28}$$

---

[17] E.g. Granade et al. 2016 and references given therein.

[18] Hradil 1997; Hradil et al. 2004.

[19] Hradil et al. 2004, Eq. (3.6).

The log-likelihood function has the following property:[20]

$$L(\rho) = -S(f\|\mathcal{M}(\rho)) - H(f) \qquad\qquad \mathcal{M}(\rho) = [\mathrm{Tr}(M_i\rho)]_{i=1}^{\mu} \tag{3.29}$$

Here, $S(q\|p) = \sum_{i=1}^{\mu} q_i[\log(q_i) - \log(p_i)]$ is the relative entropy (Kullback–Leibler divergence) and $H(q) = -\sum_{i=1}^{\mu} q_i \log(q_i)$ is the Shannon entropy. Accordingly, a state ρ with maximal (log-) likelihood is also a state with minimal relative entropy distance between the observed frequencies f_i and the probabilities p_i predicted by ρ.

Finding a state which maximizes the likelihood can be non-trivial. It is a convex optimization problem because the log-likelihood L is concave and the set of all density matrices is convex.[20] This simplifies the solution of the optimization because it implies that any local maximum of the log-likelihood L is also a global maximum.[21] A state which maximizes the likelihood has the property[20]

$$\rho = R(\rho)\rho = \rho R(\rho) \qquad\qquad R(\rho) = \sum_{i=1}^{\mu} \frac{f_i}{\mathrm{Tr}(M_i\rho)} M_i. \tag{3.30}$$

As a consequence, one may try to find the maximum with the fixed point iteration[22]

$$\rho_{k+1} = R(\rho_k)\rho R(\rho_k), \quad \rho_1 = \mathbb{1}/\mathrm{Tr}(\mathbb{1}). \tag{3.31}$$

This iteration has been shown to not decrease the likelihood.[20] Furthermore, it has been shown that the "diluted" prescription

$$\rho_{k+1} = R_\epsilon(\rho_k)\rho R_\epsilon(\rho_k) \qquad\qquad R_\epsilon(\rho) = \frac{\mathbb{1} + \epsilon R(\rho)}{1+\epsilon} \tag{3.32}$$

provides a strictly increasing likelihood if $\epsilon \in (0, 1)$ is sufficiently small.[23]

It turns out that the algorithm from Equation (3.31) can be implemented efficiently for many-body quantum systems using MPS representations.[24] Equation (3.32) is essentially a gradient descent of the likelihood function[25] and methods for accelerated gradient descent have been applied to find the likelihood's maximum with less computational time.[26] Some of these methods can be applied to the MPS-based maximum likelihood estimation algorithm as well.[27]

---

[20] Hradil et al. 2004.

[21] E.g. Boyd and Vandenberghe 2009, Sec. 4.2.2.

[22] Lvovsky 2004; Molina-Terriza et al. 2004; Hradil et al. 2004.

[23] Řeháček et al. 2007; D. S. Gonçalves et al. 2014.

[24] Baumgratz et al. 2013b.

[25] Siah Teo 2013; Shang et al. 2017.

[26] E.g. Siah Teo 2013; D. Gonçalves et al. 2016; Shang et al. 2017; Bolduc et al. 2017; Bai et al. 2017; Knee et al. 2018.

[27] Safieldeen 2016.

3.3. Efficient estimation

This section presents efficient or scalable methods for estimating quantum states. The MPS-MLE method[28] is presented and adapted to so-called locally purified matrix product states which enforce positive semidefiniteness of the density matrix. The MPS-SVT method is presented in a way[29] which enables future extensions of the original proposal.[30] MPO reconstruction is mentioned only briefly and covered in greater detail in Chapter 8.

3.3.1. Maximum likelihood estimation with MPSs

The computational resources required for maximum likelihood estimation of quantum states can often be reduced by representing quantum states with MPSs and MPOs as proposed by Baumgratz et al. (2013b). This section explains relevant details and additionally proposes to perform MLE of a mixed state by representing it as locally purified matrix product state (PMPS).

As mentioned in Section 3.2.3, the maximum likelihood estimate is a state which maximizes the log-likelihood function $L(\rho)$. One attempts to find such a state with the iteration (3.31). In the following, we discuss conditions under which the iteration may be evaluated efficiently in the sense that the necessary computation time increases only polynomially with n.

The operator $R(\rho)$ from (3.30) is constructed from the POVM elements $M_1, \dots, M_\mu$ describing the measurement and the observed relative outcome frequencies f_i. An MPO representation of the operator $R(\rho)$ can be computed in $O(\mathrm{poly}(n))$ time if the MPO bond dimension of ρ is $O(\mathrm{poly}(n))$, if the POVM is restricted to $\mu = O(\mathrm{poly}(n))$ elements and if each element admits an MPO representation of bond dimension $O(\mathrm{poly}(n))$. The restriction of the POVM to $O(\mathrm{poly}(n))$ elements implies that it is not informationally complete; therefore, the estimation methods needs to be complemented by a verification method as described at the beginning of the chapter if assumption-free QST is the goal.

If an MPO representation of ρ_k is available, an MPO representation of ρ_{k+1} can be obtained and its bond dimension is at most $D(\rho_{k+1}) = D(\rho_k)[D(R(\rho_k))]^2$. Since the bond dimension of $R(\rho_k)$ usually exceeds one, the bond dimension of ρ_k increases exponentially with k. Typically, a few hundreds or thousands of iterations are required and this necessitates approximating ρ_{k+1} by an MPO of smaller bond dimension:

---

[28] Baumgratz et al. 2013b.

[29] Cai et al. 2010.

[30] Cramer et al. 2010.

$$\rho_{k+1} = C[R(\rho_k)\rho_k R(\rho_k)]. \tag{3.33}$$

Here, C denotes compression to some given bond dimension. If the unknown state is known to be close to a pure state, one can attempt to find it with the following iteration over pure states represented as MPSs:

$$|\psi_{k+1}\rangle = C[R(|\psi_k\rangle)|\psi_k\rangle], \tag{3.34}$$

where $R(|\psi_k\rangle) := R(|\psi_k\rangle\langle\psi_k|)$. Note that compression to a smaller bond dimension can reduce the likelihood, potentially keeping the algorithm from reaching a maximum of the likelihood function. Furthermore, the set of matrix product states or matrix product operators whose bond dimension does not exceed some constant is not a convex set. In evaluating (3.33) or (3.34) we thus attempt to solve a non-convex optimization problem. As a consequence, the iterations can converge to a local maximum of the likelihood function which is not a global maximum. Nevertheless, the scheme provided satisfactory results in several applications.[31] This completes the description of scalable MLE as proposed by Baumgratz et al. (2013b).

Compressing an MPO to a smaller bond dimension also introduces another issue: It can destroy the hermiticity and positive semidefiniteness of the represented operator. Hermiticity can be enforced by representing the operator in a Hermitian operator basis with real coefficients or restored by computing $(\rho + \rho^*)/2$. Testing whether an MPO is positive semidefinite, however, can take time exponential in n.[32] Preserving positive semidefiniteness by increasing the MPO's bond dimension until a sufficiently small compression error is achieved may not be feasible if the iteration without compression traverses a region of large bond dimension before reaching the final estimate of low bond dimension. To remedy this issue, we propose to represent a mixed state as locally purified matrix product state (PMPS).[33]

MLE and purifications. To clarify the idea, we first explain how MLE without MPS can be performed with purified states. A purification of a mixed state $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$ is any pure state $|\psi\rangle \in \mathcal{H}_\Lambda \otimes \mathcal{H}_{\Lambda'}$ such that $\rho = \mathrm{Tr}_{\Lambda'}(|\psi\rangle\langle\psi|)$ and a purification exists if the dimension of Λ' is at least $\mathrm{rk}(\rho)$.[34] Consider the iteration

---

[31] Baumgratz et al. 2013b; Holzäpfel et al. 2015; Lanyon, Maier, et al. 2017. See also Chapters 4, 5 and 7.

[32] Kliesch et al. 2014a. See also Section 2.2.1.

[33] This idea came up in conversation with Daniel Suess and it has been investigated together with two Bachelor students before the submission of this thesis (Baumann 2016; Safieldeen 2016).

$$|\psi_{k+1}\rangle = R(|\psi_k\rangle)|\psi_k\rangle, \tag{3.35}$$

where all POVM elements act as the identity on the ancilla system. As a consequence, $R(|\psi_k\rangle)$ can be constructed from $\rho_k = \mathrm{Tr}_{\Lambda'}(|\psi_k\rangle\langle\psi_k|)$, $R(|\psi_k\rangle) = R(\rho_k)$, and $R(|\psi_k\rangle)$ also acts as the identity on the ancillary sites. We find

$$\rho_{k+1} = \mathrm{Tr}_{\Lambda'}\left[|\psi_{k+1}\rangle\langle\psi_{k+1}|\right] = \mathrm{Tr}_{\Lambda'}\left[R(|\psi_k\rangle)\,|\psi_k\rangle\langle\psi_k|\,R(|\psi_k\rangle)\right] \tag{3.36a}$$

$$= R(\rho_k)\;\mathrm{Tr}_{\Lambda'}\big[|\psi_k\rangle\langle\psi_k|\big]\;R(\rho_k) = R(\rho_k)\,\rho_k\,R(\rho_k)\;, \tag{3.36b}$$

which is nothing but the original iteration (3.31). In (3.31), positive semidefiniteness of ρ_k can be destroyed by numerical rounding errors,[35] which can become an issue as errors accumulate over many iterations. The iteration (3.35) preserves positive semidefiniteness of ρ_k by definition and it has been recognized early on that this is advantageous[36] e.g. to estimate the uncertainty of the estimated state.[37]

MLE and locally purified matrix product states (PMPSs). In the PMPS representation, a mixed state ρ_k of n subsystems is represented by a purification $|\psi_k\rangle$ whose ancilla system comprises n subsystems (see Section 2.2.1). Therefore, we propose the iteration

$$|\psi_{k+1}\rangle = C[R(|\psi_k\rangle)|\psi_k\rangle] \tag{3.37}$$

where POVM elements and the operators $R(|\psi_k\rangle)$ act as the identity on all ancilla systems. The last equation appears to be identical to (3.34). However, (3.34) is used to estimate a pure state of n subsystems while (3.37) is used to estimate a mixed state of n subsystems represented as a pure state on $2n$ subsystems. The symbol C in (3.37) refers to compression by approximating the state with a PMPS representation of lower bond dimension. If compression is dropped, the iterations (3.33) and (3.37) are formally equivalent (cf. (3.36b)). However, the different kinds of compression can make a significant difference: In the two cases, approximated operators are chosen from different admissible sets according to a different objective functions. In particular, compression cannot destroy the positive semidefiniteness of ρ_k if it represented as PMPS. We present simulation results of MLE with PMPS in Section 4.2. A comparison of the relative merits of MLE with MPOs and PMPSs is left for future work.

---

[34] See Section 2.2.1.

[35] This applies if ρ_k does not have full rank.

[36] Banaszek et al. 1999.

[37] Lvovsky 2004, and references therein.

3.3.2. Singular value thresholding with MPSs

Reconstructing a quantum state $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$ from an informationally incomplete vector of expectation values $y = \mathcal{M}(\rho)$ corresponds to a reconstruction from a small number of linear functionals of ρ. The problem of reconstructing an arbitrary matrix of low rank from few of its entries or other linear functionals is known as the *matrix completion* problem.[38] In this section, we present the basic properties of the singular value thresholding (SVT) algorithm for matrix completion and the application of SVT to quantum tomography of pure states represented as MPS.[39] The main purpose of this section is to show how the original MPS-SVT proposal can be applied to non-qubit systems and to measurements described by arbitrary POVMs.

Matrix completion with singular value thresholding (SVT). In the following, we summarize the results on SVT obtained by Candès and Recht (2008) and Cai et al. (2010). Following their work, we consider a real matrix $X \in \mathbb{R}^{k_1 \times k_2}$, a map $\mathcal{A} \in \mathcal{B}(\mathbb{R}^{k_1 \times k_2}; \mathbb{R}^\mu)$, a vector $b \in \mathbb{R}^\mu$ and the condition $\mathcal{A}(X) = b$. Later on, we will substitute this condition with $\mathcal{M}(\rho) = y$. The first ansatz for reconstructing a matrix from few linear functionals under the assumption that it has low rank is given by the minimization problem

$$\min_X \operatorname{rk}(X) \quad \text{subject to} \quad \mathcal{A}(X) = b. \tag{3.38}$$

Unfortunately, this minimization is NP-hard in the matrix dimension.[40] If μ is sufficiently large and if X and $\mathcal{A}$ are sufficiently random in a certain sense, then, with high probability, (3.38) has the same solution[41] as

$$\min_X \|X\|_{(1)} \quad \text{subject to} \quad \mathcal{A}(X) = b. \tag{3.39}$$

In solving (3.39), methods for convex optimization can be applied because the trace norm $\|\cdot\|_{(1)}$ is convex[42] and the constraint $\mathcal{A}(X) = b$ can be replaced by a more general constraint.[43] For (3.38) and (3.39) having the same solutions, at least $\mu = O(k^{6/5} r \log k)$ entries were required initially where $k = \max\{k_1, k_2\}$ and $r = \mathrm{rk}(X)$.[44] Later, this was improved to $\mu = O(kr \ln^2(k))$ for certain matrix bases.[45] For $k = 2^n$, both numbers are exponential in n. Instead of solving (3.39), we consider[46]

---

[38] Candès and Recht 2008; Candès and Recht 2009; Cai et al. 2010.

[39] Cramer et al. 2010.

[40] Candès and Recht 2008, Sec. II.C.

[41] Candès and Recht 2008.

[42] Boyd and Vandenberghe 2009.

$$\min_X \tau \|X\|_{(1)} + \frac{1}{2}\|X\|_{(2)}{}^2 \quad \text{subject to} \quad \mathcal{A}(X) = b \tag{3.40}$$

where $\tau > 0$ is a constant. For $\tau \to \infty$, the solution of the last equation converges to the solution of (3.39) which minimizes $\|X\|_{(2)}$.[47] In the following, we use a singular value soft-thresholding operator D_τ, which is defined in terms of an SVD $X = USV^*$:

$$D_\tau(X) = URV^*, \quad R_{ij} = \delta_{ij} \max\{0, S_{ii} - \tau\}. \tag{3.41}$$

In order to solve (3.40), we define $X_k \in \mathbb{R}^{k_1 \times k_2}$ and $Y_k \in \mathbb{R}^\mu$ via[48]

$$X_k = D_\tau(\mathcal{A}^*(Y_{k-1})), \tag{3.42a}$$

$$Y_k = Y_{k-1} + \delta_k(b - \mathcal{A}(X_k)), \quad Y_0 = 0. \tag{3.42b}$$

If suitable step sizes $\delta_k > 0$ are used, the sequence $(X_k)_k$ converges to the solution of (3.40).[49] $\mathcal{A}^*$ denotes the adjoint map of $\mathcal{A}$ (cf. (3.43) below).

SVT with matrix product states. We apply SVT to quantum state tomography by replacing $\mathcal{A}(X) = b$ with $\mathcal{M}(\rho) = y$ where $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$ is a density matrix. The map $\mathcal{M}$ is defined in terms of observables $M_i \in \mathcal{B}_H(\mathcal{H}_\Lambda)$ as $\mathcal{M}(\rho) = [\mathrm{Tr}(M_i \rho)]_{i=1}^\mu$ and its adjoint is given by[50]

$$\mathcal{M}^*(y) = \sum_{i=1}^{\mu} y_i M_i. \tag{3.43}$$

An MPO representation of $\mathcal{M}^*(y)$ with bond dimension $O(\mathrm{poly}(n))$ can be obtained if $\mu = O(\mathrm{poly}(n))$ and if each observable M_i admits an MPO representation of bond dimension $O(\mathrm{poly}(n))$. To simplify the computation of (3.42), one can replace the thresholding operator D_τ in (3.42) with an operator which preserves only the largest eigenvalue. An MPS representation of the corresponding eigenvector can often be obtained with DMRG and related algorithms.[51] This is the method for reconstructing a pure state (i.e. a rank-1 density matrix) which was proposed by Cramer et al. (2010). It could be extended by computing several eigenvalues or by computing several extreme singular values with recently proposed methods.[52]

---

[43] Cai et al. 2010.

[44] Candès and Recht 2008.

[45] D. Gross 2011; Liu 2011.

[46] Cai et al. 2010, Eq. (3.1).

[47] Cai et al. 2010, Thm. 3.1.

[48] Cai et al. 2010, Eq. (3.3).

[49] Cai et al. 2010, Cor. 4.5.

[50] Let $\mathcal{M}(\rho) = [\mathrm{Tr}(M_i \rho)]_{i=1}^\mu$. $\langle \mathcal{M}^*(y), \rho \rangle = \langle y, \mathcal{M}(\rho) \rangle = \sum_{i=1}^\mu \overline{y_i} \, \mathrm{Tr}(M_i \rho) = \mathrm{Tr}[\sum_{i=1}^\mu (y_i M_i^*)^* \rho] = \langle \sum_{i=1}^\mu y_i M_i^*, \rho \rangle$. This also holds for non-Hermitian M_i, i.e. for $\mathcal{M} \in \mathcal{B}(\mathcal{B}(\mathcal{H}); \mathbb{C}^\mu)$.

To recover the variant of (3.42) which was presented in the original MPS-SVT proposal,[53] we set $\tilde{X}_k = \mathcal{M}^*(\mathcal{M}(X_k))$ and $\tilde{Y}_k = \mathcal{M}^*(Y_k)$. With these definitions, (3.42) becomes

$$\tilde{X}_k = \mathcal{M}^*(\mathcal{M}(D_\tau(\tilde{Y}_k))), \tag{3.44a}$$

$$\tilde{Y}_k = \tilde{Y}_{k-1} + \delta_k(\mathcal{M}^*(y) - \tilde{X}_k). \tag{3.44b}$$

3.3.3. Reconstruction of matrix product operators

It was shows that a density matrix $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$ can be reconstructed from marginal density matrices on few neighbouring sites if the correlations in the state satisfy conditions which correspond to the following operator Schmidt rank conditions:[54]

$$\mathrm{OSR}(1 \dots k : k+1 \dots n)_\rho = \mathrm{OSR}(k-l+1 \dots k : k+1 \dots k+r)_\rho \tag{3.45}$$

where $l, r \geq 1$, $l + r \leq n - 2$ are fixed and the condition must hold for all $k \in \{l+1 \dots n-r-1\}$. The conditions divide a linear chain of n subsystems into a left half comprising subsystems $\{1, \dots, k\}$ and a right half comprising subsystems $\{k+1, \dots, n\}$. The conditions demand that correlations between the two halves of the chain do not decrease if all but the l sites to the left of the cut and the r sites to the right are traced out. If these conditions are satisfied, an MPO representation of ρ can be constructed from the marginal states $\rho_{k-l \dots k+r}$ where $k \in \{l+1 \dots n-r\}$. This result was obtained by Baumgratz et al. (2013a) and it is discussed and generalized in Chapter 8 (see, in particular, Example 8.20).

---

[51] E.g. Schollwöck 2011. See also Section 2.2.2.

[52] E.g. Lee and Cichocki 2015.

[53] Cramer et al. 2010, Eq. (6).

[54] Baumgratz et al. 2013a.

3.4. Efficient verification

We discuss two methods for efficient *verification* or *certification* of quantum states. The first method applies to pure states and uses a so-called parent Hamiltonian and is related to theory of MPS. The second method uses a decomposition of the von Neumann entropy and it can succeed if the state is pure or if its entropy is known. The two methods share the property that measurements can be limited to a predetermined set of neighbouring sites, allowing for a measurement of the relevant observables independently of an estimated state. Measurements which depend on an estimated state are used by the verification method *direct fidelity estimation (DFE)*,[55] which is not discussed here.

3.4.1. Parent Hamiltonian certificate

It has been shown that a large fraction of all pure states $|\psi\rangle$ with small MPS bond dimension are determined uniquely by their reduced density matrices on a small number of neighbouring sites.[56] Furthermore, the distance between such a pure state and an arbitrary state σ can be bounded if only the corresponding reduced density matrices of σ are known.[57] In the following, we discuss the technical details of these results. The condition for an MPS being uniquely determined by reduced density matrices is usually stated as injectivity of certain maps Γ which are related to the details of the MPS representation. We introduce an equivalent condition which only involves certain Schmidt ranks.

The results are derived using a so-called *parent Hamiltonian*. An observable G is called a parent Hamiltonian of a pure state $|\psi\rangle$ if $|\psi\rangle$ is a ground state of G (i.e. an eigenvector of G's smallest eigenvalue). A parent Hamiltonian with non-degenerate ground state has the following useful property:[58]

Lemma 3.5 *Let G be an observable with the two smallest eigenvalues E_0 and $E_1 > E_0$. Let $|\psi\rangle$ be a normalized eigenvector of the smallest eigenvalue E_0 and let E_0 be non-degenerate. Let σ be a density matrix. Then,*

$$1 - \langle\psi|\sigma|\psi\rangle \leq \frac{E_\sigma - E_0}{E_1 - E_0} \tag{3.46}$$

where $E_\sigma = \mathrm{Tr}(G\sigma)$.

---

[55] da Silva et al. 2011; Flammia and Liu 2011.

[56] Fannes et al. 1992; Perez-Garcia et al. 2007.

[57] Cramer et al. 2010.

[58] Cramer et al. 2010.

The lemma shows that the distance between σ and $|\psi\rangle$ can be bounded in terms of the expectation value $E_\sigma = \mathrm{Tr}(G\sigma)$ if G is a parent Hamiltonian with unique ground state $|\psi\rangle$. If G is a local Hamiltonian, i.e. a sum of terms acting on a small number of neighbouring sites, then $\mathrm{Tr}(G\sigma)$ can be obtained from the corresponding reduced density matrices of σ. The following Theorem 3.7 states a condition under which a pure state admits a local parent Hamiltonian with unique ground state. The condition is formulated in terms of ranks of reduced density matrices, which have the following property:

Lemma 3.6 *Let* $|\psi\rangle \in \mathcal{H}_\Lambda$ *be a vector and* $\rho = |\psi\rangle\langle\psi|$. *Denote Schmidt ranks by*

$$S_X := \mathrm{SR}(X : \Lambda \setminus X)_{|\psi\rangle} \qquad \textit{and} \qquad s_k := S_{\{1,\dots,k\}} \tag{3.47}$$

where $X \subset \Lambda$ *(Schmidt ranks equal* $|\psi\rangle$*'s matricization/tensor unfolding ranks, cf. Section 2.2.1). Then*

$$\mathrm{rk}(\rho_{k+1,\dots,k+r}) \;=\; S_{k+1,\dots,k+r} \;\leq\; S_{1,\dots,k}S_{k+r+1,\dots,n} \;=\; s_k s_{k+r} \tag{3.48}$$

where $r \in \{1,\dots,n\}$ *and* $k \in \{0, n-r\}$.

A proof of Lemma 3.6 is given below. Existence of a local parent Hamiltonian with unique ground state is guaranteed by the following theorem if certain reduced density matrices have maximal rank:

Theorem 3.7 *Let* $|\psi\rangle \in \mathcal{H}_\Lambda^{(1)}$ *be a pure state and* $\rho = |\psi\rangle\langle\psi|$. *Let* $r \in \{0, 1, \dots, n-2\}$. *Define the positive semidefinite* $H \in \mathcal{B}_H(\mathcal{H}_\Lambda)$ *as*

$$H = \sum_{k=1}^{n-r} \mathbb{1}_{1,\dots,k-1} \otimes h_k \otimes \mathbb{1}_{k+r+1,\dots,n} \;, \quad h_k = P_{\ker(\rho_{k,\dots,k+r})} \tag{3.49}$$

Then $H|\psi\rangle = 0$. *H's eigenvalue zero is non-degenerate if*

$$\mathrm{rk}(\rho_{k+1,\dots,k+r}) = s_k s_{k+r}, \quad s_k = \mathrm{SR}(1 \dots k : k+1 \dots n)_{|\psi\rangle} \tag{3.50}$$

holds for $k \in \{1, \dots, n-r-1\}$. *(The operator* $h_k = P_{\ker(\rho_{k,\dots,k+r})} \in \mathcal{B}_H(\mathcal{H}_{k,\dots,k+r})$ *denotes the orthogonal projection onto the null space of the reduced density matrix* $\rho_{k,\dots,k+r}$.*)*

Proof (of Lemma 3.6 and Theorem 3.7) Let $\{|\phi^{(k)}_{i_k}\rangle\}_{i_k=1}^{d_k} \subset \mathcal{H}_k$ be an orthonormal basis of the k-th quantum system. Let

$$\langle\phi^{(1)}_{i_1} \dots \phi^{(n)}_{i_n}|\psi\rangle = G_1(i_1)G_2(i_2)\dots G_n(i_n) \tag{3.51}$$

be an MPS representation of $|\psi\rangle$ whose bond dimensions equal the corresponding Schmidt ranks, i.e. $D_k = s_k$ ($k \in \{1, \dots, n-1\}$), $D_0 = D_n = 1$, $G_k(i_k) \in \mathbb{C}^{D_{k-1} \times D_k}$ and $i_k \in \{1, \dots, d_k\}$. Define a linear map

$$\Gamma_k\colon \quad \mathbb{C}^{D_{k+r} \times D_k} \to \mathcal{H}_{k+1,\dots,k+r}, \tag{3.52a}$$

$$\left\langle \phi^{(k+1)}_{i_{k+1}} \dots \phi^{(k+r)}_{i_{k+r}} \middle| \Gamma(X) \right\rangle = \operatorname{Tr}\left(G_{k+1}(i_{k+1}) \dots G_{k+r}(i_{k+r}) X \right). \tag{3.52b}$$

Inserting the MPS representation into $\rho_{k+1,\dots,k+r} = \operatorname{Tr}_{1,\dots,k,k+r+1,\dots,n}(|\psi\rangle\langle\psi|)$ yields

$$\rho_{k+1,\dots,k+r} = \Gamma_k A_k \tag{3.53}$$

where $A_k \in \mathcal{B}(\mathcal{H}_{k+1,\dots,k+r}; \mathbb{C}^{D_{k+r} \times D_k})$ is a suitably defined linear map (see Lemma 3.8 for details). The last equation and $D_k = s_k$ imply that $\operatorname{rk}(\rho_{k+1,\dots,k+r}) \leq \operatorname{rk}(\Gamma_k) \leq D_k D_{k+r} = s_k s_{k+r}$ and the remainder of Lemma 3.6 is proven as part of Lemma 3.8. Furthermore, the conditions $\operatorname{rk}(\rho_{k+1,\dots,k+r}) = s_k s_{k+r}$ imply that $\operatorname{rk}(\Gamma_k) = D_k D_{k+r}$ holds and that the map Γ_k is injective ($k \in \{1, \dots, n-r-1\}$). The proof is completed by applying results by Baumgratz (2014, Lemmata 12 and 13) which mostly go back to work by Perez-Garcia et al. (2007) and Fannes et al. (1992). ∎

If $|\psi\rangle$'s MPS matrices $G_k(i_k)$ are chosen randomly and r is chosen such that $\max_k D_k \leq d^{r/2}$ holds, the conditions $D_k = s_k$ and $\operatorname{rk}(\rho_{k+1,\dots,k+r}) = s_k s_{k+r}$ are most likely satisfied.[59] In this sense, Theorem 3.7 applies to almost all pure states of given Schmidt ranks if r is chosen such that $\max_k s_k \leq d^{r/2}$. The combination of Lemma 3.5 and Theorem 3.7 then shows that the distance between $|\psi\rangle$ and some density matrix σ can be bounded in terms of the reduced density matrices of σ on $r+1$ neighbouring sites. Whether this bound is practically useful depends on whether the difference between G's two smallest eigenvalues, which is often called *energy gap above the ground state*, is small when compared to $\|G\|_{(\infty)} \leq n - r$.[60] For pure product states which undergo time evolution under a local Hamiltonian, a closely related result is shown in Chapter 6. Theorem 6.10 constructs an approximate parent Hamiltonian whose range r is bounded in terms of evolution time and other system parameters; it applies to lattices of arbitrary dimension.

The remainder of this section presents a lemma and a Remark which provide additional technical details. It also completes the proof of Lemma 3.6.

---

[59] Under the same conditions, the maps Γ_k are most likely injective (Perez-Garcia et al. 2007). See also Lemma 3.8.

[60] See Section 6.3 for further details.

Lemma 3.8 *Let $G_k(i_k)$ be the matrices of an MPS representation of $|\psi\rangle \in \mathcal{H}_\Lambda$ (as in (3.51)).*[61] *For $l \in \{1, \dots, n\}$ and $j \in \{0, \dots, n-l\}$, set*

$$R := \{j+1, \dots, j+l\}, \qquad \big|\phi^{(R)}_{i_R}\big\rangle := \big|\phi^{(j+1)}_{i_{j+1}} \dots \phi^{(j+l)}_{i_{j+l}}\big\rangle, \tag{3.54}$$

$$i_R := (i_{j+1}, \dots, i_{j+l}), \qquad G_R(i_R) := G_{j+1}(i_{j+1}) \dots G_{j+l}(i_{j+l}) \tag{3.55}$$

and define linear maps[62]

$$\Gamma_R : \quad \mathbb{C}^{D_{j+l} \times D_j} \longrightarrow \mathcal{H}_R, \qquad \big\langle \phi^{(R)}_{i_R} \big| \Gamma(A) \big\rangle = \mathrm{Tr}(G_R(i_R) A), \tag{3.56}$$

$$\Delta_R : \quad \mathcal{H}_R \longrightarrow \mathbb{C}^{D_j \times D_{j+l}}, \qquad \big|\phi^{(R)}_{i_R}\big\rangle \mapsto G_R(i_R). \tag{3.57}$$

Choose $r \in \{1, \dots, n\}$, $k \in \{0, \dots, n-r\}$ and define the intervals

$$X := \{1, \dots, k\}, \qquad Y := \{k+1, \dots, k+r\}, \qquad Z := \{k+r+1, \dots, n\}. \tag{3.58}$$

Let $\rho_Y = \rho_{k+1,\dots,k+r}$ be a reduced density matrix of $\rho := |\psi\rangle\langle\psi|$. Then

$$\rho_Y = \Gamma_Y (\Delta_X \otimes \Delta_Z)(\Delta_X^* \otimes \Delta_Z^*) \Gamma_Y^*. \tag{3.59}$$

where $\mathbb{C}^{1 \times D_k} \otimes \mathbb{C}^{D_{k+r} \times 1}$ has been identified with $\mathbb{C}^{D_{k+r} \times D_k}$.

The relations $\mathrm{rk}(\rho_Y) = S_Y \le s_k s_{k+r}$*,* $S_Y \le \mathrm{rk}(\Gamma_Y) \le D_k D_{k+r}$*,* $s_k \le \mathrm{rk}(\Delta_X) \le D_k$ *and* $s_{k+r} \le \mathrm{rk}(\Delta_Z) \le D_{k+r}$ *hold. If $D_j = s_j$ holds for $j \in \{k, k+r\}$, then* $\mathrm{rk}(\rho_{k+1,\dots,k+r}) = s_k s_{k+r}$ *holds if and only if Γ_Y is injective.*

Proof Define the linear operator $M \in \mathcal{B}(\mathcal{H}_{XZ}; \mathcal{H}_Y)$ as

$$\big\langle \phi^{(Y)}_{i_Y} \big| M \big| \phi^{(X)}_{i_X} \phi^{(Z)}_{i_Z} \big\rangle := \big\langle \phi^{(X)}_{i_X} \phi^{(Y)}_{i_Y} \phi^{(Z)}_{i_Z} \big| \psi \big\rangle = B_X(i_X) B_Y(i_Y) B_Z(i_Z). \tag{3.60}$$

Using $\rho_Y = \mathrm{Tr}_{XZ}(|\psi\rangle\langle\psi|)$, we obtain

$$\big\langle \phi^{(Y)}_{i_Y} \big| \rho_Y \big| \phi^{(Y)}_{i'_Y} \big\rangle = \sum_{i_X i_Z} \big\langle \phi^{(X)}_{i_X} \phi^{(Y)}_{i_Y} \phi^{(Z)}_{i_Z} \big| \psi \big\rangle \big\langle \psi \big| \phi^{(X)}_{i_X} \phi^{(Y)}_{i'_Y} \phi^{(Z)}_{i_Z} \big\rangle = \big\langle \phi^{(Y)}_{i_Y} \big| M M^* \big| \phi^{(Y)}_{i'_Y} \big\rangle. \tag{3.61}$$

This shows $\rho = MM^*$, which implies $\mathrm{rk}(\rho) = \mathrm{rk}(M) = S_Y$. Note that $B_X(i_X) \in \mathbb{C}^{1 \times D_k}$, $B_Z(i_Z) \in \mathbb{C}^{D_{k+r} \times 1}$ and therefore,

$$(\Delta_X \otimes \Delta_Z) \big| \phi^{(X)}_{i_X} \phi^{(Z)}_{i_Z} \big\rangle = B_X(i_X) \otimes B_Z(i_Z) = B_Z(i_Z) B_X(i_X). \tag{3.62}$$

---

[61] $D_k = s_k$ is not required here.

[62] One may wonder whether there is a relation between the maps Γ_R and Δ_R. If the vector basis $\big|\phi^{(R)}_{i_r}\big\rangle$ and the matrix basis $E_{ij} = e_i e_j^\top$ are used, a relation is $\Gamma = \Delta^\top T$ and $\Delta = T \Gamma^\top$ where $T(A) = A^\top$.

This shows

$$\left\langle \phi^{(Y)}_{i_Y} \middle| \Gamma_Y(\Delta_X \otimes \Delta_Z) \middle| \phi^{(X)}_{i_X} \phi^{(Z)}_{i_Z} \right\rangle = \operatorname{Tr}(B_Y(i_Y) B_Z(i_Z) B_X(i_X)) = B_X(i_X) B_Y(i_Y) B_Z(i_Z),$$

which shows $M = \Gamma_Y(\Delta_X \otimes \Delta_Z)$ and $S_Y = \operatorname{rk}(M) \le \operatorname{rk}(\Gamma_Y) \le D_k D_{k+r}$. If an MPS representation with $D_j = s_j$ ($j \in \{k, k+r\}$) is used, it implies $S_Y \le s_k s_{k+r}$ and this inequality is independent of the representation used. As an alternative, $S_Y \le s_k s_{k+r}$ can also be obtained as consequence of Lemma 8.15.

We define M' by replacing, in the definition of M, the sets X and Y with $X' = \emptyset$ and $Y' = X \cup Y$. Then $M' = \Gamma_{X \cup Y} \Delta_Z$ and

$$s_{k+r} = S_{X \cup Y} = \operatorname{rk}(M') \le \operatorname{rk}(\Delta_Z). \tag{3.63}$$

In this way, we obtain $s_{k+r} \le \operatorname{rk}(\Delta_Z) \le D_{k+r}$ and $s_k \le \operatorname{rk}(\Delta_X) \le D_k$.

Note that

$$\begin{aligned} \operatorname{rk}(\Gamma_Y) \ge S_Y = \operatorname{rk}(M) &\ge \operatorname{rk}(\Gamma_Y) + \operatorname{rk}(\Delta_X \otimes \Delta_Z) - D_k D_{k+r} && (3.64a) \\ &\ge \operatorname{rk}(\Gamma_Y) + s_k s_{k+r} - D_k D_{k+r} && (3.64b) \end{aligned}$$

holds.[63] If $D_j = s_j$ holds for $j \in \{k, k+r\}$, it implies $S_Y = \operatorname{rk}(\Gamma_Y)$. In this case, Γ_Y being injective, $\operatorname{rk}(\Gamma_Y) = s_k s_{k+r}$ and $S_Y = s_k s_{k+r}$ are seen to be equivalent. ■

Remark 3.9 If a pure state $|\psi\rangle$ and an MPS representation with bond dimensions D_k are given, reduced density matrices satisfy

$$\operatorname{rk}(\rho_{k+1,\dots,k+r}) \le s_k s_{k+r} \le D_k D_{k+r}. \tag{3.65}$$

In this case, $\operatorname{rk}(\rho_{k+1,\dots,k+r}) = D_k D_{k+r}$ directly implies $s_k = D_k$, $s_{k+r} = D_{k+r}$, $\operatorname{rk}(\rho_{k+1,\dots,k+r}) = s_k s_{k+r}$ and the existence of a parent Hamiltonian with non-degenerate ground state as per Theorem 3.7. □

3.4.2. Markov entropy decomposition certificate

In this section, we present a result which upper bounds the trace distance between two density matrices in terms of von Neumann entropies of certain reduced density matrices. The von Neumann entropy of a density matrix is denoted by $S(\rho) := -\operatorname{Tr}(\rho \log \rho)$ and $S(A)_\rho := S(\rho_A)$ where $\rho_A = \operatorname{Tr}_{\Lambda \setminus A}(\rho)$ is a reduced density matrix. Conditional von Neumann entropies are denoted by $S(A|B) = S(AB) - S(B)$. The following upper bound on $\|\rho - \sigma\|_{(1)}$ was shown by Kim (2014, Eq. (12)):[64]

---

[63] We used $\operatorname{rk}(A) + \operatorname{rk}(B) - k \le \operatorname{rk}(AB)$ where A has k columns and where B has k rows. E.g. Horn and Johnson 1991a, 0.4.5(c).

Theorem 3.10 *For* $k \in \Lambda = \{1, \dots, n\}$*, choose subsets* $M_k \subset \{1, \dots, k-1\}$*. Let* $\rho, \sigma \in \mathcal{D}(\mathcal{H}_\Lambda)$*. Then*

$$S(\rho) \le S_M(\rho) := \sum_{k=1}^{n} S(k|M_k)_\rho \tag{3.66}$$

and

$$\frac{1}{4}\Big(\|\rho - \sigma\|_{(1)}\Big)^2 \le 2S\Big(\frac{\rho+\sigma}{2}\Big) - S(\rho) - S(\sigma) \le 2S_M\Big(\frac{\rho+\sigma}{2}\Big) \tag{3.67}$$

hold.

Theorem 3.10 provides an upper bound on the trace distance $\frac{1}{2}\|\rho - \sigma\|_{(1)}$ which depends only on the reduced density matrices $\rho_{M_k,k}$ and $\sigma_{M_k,k}$. The trace distance of two density matrices ρ and σ always satisfies $\frac{1}{2}\|\rho - \sigma\|_{(1)} \le 1$ and Theorem 3.10 can provide a tighter bound if ρ and σ are nearly pure (i.e. $S(\rho), S(\sigma) \ll \frac{1}{2}$) or if $S(\rho)$ and $S(\sigma)$ are known. The application of Theorem 3.10 is not limited to one-dimensional lattices because the sets M_k do not need to be contiguous. The upper bound $S_M(\rho)$ on $S(\rho)$ is called *Markov entropy* and the sets M_k are called *Markov shields*.[65] Equality in $S(\rho) \le S_M(\rho)$ implies that the density matrix ρ can be reconstructed from the reduced density matrices $\rho_{M_k k}$ with the so-called Petz recovery map.[66] This approach has been compared to MPO reconstruction (Section 3.3.3) in a previous work[67] but it is not discussed in the remainder of this work.

---

[64] The cited work further shows that $-S(\rho) \le S_{M'}(\rho)$ where $S_{M'}(\rho) := \sum_{k=1}^{n} S(k|M'_k)_\rho$ with arbitrary subsets $M'_k \subset \{k+1, \dots, n\}$. Applying the proof of $S(\rho) \le S_M(\rho)$ to a reversed chain yields $S(\rho) = \sum_{k=1}^{n} S(k|k+1, \dots, n) \le \sum_{k=1}^{n} S(k|M'_k) = S_{M'}(\rho)$ where strong subadditivity of the von Neumann entropy was used in the form $S(A|BC) \le S(A|B)$ (e.g. Nielsen and Chuang 2007, Theorem 11.15). Since $S(\rho) \ge 0$, we obtain the bound $-S(\rho) \le 0 \le S_{M'}(\rho)$.

[65] Poulin and Hastings 2011.

[66] See Petz 2003; Hayden et al. 2004; Poulin and Hastings 2011; Holzäpfel et al. 2018. Let $N_k := \{1, \dots, k-1\} \setminus M_k$. $S(\rho) = S_M(\rho)$ implies $S(k|N_k M_k) = S(k|M_k)$ or, in terms of the conditional mutual information $I(A : C|B)$, that $I(N_k : k|M_k) = S(k|M_k) - S(k|N_k M_k) = 0$ holds. This in turn implies $\rho = (\mathcal{E}_n \dots \mathcal{E}_2)(\rho_1)$ where $\mathcal{E}_k \colon \mathcal{B}(\mathcal{H}_{M_k}) \to \mathcal{B}(\mathcal{H}_{M_k k})$ are *Petz recovery maps* given by $\mathcal{E}_k(\sigma) = (\rho_{M_k k})^{1/2}(\rho_{M_k})^{-1/2}\sigma(\rho_{M_k})^{-1/2}(\rho_{M_k k})^{1/2}$.

[67] Holzäpfel et al. 2018.

Part II.

Scalable estimation and verification of quantum systems

Chapter 4.

Efficient maximum likelihood estimation (MLE)

In the following, we take a closer look at maximum likelihood estimation of quantum states represented as matrix product states by means of numerical studies. In Section 4.1, we determine the measurement effort necessary for a constant estimation error as the number of subsystems grows.[1] The measurement effort is determined as the number of repeated measurements of a fixed set of observables for a selection of time-evolved states. The following Section 4.2 presents results on MPS-MLE for mixed states represented as locally purified matrix product state (PMPS). Here, we reconstruct a state from simulated measurement data of tensor products of quadrature amplitudes of several harmonic oscillators, demonstrating successful, efficient reconstruction of a state of a system with continuous variables.

4.1. Resources for constant estimation error

Estimating a given quantum state with the MPS-MLE method can succeed or fail, depending on the correlations in the quantum state and the selection of a non-informationally-complete measurement. Insight on the method's capabilities can be gained with numerical studies involving specific states. For example, the original proposal included simulation results for mixed and pure states of 16 and 20 qubits.[2] Successful estimation from a finite number of measurements was demonstrated but the question whether a constant estimation error can be achieved

---

[1] Section 4.1 reproduces parts of the Supplementary Information of the original publication Lanyon, Maier, et al. 2017. The present author performed numerical simulations of the estimation method. MPS representations of time-evolved states were supplied by Anton S. Buyskikh (cf. Buyskikh 2017).

[2] Baumgratz et al. 2013b.

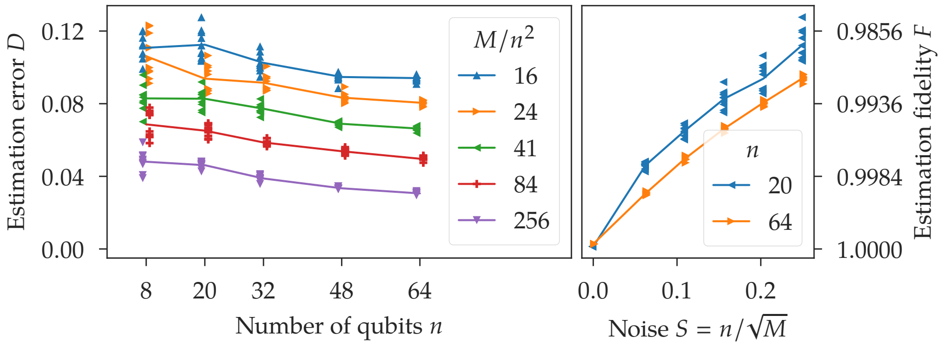


Figure 4.1: Estimation error of MPS-MLE with a finite number of measurements.[4] The estimation error D and estimation fidelity $F = |\langle\psi(t)|\psi_{\text{est}}\rangle|^2$ are related by $D = \sqrt{1-F}$. Markers show errors obtained in 10 individual reconstructions and lines connect average values. Markers are shifted slightly horizontally to enhance visibility and straight lines serve as guide to the eye.

as the number of subsystems increases was left open. Related evidence is presented in Section 5.3.2 below but the question is not directly answered there either.[3] In this section, we show, for a specific class of time-evolved states, that a total number of measurements proportional to n^3 is sufficient for estimating a pure state up to a constant error in trace distance.[1]

We consider an initial pure product state which evolves under a nearest-neighbour Hamiltonian on a linear chain. Specifically, the initial state is $|\psi(0)\rangle = |\uparrow\downarrow\uparrow\downarrow\dots\rangle$ and it evolves into $|\psi(t)\rangle = \exp(-\mathrm{i}H_{\text{Ising}}t/\hbar)|\psi(0)\rangle$ under the nearest-neighbour Hamiltonian

$$H_{\text{Ising}} = \hbar\sum_{i=1}^{n-1} 2J\,\sigma_X^{(i)}\sigma_X^{(i+1)} + \hbar\sum_{i=1}^{n}(B+B_i)\,\sigma_Z^{(i)} \tag{4.1}$$

where $\hbar$ is the reduced Planck constant. This Hamiltonian is motivated by the ion trap experiment described in Chapter 7 and we use the average nearest-neighbour coupling J and field $B+B_i$ from this setting. An MPS representation of $|\psi(t)\rangle$ can be obtained with DMRG or related numerical algorithms[1] with effort which increases exponentially with t but only polynomially with n.[5] Furthermore, $|\psi(t)\rangle$ is uniquely determined by its reduced density matrices on a number of sites which also grows exponentially with t but only polynomially with n.[6] Therefore, we attempt to estimate $|\psi(t)\rangle$ from complete information of reduced density matrices on all blocks of r neighbouring sites where $r = 3$. We simulate m measurements of each of the $(n-r+1)3^r$ POVMs $\Pi^{(r)}_{\sigma_a}$; here, $a \in \{X, Y, Z\}^r$ and $\Pi^{(r)}_{\sigma_a}$ is constructed from r-fold tensor products of Pauli matrix eigenvectors and is taken to act on sites $k, \dots, k+r-1$ (see Equation (3.15)). The outcomes of the simulated measurements are used to compute observed outcome frequencies of the POVM $\Pi^{(n,r)}_{\text{Block}}$ (Equation (3.20)), which are in turn used for MLE over pure MPS (Equation (3.34)). After each iteration, the estimated state is compressed to MPS bond dimension $D = 4$ and the iterations are started from the output from the MPS-SVT algorithm (Section 3.3.2). We quantify the estimation error with the trace distance $D = \left\| |\psi(t)\rangle\langle\psi(t)| - |\psi_{\text{est}}\rangle\langle\psi_{\text{est}}| \right\|_{(1)}/2$, which can be computed[7] from the fidelity $F = |\langle\psi(t)|\psi_{\text{est}}\rangle|^2$ via $D = \sqrt{1-F}$. The estimation error after 2000 iterations of (3.34) is shown in Figure 4.1. We observe that the estimation error decreases slightly with the number of qubits n if the number of measurements m is proportional to n^2. This shows that a constant estimation error can be achieved, for this particular family of states, with a total number of measurements given by $O(n^3)$. An open question for future work concerns the growth of the number of iterations in the SVT and MLE algorithms necessary to achieve a desired estimation error. The given, fixed number number of iterations is seen sufficient in this example but it could change for other states or larger systems.

---

[3] Originally published in Holzäpfel et al. 2015.

[4] The figure reproduces Figure S3 from the Supplementary Information of Lanyon, Maier, et al. 2017 with modifications.

4.2. Efficient maximum likelihood estimation with continuous variables

In this section, we demonstrate efficient MLE of quantum states of systems with continuous variables, i.e. systems with an infinite-dimensional Hilbert space. Our basic model is the quantum-mechanical harmonic oscillator, realized e.g. by oscillations of electromagnetic fields in photonic systems, by a molecular vibrational mode, by the motion of trapped ions or by a micromechanical oscillator.

We proceed with a short, mathematically non-rigorous introduction to the quantum mechanical harmonic oscillator. The Hamiltonian of a quantum harmonic oscillator of angular frequency ω is given by[8]

---

[5] Osborne 2006. See also Chapter 6, in particular Theorem 6.33.

[6] Lanyon, Maier, et al. 2017; Supplementary Information. See also Chapter 6, in particular Theorem 6.10.

[7] Nielsen and Chuang 2007, Eqs. (9.11), (9.60), (9.99).

$$H = \hbar\omega\left(a^*a + \frac{1}{2}\right) \tag{4.2}$$

where a^* and a are creation and annihilation operators and $\hbar$ is the reduced Planck constant. The Hamiltonian has the non-degenerate eigenvalues $E_j = \hbar\omega(j + \frac{1}{2})$ with $j \in \{0, 1, \dots\}$. The corresponding eigenvectors $|n_j\rangle$ are called the *Fock basis* and they satisfy $a^*a|n_j\rangle = j|n_j\rangle$, $j \in \{0, 1, \dots\}$. Dimensionless conjugate variables are given by $X = (a + a^*)/\sqrt{2}$ and $P = -\mathrm{i}(a - a^*)/\sqrt{2}$, describing e.g. position and momentum or electric and magnetic fields. They satisfy the commutation relation $[X, P] = +\mathrm{i}$. Independently of whether the physical system is given by a photonic mode or not, we shall call the quantum number j the *number of photons* and a harmonic oscillator *a mode*.

The essential observable of the quantum Harmonic oscillator is the so-called *quadrature amplitude*[9]

$$Q_\theta := X\cos(\theta) + P\sin(\theta) \tag{4.3}$$

where $\theta \in [0, \pi)$ is called the *phase* of the quadrature amplitude. The eigenvalues $q \in \mathbb{R}$ of the quadrature amplitude Q_θ are non-degenerate and we denote its eigenvectors by $|\phi_q^\theta\rangle$. In a photonic experiment, the quadrature amplitude can be measured with a balanced homodyne detection (BHD) experiment where θ is the phase between a to-be-detected mode and a strong, coherent field, called the local oscillator.[10] Quadrature amplitudes have also been measured for oscillations of atoms in molecules,[11] for oscillations of trapped ions[12] and for micromechanical oscillators.[13]

The quadrature amplitudes Q_θ are essential because an unknown quantum state ρ can be reconstructed if the outcome probability distributions $p_{\rho,\theta}(q) = \langle\phi_q^\theta|\rho|\phi_q^\theta\rangle$ are known for all $\theta \in [0, \pi)$.[14] If the quantum state is supported on the first d elements of the Fock basis (i.e. at most $d - 1$ photons), the distributions $p_{\rho,\theta}(q)$ for d distinct phases θ are already sufficient to reconstruct ρ.[15] The assumption of a single mode with at most $d-1$ photons allows us to restrict the density matrix to a d-dimensional subspace of the complete Hilbert space and enables the state to be estimated with the regular MLE method.[16] In the following, we consider reconstruction under the assumption that an n-mode state has at most $d-1$ photons in each mode. This assumption is satisfied e.g. by photonic modes containing few photons or by mechanical oscillators near their ground state.

---

[8] E.g. Nielsen and Chuang 2007, Section 7.3.2; Lvovsky and Raymer 2009.

[9] E.g. Lvovsky and Raymer 2009, Section II.A. Note that $Q_{\theta+\pi} = -Q_\theta$.

[10] Smithey et al. 1993; Lvovsky and Raymer 2009.

[11] Dunn et al. 1995.

[12] Poyatos et al. 1996.

[13] Aspelmeyer et al. 2014. See also e.g. Vanner et al. 2013.

[14] Vogel and Risken 1989.

[15] Leonhardt and Munroe 1996.

Measurement of a quadrature amplitude is described by the POVM

$$\Pi^Q_\theta := \left\{ P^\theta_q \colon q \in \mathbb{R} \right\}, \quad P^\theta_q := |\phi^\theta_q\rangle\langle\phi^\theta_q| \tag{4.4}$$

whose positive-semidefinite elements P^θ_q satisfy $\int_{\mathbb{R}} P^\theta_q \, \mathrm{d}q = \mathbb{1}$. In the following construction of r-mode measurements, the POVM Π^Q_θ basically assumes the role of a Pauli POVM $\Pi^{(1)}_{\sigma_a}$ from Section 3.1.3. A simultaneous or sequential measurement measurement of the quadrature amplitude Q_{θ_i} on mode $i \in \{1 \dots r\}$ is described by the POVM

$$\Pi^{(r,\theta)}_Q := \Pi^Q_{\theta_1} \otimes \cdots \otimes \Pi^Q_{\theta_r} = \left\{ P^\theta_q := P^{\theta_1}_{q_1} \otimes \cdots \otimes P^{\theta_r}_{q_r} \colon q \in \mathbb{R}^r \right\} \tag{4.5}$$

where the vector $\theta \in [0,\pi)^r \subset \mathbb{R}^r$ contains r independent phases $\theta_1, \dots, \theta_r$. We denote the outcome probabilities of $\Pi^{(r,\theta)}_Q$ by $p_{\rho,\theta}(q) := \operatorname{Tr}(P^\theta_q \rho)$ where $q \in \mathbb{R}^r$.

Assuming at most $d-1$ photons in each mode, we choose d distinct phases $\Theta := \{\theta_1, \dots, \theta_d\} \subset [0,\pi)$. The measurement of all combinations of each phase $\theta_i \in \Theta$ on each mode is described by the POVM

$$\Pi^{(n,r,k,\Theta)}_{\text{Local}} := \mathcal{U}_P\left(\left\{ \Pi^{(r,\theta)}_Q \colon \theta \in \Theta^r \right\} \right) = \left\{ \tfrac{1}{|\Theta|^r} P^{\theta_1}_{q_1} \otimes \cdots \otimes P^{\theta_r}_{q_r} \colon \theta \in \Theta^r, q \in \mathbb{R}^r \right\} \tag{4.6}$$

Assuming n modes, the POVM $\Pi^{(n,r,k,\Theta)}_{\text{Local}}$ is taken to act on modes $k, \dots, k+r-1$ where $k \in \{1, \dots, n-r+1\}$. Note that

$$\Pi^{(n,r,k,\Theta)}_{\text{Local}} = \left(\Pi_{Q,\Theta}\right)^{\otimes r} \quad \text{where} \quad \Pi_{Q,\Theta} := \mathcal{U}_P\left(\left\{ \Pi^Q_\theta \colon \theta \in \Theta \right\} \right). \tag{4.7}$$

The POVM $\Pi_{Q,\Theta}$ is IC if a single mode is restricted to $d-1$ photons; therefore, the POVM $\Pi^{(n,r,k,\Theta)}_{\text{Local}}$ is IC on r modes if there are at most $d-1$ photons in each mode. On n modes, we define the POVM

$$\Pi^{(n,r,\Theta)}_{\text{Block}} = \left\{ \tfrac{1}{n-r+1} \mathbb{1}_{k-1} \otimes P \otimes \mathbb{1}_{n-k-r+1} \colon P \in \Pi^{(n,r,k,\Theta)}_{\text{Local}},\ k \in \{1 \dots n-r+1\} \right\} \tag{4.8}$$

---

[16] Banaszek et al. 1999; Lvovsky 2004. Other approaches, such as obtaining a Wigner function representation of the density matrix using e.g. the inverse Radon transform are not covered here; cf. Lvovsky and Raymer 2009.

which collects all the elements of the POVMs from (4.6).

The operator R from MLE is straightforwardly adapted to POVMs with one or more continuous indices as

$$R(\rho) = \int_{\mathbb{R}^l} \frac{f(x)}{\mathrm{Tr}(M(x)\rho)} M(x) \,\mathrm{d}x \tag{4.9}$$

where $f(x)$ represents observed outcome frequencies and where l is a positive integer. For the corresponding POVM $\Pi = \{M(x) \colon x \in R^l\}$ we will insert $\Pi^{(n,r,k,\Theta)}_{\mathrm{Local}}$ or $\Pi^{(n,r,\Theta)}_{\mathrm{Block}}$, wherein discrete indices are formally included with an additional continuous index and delta distributions. Certainly, it is not feasible to observe the relative frequencies of $f(x)$ for all $x \in \mathbb{R}^l$ and an estimate of $f(x)$ must be used instead. Suppose that we measured Π m times and obtained the outcomes $x_1, \dots, x_m \in \mathbb{R}^l$. A basic estimate for the observed frequencies $f(x)$ is given by

$$f(x) = \sum_{i=1}^{m} \delta(x - x_m). \tag{4.10}$$

where δ is the Dirac delta distribution. More sophisticated estimates of $f(x)$ may improve the estimation method but may also increase the difficulty of obtaining $R(\rho)$. Inserting the given estimate of $f(x)$, $R(\rho)$ becomes[17]

$$R(\rho) = \sum_{i=1}^{m} \frac{1}{\mathrm{Tr}(M(x_i)\rho)} M(x). \tag{4.11}$$

As mentioned above, we assume that the to-be-reconstructed quantum state is supported on the first d elements of the Fock basis. As a consequence, it is sufficient to compute the corresponding matrix elements of $R(\rho)$ in the Fock basis. This requires the components of the quadrature amplitude eigenstates in the Fock basis, which are given by[18]

$$\langle n_j | \phi^\theta_q \rangle = \mathrm{e}^{\mathrm{i}j\theta} \psi_j(q), \qquad \psi_j(q) = \frac{1}{\pi^{1/4}} \frac{1}{\sqrt{2^j j!}} \mathrm{e}^{-q^2/2} H_j(q). \tag{4.12}$$

Here, the $H_j(q)$ are Hermite polynomials[19] and $j \in \{0, 1, 2, \dots\}$ enumerates the Fock basis.

---

[17] Lvovsky 2004.

[18] E.g. Lvovsky and Raymer 2009, Eqs. (26) and (44).

[19] $H_j(q) = j! \sum_{l=0}^{s} \frac{(-1)^{s-l}}{(s-l)!} \frac{(2x)^{j-2(s-l)}}{(j-2(s-l))!}$, $s = \lfloor j/2 \rfloor$ (Olver et al. 2010, Eq. (18.5.13), available at `https://dlmf.nist.gov/18.5.E13`).

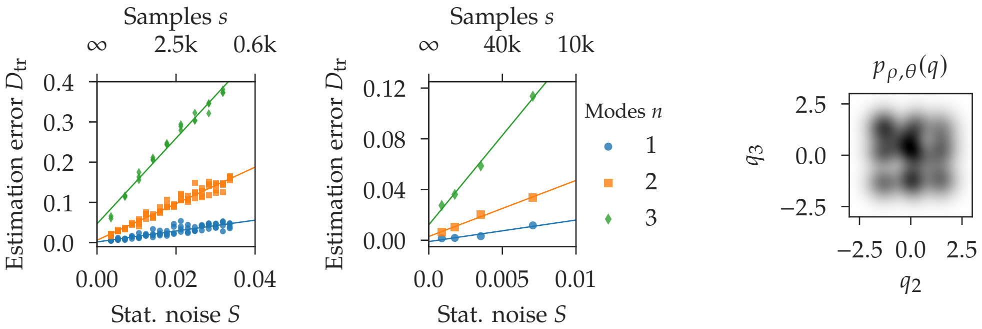


Figure 4.2: Estimation of an n-mode state from quadrature amplitude measurements on all modes. Left and middle: Estimation error as function of $S = 1/\sqrt{s}$ where s is the number of samples. Up to $s = 1.28 \times 10^6$ samples per phase setting were used. Lines serve as guide to the eye. Right: Outcome probability density of the quadrature amplitude with phase $\theta = (0, 0, 0)$ on the three-mode state with $q = (0, q_2, q_3)$. The to-be-estimated random state does not possess any obvious symmetries.

Writing the density matrix and the operator $R(\rho)$ in the Fock basis, we have everything in order to perform MLE of a quantum state from quadrature amplitude measurements on a single harmonic oscillator or mode as proposed by Lvovsky (2004). In the following, we discuss the straightforward generalization to multiple modes[20] and, at the end, efficient tomography of multiple modes from incomplete measurements.

First, we focus on the reconstruction of a state of $n \in \{1, 2, 3\}$ modes from measurements on all modes. We consider states with at most $d - 1 = 2$ photons in each mode. We set $r = n$ and choose the d equidistant phases $\Theta = \{0, \pi/d, \dots, (d - 1)\pi/d\}$.[21] For each $\theta \in \Theta^r$, we simulate s measurements of the POVM $\Pi_{\mathrm{Q}}^{(r,\theta)}$ numerically, meaning that measurements of all the POVMs contributing to $\Pi_{\mathrm{Local}}^{(n,r,k,\Theta)}$ are simulated ($n = r, k = 1$). For $r \in \{1, 2\}$, we simulated measurements by drawing from a piecewise constant approximation to the outcome distribution $p_{\rho,\theta}(q)$ and for $r \in \{1, 2, 3\}$, we sampled from the distribution with ensemble Markov chain Monte Carlo sampling implemented by the `emcee` package.[22] The full outcome

---

[20] E.g. Babichev et al. 2004.

[21] In other numerical experiments, using d phases chosen uniformly random from $[0, \pi)$ was observed to generally yield worse results. This may not be surprising because quadrature amplitude measurements with a smaller phase difference yield more similar outcome distributions.

[22] Foreman-Mackey et al. 2013.

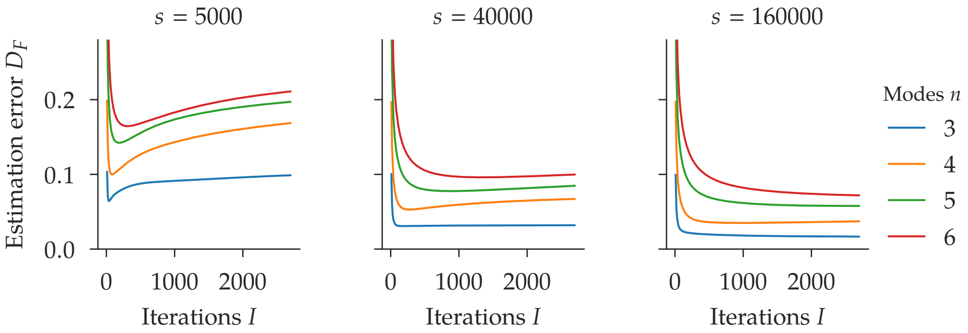


Figure 4.3: Estimation error after i iterations of (3.37). A state on $n \in \{3 \dots 6\}$ modes is estimated from measurements on $r = 3$ neighbouring modes using s samples per phase setting.

distribution of the POVM $\Pi^{(n,r,k,\Theta)}_{\text{Local}}$ is known to uniquely determine the state and we attempt reconstruction from a finite number of samples with MLE as described above.

Figure 4.2 shows results for the estimation of a random mixed state.[23] The figure shows the estimation error of states on one, two and three modes from s samples per phase setting (see above). The statistical noise caused by a finite number of samples s is heuristically quantified by $S = 1/\sqrt{s}$. The estimation error is measured by the trace distance $D_{\text{tr}}(\rho, \rho_{\text{est}}) = \|\rho - \rho_{\text{est}}\|_{(1)}/2$ and it is seen to decrease as the number of samples increases.

Next, we discuss efficient estimation of a state on $n \in \{3, 4, 5, 6\}$ modes using measurements on only $r = 3$ neighbouring modes. We perform MLE using the POVM $\Pi^{(n,r,\Theta)}_{\text{Block}}$ and simulate measurements of the POVMs $\Pi^{(r,\theta)}_{Q}$ whose elements contribute to $\Pi^{(n,r,\Theta)}_{\text{Block}}$ as described above. The total number of samples is $(n-r+1)d^r s$.

Figures 4.3 and 4.4 show results of the estimation of a random locally purified matrix product state (PMPS) of bond dimension $D = 2$ (at most $d - 1 = 2$ photons per mode). We use Equation (3.37) to perform MLE with the PMPS representation restricted to bond dimension $D = 4$. Here, we measure the estimation error with the Frobenius norm as $D_F(\rho, \rho_{\text{est}}) = \|\rho - \rho_{\text{est}}\|_{(2)}/(\|\rho\|_{(2)} + \|\rho_{\text{est}}\|_{(2)})$. Figure 4.3 shows the estimation error over 2700 iterations of (3.37). More iterations are seen to be necessary as the number of samples s and the number of modes n increase and the scaling should be investigated in future work. The estimation error at the

---

[23] $\rho = TT^*/\text{Tr}(TT^*)$ where the entries of T have normally distributed real and imaginary parts.

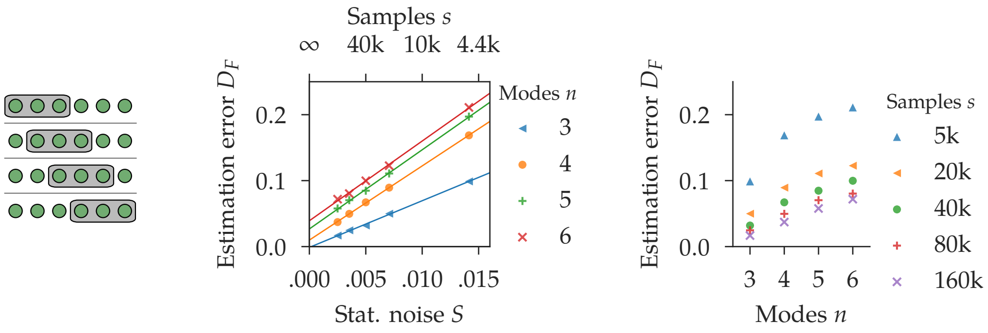


Figure 4.4: Efficient state estimation from measurements on $r = 3$ out of $n \in \{3, 4, 5, 6\}$ modes. Left: Modes are arranged in a linear chain and measurements on neighbouring modes are simulated. Middle and right: Reconstruction error of a state on n modes from measurements on $r = 3$ neighbouring modes. Statistical noise is quantified by $S = 1/\sqrt{s}$ where s is the number of samples per phase setting. Lines serve as guide to the eye.

end of the iterations is shown in Figure 4.4 as function of the number of samples s and as function of the number of modes n. The estimation error decreases with the number of samples and exhibits only a moderate, possibly linear increase with the number of modes n. As mentioned above, the total number of samples used for this estimation increases only linearly with the number of modes n. The computational time was observed to be roughly proportional to $n - r + 1$. This is consistent with a large fraction of computational time being spent on processing the measurement data, whose total size is proportional to the number of r-site blocks which is exactly $n - r + 1$. Processing the global state is comparably easy and thanks to the PMPS representation, this is not expected to change for larger mode counts n if the block size r is kept constant. In summary, we demonstrated efficient estimation of a multi-mode state with a restricted number of photons per mode from quadrature amplitude measurements.

In the following, we discuss the computational time required by our implementation and discuss possible starting points for algorithmic improvements. In each iteration of (3.37), we need to construct the operator $R(\rho)$, which requires a summation over all samples (Equation (4.11)). We replaced our original Python implementation by a faster implementation based on just-in-time compilation with

Numba[24] and a still faster C implementation interfaced to Python with Cython.[25] Estimating a three-mode state from measurements on three modes, this enabled us to perform 600 iterations with about 10^7 samples in 40 hours of CPU time.[26] This highlights the need for algorithmic improvements. As the estimated state often changes only slightly from one iteration to the next, one could update $R(\rho)$ only after a larger number of iterations instead of recomputing it after every iteration. Alternatively, one could divide the sum (4.11) into several parts, recomputing one part in each iteration. Improving our basic estimate (4.10) for the observed frequencies $f(x)$ could remove the need to consider all the samples individually in the summation (4.11). One may consider a multidimensional piecewise constant approximation to $f(x)$ with weights given by the number of samples in each interval ("binning"), which, however, can limit the achievable precision.[27] Alternatively, one could replace several samples with similar values by one sample with an appropriate weight, otherwise keeping the structure of (4.11). One can also view the computation of $R(\rho)$ as an estimation problem where the true $R(\rho)$, as defined by the exact but unknown POVM probabilities, is to be estimated. The mentioned estimates of $f(x)$ already provide three different estimators and more efficient, potentially biased, estimators may be found.

---

[24] Lam et al. 2015.

[25] Dalcin et al. 2010.

[26] Using a single core of an `Intel(R) Xeon(R) CPU E5-2670 0 @ 2.60GHz`.

[27] Lvovsky 2004.

Chapter 5.

Scalable ancilla-assisted process tomography (AAPT)

A method for efficient tomography of quantum processes acting on one-dimensional quantum many-body systems is presented in this chapter.[1] It allows for the reconstruction of unitary processes, local Hamiltonians or more general quantum operations. This is achieved by combining AAPT with efficient methods for quantum state estimation.

5.1. Local ancilla-assisted process tomography

Due to the exponential growth of Hilbert space with the number of particles n, the measurement and post-processing effort required for QST and quantum process tomography increase exponentially with the number of particles. This exponential growth refers to the task of distinguishing an arbitrary quantum state or process from all other states or processes and as mentioned above, a particularly simple state may be distinguished from all other states with much smaller effort – this is utilized by methods for efficient quantum state tomography based on MPSs.[2]

Quantum process tomography can be reformulated as a quantum state tomography problem using the Choi-Jamiołkowski isomorphism,[3] which we discuss below. The Choi-Jamiołkowski isomorphism involves a maximally entangled pure state, a state which does not immediately allow for an efficient MPS representation due to its maximal amount of entanglement. This section shows how efficient process tomography is enabled by efficient state tomography and the Choi-Jamiołkowski isomorphism despite the fact that the latter involves a maximally entangled state.

---

[1] This chapter reproduces the original publication Holzäpfel et al. 2015 (© 2015 American Physical Society). In Section 5.1, the evidently completely positive (CP) representation of quantum operations was added. Sections 5.2 and 5.3 were edited slightly.

[2] Cramer et al. 2010; Baumgratz et al. 2013a,b. See also Chapter 3.

[3] Choi 1975; Jamiołkowski 1972.

A quantum operation is a linear map $\mathcal{E} \in \mathcal{B}(\mathcal{B}(\mathcal{H}_B); \mathcal{B}(\mathcal{H}_A))$ which is completely positive and trace preserving (CPTP). If $\mathcal{H}_C$ is an arbitrary finite-dimensional Hilbert space and $\sigma \in \mathcal{D}(\mathcal{H}_B \otimes \mathcal{H}_C)$ is a density matrix, then the CPTP property ensures that $(\mathcal{E} \otimes \mathrm{id}_C)(\sigma)$ is a valid density matrix from $\mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_C)$.[4] In order to define the Choi-Jamiołkowski isomorphism, we set $C := B'$, $\mathcal{H}_{B'} := \mathcal{H}_B$ and choose an orthonormal basis $\{|\phi_b\rangle \colon b \in \{1 \dots d_B\}\}$ of $\mathcal{H}_B$. Given $\mathcal{E} \in \mathcal{B}(\mathcal{B}(\mathcal{H}_B); \mathcal{B}(\mathcal{H}_A))$, the Choi matrix $\rho_{\mathcal{E}} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is given by

$$\rho_{\mathcal{E}} := (\mathcal{E} \otimes \mathrm{id}_{B'})|\Psi\rangle\langle\Psi| \tag{5.1}$$

where $|\Psi\rangle = \sum_{b=1}^{d_B} |\phi_b \phi_b\rangle / \sqrt{d_B} \in \mathcal{H}_B \otimes \mathcal{H}'_B$ is a completely entangled state. The bijective, linear map from $\mathcal{E}$ to its Choi matrix $\rho_{\mathcal{E}}$ is called the Choi-Jamiołkowski isomorphism.[5] The map $\mathcal{E}$ is completely positive (or trace-preserving) if and only if $\rho_{\mathcal{E}}$ is positive semi-definite (or $\mathrm{Tr}_A(\rho_{\mathcal{E}}) = \mathbb{1}_{d_B}/d_B$ holds).[6] Since the linear map from $\mathcal{E}$ to $\rho_{\mathcal{E}}$ is bijective, determining $\rho_{\mathcal{E}}$ via QST provides a full description of the quantum operation $\mathcal{E}$ and this approach to quantum process tomography is known as ancilla-assisted process tomography (AAPT).[7] The "ancilla" system is the second copy B' of system B in (5.1).

In order to examine four equivalent forms of the Choi-Jamiołkowski isomorphism, we choose a basis $\{|\epsilon_a\rangle \colon a \in [d_A]\}$ of $\mathcal{H}_A$. Equation (5.1) is equivalent to

$$\rho_{\mathcal{E}} = \frac{1}{d_B} \sum_{b,b'=1}^{d_B} \mathcal{E}\Big(|\phi_b\rangle\langle\phi_{b'}|\Big) \otimes |\phi_b\rangle\langle\phi_{b'}| \tag{5.2}$$

and taking components of the last Equation shows that it is equivalent to

$$\langle\epsilon_a \phi_b|\, \rho_{\mathcal{E}}\, |\epsilon_{a'} \phi_{b'}\rangle = \frac{1}{d_B} \langle\epsilon_a|\, \mathcal{E}\Big(|\phi_b\rangle\langle\phi_{b'}|\Big)\, |\epsilon_{a'}\rangle \tag{5.3}$$

for all $a, a' \in [d_A]$, $b, b' \in [d_B]$. The last Equation is in turn equivalent to

$$\mathrm{Tr}\left(\left(|\epsilon_{a'}\rangle\langle\epsilon_a| \otimes |\phi_{b'}\rangle\langle\phi_b|\right)\, \rho_{\mathcal{E}}\right) = \frac{1}{d_B}\, \mathrm{Tr}\left(|\epsilon_{a'}\rangle\langle\epsilon_a|\, \mathcal{E}\Big(|\phi_b\rangle\langle\phi_{b'}|\Big)\right). \tag{5.4}$$

---

[4] E.g. Nielsen and Chuang 2007, Section 8.2. Specifically, complete positivity (CP) implies that $(\mathcal{E} \otimes \mathrm{id}_C)(\rho)$ is positive semidefinite for all positive semidefinite $\rho \in \mathcal{B}(\mathcal{H}_B \otimes \mathcal{H}_C)$ and $\mathcal{E}$ is called trace preserving if $\mathrm{Tr}(\mathcal{E}(\rho)) = \mathrm{Tr}(\rho)$ holds for all $\rho \in \mathcal{B}(\mathcal{H}_B)$.

[5] Choi 1975; Jamiołkowski 1972.

[6] TP: Jamiołkowski 1972, Theorem 2. CP: Choi 1975, Theorem 2.

[7] D'Ariano and Lo Presti 2001; Dür and Cirac 2001; Altepeter et al. 2003. In addition to AAPT, there are other methods for reconstructing unitary channels from a smaller number of input and output states (Baldwin et al. 2014).

Using linearity shows that last Equation for all $a, a' \in [d_A]$, $b, b' \in [d_B]$ is equivalent to

$$\operatorname{Tr}\big((A \otimes B)\ \rho_{\mathcal{E}}\big) = \frac{1}{d_B} \operatorname{Tr}(A\ \mathcal{E}(B^\top)) \tag{5.5}$$

for all $A \in \mathcal{B}(\mathcal{H}_A)$, $B \in \mathcal{B}(\mathcal{H}_B)$.[8] The transpose is defined in terms of the basis $|\phi_b\rangle$, i.e. $\langle \phi_b | B^\top | \phi_{b'} \rangle = \langle \phi_{b'} | B | \phi_b \rangle$.

In the following, we consider n particles, i.e. $\mathcal{H}_B = \mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_n$ and $\mathcal{H}_A = \mathcal{H}_{A_1} \otimes \dots \otimes \mathcal{H}_{A_n}$. We also choose product bases $|\phi_b\rangle = \big|\phi^{(1)}_{b_1}\big\rangle \otimes \dots \otimes \big|\phi^{(n)}_{b_n}\big\rangle$ and $|\epsilon_a\rangle = \big|\epsilon^{(1)}_{a_1}\big\rangle \otimes \dots \big|\epsilon^{(n)}_{a_n}\big\rangle$ where $b = (b_1, \dots, b_n)$ and $a = (a_1, \dots, a_n)$. Dimensions are denoted by $d_k := \dim(\mathcal{H}_k)$, $d = \max_k d_k$ and $d_{A_k} := \dim(\mathcal{H}_{A_k})$. The maximally entangled state $|\Psi\rangle$ is given by

$$|\Psi\rangle = \frac{1}{\sqrt{d_B}} \sum_{b_1=1}^{d_1} \cdots \sum_{b_n=1}^{d_n} \Big(\big|\phi^{(1)}_{b_1}\big\rangle \otimes \dots \otimes \big|\phi^{(n)}_{b_n}\big\rangle\Big) \otimes \Big(\big|\phi^{(1)}_{b_1}\big\rangle \otimes \dots \otimes \big|\phi^{(n)}_{b_n}\big\rangle\Big). \tag{5.6}$$

$|\Psi\rangle$ is an element of $\mathcal{H}_B \otimes \mathcal{H}_{B'} = \mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_n \otimes \mathcal{H}_{1'} \otimes \dots \otimes \mathcal{H}_{n'}$ where k' is a copy of k, $\mathcal{H}_{k'} := \mathcal{H}_k$. It is easy to verify that the bond dimension of an MPS representation of this state is maximal at every bipartition; i.e. $|\Psi\rangle$ from (5.6) does not admit an efficient MPS representation. As a consequence, the matrix $\rho_{\mathcal{E}}$ corresponding to the identity channel $\mathcal{E} = \mathrm{id}$ does not admit an efficient MPO representation and efficient MPS tomography of $\rho_{\mathcal{E}}$ cannot succeed. This changes if we switch the order of the $2n$ sites from $(1, \dots, n, 1', \dots, n')$ to $(1, 1', 2, 2', \dots, n, n')$. In this case, the state $|\Psi\rangle$ becomes

$$|\Psi\rangle = |\Phi^+_1\rangle_{1,1'} \otimes \dots \otimes |\Phi^+_n\rangle_{n,n'} \quad \in \quad \mathcal{H}_1 \otimes \mathcal{H}_{1'} \otimes \dots \otimes \mathcal{H}_n \otimes \mathcal{H}_{n'} \tag{5.7}$$

where $|\Phi^+_k\rangle_{k,k'} = \sum_{i=1}^{d_k} \big|\phi^{(k)}_i \phi^{(k)}_i\big\rangle_{k,k'} / \sqrt{d_k}$ is a completely entangled state on $\mathcal{H}_k \otimes \mathcal{H}_{k'}$. The map $\mathcal{E}$ acts on sites $1, 2, \dots, n$ of this chain of $2n$ sites, i.e. on every second site, as illustrated in Figure 5.1. Clearly, $|\Psi\rangle$ from (5.7) admits an MPS representation whose bond dimensions do not exceed $d = \max_k d_k$. As a consequence, at least the Choi matrix $\rho_{\mathcal{E}}$ of the identity channel, $\mathcal{E} = \mathrm{id}$, now admits an efficient MPO representation. In this way, the ancilla systems can enable an efficient representation of $\mathcal{E}$. However, they may represent a significant challenge in an experiment. In the following, we show how measurements can be carried out without the ancilla system and we discuss possible efficient representations of quantum operations based on their Choi matrix.

---

[8] Equation (5.5) appears e.g. in Flammia et al. 2012.

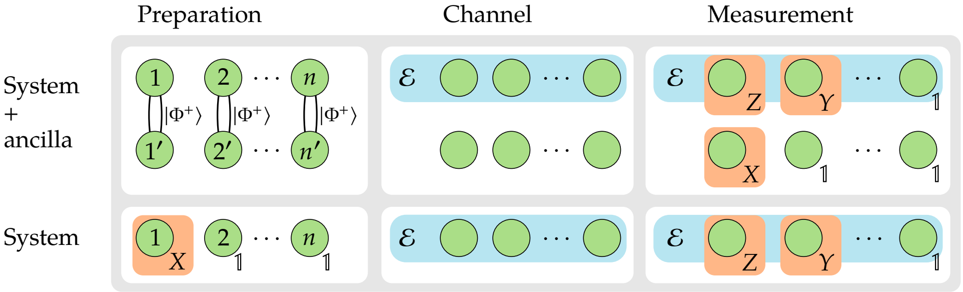


Figure 5.1: Scalable ancilla-assisted process tomography (AAPT)[9] uses tensor product measurements on few neighbouring sites of the linear chain $(1, 1', 2, 2', \dots, n, n')$ where system and ancilla sites are arranged alternatingly (top row). The quantum operation $\mathcal{E}$ acts on the sites $(1, 2, \dots, n)$. The need for the ancilla system is eliminated by preparing a product state, applying $\mathcal{E}$ and performing a local tensor product measurement (bottom row; details in the main text).

Expectation values in $\rho_\mathcal{E}$ can be obtained without the ancilla system, which is shown by (5.5), and we set out to show that this also applies to many-body systems. A general tensor product observable of the linear chain after $\mathcal{E}$ has been applied assumes the form $M = A_1 \otimes B_{1'} \otimes \cdots \otimes A_n \otimes B_{n'}$ with $A_k \in \mathcal{H}_{A_k}$ and $B_{k'} \in \mathcal{H}_{k'} = \mathcal{H}_k$. Efficient QST of $\rho_\mathcal{E}$ typically uses expectation values of tensor product observables where all but few neighbouring tensor product factors are proportional to the identity matrix (Figure 5.1 top row). We assume that $\sigma = B_{1'} \otimes \cdots \otimes B_{n'}$ is a valid quantum state (if this is not the case, it can be achieved by considering eigendecompositions of the $B_{k'}$). The expectation value of M in $\rho_\mathcal{E}$ is related to the quantum operation $\mathcal{E}$ via (use (5.5))

$$\begin{aligned} & \mathrm{Tr}[(A_1 \otimes B_{1'} \otimes \cdots \otimes A_n \otimes B_{n'})\rho_\mathcal{E}] \\ = \; & \mathrm{Tr}[(A_1 \otimes \cdots \otimes A_n)\mathcal{E}(B_{1'}^\top \otimes \cdots \otimes B_{n'}^\top)]/d_B \end{aligned} \tag{5.8}$$

where the transpose $B_{k'}^\top$ is defined in terms of the basis $|\phi_{b_k}^{(k)}\rangle$. For efficient QST, we prepare the state $\sigma^\top$, a tensor product of maximally mixed states except for a small number of consecutive sites, apply $\mathcal{E}$ and perform a tensor product measurement on few consecutive sites of $\mathcal{E}(\sigma^\top)$ (Figure 5.1 bottom row). With this procedure, we can determine the expectation value $\mathrm{Tr}(M\rho_\mathcal{E})$ efficiently and without the need for n ancilla particles.

---

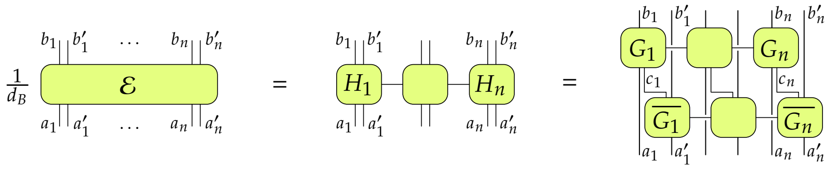
[9] The figure reproduces Figure 1 from Holzäpfel et al. 2015 (© 2015 American Physical Society) with modifications.

After $\rho_{\mathcal{E}}$ has been obtained with quantum state tomography, the quantum operation $\mathcal{E}$ is determined by (combine (5.1) and (5.7), cf. (5.3))

$$\left\langle \epsilon^{(1)}_{a_1}\phi^{(1)}_{b_1}\dots\epsilon^{(n)}_{a_n}\phi^{(n)}_{b_n}\right| \rho_{\mathcal{E}} \left|\epsilon^{(1)}_{a'_1}\phi^{(1)}_{b'_1}\dots\epsilon^{(n)}_{a'_n}\phi^{(n)}_{b'_n}\right\rangle \tag{5.9a}$$

$$= \frac{1}{d_B}\left\langle \epsilon^{(1)}_{a_1}\dots\epsilon^{(n)}_{a_n}\right| \mathcal{E}\Big(\left|\phi^{(1)}_{b_1}\dots\phi^{(n)}_{b_n}\right\rangle\!\left\langle\phi^{(1)}_{b'_1}\dots\phi^{(n)}_{b'_n}\right|\Big) \left|\epsilon^{(1)}_{a'_1}\dots\epsilon^{(n)}_{a'_n}\right\rangle. \tag{5.9b}$$

The density matrix $\rho_{\mathcal{E}}$ can be represented as a matrix product operator (MPO) or as a locally purified matrix product state (PMPS), inducing two different, possibly efficient representations of the quantum operation $\mathcal{E}$. Since positive semidefiniteness of $\rho_{\mathcal{E}}$ is equivalent to complete positivity of $\mathcal{E}$, the quantum operation $\mathcal{E}$ may or may not be completely positive if $\rho_{\mathcal{E}}$ is represented as an MPO. Moreover, complete positivity would be NP-hard to verify because this holds for determining whether an MPO is positive semidefinite.[10] If $\mathcal{E}$ is represented as a PMPS, then $\rho_{\mathcal{E}}$ is guaranteed to be positive semidefinite and this implies that $\mathcal{E}$ is completely positive. The next equation shows the result of inserting MPO and PMPS representations of $\rho_{\mathcal{E}}$ into (5.9):



(5.10)

In the representations, $\mathcal{H}_{A_k} \otimes \mathcal{H}_{k'}$ is treated as a single site. As formula,[11] the MPO representation of $\rho_{\mathcal{E}}$ provides the representation

$$\frac{1}{d_B}\left\langle \epsilon^{(1)}_{a_1}\dots\epsilon^{(n)}_{a_n}\right| \mathcal{E}\Big(\left|\phi^{(1)}_{b_1}\dots\phi^{(n)}_{b_n}\right\rangle\!\left\langle\phi^{(1)}_{b'_1}\dots\phi^{(n)}_{b'_n}\right|\Big) \left|\epsilon^{(1)}_{a'_1}\dots\epsilon^{(n)}_{a'_n}\right\rangle$$
$$= H_1(a_1, b_1, a'_1, b'_1)\dots H_n(a_n, b_n, a'_n, b'_n) \tag{5.11}$$

with $H_k(a_1, b_1, a'_1, b'_1) \in \mathbb{C}^{D_{k-1}\times D_k}$ and $D_0 = D_n = 1$. The PMPS representation of $\rho_{\mathcal{E}}$ provides the representation

$$\frac{1}{d_B}\left\langle \epsilon^{(1)}_{a_1}\dots\epsilon^{(n)}_{a_n}\right| \mathcal{E}\Big(\left|\phi^{(1)}_{b_1}\dots\phi^{(n)}_{b_n}\right\rangle\!\left\langle\phi^{(1)}_{b'_1}\dots\phi^{(n)}_{b'_n}\right|\Big) \left|\epsilon^{(1)}_{a'_1}\dots\epsilon^{(n)}_{a'_n}\right\rangle =$$
$$\sum_{c_1}\dots\sum_{c_n}\Big[G_1(a_1, b_1, c_1)\otimes\overline{G_1(a'_1, b'_1, c'_1)}\Big]\dots\Big[G_1(a_1, b_1, c_1)\otimes\overline{G_1(a'_1, b'_1, c'_1)}\Big]. \tag{5.12}$$

---

[10] Kliesch et al. 2014a. See also Section 2.2.1.

[11] Cf. Equations (2.7) and (2.9).

where $G_k(a_k, b_k, c_k) \in \mathbb{C}^{D_{k-1}\times D_k}$ and $D_0 = D_n = 1$. The indices $c_k \in [d_{C_k}]$ correspond to additional systems with Hilbert spaces $\mathcal{H}_{C_k}$ used to purify the mixed state $\rho_\mathcal{E}$ ($k \in [n]$). Their dimensions d_k are parameters of the representation. As mentioned already, the representation (5.12) guarantees that $\mathcal{E}$ is completely positive and may thus be called evidently positive. This can be an advantage over an MPO-based representation whose complete positivity can be lost if MPS compression or other operations are applied.

The representations are simplified if we restrict to $\mathcal{H}_A = \mathcal{H}_B = \mathcal{H}_\Lambda$ ($\Lambda = [n]$), $\mathcal{H}_{A_k} = \mathcal{H}_k$, $|\epsilon^{(k)}_{a_k}\rangle = |\phi^{(k)}_{a_k}\rangle$ and unitary operations $\mathcal{E}(\,\cdot\,) = U \cdot U^*$. In this case, the Choi matrix $\rho_\mathcal{E}$ is a pure state, $\rho_\mathcal{E} = |\psi_\mathcal{E}\rangle\langle\psi_\mathcal{E}|$, which can be represented as matrix product state:

$$\langle \phi^{(1)}_{a_1}\phi^{(1)}_{b_1}\dots\phi^{(n)}_{a_n}\phi^{(n)}_{b_n}|\psi_\mathcal{E}\rangle = G_1(a_1, b_1)\dots G_n(a_n, b_n) \tag{5.13}$$

where $G_k(a_k, b_k) \in \mathbb{C}^{D_{k-1}\times D_k}$ and $D_0 = D_n = 1$. As before, $\mathcal{H}_{A_k} \otimes \mathcal{H}_{k'}$ has been treated as as single site in the MPS representation. The representation (5.13) is the special case of (5.12) where all the systems C_k have unit dimension, $d_{C_k} = 1$ ($k \in [n]$). The core matrices $G_k(a_k, b_k)$ from (5.13) provide an MPO representation of U because we can derive

$$\langle \phi^{(1)}_{a_1}\dots\phi^{(n)}_{a_n}|\, U\, |\phi^{(1)}_{b_1}\dots\phi^{(n)}_{b_n}\rangle \;=\; \sqrt{d_B}\,\langle \phi^{(1)}_{a_1}\phi^{(1)}_{b_1}\dots\phi^{(n)}_{a_n}\phi^{(n)}_{b_n}|\psi_\mathcal{E}\rangle \tag{5.14}$$

in the same way as (5.9).

5.2. Hamiltonian reconstruction

In this section, we discuss a method to determine the time-independent Hamiltonian which generated a unitary evolution $U = \mathrm{e}^{-\mathrm{i}Ht}$. An MPO representation of H can be determined efficiently with this method if U is given as an MPO as well. To obtain H from the unitary U, we use the identity

$$x = \sin(x)\frac{\arccos(\cos(x))}{\sqrt{1-(\cos(x))^2}}, \qquad\qquad x \in (-\pi, \pi), \tag{5.15}$$

together with the power series

$$\frac{\arccos(z)}{\sqrt{1-z^2}} = \sum_{k=0}^{\infty} c_k (z-1)^k, \qquad\qquad c_k = \frac{(-1)^k}{2^k}\prod_{j=1}^{k}\frac{j}{j+\frac{1}{2}} \tag{5.16}$$

which converges for $|z-1| < 2$.[12] The basic idea is as follows. The relation $U = \mathrm{e}^{-\mathrm{i}Ht}$ implies

$$\sin(Ht) = \frac{1}{2\mathrm{i}}(U^* - U), \qquad \cos(Ht) = \frac{1}{2}(U^* + U). \tag{5.17}$$

With $x = Ht$, Equation (5.15) holds up to times limited by $\|Ht\|_{(\infty)} < \pi$. While this appears to limit the accessible time interval for Hamiltonian reconstruction, Section 5.3.2.2 explains how to extend this result to longer times.

For practical purposes, we only want to evaluate a finite number of terms of the series in Equation (5.16) and enforce Hermiticity. To this end, we use the expression

$$\frac{1}{2}\sin(Ht)\sum_{k=0}^{N-1} c_k(\cos(Ht) - \mathrm{id})^k + \text{h.c.} \tag{5.18}$$

to estimate Ht, with a given N and $\sin(Ht)$ and $\cos(Ht)$ as above. This approach has two advantages over using a power series expansion of the logarithm: It is valid for larger values of $\|Ht\|$ and the series converges more quickly. Note that the time evolution induced by H is just as well described by $U = \mathrm{e}^{\mathrm{i}\phi}\mathrm{e}^{-\mathrm{i}Ht}$ with an arbitrary global phase $\phi \in \mathbb{R}$. With a global phase, Equation (5.15) imposes $\|Ht - \phi\,\mathrm{id}\|_{(\infty)} < \pi$. To remedy this problem, we use $U\,\mathrm{Tr}(U)^*/|\,\mathrm{Tr}(U)|$ instead of U when computing $\cos(Ht)$ and $\sin(Ht)$.

5.3. Numerical simulations

In this section, we describe how scalable AAPT can be simulated numerically and we present results of such numerical simulations. For several exemplary channels $\mathcal{E}$, we numerically obtain $|\psi_{\mathcal{E}}\rangle$ as detailed below. The state $|\psi_{\mathcal{E}}\rangle$ is on $2n$ qubits and we simulate measurements which act on the r neighbouring qubits $(k, k+1, \dots, k+r-1)$ for $k \in \{1 \dots 2n - r + 1\}$. For each r-qubit block, we simulate M repetitions of 3^r different, projective measurements given by all r-fold tensor products of Pauli matrix eigenvectors.[13] The total number of simulated measurements is linear in n. Using the outcomes of the simulated measurements to determine an estimate $|\psi_{\mathcal{E}}^{\text{rec}}\rangle$ of $|\psi_{\mathcal{E}}\rangle$ takes statistical errors from a finite number of measurements into account and their effect is very similar in a scenario without ancilla system.[14] The estimate $|\psi_{\mathcal{E}}^{\text{rec}}\rangle$ is determined by using the singular-value-thresholding-like algorithm proposed by Cramer et al. (2010) to obtain an initial state for the scalable maximum-likelihood algorithm by Baumgratz et al. (2013b).[15] The result is an MPS representation of $|\psi_{\mathcal{E}}^{\text{rec}}\rangle$ with $2n$ sites which is easily converted to an MPS representation on n supersites as in Equation (5.13) (cf. Section 2.2).

---

[12] Derived from Olver et al. 2010, Eq. 4.24.2, available at `https://dlmf.nist.gov/4.24.E2`.

[13] M measurements of $\Pi^{(2n,r,k)}_{\text{Local}}$ for $a \in \{X, Y, Z\}^r$, see Equation (3.19). For MPS-MLE, the measurement is represented with the POVM $\Pi^{(2n,r)}_{\text{Block}}$, see Equation (3.20).

We assume that $|\psi_{\mathcal{E}}\rangle$ and $|\psi_{\mathcal{E}}^{\text{rec}}\rangle$ are normalized. With the estimate $|\psi_{\mathcal{E}}^{\text{rec}}\rangle$ and thus the corresponding operator U_{rec} at hand, we then quantify the quality of the reconstruction scheme by

$$F = F(U, U_{\text{rec}}) = |\langle\psi_{\mathcal{E}}|\psi_{\mathcal{E}}^{\text{rec}}\rangle|^2, \tag{5.19}$$

and note that this is in one-to-one correspondence to other distance measures for unitary channels used in the literature. In particular, we have the relation[16]

$$F = F_A + \frac{F_A - 1}{d_\Lambda}, \quad F_A = \int |\langle\psi|U^* U_{\text{rec}}|\psi\rangle|^2 \mathrm{d}|\psi\rangle \tag{5.20}$$

where $d_\Lambda = \dim(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n)$ and where the average fidelity F_A is defined in terms of the Haar measure.[17] In addition, $F(U, U_{\text{rec}})$ is related to the Frobenius norm distance via $\min_{\alpha\in\mathbb{R}} \|\mathrm{e}^{\mathrm{i}\alpha} U - U_{\text{rec}}\|_{(2)}{}^2 / d_\Lambda = 2(1 - F^{1/2}(U, U_{\text{rec}}))$ because $|\psi_{\mathcal{E}}\rangle$ and $|\psi_{\mathcal{E}}^{\text{rec}}\rangle$ are normalized.

In the case of Hamiltonian reconstruction, we assess our reconstructed estimates as follows: First, note that two Hamiltonians H and $H + \lambda\,\text{id}$, $\lambda \in \mathbb{R}$, are physically indistinguishable. Therefore, we measure relative distances between Hamiltonians with[18]

$$D(H, H') = \frac{\min_{\lambda\in\mathbb{R}} \|H - H' - \lambda\,\text{id}\|_{(\infty)}}{\min_{\lambda\in\mathbb{R}} \|H - \lambda\,\text{id}\|_{(\infty)}}, \tag{5.21}$$

which is independent of energy offsets in both H' and H. We choose the operator norm $\|\cdot\|$ motivated by its property[19]

---

[14] Measurement data with very similar statistics can be obtained without ancilla by ensuring that each initial state (as given by the ancilla part of a POVM element) is prepared as often as it would be obtained (on average) in the scenario with ancilla. The latter is easy to compute because the reduced state of the ancilla sites remains maximally mixed under any CPTP linear map.

[15] See also Chapter 3.

[16] Pedersen et al. 2007.

[17] See also Raginsky 2001; Gilchrist et al. 2005.

[18] The operator norm and its minimization over α can be carried out numerically for Hamiltonians given as MPOs by obtaining the largest and smallest eigenvalue using DMRG methods. See e.g. Schollwöck 2011 or Section 2.2.2.

$$\left|\langle A(t)\rangle_\varrho - \langle A'(t)\rangle_\varrho\right| \leq 2|t|\|H - H'\|_{(\infty)}\|A\|_{(\infty)},$$

where $A(t)$ and $A'(t)$ are the Heisenberg picture time evolutions of $A \in \mathcal{B}(\mathcal{H}_\Lambda)$ according to H and H', respectively. In other words, the operator norm distance defines a timescale on which two Hamiltonians can be considered equivalent.

For all results below, we repeat the whole procedure of sampling from the simulated state, reconstructing it, and assessing the quality of the reconstruction several times. All results shown are mean values of $F(U, U_{\text{rec}})$ and $D(H, H_{\text{rec}})$ over a small number of runs, with fluctuations that are, for the number of measurements per observable considered, smaller than the size of the markers. In the following, we present numerical results for the reconstruction of quantum circuits and Hamiltonians and study the performance as a function of the number of qubits n, the number M of measurements per observable, and the block size r of the subsystems on which measurements are performed. We simulate circuits and Hamiltonians on up to 32 qubits. Hence, reconstructing the unitary uses pure state reconstruction on up to 64 qubits.

5.3.1. Unitary quantum circuits

We demonstrate the feasibility of our scalable tomography scheme by considering the GHZ circuit, which maps the product state $|0\ldots0\rangle$ to an n-qubit GHZ state, and the quantum Fourier transform,[20]

$$\text{GHZ} = \text{CN}_{n-1,n}\text{CN}_{n-2,n-1}\cdots\text{CN}_{1,2}H_1, \tag{5.22a}$$

$$\text{QFT} = \prod_{k=1}^{n}\left[\left(\prod_{j=1}^{n-k}\text{CR}_{k,k+j}(\pi/2^j)\right)H_k\right], \tag{5.22b}$$

where we use the convention $\prod_{j=1}^{k} U_j = U_k \cdots U_1$ for products of non-commuting operators. Here, $H_k = \frac{1}{\sqrt{2}}\left(\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right)$ denotes the Hadamard gate acting on qubit k and $\text{CN}_{i,i+1}$ ($\text{CR}_{i,j}(\phi)$) denotes the two-qubit conditional-NOT (conditional rotation)

---

[19] Use $|\text{Tr}(X\rho)| \leq \|X\|_{(\infty)}\|\rho\|_{(1)}$ (Bhatia 1997, Exercise IV.2.12) and $\|U_t^* A U_t - U_t'^* A U_t'\|_{(\infty)} \leq 2\|U_t - U_t'\|_{(\infty)}\|A\|_{(\infty)}$ (Lemma A.4). The remaining $\|U_t - U_t'\|_{(\infty)} \leq |t|\|H - H'\|_{(\infty)}$ follows from $\partial_t U_t = -\mathrm{i}HU_t$, $\partial_t U_t' = -\mathrm{i}H'U_t$ and e.g. Osborne 2006, Eq. (11), which is also Equation (6.50) below.

[20] E.g. Nielsen and Chuang 2007, Ch. 5.

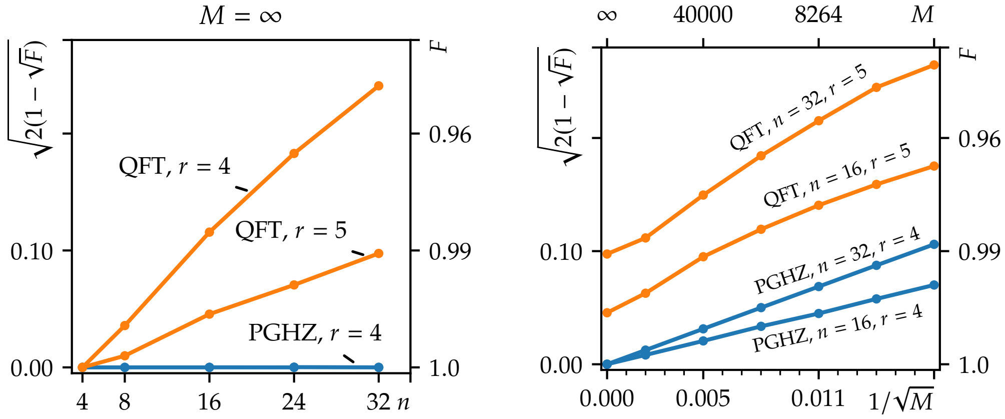


Figure 5.2: Reconstruction error of unitary circuits.[21]

gate,[22] given by

$$\mathrm{CN} = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes \sigma_X, \tag{5.23a}$$

$$\mathrm{CR}(\phi) = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes \Big(|0\rangle\langle 0| + \mathrm{e}^{\mathrm{i}\phi}|1\rangle\langle 1|\Big). \tag{5.23b}$$

where σ_X is a Pauli matrix (cf. Equation (3.9)). A simple counting argument shows that the GHZ circuit admits an exact efficient MPO representation. This counting argument can by stated in terms of the number of gates which act on each qubit.[23] A slightly improved bound on the bond dimension of an MPO representation can be obtained if the number of gates which concern each bond is counted.[24]

Omitting small conditional rotations in the quantum Fourier transform[25] provides an approximation which can be simulated classically.[26] Let QFT_c be the circuit obtained by omitting all conditional rotations with $j > c$ from Equation (5.22). The distance in operator norm satisfies $\|\mathrm{QFT} - \mathrm{QFT}_c\|_{(\infty)} \leq n\pi/2^c$,[27] i.e. $\|\mathrm{QFT} - \mathrm{QFT}_c\|_{(\infty)} \leq \epsilon$ if $2^c \geq n\pi/\epsilon$. One can also show that QFT_c admits an MPO representation with bond dimension $D = n\pi/\epsilon$.[28] In practice, we use numerical

---

[21] The figure reproduces Figure 2 from Holzäpfel et al. 2015 (© 2015 American Physical Society) with modifications.

[22] Nielsen and Chuang 2007.

[23] Jozsa 2006; a generalization to PEPOs is presented in Lemma 2.4.

[24] Holzäpfel et al. 2015.

[25] Coppersmith 1994.

[26] Yoran and Short 2007.

[27] Yoran and Short 2007.

[28] Holzäpfel et al. 2015, Appendix B.

MPO compression techniques[29] and obtain an approximate MPO representation with bond dimension 16 and error bounded by $[2(1-\sqrt{F})]^{1/2} < 2 \times 10^{-5}$ for the $n \le 32$ qubits we consider. Note that the upper bound

$$2(1-\sqrt{F(U,U')}) \le \| |\psi_U\rangle - |\psi_{U'}\rangle \|_2^2 = \frac{{\|U-U'\|_{(2)}}^2}{2^n} \le {\|U-U'\|_{(\infty)}}^2 \tag{5.24}$$

holds. Below, we use this bound for $U = \text{QFT}$ and $U' = \text{QFT}_c$.

The reconstruction results are summarized in Figure 5.2. The reconstruction of the gate GHZ performs very well. To discuss the performance of the quantum Fourier transform reconstruction, we note that the distance $\epsilon_c = [2(1-\sqrt{F})]^{1/2}$ between the exact quantum Fourier transform and its approximation QFT_c is upper bounded by $\epsilon_c \le n\pi/2^c$ (Equation (5.24)). We reconstruct the circuit QFT with high fidelity $F \approx 0.99$ from measurements on $r = 5$ consecutive qubits on the combined system + ancilla (Figure 5.2). Naively, one would expect to be able to reconstruct QFT_c only for $c \le 2$, because $r = 5$ corresponds to information about three neighbouring system qubits only. However, the upper bound on the approximation error is trivial for $c = 2$ and $8 \le n \le 32$, and numerical tests show that the numerical approximation error ϵ_c (for $c = 2$) is indeed several times larger than the reconstruction error $[2(1-\sqrt{F})]^{1/2}$ we achieve. This shows that there are non-local gates which can be reconstructed without using the corresponding non-local information.

5.3.2. Hamiltonian reconstruction

5.3.2.1. Short times

We simulate the time evolution U of time-independent local one-dimensional Hamiltonians H with well-established numerical DMRG/MPO algorithms.[30] If H is a Hamiltonian with finite range in one spatial dimension and the evolution time is fixed, then U admits an efficient, approximate MPO representation.[31] After obtaining the estimate U_{rec} of the time evolution, we determine an estimate H_{rec} of the Hamiltonian that governs the time evolution by the series given in Equation (5.18) with $N = 3$. With this, and the assumption[32] that $\min_{\lambda \in \mathbb{R}} \|H - \lambda \,\mathrm{id}\|_{(\infty)} = \|H\|_{(\infty)}$, one has

---

[29] Schollwöck 2011, see also Section 2.2.2.

[30] The time evolution $U = \mathrm{e}^{-\mathrm{i}Ht}$ of a local one-dimensional Hamiltonian can be obtained up to a certain time scale with DMRG/MPO methods (Östlund and Rommer 1995; Vidal 2004; Schollwöck 2011). We obtained U up to $t_{\max} = \frac{1}{\eta}$ from a second-order Trotter expansion of $\mathrm{e}^{-\mathrm{i}Ht}$ with 2000 Trotter steps and MPO bond dimension 128; with $\eta = \frac{1}{n-1}\sum_{k=1}^{n-1} \|h_{k,k+1}\|_{(\infty)}$, $H = \sum_{k=1}^{n-1} h_{k,k+1}$.

[31] Osborne 2006. This result is extended to higher spatial dimensions in Chapter 6. A related topic are area laws (Eisert et al. 2010) for the von Neumann and Rényi entropies (Eisert and Osborne 2006; Bravyi et al. 2006). The relation between area laws and the existence of efficient approximate representations has been analyzed by Schuch et al. (2008).

$$D(H, H_{\mathrm{rec}}) \leq \tfrac{\|H-H_{\mathrm{rec}}\|_{(\infty)}}{\|H\|_{(\infty)}} \leq \tfrac{1}{140} \|Ht\|_{(\infty)}{}^{6} + O(\|Ht\|_{(\infty)}{}^{8}). \tag{5.25}$$

Figure 5.3 shows results for three different nearest-neighbour Hamiltonians on $n \leq 32$ qubits, the isotropic Heisenberg Hamiltonian, the critical Ising Hamiltonian and a Hamiltonian with random nearest-neighbour interaction. We use $t_n = 1/\|H\|_{(\infty)}$ as a time unit. For the models considered here, we have $t_n \sim 1/n$.

The distance $D(H, H_{\mathrm{rec}})$ between the reconstructed and the exact Hamiltonian shown in Figure 5.3 displays the following features: First, the reconstruction is expected to fail for $\|Ht\|_{(\infty)} \geq \pi$ (see Equation (5.15)) and, indeed, $D(H, H_{\mathrm{rec}})$ is large in this area (indicated by the grey background). Secondly, Equation (5.25) suggests that close to $\|Ht\|_{(\infty)} = \pi$, $D(H, H_{\mathrm{rec}})$ should scale as $\|Ht\|_{(\infty)}{}^{6}/140 = t^6/(140 t_n^6)$ (thick grey line in Figure 5.3). Thirdly, we observe that for infinitely many measurements per observable, $M = \infty$, and fixed $t/t_n = \|H\|_{(\infty)} t \sim nt$, the distance $D(H, H_{\mathrm{rec}})$ decreases with system size, a behaviour inherited from the quality of the reconstruction $|\psi_{\mathcal{E}}^{\mathrm{rec}}\rangle$ (see left of Figure 5.3): The fidelity $F(U, U_{\mathrm{rec}})$ is limited by the amount of block entanglement in $|\psi_{\mathcal{E}}\rangle$. At a fixed time t, an area law[34] holds for this entanglement such that it is bounded even for arbitrarily large systems.[35] For sufficiently large systems, we hence expect $F(U, U_{\mathrm{rec}})$ at fixed t to be independent of the system size n. If we keep $t/t_n \sim t \cdot n$ fixed, we therefore expect F to increase with n. Finally, let us discuss the dependence of the distance between the exact and the reconstructed Hamiltonian in Figure 5.3 on the number M of measurements per observable. First of all, with a finite number of measurements no reconstruction will be possible at small times, because the signal of the Hamiltonian in $U \approx \mathbb{1} - \mathrm{i}Ht$ will be smaller than the noise. Furthermore, the data suggests that, for times before $t/t_n \approx \pi$,

$$D(H, H_{\mathrm{rec}}) \propto \frac{1}{t/t_n} \frac{n}{\sqrt{M}}. \tag{5.26}$$

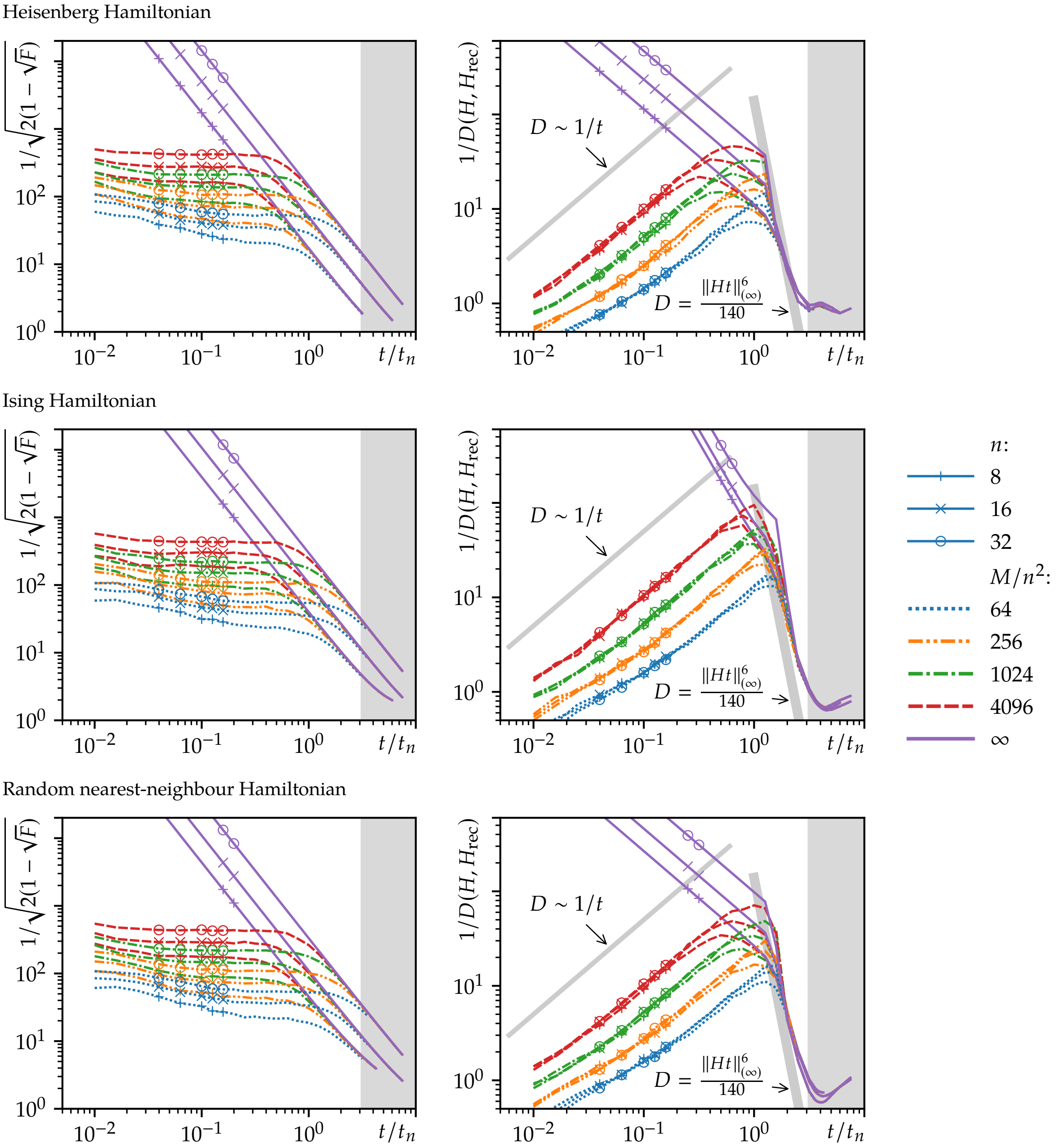
---

[32] If the assumption is violated, we get an additional time-independent prefactor.

[33] The figure reproduces Figures 3 and 5 from Holzäpfel et al. 2015 (© 2015 American Physical Society) with modifications.

[34] Eisert et al. 2010.

[35] Eisert and Osborne 2006; Bravyi et al. 2006.



Heisenberg Hamiltonian

Ising Hamiltonian

Random nearest-neighbour Hamiltonian

Figure 5.3: Reconstruction quality (inverse reconstruction error) of local Hamiltonians. The right (left) column shows the reconstruction error of the Hamiltonian H (induced unitary evolution $U = \mathrm{e}^{-\mathrm{i}Ht}$).[33]

This is the behaviour one would expect if one assumes that the relative error $D(H, H_{\text{rec}})$ is proportional to the ratio R/S of a noise amplitude R and the strength of the signal $S = \|Ht\| = t/t_n$, in which $R = n/\sqrt{M}$ is motivated by the fact that we have measured $\propto n$ observables, each of which has been estimated to within a standard deviation which is, for sufficiently large M, proportional to $1/\sqrt{M}$. The explanation of the scaling properties together with the fact that all properties described are the same in the three different local Hamiltonians investigated suggests that the scaling laws apply for many local Hamiltonians on a linear chain, for a large range of system sizes and any sufficiently large (as indicated by the examples) number of measurements.

To summarize, measuring at larger times gives a larger signal and a smaller error, but we are limited by the condition $t/t_n < \pi$ imposed by Equation (5.15). Solving Equation (5.26) for the number of measurements per observable, we obtain $M \propto n^2 (t/t_n)^2/D^2$: A constant relative error D at a fixed $t/t_n < \pi$ requires $M \propto n^2$ measurements per observable, resulting in a total number of measurements proportional to n^3 .

5.3.2.2. Long times

In its present formulation, the reconstruction scheme is limited to $t/t_n < \pi$, a restriction that may be overcome by measuring at two different times t and t': The times up to which the fidelity $F(U, U_{\text{rec}})$ is sufficiently high is only limited by r – increasing r will increase the time up to which full information about U may be obtained by measuring on r consecutive qubits. In fact, as can be seen on the left of Figure 5.3, for $n = 32$ and the relatively small $r = 3$, the fidelity $F(U, U_{\text{rec}})$ is still quite high at $t/t_n = \pi$ while the reconstruction of H fails for these times. Measuring at t, t' and obtaining $U = \mathrm{e}^{-\mathrm{i}Ht}$, $U' = \mathrm{e}^{-\mathrm{i}Ht'}$ by reconstruction, we are only limited by $|t' - t| < \pi t_n$ when reconstructing H from $U^* U' = \mathrm{e}^{\mathrm{i}H(t-t')}$.

Figure 5.4 shows results of this reconstruction scheme with $t/t_n = 3.51$ and $t' > t$. Reconstructing the Hamiltonian from $U(t' - t) = U^* U'$, the time difference $t' - t$ clearly assumes the role of the time t when reconstructing the Hamiltonian from U at time t alone. Therefore, all scaling properties carry over as long as U_{rec} and U'_{rec} can be obtained with sufficiently high fidelity. We simulated measurements on blocks of $r = 5$ consecutive sites to satisfy this requirement. For exact measurements, $M = \infty$, the relative error $D(H, H_{\text{rec}})$ does not approach zero for $t' - t \to 0$. The reason is that the error in $U_{\text{rec}}(t' - t)$ remains non-zero as $t' - t \to 0$ because $U_{\text{rec}}(t)$ has a fixed non-zero error at fixed t. This non-zero error may also become larger

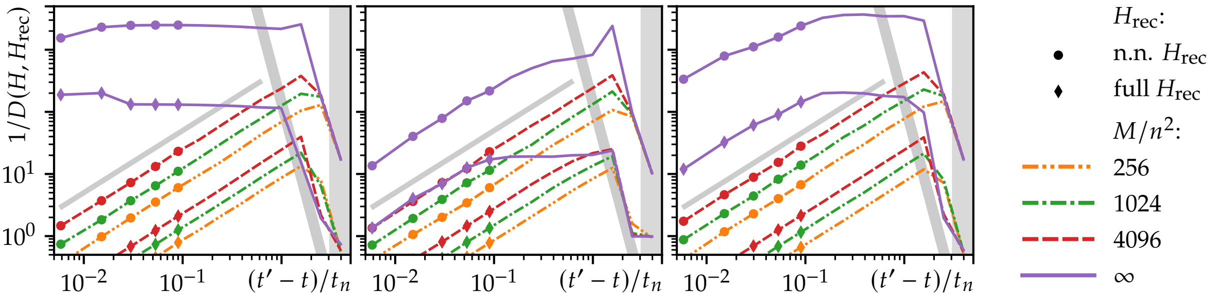


Figure 5.4: Reconstruction quality (inverse reconstruction error) of local Hamiltonians reconstructed from unitary evolution at two times t and t'.[36]

than the signal amplitude $\|Ht\|$, explaining the increasing error as $t' - t \longrightarrow 0$ for some of the Hamiltonians.

Note that, from $U(t' - t)$, we can also reconstruct Hamiltonians that are time-dependent for times before t and nearly constant between t and t'. In this way, stroboscopic reconstructions of a time-dependent Hamiltonian may be obtained after large propagation times. Furthermore, $t/t_n < \pi$ becomes more restrictive as n increases, thus the usefulness of taking measurements at two times increases for larger systems.

5.3.2.3. Enforcing a local reconstruction

Of course, making use of additional information can only improve the scheme. As an example, suppose that we know that the Hamiltonian is nearest-neighbour only. One may then project the reconstructed H_{rec} onto a nearest-neighbour Hamiltonian. As can be seen in Figure 5.4 this reduces the reconstruction error.

---

[36] The figure reproduces Figure 4 from Holzäpfel et al. 2015 (© 2015 American Physical Society) with modifications.

Chapter 6.

Efficient verification and simulation of local time evolution

6.1. Summary

In this chapter, we discuss efficient simulation and certification of the dynamics induced by a quantum many-body Hamiltonian H with short-ranged interactions. Here, we extend prior results for one-dimensional systems[1] to lattices in arbitrary spatial dimensions.[2] We consider $n < \infty$ lattice sites and Hamiltonians whose interactions have a strictly finite range.

We present a method which can certify the fact that an unknown quantum system evolves according to a certain Hamiltonian. Suppose that the evolution time grows at most poly-logarithmically[3] with n. We prove that the necessary measurement effort scales quasi-polynomially in the number of particles n. It also scales quasi-polynomially in the inverse tolerable error $1/\mathcal{I}$. In addition, we show that a projected entangled pair state (PEPS) representation of a time-evolved state can be obtained efficiently in the following sense. Suppose that the evolution time t grows at most poly-logarithmically with n. We prove that the necessary computation time and the PEPS bond dimension of the representation scale quasi-polynomially in the number of particles n and the inverse approximation error $1/\epsilon$.

For certification of a time-evolved state, we consider an initial product state $|\psi(0)\rangle$,

---

[1] Osborne 2006; Lanyon, Maier, et al. 2017.

[2] This chapter reproduces the preprint Holzäpfel and Plenio 2017 with the following changes: The formulation of the main results in Section 6.3, Theorem 6.10 and Lemma 6.15, as well as Lemma 6.5 was improved. A missing factor η was added in Equation (6.31). A missing exponential function was added in Corollary 6.20. Definition 6.25 was added to improve the presentation.

[3] Recall the definitions from Section 1.2.

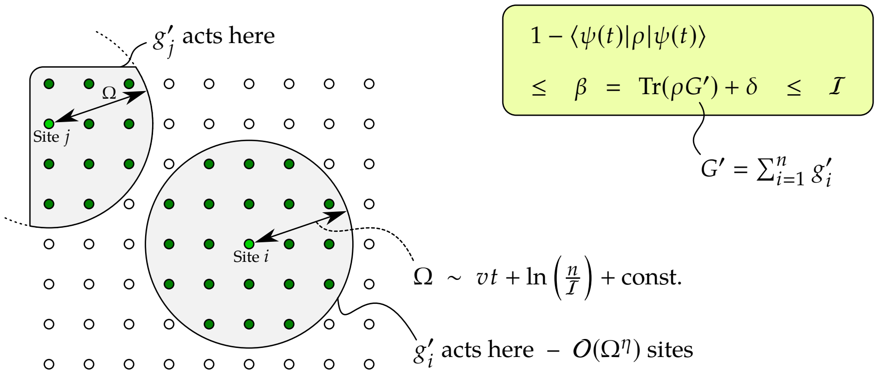

g'_j acts here
Ω
Site j
Site i
$1 - \langle\psi(t)|\rho|\psi(t)\rangle$
$\leq \beta = \mathrm{Tr}(\rho G') + \delta \leq \mathcal{I}$
$G' = \sum_{i=1}^{n} g'_i$
$\Omega \sim vt + \ln\left(\frac{n}{\mathcal{I}}\right) + \text{const.}$
g'_i acts here – $O(\Omega^\eta)$ sites

Figure 6.1: Certifying time-evolved states: The local terms g'_i of $G' = \sum_i g'_i$ act on regions whose diameter is proportional to Ω, i.e. on $O(\Omega^\eta)$ sites if the lattice has η dimensions. The Lieb–Robinson velocity v determines the growth of Ω with time. The expectation value $\mathrm{Tr}(\rho G')$, which provides an upper bound on the distance $I(\rho, |\psi(t)\rangle) = 1 - \langle\psi(t)|\rho|\psi(t)\rangle$, can be determined from complete measurements on n regions of size $O(\Omega^\eta)$ (Theorem 6.10 and Equation (6.29)).

the time-evolved state $|\psi(t)\rangle$ and an unknown state ρ. We measure the distance between a pure and a mixed state by the infidelity

$$I(\rho, |\psi\rangle) := 1 - \langle\psi|\rho|\psi\rangle. \tag{6.1}$$

In order to *verify* or *certify* that the unknown state ρ is almost equal to the time-evolved state $|\psi(t)\rangle$, we provide an upper bound β on the infidelity of the two states, i.e.

$$I(\rho, |\psi(t)\rangle) \leq \beta. \tag{6.2}$$

We prove that the bound β can be obtained from the expectation values of complete sets of observables on regions whose diameter is proportional to some distance Ω (Figure 6.1). If the unknown state ρ is exactly equal to the time-evolved state $|\psi(t)\rangle$, then a bound β which is no larger than a tolerable error $\mathcal{I}$ can be obtained if Ω grows linearly with $\log(n/\mathcal{I})$ and if it also grows linearly with the evolution time t. If we assume a spatial dimension $\eta \geq 1$, a region of diameter $\sim \Omega$ contains $\sim \Omega^\eta$ sites. Since there are n regions of diameter Ω and since $O(\exp(c\Omega^\eta))$ observables are sufficient for a complete set on a single region, the total measurement effort is $O(n \exp(c \log(n)^\eta)) = O(n^{1+(c(\log n)^{\eta-1})})$, i.e. it increases quasi-polynomially with n. This scaling reduces to polynomial in n if the system is one-dimensional ($\eta = 1$). In addition, we show that the upper bound β increases only slightly if ρ has a finite distance from $|\psi(t)\rangle$ or if the bound is obtained from expectation values which are not known exactly, e.g. due to a finite number of measurements per observable.

Suppose that the Hamiltonian is a nearest-neighbour Hamiltonian in one spatial dimension and that the evolution time t grows at most logarithmically with the number of particles n. In this case, an approximate matrix product state (MPS) representation of the time-evolved state $|\psi(t)\rangle$ can be obtained efficiently, i.e. the computational time grows at most polynomially with n/ϵ where ϵ is the approximation error.[4] PEPSs are a generalization of MPSs to higher spatial dimensions. It has been demonstrated that MPS-based numerical algorithms for computing time evolution can also be applied to PEPSs.[5] However, the computational time required by these algorithms has not been determined in general. Here, we show that an approximate PEPS representation of the time-evolved state $|\psi(t)\rangle$ can be obtained efficiently for poly-logarithmic times (in n). Suppose that the evolution time t grows at most poly-logarithmically with n (i.e. $t \sim (\log n)^c$). We prove that the necessary computational time and the PEPS bond dimension of the representation scale quasi-polynomially in the number of particles n and the inverse approximation error $1/\epsilon$. Furthermore, we show that there is an efficient projected entangled pair operator (PEPO) representation of the unitary evolution generated by the Hamiltonian. This representation is structured in a way which guarantees efficient computation of expectation values of single-site observables in $|\psi(t)\rangle$, an operation which can be computationally difficult for a general PEPS.[6]

In Section 6.2, existing results on Lieb–Robinson bounds are mentioned and some corollaries are derived. In Section 6.3, parent Hamiltonians and their use as fidelity witnesses are introduced.[7] Parent Hamiltonians are then used to efficiently certify time-evolved states. In Section 6.4, we construct efficient representations of a unitary time evolution operator U_t. The first subsection discusses the Trotter decomposition. The remaining two subsections construct an efficient representation of U_t, first for an arbitrary lattice and and then for a hypercubic lattice. In the special case of a hypercubic lattice, a representation with improved properties is achieved. Section 6.5 concludes.

---

[4] Osborne 2006.

[5] Murg et al. 2007; Verstraete et al. 2008.

[6] Computing the expectation value of a single-site observable in an arbitrary PEPS has been shown to be #P-complete and it is widely assumed that a polynomial-time solution for such problems does not exist (Schuch et al. 2007).

[7] Cramer et al. 2010.

6.2. Lieb–Robinson bounds

Suppose that H is a nearest-neighbour Hamiltonian on a lattice. The time evolution of an observable A under a Hamiltonian H is given by $\tau_t^H(A) = \mathrm{e}^{\mathrm{i}Ht} A \mathrm{e}^{-\mathrm{i}Ht}$ (assuming that H is time-independent). Even if A acts non-trivially only on a small part of the system, $\tau_t^H(A)$ acts on the full system for any $t > 0$ because the exponential functions contain arbitrarily large powers of H. (We shall assume that no part of the system is decoupled from the rest.) However, $\tau_t^H(A)$ can be approximated by an observable which acts non-trivially on a small region around the original A. The approximation error is exponentially small in the diameter of the region and the error remains constant if the diameter increases linearly with time (see also Figure 6.2 on Page 94). In this sense, information propagates at a finite velocity in a quantum lattice system. A *Lieb–Robinson bound* is an upper bound on the norm of the commutator $[\tau_t^H(A), B]$ and provides a means to bound the error of the named approximation. The first bound on the commutator $[\tau_t^H(A), B]$ has been given by Lieb and Robinson (1972) for a regular lattice. More recently, these bounds have been extended to lattices described by means of graphs or metric spaces.[8] For interactions which decay exponentially (polynomially) with distance, Lieb–Robinson bounds have been proved which are exponentially (polynomially) small in distance;[9] here, the distance is between the regions on which A and B act non-trivially.

Basic notation is defined in Chapter 1. The time-evolved observable $\tau_t^H(A)$ can be approximated by $\tau_t^{H'}(A)$ where the Hamiltonian H' contains only the interaction terms which act on a given region R of the system and this has been proven for a one-dimensional system by Osborne (2006). An explicit bound on the approximation error $\|\tau_t^H(A) - \tau_t^{H'}(A)\|_{(\infty)}$ for a lattice with a metric has been given by Barthel and Kliesch (2012) for the case of a local Liouvillian evolution.[10] Their result is limited to interactions with a strictly finite range but this restriction also enables an explicit definition of all constants. In the remainder of this section, we introduce their result and derive corollaries used below.

The time evolution from time s to time t under a time-dependent Hamiltonian $H(t)$ is described by the unitary $U_{ts} = [U_{st}]^*$ given by the unique solution of $\partial_t U_{ts} = -\mathrm{i}H(t)U_{ts}$ where $U_{ss} = \mathbb{1}$, $s, t \in \mathbb{R}$ and $H(t)$ is assumed to be continuous except for finitely many discontinuities in any finite interval. The unitary satisfies

---

[8] Nachtergaele and Sims 2006; Hastings and Koma 2006; Nachtergaele et al. 2006.

[9] Hastings and Koma 2006.

[10] See also Kliesch et al. 2014b for an introduction.

$\partial_s U_{ts} = +\mathrm{i} U_{ts} H(s)$ and, if H is time-independent, it is given by $U_{ts} = \exp(-\mathrm{i}H(t-s))$. To distinguish time evolutions under different Hamiltonians, we use the notation $U^H_{ts} = U_{ts}$. The time evolution of a pure state $|\psi(s)\rangle$ and a density matrix $\rho(s)$ are given by $|\psi(t)\rangle = U_{ts}|\psi(s)\rangle$ and $\rho(t) = U_{ts}\,\rho(s)\,U_{st} =: \tau^H_{ts}(\rho(s))$. If we omit the second time argument s, it is equal to zero: $U_t := U_{t0}$ and $\tau^H_t := \tau^H_{t0}$.

We consider a system of $n < \infty$ sites and Λ denotes the set of all sites. Associated to each site $x \in \Lambda$, there is a Hilbert space $\mathcal{H}_x$ of finite dimension $d(x) \geq 2$. In this chapter, quantum states ρ, ψ generally are density matrices (not state vectors), $\rho, \psi \in \mathcal{D}(\mathcal{H}_\Lambda)$. We assume that there is a metric $d(x, y)$ on Λ. The diameter of a set $X \subset \Lambda$ is given by $\mathrm{diam}(X) = \max_{x,y\in X} d(x, y)$. Distances between sets are given by $d(x, Y) = \min_{y\in Y} d(x, y)$ and $d(X, Y) = \min_{x\in X, y\in Y} d(x, y)$ where $X, Y \subset \Lambda$. The Hamiltonians H_V and H of a subsystem $V \subset \Lambda$ and of the whole system, respectively, are given by

$$H_V := \sum_{Z\subset V} h_Z, \qquad\qquad H := H_\Lambda. \tag{6.3}$$

The local terms $h_Z(t)$ can be time-dependent but we often omit the time argument. At a given time, each local term $h_Z(t)$ is either zero or acts non-trivially at most on Z. In the following, we define several properties of the Hamiltonian, including the interaction strength J, the interaction range a, the maximal number $\mathcal{Y}$ of sites per interaction term, the maximal number $\mathcal{Z}$ of nearest neighbours and the lattice's spatial dimension $\eta := \kappa + 1$. These parameters are constants which are independent of the evolution time t and the number of sites n. The interaction strength and range are given by

$$J := 2 \sup_{t, Z\subset\Lambda} \|h_Z(t)\|_{(\infty)}, \qquad\qquad a := \sup_{Z\colon h_Z\neq 0} \mathrm{diam}(Z). \tag{6.4}$$

Terms which act non-trivially only on a single site, which may unduly enlarge the maximal norm J, can be eliminated from our discussion by employing a suitable interaction picture as described in Appendix A.1. The maximal number of nearest neighbours is given by

$$\mathcal{Z} := \max_{Z\colon h_Z\neq 0} |\{Z' \subset \Lambda\colon h_{Z'} \neq 0, Z' \cap Z \neq \emptyset\}|. \tag{6.5}$$

This restricts the number of local terms in the Hamiltonian to $|\{Z \subset \Lambda\colon h_Z \neq 0\}| \leq \mathcal{Z}n = O(n)$.[11] The number of local terms at a certain distance r is given by the

---

[11] Write $\{Z \subset \Lambda\colon h_Z \neq 0\} = \bigcup_{x\in\Lambda}\{Z \subset \Lambda\colon h_Z \neq 0, x \in Z\}$.

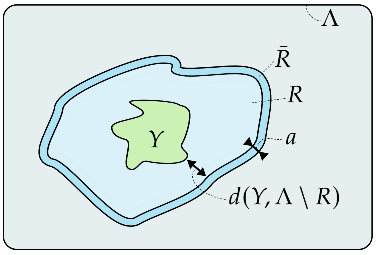


Figure 6.2: An observable A at time zero acts non-trivially on Y and its time evolution $\tau_t^H(A)$ is approximated by $\tau_t^{H_{\bar{R}}}(A)$ on $\bar{R}$. The approximation error is determined by the evolution time t and the distance $d(Y, \Lambda \setminus R)$ (Theorem 6.1; Barthel and Kliesch 2012).

number of elements in the set

$$R_{r,y} := \left\{ Z \subset \Lambda \colon h_Z \neq 0, d(y, Z)/a \in [r, r+1) \right\} \tag{6.6}$$

and we assume that it is bounded by a power law:

$$|R_{r,y}| \leq M r^{\kappa} \quad \forall\, y \in \Lambda, r \in \{0, 1, 2, \dots\}, \tag{6.7}$$

where M and κ are constants. A regular lattice in an Euclidean space of dimension η satisfies this condition with $\kappa = \eta - 1$. Equation (6.7) restricts the number of local terms $h_Z \neq 0$ within a certain distance in terms of the metric but the number of sites on which a local term may act remains unbounded. We demand that this number of sites is bounded by a finite

$$\mathcal{Y} := \sup_{Z \colon h_Z \neq 0} |Z|. \tag{6.8}$$

We assume that for each $x \in \Lambda$, there is a $Z \subset \Lambda$ with $x \in Z$ and $h_Z \neq 0$. Together with Equations (6.7) and (6.8), this assumption implies that $|B_r^{o,c}(\{x\})| = O(r^{\eta})$ where $\eta = \kappa + 1$, $x \in \Lambda$, $B_r^o(X) := \{y \in \Lambda \colon d(X, y) < r\}$, $B_r^c(X) := \{y \in \Lambda \colon d(X, y) \leq r\}$ and $X \subset \Lambda$. The extension of a volume $V \subset \Lambda$ in terms of the Hamiltonian is given by

$$\bar{V} := \bigcup_{\substack{Z \colon h_Z \neq 0, \\ Z \cap V \neq \emptyset}} Z. \tag{6.9}$$

The following theorem has been shown by Barthel and Kliesch (2012, Theorem 2) and they have called it *quasilocality*:

Theorem 6.1 *Let a, $\mathcal{Z}$ and J be finite and $t \in \mathbb{R}$. Let $Y \subset R \subset \Lambda$ and let A act on Y (Figure 6.2). Let $d_a := d(Y, \Lambda \setminus R)/a$ and $\lceil d_a \rceil > 2\kappa + 1$. Then*

$$\|\tau_t^H(A) - \tau_t^{H_{\bar{R}}}(A)\|_{(\infty)} \leq \frac{2M}{\mathcal{Z}} \|A\|_{(\infty)} \lceil d_a \rceil^\kappa \exp(v|t| - \lceil d_a \rceil) \tag{6.10}$$

holds. The Lieb–Robinson velocity is given by $v := J\mathcal{Z} \exp(1)$.

The upper bound from Equation (6.10) can be simplified as $x^\kappa \exp(-x) \leq \exp(-(1-q)x)$ holds for any $q \in (0,1)$ if x is large enough. The following lemma provides a precise formulation of this fact and Corollary 6.3 applies it to Equation (6.10).

Lemma 6.2 *Choose $q \in (0,1)$ and set $D := (1-q)d_a$ with $d_a \geq 0$. We say that $\lceil d_a \rceil$ is* large enough *if it satisfies $\lceil d_a \rceil > 2\kappa + 1$ and $\lceil d_a \rceil \geq \frac{2\kappa}{q} \ln(\frac{\kappa}{q})$; let $\lceil d_a \rceil$ be large enough.*[12] *Set $\alpha_q := \exp[-(1-q)(\lceil d_a \rceil - d_a)]$. Then $\frac{1}{\mathrm{e}} < \frac{1}{\exp(1-q)} < \alpha_q \leq 1$ and $\lceil d_a \rceil^\kappa \exp(-\lceil d_a \rceil) \leq \alpha_q \exp(-D)$ hold.*

Proof We have $\alpha_q = \exp(D - (1-q)\lceil d_a \rceil)$. Because $\lceil d_a \rceil$ was assumed to be large enough, we can use Lemma A.5 to obtain $\lceil d_a \rceil^\kappa \exp(-\lceil d_a \rceil) \leq \exp(-(1-q)\lceil d_a \rceil) = \alpha_q \exp(-D)$. This completes the proof. ∎

We simplify the bound from Equation (6.10) by applying Lemma 6.2:

Corollary 6.3 *Let a, $\mathcal{Z}$ and J be finite and $t \in \mathbb{R}$. Let $Y \subset R \subset \Lambda$ and let A act on Y. Choose $q \in (0,1)$ and set $D := (1-q)d_a$ where $d_a := d(Y, \Lambda \setminus R)/a$ and where $\lceil d_a \rceil$ is large enough (Lemma 6.2). Then*

$$\|\tau_t^H(A) - \tau_t^{H_{\bar{R}}}(A)\|_{(\infty)} \leq \frac{2M\alpha_q}{\mathcal{Z}} \|A\|_{(\infty)} \exp(v|t| - D) \tag{6.11}$$

holds. The Lieb–Robinson velocity is given by $v = J\mathcal{Z} \exp(1)$ and $\alpha_q \in (\exp(-(1-q)), 1]$. Specifically, $\alpha_q = \exp(-(1-q)(\lceil d_a \rceil - d_a))$.

The upper bound from Equation (6.11) is at most ϵ if D is large enough:

Corollary 6.4 *Let a, $\mathcal{Z}$ and J be finite and $t \in \mathbb{R}$. Let $Y \subset R \subset \Lambda$ and let A act on Y. Choose $q \in (0,1)$ and set $D = (1-q)d_a$ where $d_a = d(Y, \Lambda \setminus R)/a$ and where $\lceil d_a \rceil$ is large enough (Lemma 6.2). If D satisfies*

$$D \geq v|t| + \ln\left(\frac{1}{\epsilon}\right) + \ln\left(\|A\|_{(\infty)}\right) + c_1, \qquad c_1 := \ln(2M/\mathcal{Z}), \tag{6.12}$$

---

[12] The slightly simpler conditions $D \geq 2\kappa + 1$ and $D \geq \frac{2\kappa}{q} \ln(\frac{\kappa}{q})$ are stricter and can also be used.

then[13]

$$\|\tau_t^H(A) - \tau_t^{H_{\bar{R}}}(A)\|_{(\infty)} \le \alpha_q \epsilon \le \epsilon. \tag{6.13}$$

The Lieb–Robinson velocity is given by $v = J\mathcal{Z}\exp(1)$. *Refer to Corollary 6.3 for* α_q.

Corollary 6.4 states that the time evolution $A(t)$ of a local observable $A(0)$ acting on Y can be approximated by another local observable $A'(t)$ which acts on a certain region $\bar{R}$ around Y. This is possible with high accuracy if the region $\bar{R}$ is large enough. Suppose that G is a sum of time-evolved local observables and G' is obtained by taking the sum of corresponding approximated observables. The next lemma compares the expectation value $\mathrm{Tr}(\rho G')$ of the approximated observable G' with the expectation values $\mathrm{Tr}(\rho G)$ and $\mathrm{Tr}(\psi G)$ where the quantum state ψ has small distance from ρ (in trace norm).

Lemma 6.5 *Let* $g_i(0)$ *be observables with* $\|g_i(0)\|_{(\infty)} \le 1$ *which act non-trivially on* Y_i, $Y_i \subset R_i \subset \Lambda$. *Choose a fixed time* $t \in \mathbb{R}$, *let* $G = \sum_{i=1}^{\Gamma} g_i(t)$ *and let* G' *be the sum of* $g_i'(t) = \tau_t^{H_{\bar{R}_i}}(g_i(0))$. *Let* ψ *and* ρ *be quantum states. Choose* $\mathcal{I}$ *and* γ *such that* $0 \le \Gamma\gamma < \mathcal{I}$. *Choose* $q \in (0,1)$ *and set* $D := (1-q)d_a$ *where* $d_a := \frac{1}{a}\max_i d(Y_i, \Lambda \setminus R_i)$. *Let* $\lceil d_a \rceil$ *be large enough (Lemma 6.2), let* D *satisfy*

$$D \ge v|t| + \ln\left(\frac{2\Gamma}{\mathcal{I} - \Gamma\gamma}\right) + c_1 \tag{6.14}$$

and set $\delta := \frac{1}{2}(\mathcal{I} - \Gamma\gamma)$. *Then*

$$\mathrm{Tr}(\rho G) \le \mathrm{Tr}(\rho G') + \delta \le \mathrm{Tr}(\psi G) + \mathcal{I} \tag{6.15}$$

where the second inequality holds if $\|\rho - \psi\|_{(1)} \le \gamma$.

Proof Set $\epsilon := \frac{\delta}{\Gamma} = \frac{1}{2\Gamma}(\mathcal{I} - \Gamma\gamma)$. Applying Corollary 6.4 provides

$$\|G - G'\|_{(\infty)} \le \sum_{i=1}^{\Gamma} \|g_i(t) - g_i'(t)\|_{(\infty)} \le \Gamma\epsilon = \delta. \tag{6.16}$$

Using the bound[14] $|\mathrm{Tr}(\rho(G - G'))| \le \|\rho\|_{(1)}\|G - G'\|_{(\infty)}$ provides

$$\mathrm{Tr}(\rho G) \le \mathrm{Tr}(\rho G') + \delta \le \mathrm{Tr}(\rho G) + 2\delta. \tag{6.17}$$

Using $|\mathrm{Tr}\,(\rho - \psi)G| \le \|\rho - \psi\|_{(1)}\|G\|_{(\infty)}$ provides

$$\mathrm{Tr}(\rho G) - \mathrm{Tr}(\psi G) \le \|\rho - \psi\|_{(1)}\|G\|_{(\infty)} \le \gamma\Gamma. \tag{6.18}$$

Inserting (6.18) into (6.17) completes the proof. ■

---

[13] This holds for all D which satisfy (6.12); it holds e.g. if D is equal to the lower bound stated in (6.12).

[14] Bhatia 1997, Exercise IV.2.12.

6.3. Efficient verification of local time evolution

An observable G is called a parent Hamiltonian of a pure state $|\psi\rangle$ if $|\psi\rangle$ is a ground state of G (i.e. an eigenvector of G's smallest eigenvalue). If such a ground state is non-degenerate, the expectation value $\mathrm{Tr}(\rho G)$ in an arbitrary state ρ provides a lower bound on the fidelity of ρ and the ground state $|\psi\rangle$:[15]

Lemma 6.6 *Let G be an observable with the two smallest eigenvalues E_0 and $E_1 > E_0$. Let $|\psi\rangle$ be an eigenvector of the smallest eigenvalue E_0 and let E_0 be non-degenerate. Let ρ be some quantum state. Then,*

$$1 - \langle\psi|\rho|\psi\rangle \leq \beta := \frac{E_\rho - E_0}{E_1 - E_0} \tag{6.19}$$

where $E_\rho = \mathrm{Tr}(\rho G)$.[16] *The value of the right hand side is bounded by*

$$\beta = \frac{E_\rho - E_0}{E_1 - E_0} \leq \frac{\big\|\rho - |\psi\rangle\langle\psi|\big\|_{(1)} \big\|G\big\|_{(\infty)}}{E_1 - E_0}. \tag{6.20}$$

Proof Proofs of Equation (6.19) have been given by Cramer et al. (2010) and Baumgratz (2014). Equation (6.20) follows from[17]

$$\mathrm{Tr}(\rho G) - E_0 = |\mathrm{Tr}([\rho - \psi]G)| \leq \|\rho - \psi\|_{(1)} \|G\|_{(\infty)} \tag{6.21}$$

where $\psi := |\psi\rangle\langle\psi|$, which completes the proof. ∎

Remark 6.7 Suppose that the expectation value $E_\rho = \mathrm{Tr}(\rho G)$ is not exactly known e.g. because it has been estimated from a finite number of measurements. The resulting uncertainty about the value of β is given by the uncertainty about E_ρ multiplied by the inverse of the energy gap $\Delta = E_1 - E_0$ above the ground state. For robust certification, this energy gap must be sufficiently large.

Suppose that ρ is the unknown quantum state of some experiment which attempts to prepare the state $|\psi\rangle$. If the experiment succeeds, ρ will be close to the ideal state $|\psi\rangle$ (e.g. in trace distance) but the two states will not be equal. The maximal value of the infidelity upper bound β from Equation (6.19) is provided by (6.20). In the worst case, β is given by the trace distance of ρ and $|\psi\rangle$, multiplied by the ratio of the Hamiltonian's largest eigenvalue and its energy gap Δ.

---

[15] Cramer et al. 2010.

[16] Cramer et al. 2010.

[17] Bhatia 1997, Exercise IV.2.12.

In a typical application, the expectation value E_ρ is not exactly known and the states ρ and $|\psi\rangle$ are not exactly equal. In order to obtain a useful certificate, it is necessary that both the energy gap Δ is sufficiently large and that the largest eigenvalue $\|G\|_{(\infty)}$ is sufficiently small. □

The following simple Lemma shows that pure product states admit a parent Hamiltonian that has unit gap and only single-site local terms. This result is a simple special case of prior work involving matrix product states.[18]

Lemma 6.8 *Let* $|\phi\rangle = |\phi_1\rangle \otimes |\phi_2\rangle \otimes \dots \otimes |\phi_n\rangle$ *be a product state on* n *systems of dimension* $d_i \geq 2$, $\langle\phi_i|\phi_i\rangle = 1$, $i \in \{1 \dots n\}$. *Define*

$$G = \sum_{i=1}^{n} h_i, \quad h_i = \mathbb{1}_{1,\dots,i-1} \otimes P_{\ker(\rho_i)} \otimes \mathbb{1}_{i+1,\dots,n} \tag{6.22}$$

where $P_{\ker(\rho_i)} = \mathbb{1} - |\phi_i\rangle\langle\phi_i|$ *is the orthogonal projection onto the null space of the reduced density operator* $\rho_i = |\phi_i\rangle\langle\phi_i|$ *of* $|\phi\rangle$ *on site* i. *The eigenvalues of* G *are given by* $\{0, 1, 2, \dots, n\}$, *the smallest eigenvalue zero is non-degenerate and* $|\phi\rangle$ *is an eigenvector of eigenvalue zero.*

Proof Let $|\mu^{(k)}_{i_k}\rangle$ ($i_k \in \{1, \dots, d_i\}$) an orthonormal basis of system k with $|\mu^{(k)}_1\rangle = |\phi_k\rangle$ ($k \in \{1, \dots, n\}$). The product basis constructed from these bases is an eigenbasis of H:

$$G|\mu\rangle = \lambda|\mu\rangle, \quad |\mu\rangle = |\mu^{(1)}_{i_1}\rangle \otimes \dots \otimes |\mu^{(n)}_{i_n}\rangle, \quad \lambda = |\{k \in \{1 \dots n\} : i_k > 1\}|.$$

As we required $d_i \geq 2$, the eigenvalues of H are given by $\{0, 1, 2, \dots, n\}$. We also see that the smallest eigenvalue zero is non-degenerate and that $|\phi\rangle$ is an eigenvector of eigenvalue zero. This completes the proof. ■

In Lemma 6.8, a parent Hamiltonian G is constructed from projectors onto null spaces of single-site reduced density matrices. One projection is required for each of the n sites and this determines the value of the operator norm $\|G\|_{(\infty)} = n$. In Lemma 6.6, a smaller operator norm was seen to be advantageous for robust certification. By projecting onto null spaces of multi-site reduced density matrices, the following Lemma obtains a parent Hamiltonian with smaller operator norm. More importantly, it also provides a parent Hamiltonian for the time-evolved state $|\psi(t)\rangle$.

---

[18] Perez-Garcia et al. 2007; Cramer et al. 2010; Baumgratz 2014. See also Section 3.4.1.

Lemma 6.9 *Let* $|\psi(0)\rangle = |\phi_1\rangle \otimes \cdots \otimes |\phi_n\rangle$ *be a product state on the lattice* Λ. *Let* $Y_1, \dots, Y_\Gamma$ *a partition of the set of sites* Λ. *For a subset* $Y \subset \Lambda$, *define* $|\phi_Y\rangle = \bigotimes_{k \in Y} |\phi_k\rangle$. *Set* $g_i(0) = (\mathbb{1} - |\phi_{Y_i}\rangle\langle\phi_{Y_i}|) \otimes \mathbb{1}_{\Lambda \setminus Y_i}$. *Choose a fixed time* $t \in \mathbb{R}$ *and let* $g_i(t) = \tau_t^H(g_i(0))$ *and* $G = \sum_{i=1}^{\Gamma} g_i(t)$.

The time-evolved state $|\psi(t)\rangle = U_{t0}|\psi(0)\rangle$ *is an eigenvector of G's non-degenerate eigenvalue zero and the eigenvalues of G are given by* $\{0, 1, \dots, \Gamma\}$.

Proof Let $G_0 = \sum_{i=1}^{\Gamma} g_i(0)$. G_0's eigenvalues are given by $\{0, \dots, \Gamma\}$ and $|\psi(0)\rangle$ is a non-degenerate eigenvector of G_0's eigenvalue zero (Lemma 6.8; group sites into supersites as specified by the sets Y_i). The operators G and G_0 are related by the unitary transformation $G = U_{t0} G_0 U_{0t}$, which implies that they have the same eigenvalues including degeneracies and also that $G|\psi(t)\rangle = 0$. This completes the proof. ■

The parent Hamiltonian of $|\psi(t)\rangle$ from the last Lemma is not directly useful for certification because it is a sum of terms $g_i(t)$ which all act on the full system (for $t \neq 0$). However, these terms can be approximated by terms which act on smaller regions, as described in the next theorem. The theorem is illustrated in Figure 6.1 on Page 90 for $Y_i = \{i\}$ $(i \in \Lambda)$.

Theorem 6.10 *Consider the setting of Lemma 6.9 which includes a fixed time* $t \in \mathbb{R}$. *Choose sets* R_i *such that* $Y_i \subset R_i \subset \Lambda$ *and let* G' *be the sum of* $g'_i(t) = \tau_t^{H_{\bar{R}_i}}(g_i(0))$. *Choose* $\mathcal{I} > 0$ *and* $\gamma \geq 0$ *such that*

$$\delta := \frac{\mathcal{I} - \Gamma\gamma}{2} > 0. \tag{6.23}$$

Set $D := (1-q)d_a$ *where* $d_a := \frac{1}{a}\min_i d(Y_i, \Lambda \setminus R_i)$ *and* $q \in (0,1)$. *Let* $\lceil d_a \rceil$ *be large enough (Lemma 6.2), let D satisfy*

$$D \geq v|t| + \ln\left(\frac{2\Gamma}{\mathcal{I} - \Gamma\gamma}\right) + c_1, \qquad c_1 = \ln(2M/\mathcal{Z}), \tag{6.24}$$

let ρ *be a quantum state and set* $I(\rho, \psi(t)) = 1 - \langle\psi(t)|\rho|\psi(t)\rangle$. *Then*

$$I(\rho, \psi(t)) \leq \operatorname{Tr}(\rho G') + \delta \leq \mathcal{I} \tag{6.25}$$

where the second inequality holds if $\big\| \rho - |\psi(t)\rangle\langle\psi(t)| \big\|_{(1)} \leq \gamma$ *is satisfied.*

Proof Using Lemma 6.6, the properties of G from Lemma 6.9 imply that

$$1 - \langle\psi(t)|\rho|\psi(t)\rangle \leq \operatorname{Tr}(\rho G). \tag{6.26}$$

Inserting $G|\psi(t)\rangle = 0$, Lemma 6.5 completes the proof. ■

The next lemma simplifies the premise of Theorem 6.10 by eliminating Γ:

Lemma 6.11 *Let $f(D)$ a function with $1 \leq f(D) \leq \frac{n}{\Gamma}$. Assuming $n\gamma < \mathcal{I}$, the inequality*

$$D + \ln(f(D)) \geq v|t| + \ln\left(\frac{2n}{\mathcal{I} - n\gamma}\right) + c_1 \tag{6.27}$$

is sufficient for (6.24).

Proof The premise implies $\mathcal{I} - n\gamma \leq \mathcal{I} - \Gamma\gamma$ and

$$D \geq v|t| + \ln\left(\frac{2\Gamma}{\mathcal{I} - n\gamma}\right) + c_1 \geq v|t| + \ln\left(\frac{2\Gamma}{\mathcal{I} - \Gamma\gamma}\right) + c_1, \tag{6.28}$$

which completes the proof. ■

The following lemma bounds the measurement effort if $\mathrm{Tr}(\rho G')$ is estimated from finitely many measurements:

Lemma 6.12 *Let $R := \max_i |\bar{R}_i|$ be the maximal number of sites on which any of the local terms of G' from Theorem 6.10 act. On each region $\bar{R}_i$, choose an IC POVM (examples are provided in Remark 6.14). Let "one measurement" refer to one outcome of one of the POVMs. The upper bound $\mathrm{Tr}(\rho G') + \delta$ from Equation (6.25) can be estimated with standard error ϵ from $M = O(\exp(R)n^3/\epsilon^2)$ such measurements.*

Proof The individual $\mathrm{Tr}(\rho g_i'(t))$ can be estimated independently by carrying out separate measurements for the estimation of each $\mathrm{Tr}(\rho g_i'(t))$. By the central limit theorem, M' measurements are sufficient to estimate a single $\mathrm{Tr}(\rho g_i'(t))$ with standard error $\epsilon' = c/\sqrt{M'}$. Here, $c \leq \exp(\tilde{c}R) = O(\exp(R))$ where $\tilde{c}$ is a constant. To achieve standard error ϵ for $\mathrm{Tr}(\rho G')$, we set $\epsilon' = \epsilon/n$ and obtain $M' = c^2n^2/\epsilon^2$. As separate measurements for each $g_i'(t)$ were assumed, the total number of measurements is at most $M = nM' = c^2n^3/\epsilon^2$. ■

Remark 6.13 (Discussion of Theorem 6.10) Theorem 6.10 provides a means to verify that an unknown state ρ is close to an ideal time-evolved state $\psi(t)$ with the expectation values of few observables. Specifically, the theorem warrants that the infidelity $I(\rho, \psi(t))$ is at most $\beta = \mathrm{Tr}(\rho G') + \delta$ where G' is a sum of observables which act non-trivially only on small parts of the full system. Furthermore, the theorem guarantees $\beta \leq \mathcal{I}$ and we can choose any desired $\mathcal{I} > 0$. To simplify the discussion, we restrict to $\left\|\rho - |\psi(t)\rangle\langle\psi(t)|\right\|_{(1)} \leq \gamma = \frac{\mathcal{I}}{2n}$. For larger systems or smaller certified infidelities, the unknown state ρ has to be closer to the ideal state $\psi(t)$.

Let $\dot{\bigcup}_{i=1}^{\Gamma} Y_i = \Lambda$ be a partition of Λ with $\operatorname{diam}(Y_i) \le r'$ for some $r' > 0$. Note that $Y_i \subset B^c_{r'}(\{y_i\})$ holds for all $y_i \in Y_i$ (Lemma A.12). Let $r > 0$ and set $R_i = B^o_r(Y_i)$, then $\bar{R}_i \subset B^o_{r+r'+a}(\{y_i\})$ (Lemma A.10). We assume $r' = O(r)$ and obtain $|\bar{R}_i| = O((r + r' + a)^\eta) = O(r^\eta)$ (a is independent of n). Note that $d(Y_i, \Lambda \setminus R_i) \ge r$ (Lemma A.10), i.e. $D \ge (1 - q)r/a$ where $q \in (0, 1)$ is a constant.

A particularly simple partition which works for any lattice is $Y_i = \{i\}$ with $i \in \Lambda = \{1 \dots n\}$, $\Gamma = n$, $r' = 0$ and $|\bar{R}_i| = O(D^\eta)$. We choose D according to Lemma 6.11 using $f(D) = n/\Gamma = 1$:

$$D = v|t| + \ln\left(\frac{4n}{\mathcal{I}}\right) + c_1. \tag{6.29}$$

The length scale D grows linearly in time and logarithmically in $n/\mathcal{I}$. As discussed in Lemma 6.12, the measurement effort to estimate $\operatorname{Tr}(\rho G')$ with standard error ϵ is

$$O(n^3 \exp(D^\eta)/\epsilon^2) = O\left(n^3 \exp\left(\left[v|t| + \ln\left(\frac{4n}{\mathcal{I}}\right) + c_1\right]^\eta\right)/\epsilon^2\right). \tag{6.30}$$

The measurement effort grows exponentially with time but only quasi-polynomially with n and with $\frac{1}{\mathcal{I}}$. For one-dimensional systems, $\eta = 1$, this quasi-polynomial scaling reduces to a polynomial scaling.

Finally, we explore what can be gained by choosing a coarser partition $\Lambda = Y_1 \dot\cup \cdots \dot\cup Y_\Gamma$ of a cubic lattice $\Lambda = \{1 \dots L\}^\eta$ with the metric $d(x, y) = \max_i |x_i - y_i|$.[19] Let $\Omega = \lfloor Da \rfloor \in \{1 \dots L\}$ and $B = \lceil L/\Omega \rceil$. The cubic lattice can be divided into $\Gamma = B^\eta$ smaller cubes of maximal diameter $r' = \Omega = O(D)$ and we still have $|\bar{R}_i| = O((r + r' + a)^\eta) = O(D^\eta)$. We set $f(D) = (\lfloor Da \rfloor/2)^\eta$ which satisfies $f(D) \le n/\Gamma$.[20] Inserting $f(D)$ into Equation (6.27) provides

$$D + \eta \ln(\lfloor Da \rfloor) \ge v|t| + \ln\left(\frac{4n}{\mathcal{I}}\right) + c_1 + \eta \ln(2). \tag{6.31}$$

We have increased the radius of $\bar{R}_i$ from $(D + 1)a$ to about $(2D + 1)a$. However, the last equation shows that it is then already sufficient if D grows slightly less than linearly in the right hand side, i.e. slightly less than mentioned above, as described by the additional logarithmic term. □

Remark 6.14 (Examples of IC POVMs) In this remark, we discuss measurements on a region $\bar{R}_i$ where i is fixed. Recall that a set of operators $\{M_k \colon k\}$ on the Hilbert space $\mathcal{H}_{\bar{R}_i}$ is a POVM if each M_k is positive semidefinite and $\sum_k M_k = \mathbb{1}$.[21] The POVM is IC if the operators M_k span $\mathcal{H}_{\bar{R}_i}$.

---

[19] Cubic lattices are also discussed in more detail in Section 6.4.3.

[20] $\Omega \le L$ implies $L/\lceil \frac{L}{\Omega} \rceil > L/(\frac{L}{\Omega} + 1) = \Omega/(1 + \frac{\Omega}{L}) \ge \Omega/2$, i.e. $n/\Gamma = (L/\lceil \frac{L}{\Omega} \rceil)^\eta > (\Omega/2)^\eta = f(D)$.

Measurement outcomes of an IC POVM on R_i can be obtained in several different ways in an experiment. For example, a measurement of a tensor product observable $A = A_1 \otimes \cdots \otimes A_{|\bar{R}_i|}$ on $\mathcal{H}_{\bar{R}_i}$ returns one of the eigenvalues of A as measurement outcome. Access to measurement outcomes of a set of observables which spans $\mathcal{H}_{\bar{R}_i}$ allows sampling outcomes of an IC POVM on $\bar{R}_i$. Alternatively, one can measure the single-site observables A_j ($j \in \{1 \dots |\bar{R}_i|\}$) in any order or simultaneously. Here, the measurement outcome is given by a vector $(\lambda_1, \dots, \lambda_{|\bar{R}|})$ where λ_j is an eigenvalue of A_j. Access to this type of measurement outcomes of a set of observables which spans $\mathcal{H}_{\bar{R}_i}$ provides another way to sample outcomes of an IC POVM on $\bar{R}_i$. □

Lemma 6.9 provides a parent Hamiltonian G of the time-evolved state $|\psi(t)\rangle$ at a fixed time t. Theorem 6.10 provides an upper bound on the distance between an unknown state and the time-evolved state in terms of G' which is an approximation of G. The next Lemma shows that G' is the parent Hamiltonian of a state $|\psi'\rangle$ which is approximately equal to the time-evolved state. As a consequence, an upper bound on the distance between an unknown state and $|\psi'\rangle$ can also be obtained.

Lemma 6.15 *In the setting of Theorem 6.10, let*

$$\delta' := \frac{\mathcal{I} - \gamma\Gamma}{2(1+\mathcal{I})}, \qquad 0 < \delta' < \frac{1}{2}. \tag{6.32}$$

Let the length D be at least

$$D \geq v|t| + \ln\left(\frac{\Gamma}{\delta'}\right) + c_1. \tag{6.33}$$

The operator G' has a non-degenerate ground state and the difference between its two smallest eigenvalues is at least $E'_1 - E'_0 \geq 1 - 2\delta'$. The ground state $|\psi'\rangle$ of G' satisfies

$$|\langle\psi(t)|\psi'\rangle| \geq 1 - \frac{\delta'}{1-\delta'} \geq 1 - 2\delta', \quad \|\psi(t) - \psi'\|_{(1)} \leq 2\sqrt{\frac{2\delta'}{1-\delta'}} \leq 4\sqrt{\delta'}. \tag{6.34}$$

where $\psi(t) := |\psi(t)\rangle\langle\psi(t)|$ and $\psi' := |\psi'\rangle\langle\psi'|$. Then

$$1 - \langle\psi'|\rho|\psi'\rangle \leq \frac{\mathrm{Tr}(\rho G') + \delta'}{1 - 2\delta'} \leq \mathcal{I} \tag{6.35}$$

where the second inequality holds if $\left\|\rho - |\psi(t)\rangle\langle\psi(t)|\right\|_{(1)} \leq \gamma$.

---

[21] E.g. Nielsen and Chuang 2007. See also Sections 1.4 and 3.1.2.

Proof Set $\epsilon = \frac{\delta'}{\Gamma}$. Applying Corollary 6.4 provides

$$\|G - G'\|_{(\infty)} \leq \sum_{i=1}^{\Gamma} \|g_i(t) - g_i'(t)\|_{(\infty)} \leq \Gamma\epsilon = \delta'. \tag{6.36}$$

All eigenvalues change by at most $\|G - G'\|_{(\infty)} \leq \delta'$.[22] Accordingly, the two smallest eigenvalues of G' satisfy $E_0' \in [-\delta', \delta']$, $E_1' \in [1 - \delta', 1 + \delta']$ and $\delta' < \frac{1}{2}$ ensures that the ground state remains non-degenerate. In addition, we have $E_1' - E_0' \geq 1 - 2\delta$. Lemma 6.6 provides

$$1 - \langle\psi'|\rho|\psi'\rangle \leq \frac{\mathrm{Tr}(\rho G') + \delta'}{1 - 2\delta'}. \tag{6.37}$$

We bound (cf. proof of Lemma 6.5)

$$\mathrm{Tr}(\rho G') \leq \mathrm{Tr}(\rho G) + \delta' \leq \|\rho - \psi(t)\|_{(1)}\|G\|_{(\infty)} + \delta' \leq \gamma\Gamma + \delta'. \tag{6.38}$$

Combining the last two equations provides

$$1 - \langle\psi'|\rho|\psi'\rangle \leq \frac{\mathrm{Tr}(\rho G') + \delta'}{1 - 2\delta'} \leq \frac{\gamma\Gamma + 2\delta'}{1 - 2\delta'} = \frac{(1 + \mathcal{I})\gamma\Gamma + \mathcal{I} - \gamma\Gamma}{1 + \mathcal{I} - \mathcal{I} + \gamma\Gamma} = \mathcal{I} \tag{6.39}$$

where the second inequality depends on $\|\rho - \psi(t)\|_{(1)} \leq \gamma$. To quantify the change in the ground state, we use[23]

$$\|EF\|_{(\infty)} \leq \frac{1}{\Delta}\|G - G'\|_{(\infty)} \tag{6.40}$$

where $E = P_G(S_1)$ and $F = P_{G'}(S_2)$ are projectors onto eigenspaces of G and G' with eigenvalues from S_1 and S_2. The sets S_1 and S_2 must be separated by an annulus or infinite strip of width Δ in the complex plane. We set $S_1 = [1, \Gamma]$, $S_2 = [-\delta', \delta']$ and $\Delta = 1 - \delta'$. We denote by $|\psi\rangle := |\psi(t)\rangle$ and $|\psi'\rangle$ the (normalized) ground states of G and G'. Therefore, we have $E = \mathbb{1} - |\psi\rangle\langle\psi|$, $F = |\psi'\rangle\langle\psi'|$ and

$$\begin{aligned} 1 - |\langle\psi|\psi'\rangle| &= \Big\| |\psi'\rangle\langle\psi'| \Big\|_{(\infty)} - |\langle\psi|\psi'\rangle| \Big\| |\psi\rangle\langle\psi'| \Big\|_{(\infty)} \\ &\leq \Big\| |\psi'\rangle\langle\psi'| - |\psi\rangle\langle\psi|\psi'\rangle\langle\psi'| \Big\|_{(\infty)} \\ &= \|EF\|_{(\infty)} \leq \frac{\delta'}{1 - \delta'} \leq 2\delta' \end{aligned} \tag{6.41}$$

---

[22] Bhatia 1997, Theorem VI.2.1.

[23] Bhatia 1997, Theorem VII.3.1.

where the very last inequality holds for $\delta' \leq \frac{1}{2}$. The change in the ground state is at most $1 - |\langle\psi|\psi'\rangle| \leq \delta'/(1-\delta') \leq 1$ (using $\delta' \leq \frac{1}{2}$). This implies (Lemma A.7)

$$\|\psi - \psi'\|_{(1)} \leq 2\sqrt{\frac{2\delta'}{1-\delta'}} \leq 4\sqrt{\delta'} \tag{6.42}$$

and completes the proof. ■

Remark 6.16 The certificates provided by Theorem 6.10 and Lemma 6.15 differ in that the former certifies the fidelity with the time-evolved state $|\psi\rangle$ while the latter certifies the fidelity with its approximation $|\psi'\rangle$. The value of the infidelity upper bound provided by Lemma 6.15 is slightly larger than that provided by Theorem 6.10, but in the limit $\mathcal{I} \to 0$ both results have the same scaling including all constants. □

6.4. Efficient simulation of local time evolution

In this section, we construct a unitary circuit which approximates the unitary evolution U_t induced by a local Hamiltonian $H(t)$ on n quantum systems; the circuit approximates U_t up to operator norm distance ϵ. For times poly-logarithmic in n, the circuit is seen to admit an efficient PEPS representation. Hence, the circuit shows that U_t can be approximated by an efficient PEPS.

Note that the following line of argument also provides an efficient PEPS representation of U_t. Time evolution under an arbitrary few-body Hamiltonian can be efficiently simulated with a unitary quantum circuit and the Trotter decomposition.[24] This unitary circuit is efficiently encoded as a measurement-based quantum computation (MBQC). In turn, a PEPS of the smallest non-trivial bond dimension two is sufficient to encode an arbitrary MBQC efficiently.[25] The PEPS representation from this construction is efficient but it is supported on a larger lattice than the original Hamiltonian: For example, the lattice grows as $O(nt^2/\epsilon)$ if the first-order Trotter decomposition is used (cf. Section 6.4.1). Application of the Trotter formula also leads to an efficient representation of U_t as a tensor network state (TNS) but the lattice of this construction grows in the same way.[26] In this work, we construct an efficient PEPS representation of U_t which lives on the same lattice as the Hamiltonian and which has another advantageous property: Computing the expectation value of a local observable in an arbitrary PEPS is assumed to be impossible in polynomial time[27] but the unitary circuit from which we construct our PEPS representation always enables efficient computation of such local expectation values. This property is shared e.g. with the class of so-called block sequentially generated states (BSGSs),[28] a subclass of all PEPS, where a state is also represented by a sequence of local unitary operations (albeit arranged differently).

---

[24] Nielsen and Chuang 2007, Chapter 4.7.2.

[25] Schuch et al. 2007.

[26] Hübener et al. 2010.

The limitations of the first-order Trotter decomposition become apparent already in one spatial dimension as discussed in Section 6.4.1. Section 6.4.2 presents an efficient representation of U_t for an arbitrary graph. This representation is non-optimal in that it evolves local observables into observables which seemingly act non-trivially on a region whose diameter grows polynomially with time. Lieb–Robinson bounds already tell us that this diameter should grow only linearly with time (Section 6.2). An improved representation which fulfills this property is presented in Section 6.4.3 for a hypercubic lattice of spatial dimension $\eta \geq 1$. In both Sections, representations in terms of unitary circuits are converted to PEPS representations using the results from Section 2.3.

6.4.1. Properties of the Trotter decomposition

The Trotter decomposition is the key ingredient of many numerical methods for the computation of U_t with MPSs or PEPSs.[29] As discussed above, it also enables various efficient representations of U_t. The following lemma presents the well-known first-order Trotter decomposition:

Lemma 6.17 (Trotter decomposition in 1D) *Let H a time-independent[30] nearest neighbour Hamiltonian on a linear chain of n sites, $H = \sum_{j=1}^{n-1} h_{j,j+1}$. Let the operator norm of the local terms be uniformly bounded, i.e. $\|h_{j,j+1}\|_{(\infty)} \leq J$ ($j \in \{1 \dots n-1\}$). Take H_1 and H_2 to be the sum of the terms with even and odd j, respectively: Set $H_1 := \sum_{j=1}^{\lfloor (n-1)/2 \rfloor} h_{2j,2j+1}$ and $H_2 := \sum_{j=1}^{\lfloor n/2 \rfloor} h_{2j-1,2j}$. The time evolution induced by H is given by $U_t = \mathrm{e}^{-\mathrm{i}Ht}$ and its Trotter approximation is given by $U_t^{(T)} = \left(\mathrm{e}^{-\mathrm{i}H_1\tau}\mathrm{e}^{-\mathrm{i}H_2\tau}\right)^L$ where L is a positive integer and $\tau = t/L$. The approximation error is at most ϵ, i.e.*

$$\|U_t - U_t^{(T)}\|_{(\infty)} \leq \epsilon \tag{6.43}$$

if L is at least $L \geq \tilde{c}t^2 n/(2\epsilon)$ where $\tilde{c} > 0$ is some constant which depends only on J.

---

[27] Schuch et al. 2007.

[28] Bañuls et al. 2008.

[29] E.g. Vidal 2004; Murg et al. 2007; Verstraete et al. 2008 and references given in Schollwöck 2011.

[30] The time-dependent case is discussed e.g. in Poulin et al. 2011.

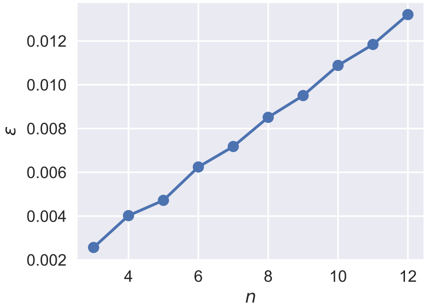


Figure 6.3: Operator norm error $\epsilon = \|U_t - U_t^{(\mathrm{T})}\|_{(\infty)}$ of the first order Trotter decomposition as a function of the number of sites n. The figure shows data for $L = 200$ Trotter steps and the 1D nearest-neighbour isotropic Heisenberg Hamiltonian at $t = \frac{11}{9J}$ where $J = \|h_{i,i+1}\|$ is the operator norm of a coupling term.

Proof For any division of H into $H = H_1 + H_2$ and any $\tau \geq 0$, the following inequality holds:[31]

$$\left\| \mathrm{e}^{-\mathrm{i}H\tau} - \mathrm{e}^{-\mathrm{i}H_1\tau}\mathrm{e}^{-\mathrm{i}H_2\tau} \right\|_{(\infty)} \leq \frac{\tau^2}{2} \left\| [H_1, H_2] \right\|_{(\infty)}. \tag{6.44}$$

Using the triangle inequality (as in Lemma A.4) and $\tau = t/L$, we obtain

$$\left\| \mathrm{e}^{-\mathrm{i}Ht} - \left(\mathrm{e}^{-\mathrm{i}H_1\tau}\mathrm{e}^{-\mathrm{i}H_2\tau} \right)^L \right\|_{(\infty)} \leq \frac{L\tau^2}{2} \left\| [H_1, H_2] \right\|_{(\infty)}. \tag{6.45}$$

It is simple to show that $\|[H_1, H_2]\|_{(\infty)} \leq \tilde{c} n$ holds for some constant $\tilde{c} > 0$ which depends only on J. This provides

$$\|U_t - U_t^{(\mathrm{T})}\|_{(\infty)} \leq \frac{\tilde{c} t^2 n}{2L} \tag{6.46}$$

which completes the proof. ∎

Figure 6.3 shows the approximation error $\epsilon = \|U_t - U_t^{(\mathrm{T})}\|_{(\infty)}$ of a particular Hamiltonian as function of n at fixed t and L. The approximation error appears to grow linearly with n and this suggests that the bound (6.46) is optimal in n up to constants; in this case, the scaling $L \geq \tilde{c} t^2 n / \epsilon$ is optimal in n up to constants as well.

Lemma 6.17 provides an approximate decomposition of U_t into $O(n^2 t^2/\epsilon)$ two-body unitaries and it has been recognized before that this constitutes an approximate, efficient decomposition of U_t by a tensor network on a two-dimensional lattice with $O(n^2t^2/\epsilon)$ sites.[32] However, the lattice of the Hamiltonian is only one-dimensional. The bond dimension of a one-dimensional MPO representation of the circuit $U_t^{(\mathrm{T})}$ can grow exponentially with n.[33] It has been shown that U_t indeed admits a smaller bond dimension[34] but this is not visible from the circuit $U_t^{(\mathrm{T})}$ provided by the Trotter decomposition and needs additional arguments based on Lieb–Robinson bounds. Since the first-order Trotter decomposition does not provide an efficient MPO representation of U_t with H on a one-dimensional lattice, it does not provide an efficient PEPS representation on the same lattice as the Hamiltonian in higher dimensions either.

---

[31] This is De Raedt 1987, Eq. (A.15a). See also Suzuki 1985.

Another important property of representations of the time evolution U_t concerns the growth of the region on which a time-evolved, initially local observable appears to act non-trivially. If an initially local observable A is evolved with the Trotter decomposition $U_t^{(\mathrm{T})}$ into $(U_t^{(\mathrm{T})})^* A U_t^{(\mathrm{T})}$, it appears to act non-trivially on a region of diameter $O(L) = O(nt^2/\epsilon)$. In the following Sections 6.4.2 and 6.4.3, we construct circuits under which this diameter grows only poly-logarithmically with n/ϵ. This is an improvement over the Trotter circuit but it does not reach the ideal case from Corollary 6.4 (no growth with n).

6.4.2. Efficient simulation of time evolution: Arbitrary lattice

Suppose that a local Hamiltonian $H(t)$ is perturbed by a spatially local and possibly time-dependent perturbation $A(t)$. The following lemma states that there is a spatially local unitary V' such that $\|V' U_{ts}^{H-A} - U_{ts}^{H}\|_{(\infty)}$ is small; the lemma has been proven for one-dimensional systems by Osborne (2006). His proof also works for higher-dimensional systems if combined with Theorem 6.1 (proven by Barthel and Kliesch 2012). We pretend to extend the existing proof by accounting for time-dependent Hamiltonians explicitly.

Lemma 6.18 *Let a, $\mathcal{Z}$ and J be finite and $t \in \mathbb{R}$. Let $Y \subset R \subset \Lambda$ and let $A(s)$ act on Y. Let $A(s)$ be continuous except for finitely many discontinuities in any finite interval. Choose $q \in (0,1)$ and set $D := (1-q)d_a$ where $d_a := d(Y, \Lambda \setminus R)/a$. Let $\lceil d_a \rceil$ large enough (Lemma 6.2). Let $V_s'(t)$ on $\bar{R}$ be the solution of $\partial_s V_s'(t) = \mathrm{i} L_t'(s) V_s'(t)$ where*

---

[32] Hübener et al. 2010.

[33] Osborne 2006. This can be seen by applying the counting argument from Jozsa 2006, which was generalized to PEPS in Lemma 2.2.

[34] Osborne 2006.

$L'_t(s) = \tau_{ts}^{H_{\bar{R}}}(A(s))$ *and* $V'_t(t) = \mathbb{1}$ *(*$s \in \mathbb{R}$*). Then*

$$\|V'_s(t)U_{ts}^{H-A} - U_{ts}^{H}\|_{(\infty)} \le \frac{2M\alpha_q}{v\mathcal{Z}}|A| \exp(v|t-s| - D) \tag{6.47}$$

where $|A| = \max_{r\in[s',t']} \|A(r)\|_{(\infty)}$*,* $s' := \min\{s,t\}$ *and* $t' := \max\{s,t\}$*. The Lieb–Robinson velocity is given by* $v = J\mathcal{Z}\exp(1)$ *and* $\alpha_q \in (1/\mathrm{e}^{1-q}, 1]$*. Specifically,* $\alpha_q = \exp(-(1-q)(\lceil d_a \rceil - d_a))$*.*

Proof Let $V_s(t) = U_{ts}^{H}U_{st}^{H-A}$.[35] Due to unitary invariance of the operator norm, we have

$$\|V'_s(t)U_{ts}^{H-A} - U_{ts}\|_{(\infty)} = \|(V'_s(t)U_{ts}^{H-A} - U_{ts})U_{st}^{H-A}\|_{(\infty)} = \|V'_s(t) - V_s(t)\|_{(\infty)}.$$

For fixed $t \in \mathbb{R}$, $V_s(t)$ satisfies the differential equation

$$\partial_s V_s(t) = \mathrm{i}U_{ts}^{H}(H(s) - H(s) + A(s))U_{st}^{H-A} = \mathrm{i}L_t(s)V_s(t) \tag{6.48}$$

where $L_t(s) := U_{ts}^{H}A(s)U_{st}^{H} = \tau_{ts}^{H}(A(s))$ and $V_t(t) = \mathbb{1}$. The Trotter decomposition of V is given by[36]

$$V_t(s) = \lim_{m\to\infty} \prod_{j=1}^{m} \mathrm{e}^{-\mathrm{i}L_t(s+j\delta_m)\delta_m} \tag{6.49}$$

where $\delta_m := (t-s)/m$. The operator norm is unitarily invariant, therefore the triangle inequality implies $\|U_1U_2 - V_1V_2\|_{(\infty)} \le \|U_1 - V_1\|_{(\infty)} + \|U_2 - V_2\|_{(\infty)}$ (Lemma A.4). We obtain

$$\|V_s(t) - V'_s(t)\|_{(\infty)} \le \lim_{m\to\infty} \sum_{j=1}^{m} \left\|\mathrm{e}^{-\mathrm{i}L_t(s+j\delta_m)\delta_m} - \mathrm{e}^{-\mathrm{i}L'_t(s+j\delta_m)\delta_m}\right\|_{(\infty)} \tag{6.50a}$$

$$\le \lim_{m\to\infty} \sum_{j=1}^{m} |\delta_m| \left\|L_t(s+j\delta_m) - L'_t(s+j\delta_m)\right\|_{(\infty)} \tag{6.50b}$$

$$= \int_{s'}^{t'} \left\|L_t(r) - L'_t(r)\right\|_{(\infty)} \mathrm{d}r. \tag{6.50c}$$

---

[35] Alternatively, one can obtain an approximation of the form $U_{ts}^{H} \approx U_{ts}^{H-A}W'_t(s)$ where $W'_t(s)$ is the solution of $\partial_t W'_t(s) = -\mathrm{i}W'_t(s)\tau_{st}^{H_{\bar{R}}}(A_t)$, $W'_s(s) = \mathbb{1}$. $W'_t(s)$ is an approximation of $W_t(s) = U_{st}^{H-A}U_{ts}^{H}$. This approach is a bit more similar to the original proof by Osborne (2006).

[36] See e.g. Theorem 1.1 and 1.2 by Dollard and Friedman (1979b) or Theorem 3.1 and 4.3 by Dollard and Friedman (1979a).

For all $r, r' \in [s', t']$, Corollary 6.3 provides the bound

$$\left\| \tau_{tr}^{H}(A(r')) - \tau_{tr}^{H_{\bar{R}}}(A(r')) \right\|_{(\infty)} \leq \frac{2M\alpha_q}{\mathcal{Z}} \|A(r')\|_{(\infty)} \exp(v|t-r| - D). \tag{6.51}$$

Inserting $r' = r$ provides a bound on $\left\| L_t(r) - L'_t(r) \right\|_{(\infty)}$; inserting this bound into (6.50c) completes the proof. ∎

In the following lemma, we decompose the global evolution U_{ts} into a sequence of local unitaries by removing all local terms of the Hamiltonian which involve site n, then removing those which involve site $n-1$ and so on. Here, the order of the sites does not matter and the geometry of the lattice enters only via the constants defined above. However, the subsequent Theorem 6.21 shows that ordering the sites of the system in a certain way improves the properties of the resulting unitary circuit.

Lemma 6.19 *Let $H_j = \sum_{Z \subset \Lambda_j} h_Z$ denote the sum of all terms which act at most on the first j sites $\Lambda_j = \{1 \dots j\}$. Denote by $Y_j \subset \Lambda_j$ the set of sites on which $F_j := H_j - H_{j-1}$ acts non-trivially. Choose $q \in (0,1)$. Let r be such that $\lceil r \rceil > 2\kappa + 1$ and $\lceil r \rceil \geq \frac{2\kappa}{q} \ln(\frac{\kappa}{q})$. Let $R := (1-q)r$ satisfy*

$$R \geq v|t-s| + \ln\left(\frac{n}{\epsilon}\right) + c_2 \tag{6.52}$$

where $c_2 = \ln\left(\frac{M}{\mathcal{Z}\exp(1)}\right) + 2(1-q)$. Set $R_j := B^o_{da}(Y_j) \cap \Lambda_j$ where $d := r - 2$. Let $\bar{R}_j$ be the extension of R_j in terms of the Hamiltonian H_j, i.e. $\bar{R}_j \subset \Lambda_j$. Then $\bar{R}_j \subset B^o_{ra}(\{j\}) \cap \Lambda_j$. Let $V'_{js}(t)$ on $\bar{R}_j$ be the solution of $\partial_s V'_{js}(t) = \mathrm{i}\tau_{ts}^{G_j}(F_j(s)) V'_{js}(t)$ where $G_j := H_{\bar{R}_j}$ and $V'_{jt}(t) = \mathbb{1}$. Then

$$\|U_{ts}^{H} - V'_n \dots V'_2 V'_1\|_{(\infty)} \leq \epsilon \tag{6.53}$$

holds where $V'_j := V'_{js}(t)$.

Proof There are at most $\mathcal{Z}$ non-zero local terms h_Z with $j \in Z$. As a consequence, $\|F_j(s)\|_{(\infty)} \leq \mathcal{Z}J/2$ holds. In addition, $Y_j \subset B^c_a(\{j\})$ holds and implies $\bar{R}_j \subset B^c_a(R_j) \cap \Lambda_j \subset B^c_a(B^o_{da}(B^c_a(\{j\}))) \cap \Lambda_j \subset B^o_{ra}(\{j\}) \cap \Lambda_j$ (Lemma A.10). The definitions imply that $d(Y_j, \Lambda_j \setminus R_j)/a \geq d$ (Lemma A.10).[37] Set $D = (1-q)d$. Note that $D = R - 2(1-q)$. Therefore, Lemma 6.18 implies that

$$\|U_{ts}^{H_j} - V'_j U_{ts}^{H_j - F_j}\|_{(\infty)} \leq \frac{JM}{v} \exp(v|t-s| - D) \leq \frac{JM}{v} \frac{\epsilon}{n} \exp(2(1-q) - c_2) = \frac{\epsilon}{n} \tag{6.54}$$

---

[37] Note that we restrict to the sublattice Λ_j.

holds for all $j \in \{1 \dots n\}$. Note that we have

$$U_{ts}^{H} - V_n' \dots V_2' V_1' \quad = \quad \sum_{j=1}^{n} V_n' \dots V_{j+1}' U_{st}^{H_j} - V_n' \dots V_j' U_{ts}^{H_{j-1}} \tag{6.55}$$

where $H = H_n$ and $U_{ts}^{H_0} = \mathbb{1}$. The triangle inequality and unitary invariance of the operator norm imply

$$\| U_{ts}^{H} - V_n' \dots V_2' V_1' \|_{(\infty)} \leq \sum_{j=1}^{n} \| U_{ts}^{H_j} - V_j' U_{ts}^{H_{j-1}} \|_{(\infty)} \leq \epsilon \tag{6.56}$$

where we have used $H_{j-1} = H_j - F_j$. This completes the proof of the Lemma. ∎

Corollary 6.20 *Let d be the graph metric*[38] *of the PEPS graph and let t be poly-logarithmic in n. The operator $V = V_n' \dots V_1'$ from Lemma 6.19 provides an efficient, approximate PEPS representation of the time evolution U_{ts}^{H} because it admits a PEPS representation with bond dimension $D = O(\exp(\mathrm{poly}(R)))$.*

Specifically, the bond dimension is $D = d^{2n_r^2}$ where $n_r = \max_{j \in \Lambda} |B_{ar}^{o}(\{j\})|$ is the maximal number of sites in a ball of radius ar (where $d = \max_{x \in \Lambda} d(x)$ is the maximal local dimension).

Proof For a set $Z \subset \Lambda$ which is connected in terms of the PEPS graph, the open ball $B_k^{o}(Z)$ ($k \geq 0$) is also connected in terms of the PEPS graph because d is the graph metric of that graph. The unitary V_j' acts as the identity outside the connected set $B_{ar}^{o}(\{j\})$ which contains at most $n_r = \max_{j \in \Lambda} |B_{ar}^{o}(\{j\})| = \mathrm{poly}(R)$ sites. At most $|B_{ar}^{o}(\{j\})| \leq n_r$ of the n operators $V_1', \dots, V_n'$ act non-trivially on a given, arbitrary site j. Applying Lemma 2.4 with $K = L = n_r$ completes the proof. ∎

Lemma 6.19 provides an efficient, approximate representation of the time evolution U_{ts}^{H}. However, this representation may not be particularly useful: Consider a one-dimensional setting where V_j' acts only on $\{j, j-1\}$ and A is an observable which acts on site n.[39] We want to compute the expectation value $\mathrm{Tr}(\tau_{st}^{H}(A)\rho(s))$ where the initial state $\rho(s)$ is a product state. The time-evolved observable is given by $\tau_{st}^{H}(A) = U_{st}^{H} A U_{ts}^{H}$. We could obtain an approximation from $\tau_{st}^{H}(A) \approx (V_1')^* \dots (V_n')^* A V_n' \dots V_1'$, but the latter operator can act non-trivially on the full system. The structure of the approximation does not convey the fact that operators propagate with the finite Lieb–Robinson velocity, as shown e.g. by Theorem 6.1. The next theorem shows how the representation can be improved by reordering the sites of the system before applying Lemma 6.19.

---

[38] The distance between $x, y \in \Lambda$ in the graph metric is given by the number of edges in a shortest path from x to y.

[39] Indeed, the operators V_j' would need to act on larger numbers of neighbouring sites to achieve a non-zero value of R if the Hamiltonian contains any interactions.

Theorem 6.21 *Choose* $R > 0$, $q \in (0, 1)$ *and set* $r := R/(1-q)$. *Let* $L := \max_{j\in\Lambda} |B^o_{2ar}(\{j\})|$. *There is an efficiently computable colouring function* $C\colon \Lambda \longrightarrow \{1 \dots L\}$ *which has the property that* $C(x) = C(y)$ *implies* $d(x, y)/a \geq 2r$. *Suppose that the sites of the system are ordered such that there are integers* a_k *with* $1 = a_0 \leq a_1 \leq \cdots \leq a_L = n$ *in terms of which the consecutive sites* $\{a_{k-1} + 1 \dots a_k\}$ *have the same colour* $k \in \{1 \dots L\}$. *In this case,* $V = V'_n \dots V'_1$ *from Lemma 6.19 can be expressed as* $V = W_L \dots W_1$ *where* $W_k = V'_{a_k} \dots V'_{a_{k-1}+1}$. V'_j *and* $V'_{j'}$ *do not act non-trivially on the same site if* j *and* j' *have the same colour. At most* L *of the* n *operators* $V'_1, \dots, V'_n$ *act non-trivially on any given site* $j \in \Lambda$.

Proof Consider a graph with sites given by Λ and edges given by $E_C = \{\{x, y\}\colon x, y \in \Lambda, 0 < d(x, y) < 2ar\}$. The number of nearest neighbours (degree) of this graph is $L - 1$. A so-called greedy colouring of the graph (Λ, E_C), which can be computed in $\mathcal{O}(nL)$ time,[40] has the property $d(x, y) < 2ar \Rightarrow C(x) \neq C(y)$. In other words, a greedy colouring already has the necessary property $C(x) = C(y) \Rightarrow d(x, y) \geq 2ar$. Note that $B^o_{ar}(\{j\})$ and $B^o_{ar}(\{j'\})$ have an empty intersection if $d(j, j') \geq 2ar$ (Lemma A.10). Therefore, in this case, at most one of V'_j and $V'_{j'}$ act non-trivially on any site. ∎

Remark 6.22 The operator V from Theorem 6.21 admits a PEPS representation with the bond dimension mentioned in Corollary 6.20.

Note that Theorem 6.21 states that at most $L = \max_{j\in\Lambda} |B^o_{2ar}(\{j\})|$ unitary operations act on a given site while we already know that this number is at most $n_r = \max_{j\in\Lambda} |B^o_{ar}(\{j\})|$ (proof of Corollary 6.20). This difference enables efficient computation of the colouring function which arranges the operations V'_j into L groups of non-overlapping operations.

Let A act non-trivially only on site j. The advantage of Theorem 6.21 over Lemma 6.19 is that $V^* A V$ now acts non-trivially at most on $n_A = |B^o_s(\{j\})|$ sites where $s = 2arL = \text{poly}(R)$, i.e. at most on $n_A = \text{poly}(R)$ sites (use Lemma 6.23 and

---

[40] A greedy colouring is obtained by picking a vertex which has not been assigned a colour and assigning the first colour which has not been assigned to any neighbour of the given vertex (neighbour in terms of E_C). See e.g. Bondy and Murty (2008, Sec. 14.1, Heuristic 14.3, p. 363) or J. L. Gross et al. (2014, Sec. 5.1.2, Fact F13).

Lemma A.10). This is an improvement over Lemma 6.19 alone where V^*AV can (appear to) act non-trivially on the full system. The radius s increases polynomially with R, i.e. polynomially with time. Below, we construct an improved representation where s increases linearly with time (Corollary 6.34), which matches what is already known from Lieb–Robinson bounds (e.g. Theorem 6.1). □

Lemma 6.23 *Let A be an operator which acts non-trivially on $Y \subset \Lambda$. Let $V = V_1 V_2 \dots V_b$ where the V_k ($k \in \{1 \dots b\}$) are unitary and V_k acts non-trivially (at most) on $B^o_r(\{j_k\})$ with some $j_k \in \Lambda$ and $r > 0$. Let the V_k commute pairwise, i.e. $[V_k, V_l] = 0$ for all $k, l \in \{1 \dots b\}$. Then $B = V^*AV$ acts non-trivially at most on $B^o_{2r}(Y)$.*

Proof In the expression $B = V^*_b \dots V^*_1 A V_1 \dots V^*_b$, all V_k which commute with A can be omitted (because a given V_k commutes with all other V_l). In particular, all V_k which do not act non-trivially on Y can be omitted without changing B. Let $x \in \Lambda$ be a site on which B acts non-trivially. If $x \in Y$ holds, $x \in B^o_{2r}(Y)$ holds as well and we are finished. In the following, let $x \notin Y$. Then, there is a $k \in \{1 \dots b\}$ such that $x \in B^o_r(\{j_k\})$. In addition, there is a $y \in B^o_r(\{j_k\}) \cap Y$ (otherwise, V_k and A commute and V_k can be omitted from B). Note that $d(x, y) < 2r$ (the diameter of the given open ball). As a consequence, $x \in B^o_{2r}(Y)$ holds, which completes the proof. ■

6.4.3. Efficient simulation of time evolution: Hypercubic lattice

In this section, we construct a representation of time evolution under a local Hamiltonian which has a smaller bond dimension than the representation presented above. In order to split the complete time evolution into independent parts in a more efficient way, we consider a cubic lattice Λ of finite dimension η with L sites in each direction:

$$\Lambda := \left\{(x_1, \dots x_\eta) \colon x_i \in [1 : L], i \in [1 : \eta]\right\} \tag{6.57}$$

Here, we used the notation $[1 : L] = \{1, 2, \dots, L\}$ to denote a set of consecutive integers. The total number of sites is $n = |\Lambda| = L^\eta$. In this Section, $\dot{a} = \lfloor a \rfloor$ denotes the interaction range rounded down.

The Cartesian product (Section 1.2) has the basic property $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$. Powers of sets are given by the Cartesian product, e.g. $[1 : L]^2 = [1 : L] \times [1 : L]$, and this allows us to write $\Lambda = [1 : L]^\eta \subset \mathbb{Z}^\eta$ where $\mathbb{Z}$ is the set of all integers.

We assume a metric d on Λ which satisfies the property

$$|x_i - y_i| \le d(x, y) \quad \forall\, i \in [1 : \eta] \tag{6.58}$$

For example, the metric induced by the vector-p norm, $d(x, y) = [\sum_{i=1}^{\eta} |x_i - y_i|^p]^{1/p}$ with $p \in [1, \infty]$, has this property. Below, we partition the lattice into cubic sets defined as follows:

Definition 6.24 Two points $x, y \in \mathbb{Z}^\eta$ define the cube[41] $C(x, y) := \Lambda \cap \bigtimes_{i=1}^{\eta} [x_i : y_i]$. For a non-negative integer r, the enlarged cube is defined as $C_r(C(x, y)) := C(x - rv, y + rv)$ where $v := (1, 1, \dots, 1) \in \mathbb{Z}^\eta$. □

Definition 6.25 For PEPS representations, we use the graph in which $x, y \in \Lambda$ are connected if and only if $\sum_{i=1}^{\eta} |x_i - y_i| = 1$, such that each site has at most 2η nearest neighbours. In this graph, the cube $C(x, y)$ is a connected set for all $x, y \in \mathbb{Z}^\eta$. □

We employ the following notation for Cartesian products: Let $c, d \in \mathbb{Z}$ and $x, y \in \mathbb{Z}^{\eta-1}$, then

$$\begin{aligned}&[c : d]_i \times C(x, y) = \\ &[x_1 : y_1] \times \cdots \times [x_{i-1} : y_{i-1}] \times [c : d] \times [x_i : y_i] \times \cdots \times [x_{\eta-1} : y_{\eta-1}].\end{aligned} \tag{6.59}$$

We partition the full lattice Λ into cubes Q_m of edge length Ω and aim at splitting the full time evolution into independent evolutions on the cubes Q_m. Figure 6.4 illustrates the partition $\Lambda = \dot{\bigcup}_m Q_m$ and outlines the way forward. The next lemma identifies all local terms h_Z which couple at least two cubes Q_m and $Q_{m'}$:

Lemma 6.26 *Let Ω be a positive integer and $B = \lceil L/\Omega \rceil$. For $m \in [1 : B]^\eta$, set*

$$Q_m := C(x_m, y_m), \qquad x_m := [\Omega(m_i - 1) + 1]_{i=1}^{\eta}, \qquad y_m := [\Omega m_i]_{i=1}^{\eta}. \tag{6.60}$$

These cubes partition the lattice, $\Lambda = \dot{\bigcup}_m Q_m$. For $i \in [1 : \eta]$ and $j \in [1 : B - 1]$, set

$$A_{ij} := [1 : \Omega j]_i \times [1 : L]^{\eta-1}, \qquad B_{ij} := [\Omega j + 1 : L]_i \times [1 : L]^{\eta-1} \tag{6.61}$$

and

$$S_{ij} := \left\{ Z \subset \Lambda\colon\ h_Z \neq 0,\ Z \cap A_{ij} \neq \emptyset,\ Z \cap B_{ij} \neq \emptyset \right\}. \tag{6.62}$$

The complete Hamiltonian is given by $H = H_Q + H_S$ where H_Q contains all terms which act within one of the cubes Q_m and H_S contains all terms which couple at least two cubes:

$$H_Q := \sum_{m \in [1:B]^\eta} \sum_{Z \subset Q_m} h_Z, \qquad H_S := \sum_{Z \in S} h_Z, \qquad S := \bigcup_{i=1}^{\eta} \bigcup_{j=1}^{B-1} S_{ij}. \tag{6.63}$$

---

[41] To be precise, $C(x, y)$ is a hyperrectangle.

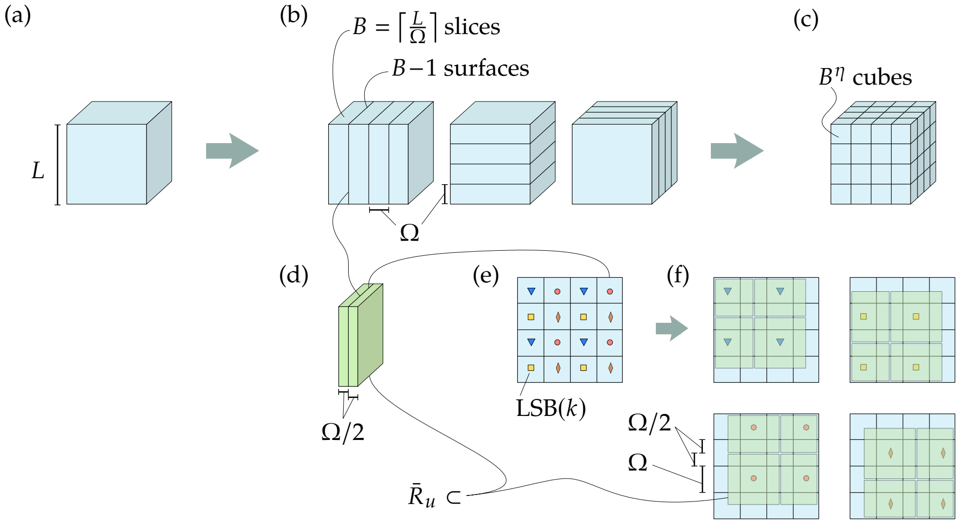


Figure 6.4: Decomposition of an η-dimensional hypercube, illustrated for $\eta = 3$. (a) The lattice Λ lives inside a cube of edge length L. (b) Along each direction, the cube is split into $B = \lceil L/\Omega \rceil$ slices of width Ω (Lemma 6.26). (c) As a result, the cube is split into B^η smaller cubes Q_m of edge length Ω (Equation (6.60)). (d) Centered around each surface from (b), there is a slice of width Ω. (e) Each surface from (b) is split into $B^{\eta-1}$ surface segments $\tilde{Q}_k$ of edge length Ω (Lemma 6.28). The surface segments are divided into $2^{\eta-1}$ groups of non-neighbouring surfaces as specified by LSB(k) and indicated by the symbols (Lemma 6.32). (f) The sets $\bar{R}_u$ do not overlap for surface segments with the same symbol (the same value of LSB(k); Lemma 6.31).

Proof The definition directly implies that the cubes Q_m partition the lattice Λ (any two cubes do not intersect and the union of all cubes equals the complete lattice). As the sets Q_m are disjoint, H_Q contains each local term from H at most once. It remains to show that H_S contains exactly once all local terms which are not in H_Q. Let $Z \subset \Lambda$ be such that $h_Z \neq 0$ is not in H_Q, i.e. there are $m, m' \in [1 : B]^\eta$ with $m \neq m'$ such that both $Z \cap Q_m \neq \emptyset$ and $Z \cap Q_{m'} \neq \emptyset$. There is an $i \in [1 : \eta]$ such that $m_i \neq m'_i$. Without loss of generality, assume that $m_i < m'_i$ (otherwise, exchange m and m'). Let $a \in Z \cap Q_m$, then $a_i \leq m_i \Omega$ holds. Let $b \in Z \cap Q_{m'}$, then $b_i \geq \Omega(m'_i - 1) + 1 \geq \Omega m_i + 1$ holds. Set $j = m_i$, then $a \in Z \cap A_{ij}$ and $b \in Z \cap B_{ij}$ and this shows that both intersections are non-empty, i.e. $Z \in S_{ij} \subset S$. This shows that the local term h_Z, which is not in H_Q, appears in H_S exactly once. ∎

The last lemma has identified the local terms which we want to remove from H.

The next lemma determines the possible extent of these local terms:

Lemma 6.27 *Let $Z \in S_{ij}$. Then $Z \subset I_{ij} \times [1 : L]^{\eta-1}$ where the interval $I_{ij} := [\Omega j - \dot{a} + 1 : \Omega j + \dot{a}]_i$ is along dimension i (cf. Equation (6.59)).*

Proof Recall that $\operatorname{diam}(Z) \leq a$ as $h_Z \neq 0$. The property $Z \cap A_{ij} \neq \emptyset$ implies $Z \subset B^c_a(A_{ij}) \subset C_{\dot{a}}(A_{ij}) = [1 : \Omega j + \dot{a}]_i \times [1 : L]^{\eta-1}$ (Lemmata A.11 and A.12). In the same way, $Z \cap B_{ij} \neq \emptyset$ implies $Z \subset B^c_a(B_{ij}) \subset C_{\dot{a}}(B_{ij}) = [\Omega j - \dot{a} + 1 : L]_i \times [1 : L]^{\eta-1}$. Combining both provides $Z \subset B^c_a(A_{ij}) \cap B^c_a(B_{ij}) \subset I_{ij} \times [1 : L]^{\eta-1}$ (Lemma A.13). ■

The local terms $Z \in S_{ij} \subset S$, which we aim at removing, generally cover the full volume described in the last lemma; if we removed all $Z \in S_{ij}$ with a single application of Lemma 6.18, the resulting correction V' would act on a large fraction of the lattice, which we want to avoid. In addition, a given local term Z may be a member of more than one of the sets S_{ij}. We construct a partition of the set S which addresses these issues:

Lemma 6.28 *Let $[1 : L]^{\eta-1} = \dot{\bigcup}_{k\in[1:B]^{\eta-1}} \tilde{Q}_k$ a partition into cubes as in Equation (6.60).*[42] *For $i \in [1 : \eta]$, $j \in [1 : B - 1]$ and $k \in [1 : B]^{\eta-1}$, let*

$$S_{ijk} = \{Z \in S_{ij} \colon Z \cap Q_{ijk} \neq \emptyset\}, \qquad Q_{ijk} = I_{ij} \times \tilde{Q}_k \tag{6.64}$$

where I_{ij} is from Lemma 6.27. Then $S_{ij} = \bigcup_{k\in[1:B]^{\eta-1}} S_{ijk}$ holds and $Z \in S_{ijk}$ implies $Z \subset I_{ij} \times C_{\dot{a}}(\tilde{Q}_k)$. Subsets $S'_{ijk} \subset S_{ijk}$ which partition S, $S = \dot{\bigcup} S'_{ijk}$, can be chosen in $O(n^2)$ computational time.

Proof The equality $S_{ij} = \bigcup_{k\in[1:B]^{\eta-1}} S_{ijk}$ holds because the Q_{ijk} partition $I_{ij} \times [1 : L]^{\eta-1}$, which is a superset of all $Z \in S_{ij}$ (Lemma 6.27); this equality also implies $S = \bigcup_{ijk} S_{ijk}$.

Let $Z \in S_{ijk}$. This implies $h_Z \neq 0$ and $\operatorname{diam}(Z) \leq a$. We have $Z \subset B^c_a(Q_{ijk}) \subset C_{\dot{a}}(Q_{ijk}) = C_{\dot{a}}(I_{ij}) \times C_{\dot{a}}(\tilde{Q}_k)$ (Lemmata A.11 and A.12, Definition 6.24). Combining this with $Z \subset I_{ij} \times [1 : L]^{\eta-1}$ provides $Z \subset [C_{\dot{a}}(I_{ij}) \times C_{\dot{a}}(\tilde{Q}_k)] \cap [I_{ij} \times [1 : L]^{\eta-1}] = I_{ij} \times C_{\dot{a}}(\tilde{Q}_k)$.

In order to obtain suitable subsets $S'_{ijk} \subset S_{ijk}$, choose any fixed order for the sets S_{ijk} and remove all elements from S_{ijk} which are already an element of a previous S_{ijk}. This takes computational time $O(n^2)$ where $n = L^\eta = |\Lambda|$. ■

We aim at removing all interactions in a set S'_{ijk} with a single application of Lemma 6.18. For this purpose, we define a sequence $H_0, \dots, H_\Xi$ of Hamiltonians where $H_0 = H_Q$, $H_\Xi = H$. Consecutive Hamiltonians in this sequence differ precisely by the local terms contained in one of the sets S'_{ijk}. In order to define this sequence of Hamiltonians, we define a specific order of the sets S'_{ijk} which also proves to be advantageous below.

---

[42] I.e. $\tilde{Q}_k = C(x_k, y_k)$, $x_k = [\Omega(k_i - 1) + 1]_{i=1}^{\eta-1}$ and $y_k = [\Omega k_i]_{i=1}^{\eta-1}$.

Definition 6.29 For $k \in [1 : B]^{\eta-1}$, let $b = \mathrm{LSB}(k) \in [0 : 1]^{\eta-1}$ be the vector whose component b_i is the least significant bit of k_i; i.e. $b_i = 1$ ($b_i = 0$) if k_i is odd (even).□

Lemma 6.30 *Let $i \in [1 : \eta]$, $j \in [1 : B-1]$, $k \in [1 : B]^{\eta-1}$ and $\Xi := \eta(B-1)B^{\eta-1}$. Let $\omega \colon [1 : \Xi] \to [1 : \eta] \times [1 : B-1] \times [1 : B]^{\eta-1}$ be a bijective function such that its inverse ω^{-1} maps all (i, j, k) with the same value of $(i, \mathrm{LSB}(k))$ to consecutive integers from $[1 : \Xi]$.*[43] *For $u \in [1 : \Xi]$, set*

$$\Sigma_u := \Sigma_{u-1} \,\dot\cup\, S'_{\omega(u)}, \qquad F_u := \sum_{Z \in S'_{\omega(u)}} h_Z \tag{6.65}$$

where $\Sigma_0 := \dot\bigcup_{m \in [1:B]^\eta} \{Z \subset Q_m \colon h_Z \neq 0\}$. For $u \in [0 : \Xi]$ and subsets $R_u \subset \Lambda$, set

$$H_u := \sum_{Z \in \Sigma_u} h_Z, \qquad H'_u := \sum_{\substack{Z \in \Sigma_u \\ Z \subset \bar{R}_u}} h_Z \tag{6.66}$$

Then, $H_0 = H_Q$, $H_\Xi = H$ and $H_u - H_{u-1} = F_u$ ($u \in [1 : \Xi]$).

Proof The sets $\{Z \subset Q_m \colon h_Z \neq 0\}$ are disjoint because the sets Q_m are disjoint (Lemma 6.26). Let $E_H = \{Z \subset \Lambda \colon h_Z \neq 0\}$. Lemma 6.26 implies $H_0 = H_Q$ and $E_H = S \,\dot\cup\, \Sigma_0$. $S = \dot\bigcup_{u=1}^{\Xi} S'_{\omega(u)}$ is provided by in Lemma 6.28 and implies $E_H = \Sigma_0 \,\dot\cup\, \left(\dot\bigcup_u S'_{\omega(u)}\right)$, $\Sigma_{u-1} \cap S'_{\omega(u)} = \emptyset$ as well as $H_\Xi - H_0 = H_S$, i.e. $H_\Xi = H_Q + H_S = H$. $F_u = H_u - H_{u-1}$ is implied by the definitions. ■

The correction for removing the interactions from $S'_{\omega(u)}$ is to be supported on $\bar{R}_u$ and the choice of $R_u \subset \Lambda$ is still open. The next lemma defines the sets R_u and discusses whether two given $\bar{R}_u$ overlap.

Lemma 6.31 *Let Ω be an even integer and $\Omega > 4\dot{a}$. Let $u \in [1 : \Xi]$ and set*

$$Y_u = \bigcup_{Z \in S'_{\omega(u)}} Z, \qquad R_u = B^o_r(Y_u), \qquad r = \Omega/2 - 2\dot{a}. \tag{6.67}$$

Let $i \in [1 : \eta]$, $j, j' \in [1 : B-1]$, $k, k' \in [1 : B]^{\eta-1}$, $u = \omega^{-1}(i, j, k)$ and $u' = \omega^{-1}(i, j', k')$. The set $\bar{R}_u$ is at most $\bar{R}_u \subset [\Omega(j - \frac{1}{2}) + 1 : \Omega(j + \frac{1}{2})]_i \times C_{\Omega/2}(\tilde{Q}_k)$. $\bar{R}_u \cap \bar{R}_{u'} = \emptyset$ holds if (i) $j \neq j'$ or (ii) $k \neq k'$ and $\mathrm{LSB}(k) = \mathrm{LSB}(k')$.

---

[43] For example, $\omega^{-1}(i, j, k)$ can be defined as position the of $(i, \mathrm{LSB}(k), j, k)$ within the lexicographically ordered sequence of all $(i, \mathrm{LSB}(k), j, k)$.

Proof Lemma 6.28 implies $Y_u \subset I_{ij} \times C_{\dot{a}}(\tilde{Q}_k)$. We have $\bar{R}_u \subset B^c_a(B^o_r(Y_u)) \subset B^o_{r+a}(Y_u) \subset C_{r+\dot{a}}(I_{ij}) \times C_{r+2\dot{a}}(\tilde{Q}_k)$ and the same for $\bar{R}_{u'}$ and (i, j', k') (Lemmata A.10 and A.11 and Definition 6.24). Note that $r + 2\dot{a} = \Omega/2$.

Assume that $j \neq j'$ holds. $C_{r+\dot{a}}(I_{ij}) = [\Omega j - (r + 2\dot{a}) + 1 : \Omega j + r + 2\dot{a}]_i = [\Omega(j - \frac{1}{2}) + 1 : \Omega(j + \frac{1}{2})]_i$. This set does not intersect with the same set for j' if $j \neq j'$. As a consequence, $\bar{R}_u$ and $\bar{R}_{u'}$ do not intersect (use Lemma A.13).

Assume that $k \neq k'$ and $\mathrm{LSB}(k) = \mathrm{LSB}(k')$ hold. Let $\mu \in [1 : \eta{-}1]$ such that $k_\mu \neq k'_\mu$. Without loss of generality, assume that $k_\mu < k'_\mu$ (exchange k and k' if necessary). Note that this implies $k'_\mu - k_\mu \geq 2$ because k_μ and k'_μ are both even or both odd (which follows from $\mathrm{LSB}(k) = \mathrm{LSB}(k')$). Note that $C_{\Omega/2}(\tilde{Q}_k) = C(x_k - \frac{\Omega}{2}v, y_k + \frac{\Omega}{2}v)$ where $v = (1, 1, \dots, 1) \in \mathbb{Z}^\eta$ and the same for k'. We have

$$[x_{k'} - \tfrac{\Omega}{2}v]_\mu - [y_k + \tfrac{\Omega}{2}v]_\mu = \Big(\Omega(k'_\mu - 1 - \tfrac{1}{2}) + 1\Big) - \Omega(k_\mu + \tfrac{1}{2}) \geq 1 \qquad (6.68)$$

where we have used $k'_\mu - k_\mu \geq 2$. As a consequence, $C_{\Omega/2}(\tilde{Q}_k)$ does not overlap with the same set for k' and this implies that $\bar{R}_u$ and $\bar{R}_{u'}$ do not overlap (use Lemma A.13). ∎

The next lemma provides the necessary definitions for applying Lemma 6.18, taking advantage of the particular ordering function ω (Lemma 6.30) and of non-overlapping sets $\bar{R}_u$ (Lemma 6.31):

Lemma 6.32 *Let Ω be an even integer and $\Omega > 4\dot{a}$. For $u \in [1 : \Xi]$ and $s, t \in \mathbb{R}$, let $V'_{us}(t)$ on $\bar{R}_u$ be the solution of $\partial_s V'_{us}(t) = \mathrm{i}L'_{ut}(s)V'_{us}(t)$ where $L'_{ut}(s) = \tau^{H'_u}_{ts}(F_u(s))$ and $V'_{ut}(t) = \mathbb{1}$. Set $V'_u = V'_{us}(t)$ and $V' := V'_\Xi \dots V'_2 V'_1$. Then, V' is given by*

$$V' := \prod_{i=1}^{\eta} \prod_{l \in [0:1]^{\eta-1}} V'_{il}, \qquad\qquad V'_{il} := \bigotimes_{j=1}^{B-1} \bigotimes_{\substack{k \in [1:B]^{\eta-1} \\ \mathrm{LSB}(k)=l}} V'_{ijk}. \qquad (6.69)$$

where $V'_{ijk} := V'_u$ with $u = \omega^{-1}(i, j, k)$.[44] *In addition, set $V := V' U^{H_Q}_{ts}$.*

Proof Use Lemmata 6.30 and 6.31 recalling that all $(i, j, k) = \omega(u)$ with the same value of $(i, \mathrm{LSB}(k))$ appear consecutively as u proceeds from 1 to Ξ. ∎

Finally, we have completed the preparations for applying Lemma 6.18:

Theorem 6.33 *Let $\Lambda = [1 : L]^\eta$, $n = |\Lambda| = L^\eta$ and let d be a metric on Λ which satisfies (6.58). Let Ω be an even integer and $B = \lceil L/\Omega \rceil$. Choose $q \in (0, 1)$ and let Ω be such that*

---

[44] The order of the terms V'_{il} in (6.69) is specified by the function ω.

$r := \Omega/2 - 2\dot{a}$ satisfies $r > 0$, $\lceil r/a \rceil > 2\kappa + 1$ and $\lceil r/a \rceil \geq \frac{2\kappa}{q} \ln(\frac{\kappa}{q})$ where $\kappa = \eta - 1$. The distance between $V = V' U_{ts}^{H_Q}$ from Lemma 6.32 and the exact time evolution U_{ts}^{H} is at most

$$\left\| V' U_{ts}^{H_Q} - U_{ts}^{H} \right\|_{(\infty)} \leq \epsilon \tag{6.70}$$

if

$$\Omega \geq \frac{2a}{1-q} \left[v|t-s| + \ln\left(\frac{n}{\epsilon}\right) + \ln(c_3) \right] \tag{6.71}$$

where $c_3 := 2\eta a M \exp(1)/\mathcal{Z}$.

The operator $U_{ts}^{H_Q}$ is the tensor product of $B^\eta < n$ independent time evolutions on Ω^η sites. The operator V' consists of $\Xi = \eta(B-1)B^{\eta-1} < \eta n$ independent time evolutions on $\Omega(2\Omega)^{\eta-1}$ sites. All constituents of the two operators can be computed in $O(n\eta\exp(\Omega^\eta))$ computational time. The operator V admits a PEPS representation of bond dimension $D = O(\exp(\eta 4^\eta \Omega^\eta \ln(d)))$ where $d = \max_{x\in\Lambda} d(x)$ is the maximal local dimension. The PEPS representation is defined in terms of the graph from Definition 6.25.

Proof Let $u \in [1:\Xi]$ and $(i,j,k) = \omega(u)$. Note that $\dot{\bigcup}_{k\in[1:B]^{\eta-1}} Q_{ijk} = I_{ij} \times [1:L]^{\eta-1}$ (cf. Lemma 6.28), which implies $\sum_k |Q_{ijk}| \leq 2\dot{a}L^{\eta-1}$. The operator F_u is the sum of a subset of all terms which intersect with Q_{ijk} (Lemma 6.28 and Equation (6.65)); i.e. F_u is the sum of at most $|Q_{ijk}|\mathcal{Z}$ local terms. As a consequence, $\|F_u\|_{(\infty)} \leq (J/2)|Q_{ijk}|\mathcal{Z} = v|Q_{ijk}|/2\mathrm{e}$. We have $R_u = B_r^o(Y_u)$ with $r = \Omega/2 - 2\dot{a}$ (cf. Lemma 6.31), therefore $d(Y_u, \Lambda \setminus R_u)/a \geq r/a = \Omega/(2a) - 2\lfloor a \rfloor/a \geq \Omega/(2a) - 2$ (Lemma A.10). We have (use Lemma 6.18)

$$\left\| V'_{ijk} U_{ts}^{H_{u-1}} - U_{ts}^{H_u} \right\|_{(\infty)} \leq \frac{M}{\mathcal{Z}\mathrm{e}} |Q_{ijk}| \exp(v|t-s| - (1-q)\Omega/2a + 2). \tag{6.72}$$

The total distance is at most the sum of such terms for all $u \in [1:\Xi]$ or all (i,j,k), respectively.[45] We evaluate

$$\sum_{ijk} |Q_{ijk}| \leq \sum_{ij} 2\dot{a}L^{\eta-1} = 2\dot{a}\eta(B-1)L^{\eta-1} < 2\eta a n. \tag{6.73}$$

This provides

$$\left\| V' U_{ts}^{H_Q} - U_{ts}^{H} \right\|_{(\infty)} \leq n c_3 \exp(v|t-s| - (1-q)\Omega/2a) \leq \epsilon \tag{6.74}$$

---

[45] Completely analogous to the proof of Lemma 6.19.

where $c_3 = 2\eta a M \exp(1)/\mathcal{Z}$.

In order to determine the maximal bond dimension of a PEPS representation, recall that hypercubes are connected sets (Definition 6.25) and note that all unitaries in the decomposition act on hypercubes Q_m (Lemma 6.26) or on sets $\bar{R}_u$ (Lemma 6.32) which are subsets of hypercubes (Lemmata 6.28 and 6.31). Therefore, Q_m is a connected set whose number of sites is $|Q_m| \le \Omega^\eta$ and $\bar{R}_u$ is a subset of a connected set whose number of sites is $|\bar{R}_u| \le \Omega(2\Omega)^{\eta-1}$ (Lemma 6.31). Using Lemmata 2.1 and 2.2 as well as Equation (6.69) shows that the bond dimension of a PEPS representation of $V' U_{ts}^{H_Q}$ is at most $D \le \exp[(\Omega^\eta + \eta 2^{\eta-1}\Omega(2\Omega)^{\eta-1})\ln(d^2)]$. ■

Corollary 6.34 *Let A be an operator which acts non-trivially on a single site x. Then $V^* A V$ with $V = V' U_{ts}^{H_Q}$ acts non-trivially at most on $C_r(\{x\})$ and the radius $r = \lambda\Omega$ increases linearly with time and with $\ln(n/\epsilon)$ ($\lambda = \eta 2^\eta + 3/2$). (Proof: Analogous to Lemma 6.23.)*

Remark 6.35 The radius in the last corollary is proportional to Ω; using the representation for an arbitrary lattice, this radius is proportional to Ω^η (Remark 6.22). □

6.5. Discussion

In this work, we have discussed the unitary time evolution operator U_t induced by a time-dependent finite-range Hamiltonian on an arbitrary lattice with n sites. In addition, we have discussed time-evolved states $|\psi(t)\rangle = U_t|\psi(0)\rangle$ where the initial state $|\psi(0)\rangle$ is a product state. We have shown that such a time-evolved state can be certified or verified efficiently, i.e. there is an efficient method to determine an upper bound β on the infidelity of the time-evolved state $|\psi(t)\rangle$ and an arbitrary, unknown state ρ. We presented a method where the measurement effort for obtaining the upper bound β was only $O(n^3 \exp[(v|t| + \ln(n/\mathcal{I}))^\eta])$ instead of $O(\exp(n))$. If the time-evolved state $|\psi(t)\rangle$ and the unknown state ρ are sufficiently close, the upper bound β is guaranteed to not exceed $\mathcal{I}$. The measurement effort is seen to increase quasi-polynomially with n if the spatial dimension η is two or larger and polynomially with n in one spatial dimension. The scaling in a single spatial dimension matches results from previous work.[46] The complete time evolution operator U_t can be encoded into a time-evolved state $|\psi(t)\rangle$ if each site of the lattice is augmented by a second site of the same dimension and the initial state is one where each pair of sites is maximally entangled.[47] A certificate for this time-evolved state then also provides a certificate for the time evolution operator U_t. This enables assumption-free verification of the output of methods which, under the assumption that it is a finite-ranged Hamiltonian, determine the unknown Hamiltonian of a system.[48]

---

[46] Lanyon, Maier, et al. 2017; Supplementary Information.

[47] Holzäpfel et al. 2015. See also Section 5.1.

We have also shown that the time evolution operator U_t admits an efficient PEPO representation on the same lattice as the Hamiltonian, implying that the time-evolved state $|\psi(t)\rangle$ admits an efficient PEPS representation. This holds if time t is at most poly-logarithmic in the number of sites n. An efficient representation on the same lattice is different from efficient PEPO representations of U_t based on the Trotter decomposition, which use a lattice of a larger dimension than the Hamiltonian itself. Our result provides guidelines on the necessary resources for numerically computing the time-evolved state $|\psi(t)\rangle$ with PEPSs (or a suitable subclass thereof); such methods typically attempt to represent the time-evolved state $|\psi(t)\rangle$ on the same lattice as the Hamiltonian. We construct an efficient representation of U_t which approximates U_t up to an error ϵ and which is based on a unitary circuit which propagates a local observable to a region whose diameter grows only linearly[49] with $v|t| + \ln(n/\epsilon)$. This highlights that U_t is approximated by a PEPO with a very specific structure; a general PEPO might e.g. displace local observables by arbitrarily large distances. This property can also be used for an alternative proof of efficient certification of time-evolved states $|\psi(t)\rangle$, following the original approach pursued in one spatial dimension.[50]

We have shown that time-evolved states of finite-range Hamiltonians can be certified and represented efficiently. At this point, it remains an open question whether these results can be extended to Hamiltonians with exponentially decaying couplings.

---

[48] da Silva et al. 2011; Holzäpfel et al. 2015. See also Chapter 5.

[49] This holds for a hypercubic lattice. For an arbitrary lattice, the diameter increases polynomially instead of linearly with the given expression.

[50] Lanyon, Maier, et al. 2017; Supplementary Information.

Chapter 7.

Tomography of an ion trap quantum simulator

This chapter discusses results on efficient tomography of an ion trap quantum simulator experiment with 14 qubits realized by Lanyon, Maier, et al. (2017).[1] We focus on the description of the basic method, which could be applied to any n-qubit system which allows for control of individual qubits. For example, the quantum state of four photonic qubits has been estimated from complete information on two or three neighbouring qubits in another recent experiment.[2]

After a short review of the experiment's capabilities, we describe the three basic steps of efficient quantum state tomography with reference to the experiment:

1. **Measure:** Take informationally complete measurements on blocks of r neighbouring qubits, i.e. on qubits $k, \dots, k+r-1$ for all $k \in \{1 \dots n-r+1\}$.

2. **Estimate:** Find an n-qubit pure state $|\psi\rangle$ compatible with the measurements.

3. **Certify:** Verify that the estimate $|\psi\rangle$ is close to the unknown, true state ρ.

While measurements contain complete information on r neighbouring qubits, we additionally perform steps 2 and 3 using complete information on only $\tilde{r}$ qubits where $\tilde{r} \leq r$.

---

[1] This chapter reproduces parts of the original publication Lanyon, Maier, et al. 2017 (published by Macmillan Publishers Limited, part of Springer Nature), which was subsequently also discussed by coauthor Buyskikh (2017). The present author contributed data analysis and numerical simulations. This chapter presents selected aspects and adds Figures 7.1, 7.2 left and 7.3.

[2] Zhao et al. 2017.

7.1. The ion trap quantum simulator experiment

Here, we give a schematic overview of the capabilities of the experiment presented by Lanyon, Maier, et al. (2017). Most technical information, such as the frequencies of all the different involved lasers, is not mentioned here and can be found in the original publication.

Qubits One qubit in the experiment is provided by two electronic states of a single ion in an ion trap. A linear Paul trap, which confines charged particles using oscillating and constant electric fields, contains n ions in a linear chain. The position of each ion is described in terms of the axial direction along the chain and the two radial directions perpendicular to the chain. The ion trap contains $^{40}\mathrm{Ca}^{+}$ ions and one qubit is provided by the two states $|{\downarrow}\rangle = |S_{1/2}, m = 1/2\rangle$ and $|{\uparrow}\rangle = |D_{5/2}, m = 1/2\rangle$ of the remaining valence electron of a single ion.[3] The states of the qubits are manipulated and measured by shining continuous wave (CW) laser light on the ions in a radial direction and detecting fluorescence light from the ions in a perpendicular radial direction. Fluorescence light is detected with an electron-multiplying CCD camera.

Single-qubit rotations The single-qubit rotations used were described by Schindler et al. (2013). A σ_Z rotation of qubit k is implemented by shining laser light on the corresponding ion, implementing the operation $\exp(-\mathrm{i}\theta\sigma_Z^{(k)})$, where θ is proportional to the illumination time. Collective σ_X and σ_Y rotations are implemented by shining laser light on all ions. Specifically, this implements the operation $\exp(-\mathrm{i}\theta S_\phi)$ where $S_\phi = \sum_{k=1}^{n}[\sigma_X^{(k)}\cos(\phi) + \sigma_Y^{(k)}\sin(\phi)]$, ϕ depends on the phase of the light and θ is again proportional to the illumination time. Both operations together allow arbitrary single-qubit rotations.[4]

State preparation The qubit state $|{\downarrow\downarrow}\dots{\downarrow}\rangle$ is prepared by Doppler cooling and optical pumping. After state preparation, the $2N$ radial motional modes are also cooled to their ground states via frequency-resolved sideband cooling and optical pumping. After that, single-qubit rotations are used to flip every second spin, obtaining the Néel-ordered state $|\uparrow\downarrow\uparrow\downarrow\ldots\rangle$.

---

[3] Recall that l, j and m denote the quantum numbers of orbital angular momentum, total angular momentum and the z component of orbital angular momentum, respectively. The letters S, P, D indicate $l = 0$, $l = 1$ and $l = 2$ and their subscript indicates j. The two states have the quantum numbers $(l = 0, j = \frac{1}{2}, m = \frac{1}{2})$ and $(l = 2, j = \frac{5}{2}, m = \frac{1}{2})$.

[4] Schindler et al. 2013.

Time evolution A global beam which contains three frequencies induces the Ising Hamiltonian

$$H_{\text{Ising}} = \hbar \sum_{i,j=1}^{n} J_{ij}\, \sigma_X^{(i)} \sigma_X^{(j)} + \hbar \sum_{i=1}^{n} (B + B_i)\, \sigma_Z^{(i)} \tag{7.1}$$

where $\hbar$ is the reduced Planck constant. The Hamiltonian is turned on after state preparation for some time t and it is turned off before measurement. Parameters J_{ij} and B_i which match the experiment were derived from other experimentally measured parameters.[5] The coupling roughly decays with a power-law $J_{ij} \sim 1/|i-j|^{\alpha}$ with $1.1 < \alpha < 1.6$. The time unit we use is J^{-1} where $J = \sum_{i=1}^{n-1} J_{i,i+1}/(n-1)$ is the average nearest-neighbour coupling. The parameters satisfy $|B + B_i| \gg |J_{ij}|$, which implies that the Hamiltonian is approximately equivalent to an XY model in a transverse field.[6] With the given initial state and Hamiltonian, correlations increase with time as shown by the growing von Neumann entropies in column 1 of Figure 7.1 on page Page 130.

Measurement A single-qubit projective measurement of σ_Z could be implemented by shining light on the corresponding ion and collecting fluorescence light in a perpendicular direction. A simultaneous single-qubit projective measurement of σ_Z on each qubit is implemented by shining light on all ions and collecting the fluorescence light of all ions with a CCD camera which resolves single ions. Other measurements are implemented by applying suitable rotations, followed by a simultaneous σ_Z measurement of all qubits.

7.2. Measurements

For $a = (Z, \ldots, Z)$, the following POVM describes a simultaneous measurement of σ_Z on each qubit (cf. Equation (3.15)):

$$\Pi_{\sigma_a}^{(n)} = \left\{ P_{\sigma_{a_1}, s_1} \otimes \cdots \otimes P_{\sigma_{a_n}, s_n} : s \in \{+1, -1\}^n \right\}. \tag{7.2}$$

---

[5] Lanyon, Maier, et al. 2017.

[6] Lanyon, Maier, et al. 2017; Jurcevic et al. 2015.

Recall that the POVM is defined for all $a \in \{X, Y, Z\}^n$ and that its elements are $P_{\sigma_{a_k}, s_k} = |\Psi_{\sigma_{a_k}, s_k}\rangle\langle\Psi_{\sigma_{a_k}, s_k}|$ where $|\Psi_{\sigma_{a_k}, s_k}\rangle$ is the eigenvector of the eigenvalue s_k of the Pauli matrix σ_{a_k}.

In the experiment, the following POVMs on n qubits were measured for $r = 3$:

$$\mathcal{P} = \left\{\Pi^{(n)}_{\sigma_a} \colon a \in \mathcal{A}\right\}, \quad \mathcal{A} = \{a \in \{X, Y, Z\}^n \colon a_k = a_{k+r}, k \in \{1 \dots n - r\}\} \tag{7.3}$$

The set $\mathcal{P}$ contains 3^r POVMs. Each POVM from $\mathcal{P}$ represents a projective measurement which was implemented in the experiment by a sequence of unitary rotations followed by a global, projective measurement of σ_Z on each qubit. In a numerical simulation, measurements of the POVMs from $\mathcal{P}$ are simulated efficiently by simulating this sequence of projective measurements.[7] In numerical simulations, $m = 1000$ measurements of each POVM from $\mathcal{P}$ are simulated for $r \in \{3, 4, 5\}$. In the experiment, $m = 1000$ measurements of each POVM were taken for $r = 3$.

The elements of all POVMs from $\mathcal{P}$ give rise to the following POVM with $3^r 2^n$ elements:

$$\Pi^{(n,r)}_{\text{Cyc}} = \left\{\frac{1}{3^r} P \colon P \in \Pi, \Pi \in \mathcal{P}\right\}. \tag{7.4}$$

All probabilities of the POVM $\Pi^{(n,r)}_{\text{Cyc}}$ can be estimated from the measurements taken in the experiment. For $\tilde{r} \in \{1 \dots r\}$, consider the $\tilde{r}$-qubit Pauli POVMs

$$\Pi^{(n,\tilde{r},k)}_{\text{Local}} = \left\{\frac{1}{3^{\tilde{r}}} P \colon P \in \Pi^{(\tilde{r})}_{\sigma_a}, a \in \{X, Y, Z\}^{\tilde{r}}\right\}, \quad k \in \{1 \dots n - \tilde{r} + 1\} \tag{7.5}$$

where the superscript k indicates that the elements of the POVM act on qubits $k, \dots, k + \tilde{r} - 1$. All probabilities of these POVMs can be estimated from the measurement data by considering the partial measurement results for the corresponding qubits as long as we keep $\tilde{r} \leq r$. As the POVMs $\Pi^{(n,\tilde{r},k)}_{\text{Local}}$, which act only on $\tilde{r}$ qubits, are IC,[8] they can be used to reconstruct the corresponding local reduced density matrices on $\tilde{r}$ neighbouring sites. We conduct tomography and certification of the n-qubit state on the basis of the estimated probabilities of the POVMs $\Pi^{(n,\tilde{r},k)}_{\text{Local}}$, which include complete information on $\tilde{r}$ neighbouring sites. These probabilities are estimated from the measurement data of the 3^r POVMs from $\mathcal{P}$ (or, equivalently, the data of $\Pi^{(n,r)}_{\text{Cyc}}$). The elements of the POVMs from Equation (7.5) give rise to the following POVM on n qubits:[9]

---

[7] Details can be found in the function `mpnum.povm.MPPovm.sample()` of the open source library `mpnum` (Suess and Holzäpfel 2017). A similar problem has been discussed in da Silva et al. 2011, Supplemental Material.

[8] See Section 3.1.3.

$$\Pi^{(n,\tilde{r})}_{\text{Block}} = \left\{ \frac{1}{n-\tilde{r}+1} \mathbb{1}_{k-1} \otimes P \otimes \mathbb{1}_{n-k-\tilde{r}+1} : P \in \Pi^{(n,\tilde{r},k)}_{\text{Local}},\ k \in \{1 \dots n-\tilde{r}+1\} \right\} \tag{7.6}$$

The measurement data, which comprise estimated probabilities for all POVMs from $\mathcal{P}$, are sufficient to estimate all probabilities of the POVM from the last equation. These are also exactly the probabilities which we use for state estimation and certification below.

7.3. State estimation

We use the scheme for scalable maximum likelihood estimation (MLE)[10] to reconstruct a pure state $|\psi_{\text{est}}\rangle$ from the estimated probabilities of the POVM (7.6). In future work, one could consider maximum likelihood estimation with more information from the measurements, such as complete information about all pairs of qubits or the full POVM (7.4). With these two options, the estimation scheme remains efficient in principle but its practical usefulness has to be evaluated.[11]

Convergence of the iterative likelihood maximization may require many iterations and a lot of computation time if it is started from a random pure MPS. To reduce computation time, we start likelihood maximization from the output of the MPS-SVT algorithm.[12] The estimated probabilities of the POVM (7.6) are used as input to the MPS-SVT reconstruction algorithm.

7.4. Certificate

Let H be an operator with two smallest eigenvalues $E_0 < E_1$ and non-degenerate ground state $|\psi_{\text{GS}}\rangle$ (H is called a *parent Hamiltonian* of $|\psi_{\text{GS}}\rangle$). For any mixed state

---

[9] Cf. Equation (3.20).

[10] Baumgratz et al. 2013b. See also Section 3.3.1.

[11] The scheme may remain efficient if the number of measurements M increases at most polynomially with the number of qubits n. For sufficiently large n, the total number of outcomes is smaller than $|\Pi^{(n,r)}_{\text{Cyc}}| = 3^r 2^n$ and it is not possible to estimate all probabilities of this POVM accurately. Nevertheless, it is possible that the resulting estimate is more accurate than an estimate based on $\Pi^{(n,r)}_{\text{Block}}$. For continuous variables, one has the similar situation that the total number of outcomes is much smaller than the (infinite) number of elements of the POVM in question, and one still obtains useful estimates; see Section 4.2.

[12] Cramer et al. 2010. See also Section 3.3.2.

ρ, we have[13]

$$\langle \psi_{\mathrm{GS}} | \rho | \psi_{\mathrm{GS}} \rangle \geq 1 - \frac{E - E_0}{E_1 - E_0} =: F_{\mathrm{LB}} \tag{7.7}$$

where $E := \mathrm{Tr}(H\rho)$ is the "energy" of ρ in terms of H. Note that H usually is artificial and unrelated to any energy in the physical system in the laboratory. Suppose that $H = \sum_{k=1}^{n-\tilde{r}+1} h_k$ where h_k acts non-trivially only on qubits $k, \dots, k + \tilde{r} - 1$. In this case, we can obtain E and the fidelity lower bound F_{LB} from the probabilities of the POVM (7.6), which can be estimated from the measurement data. In Section 7.4.1, a suitable parent Hamiltonian is constructed and in Section 7.4.2, we derive the measurement uncertainty of the fidelity lower bound F_{LB}.[14]

7.4.1. Empirical parent Hamiltonian

The following Hamiltonian is of the desired form:

$$H = \sum_{k=1}^{n-\tilde{r}+1} \mathbb{1}_{1,\dots,k-1} \otimes P_{\ker(T_\tau(\sigma_k))} \otimes \mathbb{1}_{k+\tilde{r},\dots,N}. \tag{7.8}$$

Here,

$$\sigma_k = \mathrm{Tr}_{1,\dots,k-1,k+\tilde{r},\dots,n}(|\psi_{\mathrm{est}}\rangle\langle\psi_{\mathrm{est}}|) \tag{7.9}$$

are the local reduced states of the initial estimate $|\psi_{\mathrm{est}}\rangle$. The thresholding function T_τ replaces eigenvalues of σ_k smaller than or equal to some threshold τ by zero and $P_{\ker(A)}$ denotes the orthogonal projection onto the null space of the operator A. For $\tau = 0$, $|\psi_{\mathrm{est}}\rangle$ is a ground state of H from Equation (7.8) and this ground state is non-degenerate if $|\psi_{\mathrm{est}}\rangle$ is an injective MPS.[15] A positive threshold τ can make the ground state non-degenerate even if this condition is not met.

The ground states of H depend on τ. If H has a non-degenerate ground state, we denote it by $|\psi_{\mathrm{GS}}\rangle$. We consider all possible values $\tau \geq 0$ and select all Hamiltonians with non-degenerate ground state. Among these Hamiltonians, we choose one which minimizes

$$cD(|\psi_{\mathrm{est}}\rangle, |\psi_{\mathrm{GS}}\rangle) - (E_1 - E_0) \tag{7.10}$$

---

[13] Cramer et al. 2010.

[14] The same method for deriving the measurement uncertainty of a sum of local observables from correlated measurement outcomes was subsequently used in Friis et al. 2018.

[15] Perez-Garcia et al. 2007; Baumgratz 2014. See also Section 3.4.1.

where $c > 0$ is some constant and

$$D(|\psi\rangle, |\phi\rangle) = \big\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \big\|_{(1)}/2 = \sqrt{1 - |\langle\psi|\phi\rangle|^2} \tag{7.11}$$

is the trace distance.[16] The motivation for this choice is twofold: (i) We aim for a large energy gap $E_1 - E_0$ to avoid a large measurement uncertainty in F_{LB}. (ii) The desired $F_{\mathrm{LB}} \approx 1$ requires $\rho \approx |\psi_{\mathrm{GS}}\rangle\langle\psi_{\mathrm{GS}}|$ which implies, if we assume that estimation was successful, $|\psi_{\mathrm{est}}\rangle\langle\psi_{\mathrm{est}}| \approx \rho \approx |\psi_{\mathrm{GS}}\rangle\langle\psi_{\mathrm{GS}}|$. Therefore, we aim for a small distance $D(|\psi_{\mathrm{est}}\rangle, |\psi_{\mathrm{GS}}\rangle)$ (if estimation was not successful, we do not attempt to obtain a useful certificate). We obtain a valid fidelity lower bound for any value of the constant c. However, we may obtain a very small lower bound or a lower bound associated with a large measurement uncertainty for some choices of this constant. We have used the value $c = 5$ and we do not observe significantly higher fidelity lower bounds for other values of this constant.

7.4.2. Measurement uncertainty of the fidelity lower bound

Let $\rho_k = \mathrm{Tr}_{1\dots k-1,k+\tilde{r}\dots n}(\rho)$ be the local reduced state on qubits $k, \dots, k + \tilde{r} - 1$ of the unknown state ρ. We use the following map $\mathcal{M}_k$ from a quantum state on r sites to the probabilities of the POVM $\Pi^{(n,\tilde{r},k)}_{\mathrm{Local}}$ (Equation (7.5)):

$$\mathcal{M}_k(\rho_k) := [\mathrm{Tr}(P_{\tilde{a}\tilde{s}}\rho_k)]_{P_{\tilde{a}\tilde{s}} \in \Pi^{(n,\tilde{r},k)}_{\mathrm{Local}}} \tag{7.12}$$

In this section, Pauli eigenprojectors are denoted by $P_{as} := P_{\sigma_a,s}$. As the POVMs $\Pi^{(n,\tilde{r},k)}_{\mathrm{Local}}$ are IC, the identity $\mathcal{M}_k{}^{+}\mathcal{M}_k(\rho_k) = \rho_k$ holds; here, $\mathcal{M}_k{}^{+}$ denotes the Moore–Penrose pseudoinverse of $\mathcal{M}_k$. This relation provides

$$E = \mathrm{Tr}(H\rho) = \sum_{k=1}^{N-\tilde{r}+1} \mathrm{Tr}(h_k \mathcal{M}_k{}^{+}(\mathcal{M}_k(\rho_k))) = \sum_{k=1}^{N-\tilde{r}+1} \sum_{\tilde{a}\tilde{s}} c_{k\tilde{a}\tilde{s}} p_{k\tilde{a}\tilde{s}}, \tag{7.13a}$$

$$c_{k\tilde{a}\tilde{s}} := \mathrm{Tr}(h_k \mathcal{M}_k{}^{+}(e_{\tilde{a}\tilde{s}})), \quad p_{k\tilde{a}\tilde{s}} := \mathrm{Tr}(P_{\tilde{a}\tilde{s}}\rho_k). \tag{7.13b}$$

where the sum runs over $\tilde{a} \in \{X, Y, Z\}^{\tilde{r}}$ and $\tilde{s} \in \{+1, -1\}^{\tilde{r}}$. The vectors $e_{\tilde{a}\tilde{s}} \in \mathbb{R}^{6^{\tilde{r}}}$ form a standard basis. The last equation provides a decomposition of $E = \mathrm{Tr}(H\rho)$ in terms of the probabilities $p_{k\tilde{a}\tilde{s}}$ of the POVM $\Pi^{(n,\tilde{r})}_{\mathrm{Block}}$ (Equation (7.6)).

We now spell out how to estimate the probabilities $p_{k\tilde{a}\tilde{s}}$ from the data of the POVM $\Pi^{(n,r)}_{\mathrm{Cyc}}$. In order to use the machinery from Section 1.1 for this task, we need to make some definitions. Recall that $\mathcal{P}$ contains one POVM for each

---

[16] Nielsen and Chuang 2007, Eqs. (9.60), (9.99).

$a \in \mathcal{A} = \{a \in \{X, Y, Z\}^n \colon a_k = a_{k+r}, k \in \{1 \dots n - r\}\}$ and that the POVM $\Pi^{(n,r)}_{\text{Cyc}}$ is constructed from the elements of all these POVMs. An outcome of one measurement of one POVM is given by $s \in \{+1, -1\}^n$, which is treated as integer from the set $\{1 \dots 2^n\}$ where necessary. The complete measurement data is represented by $x = (x_1, \dots, x_m)$ where m is the number of measurements per POVM. Each x_i contains one outcome of each POVM from $\mathcal{P}$, i.e. $x_i = (x_{ia})_{a \in \mathcal{A}}$ with $x_{ia} \in \{+1, -1\}^n$ or, equivalently, $x_{ia} \in \{1 \dots 2^n\}$.

The sampling distribution $p_m(x) = p(x_1) \dots p(x_m)$ is i.i.d. The outcome distribution of a single measurement of all POVMs is $p(x_i) = \prod_{a \in \mathcal{A}} p_a(x_{ia})$. The outcome of a single POVM is distributed according to $p_a(x_i) = \sum_{s=1}^{2^n} \delta(x_{ia} - s) \operatorname{Tr}(P_{as} \rho)$, where δ denotes the Dirac delta distribution.

In the following, let $y = (y_b)_{b \in \mathcal{A}}$, $y_b \in \{+1, -1\}^n$ denote one outcome from each POVM from $\mathcal{P}$. As before, let $a \in \mathcal{A}$ and $\tilde{s} \in \{+1, -1\}^{\tilde{r}}$. Define the comparison function

$$\theta_{ak\tilde{s}}(y) := \begin{cases} 1, & \text{if } (s_k, \dots, s_{k+\tilde{r}-1}) = (\tilde{s}_1, \dots, \tilde{s}_{\tilde{r}}) \\ & \text{with } y = (y_b)_{b \in \mathcal{A}} \text{ and } y_a =: (s_1, \dots, s_n), \\ 0, & \text{otherwise.} \end{cases} \tag{7.14}$$

The set of POVMs from $\mathcal{P}$ which can be used to estimate a given $p_{k\tilde{a}\tilde{s}}$ is indexed by

$$\mathcal{A}_{k\tilde{a}} := \left\{ a \in \mathcal{A} \colon a_{k+i-1} = \tilde{a}_i, i \in \{1 \dots \tilde{r}\} \right\}. \tag{7.15}$$

The set $\mathcal{A}_{k\tilde{a}}$ has $3^{r-\tilde{r}}$ elements, i.e. it has at least one element. Let $a \in A_{k\tilde{a}}$, then

$$\mathbb{E}_{p_m}(\mathbb{E}_x(\theta_{ak\tilde{s}})) = \mathbb{E}_p(\theta_{ak\tilde{s}}) = \sum_{s \in \mathcal{S}_{\tilde{s}}} \operatorname{Tr}(\rho P_{as}) = \operatorname{Tr}(\rho_k P_{\tilde{a}\tilde{s}}) = p_{k\tilde{a}\tilde{s}} \tag{7.16}$$

where $\mathcal{S}_{\tilde{s}} = \{s \in \{+1, -1\}^n \colon s_{k+i-1} = \tilde{s}_i, i \in \{1 \dots \tilde{r}\}\}$. In the four steps, we used (i) Equation (1.14a) (ii) definitions, (iii) the definition of P_{as} and $a \in \mathcal{A}_{k\tilde{a}}$ and (iv) definitions. Further, we set

$$q_{k\tilde{a}\tilde{s}}(y) := \frac{1}{|\mathcal{A}_{k\tilde{a}}|} \sum_{a \in \mathcal{A}_{k\tilde{a}}} \theta_{ak\tilde{s}}(y), \tag{7.17}$$

which preserves the property $\mathbb{E}_{p_m}(\mathbb{E}_x(q_{k\tilde{a}\tilde{s}}(y))) = p_{k\tilde{a}\tilde{s}}$. Finally, we are able to define an estimator for $E = \operatorname{Tr}(H\rho)$ (the coefficients are defined in Equation (7.13b)):

$$\epsilon(x) := \mathbb{E}_x(f_\epsilon), \quad f_\epsilon(y) := \sum_{k=1}^{n-r+1} \sum_{\tilde{a}\tilde{s}} c_{k\tilde{a}\tilde{s}} q_{k\tilde{a}\tilde{s}}(y). \tag{7.18}$$

Table 7.1: Estimators of the energy E which enters the fidelity lower bound (Equation (7.7)) and of the variance/standard deviation of the estimated energy. f_ϵ is defined in Equation (7.18).

Estimand	Estimator	Name	Bias
$E = \operatorname{Tr}(H\rho)$	$\epsilon(x) = \mathbb{E}_x(f_\epsilon)$	Mean	Unbiased
$v(\epsilon) = \mathbb{V}_{p_m}(\epsilon)$	$v_\epsilon(x) = \frac{1}{m-1}\mathbb{V}_x(f_\epsilon)$	Variance (of the mean)	Unbiased
$\sqrt{v(\epsilon)}$	$\sqrt{v_\epsilon(x)}$	Standard deviation (of the mean)	Biased

$\mathbb{E}_{p_m}(\epsilon) = E$ immediately follows because expectation values are linear. $\frac{1}{m-1}\mathbb{V}_x(f_\epsilon)$ provides an unbiased estimator of our estimate's variance (Section 1.1). The results are summarized in Table 7.1. Naturally, the standard deviation of the fidelity lower bound F_{LB} is obtained from that of E by dividing by the energy gap $E_1 - E_0$.

7.5. Results

In this section, we discuss results from the ion trap quantum simulator experiment presented by Lanyon, Maier, et al. (2017) and compare the experimental results to numerical simulations. Recall that the ideal time-evolved state $|\psi(t)\rangle$ is given by $|\psi(t)\rangle = \exp(-\mathrm{i}H_{\mathrm{Ising}}t/\hbar)|\psi(0)\rangle$ where $|\psi(0)\rangle = |{\uparrow\downarrow\uparrow\downarrow}\dots\rangle$ is a Néel-ordered state. For simulations of 8 and 14 qubits, the Hamiltonian was constructed as a sparse matrix and time evolution was simulated with a library function.[17] Measurements on the time-evolved state $|\psi(t)\rangle$ were simulated as discussed above and provide complete information on r neighbouring qubits. State estimation and certification was performed using complete information on $\tilde{r}$ qubits where $\tilde{r} \le r$. We set $r = 3$ for $\tilde{r} \in \{1, 2, 3\}$ and $r = \tilde{r}$ for $\tilde{r} \in \{3, 4, 5\}$. In order to ensure that our estimate of the fidelity lower bound F_{LB} is unbiased, we split the $m = 1000$ measurement outcomes into two parts of 500 outcomes, of which the first (second) part is used for estimation (certification). From measurement data, we obtain the initial estimate $|\psi_{\mathrm{est}}\rangle$ (Section 7.3) and the parent Hamiltonian ground state $|\psi_{\mathrm{GS}}\rangle$ (Section 7.4). As explained above, the parent Hamiltonian provides the lower bound

$$\langle\psi_{\mathrm{GS}}|\rho|\psi_{\mathrm{GS}}\rangle \ge F_{\mathrm{LB}} \tag{7.19}$$

---

[17] `scipy.sparse.linalg.expm_multiply` from SciPy (Jones et al. 2001).

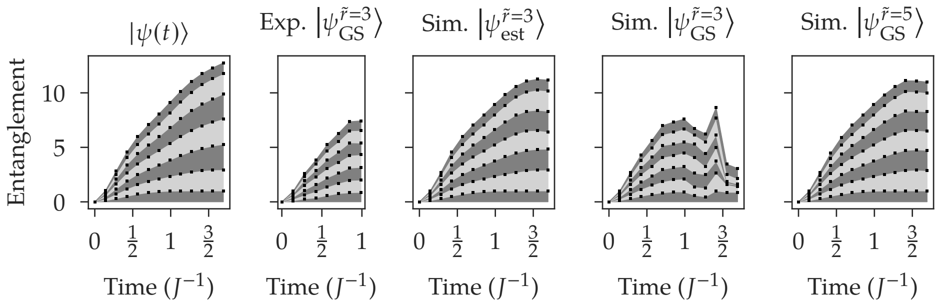


Figure 7.1: Bipartite entanglement of the ideal time-evolved state $|\psi(t)\rangle$ (column 1) and certified/initial estimates (columns 2–5). The width of each stripe indicates the entanglement across a bipartition $1 \dots k|k+1 \dots n$, as quantified by the von Neumann entropy of a corresponding reduced state (binary logarithm). Columns 2–5 show the entropy of certified estimates $|\psi_{\mathrm{GS}}\rangle$, for which the fidelity lower bounds presented below hold, and of initial estimates $|\psi_{\mathrm{est}}\rangle$.

on the fidelity of its ground state with the unknown state ρ in the laboratory. Therefore, we call $|\psi_{\mathrm{GS}}\rangle$ the *certified estimate* of the unknown state.

Figure 7.1 compares the bipartite entanglement found in the ideal time-evolved state $|\psi(t)\rangle$ with that of initial and certified estimates. Entanglement grows with time in the ideal state (column 1 in the figure) and the certified estimated $|\psi^{\tilde{r}=3}_{\mathrm{GS}}\rangle$ based on experimental data reproduces this growth except for the last time step where experimental data is available (column 2). In the numerical simulation which uses complete information on $\tilde{r} = 3$ neighbouring sites, growing entanglement can be estimated up to about $t \approx \frac{3}{2} J^{-1}$, but it can be certified only up to $t \approx J^{-1}$ ($|\psi^{\tilde{r}}_{\mathrm{est}}\rangle$ and $|\psi^{\tilde{r}=3}_{\mathrm{GS}}\rangle$ in columns 3 and 4). Complete information on $\tilde{r} = 5$ sites is sufficient to certify growing entanglement for longer times ($|\psi^{\tilde{r}=3}_{\mathrm{GS}}\rangle$ in column 5).

Figure 7.2 shows fidelity lower bounds F_{LB} obtained in simulations and experiments with 8 qubits. Solid lines show lower bounds obtained with simulations assuming exact knowledge of probabilities (of the POVM (7.6), which has $(n-\tilde{r}+1)6^{\tilde{r}}$ elements). Shaded areas show the corresponding expected standard deviation of the fidelity lower bound if estimation is performed with exact probabilities and certification is performed with 500 measurement outcomes. This standard deviation accounts for imperfect knowledge of the energy in terms of the parent Hamiltonian but it does not account for changes in the parent Hamiltonian due to imperfect knowledge of measured probabilities used for state estimation. To also

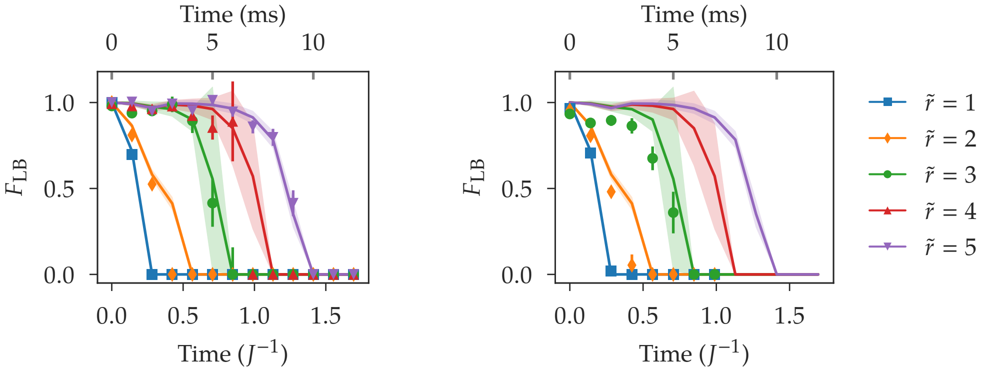


Figure 7.2: Certified tomography of 8 qubits.[18] Symbols: Tomography and certification with 500 measurements each in simulation (left) and experiment (right). Solid lines: Simulation of certified tomography with exactly known probabilities. Shaded areas: Expected standard deviation of the fidelity lower bound if estimated from 500 measurements.

incorporate the latter, symbols show fidelity lower bounds where estimation and certification have been performed with 500 measurement outcomes each. Here, the error bars indicate an estimated standard deviation for repetitions of certification with *the same* parent Hamiltonian, *without* repeating state estimation. The symbols and error parts in the left and right parts of Figure 7.2 show results from numerical simulations and from the experiment, respectively.

In Figure 7.2, fidelity stays close to unity for some time and then drops to zero rather quickly; the time at which it drops increases with $\tilde{r}$, i.e. as complete information on more neighbouring sites is used. The bond dimension of the estimated state was constrained to $D = 4$. Non-zero fidelity lower bounds can be obtained for longer times from $\tilde{r} = 5$ if this bond dimension is increased: In this case, certification was not limited by the available measurements but by the constraint imposed during state estimation. The experimental results, which are indicated by the symbols in Figure 7.2 right, are quite similar to the results of the numerical simulation.

Figure 7.3 shows results of a numerical simulation involving 14 qubits and complete measurements on $r = \tilde{r} = 3$ qubits. Lines (solid, dotted or dashed), shaded areas and symbols have the same meaning as before. The figure shows results from simulations based on two different Hamiltonians and initial states. The

---

[18] The right part of the figure reproduces Figure 3 from Lanyon, Maier, et al. 2017 with modifications. The left part of the figure is based on data from the same work.

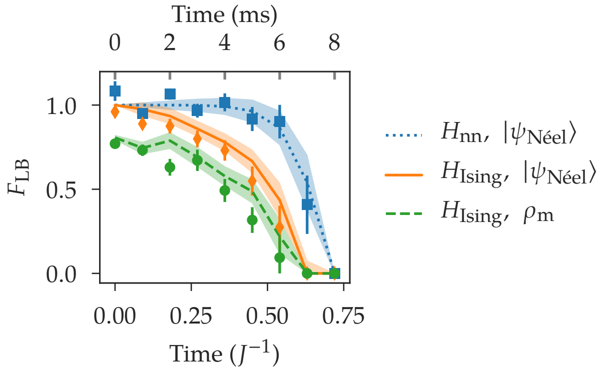


Figure 7.3: Certified tomography of 14 qubits (numerical simulation, $\tilde{r} = 3$). The figure compares tomography of time-evolved states under different Hamiltonians (H_{Ising}, H_{nn}) and starting from different initial states ($|\psi_{Néel}\rangle$, ρ_m). Dotted, solid and dashed lines: Certified tomography with exactly known probabilities. Shaded areas: Expected variance of the fidelity lower bound if estimated from 500 measurements. Symbols: Tomography and certification with 500 measurements each.

Hamiltonian H_{nn} is obtained from H_{Ising} by removing all non-nearest-neighbour couplings (i.e. by zeroing all J_{ij} with $|i - j| > 1$). The mixed initial state ρ_m is a mixture of the Néel-ordered state $|\uparrow\downarrow\uparrow\downarrow\dots\rangle$ and the same state with one or two spins flipped[19] where the weights are given by the experimental measurement outcomes of $\Pi^{(n)}_{\sigma_a}$, $a = (Z, Z, \dots, Z)$ at $t = 0$: Zero, one and two spin flips were observed 893, 93 and 12 times. Results for the nearest-neighbour Hamiltonian and the Néel-ordered initial state are quite similar to the results for 8 qubits (Figure 7.2) in that the fidelity lower bound stays close to unity before dropping to zero after some time. Results for the full Hamiltonian differ in that fidelity starts dropping earlier at a slower rate. Under the mixed initial state, fidelities do not exceed about 0.8 at any time. Out of the three models considered here, the simulations involving the mixed initial state most closely resembled the experimental results. In the experiment, the fidelity lower bound $F_{LB} = 0.39 \pm 0.08$ was achieved at $t = 4$ ms. The fidelity of the estimated state at $t = 4$ ms was also estimated with direct fidelity estimation (DFE),[20] obtaining a result of 0.74 ± 0.05.[21] Here, DFE required a much larger number of measurements not constrained to three neighbouring qubits, illustrating that measurements on three neighbouring qubits are, in this case, sufficient for reasonable state estimation but may not be sufficient for certification.

---

[19] The time evolution of each pure state involved in the mixture was simulated with the library function mentioned above.

[20] da Silva et al. 2011; Flammia and Liu 2011.

[21] Lanyon, Maier, et al. 2017.

7.6. Conclusion

The results presented above demonstrate that quantum state tomography of certain states can be achieved without complete information. Certifying the estimated state eliminates the need to assume that the unknown state belongs to a suitable class. Specifically, this was demonstrated for product states evolved in time under a Hamiltonian with decaying couplings. Single-qubit observables were seen to be sufficient to estimate and certify the product state at $t = 0$ and measurements on more neighbouring qubits were seen to admit tomography for longer times; this is in qualitative agreement with the predictions from Chapter 6 based on Lieb–Robinson bounds. There is good agreement between theory and experiment for 8 qubits but the results for 14 qubits also show room for improvement in theory, data analysis and experiment. E.g. regarding data analysis, the certified fidelity of around 0.8 at $t = 0$ is unsatisfactory because a fidelity of about 0.893 with the ideal initial state can be inferred directly from the named measurement results.

There are several avenues for future work. In order to obtain an unbiased estimate of the fidelity lower bound, we split the data into two halves, where one is used for state estimation and one is used for certification. A more precise estimate can be obtained by using the same data for state estimation and certification, necessitating an investigation of the statistical bias of the resulting estimator. In the case of a nearest-neighbour Hamiltonian, the results from Chapter 6 guarantee successful certification if the block size scales logarithmically with system size and linearly with time. Future work could determine how the parameter range where the numerical method from this chapter is successful compares with the parameter range with guaranteed certification from Chapter 6. In an experiment, the unknown state is always slightly or less slightly mixed. Estimating a mixed state instead of a perfectly pure state, together with a suitable certificate, can increase the overall performance of the method. State estimation of mixed states can be performed with MPO-MLE, PMPS-MLE or density matrix reconstruction.[22] One can then certify the main component of the mixed state with the above method, attempt certification with a degenerate parent Hamiltonian[23] or use other methods.[24]

---

[22] Baumgratz et al. 2013a,b; Holzäpfel et al. 2018. See also Section 3.3.1 and Chapter 8.

[23] Flammia et al. 2010.

[24] Kim 2014. See also Section 3.4.2.

Chapter 8.

Tensor reconstruction

The number of entries in a tensor grows exponentially with the number of tensor indices. In this chapter, we reconstruct a tensor from few entries or other linear functionals.[1] The necessary number of entries or linear functionals is determined by certain tensor unfolding ranks and can be much smaller than the total number of entries. The necessary computation time and storage complexity are reduced substantially because they are determined by the number of entries used for reconstruction. Combining aspects of several previous contributions,[2] we derive a framework which allows for reconstruction of the tensor in the Tucker, hierarchical Tucker and tensor train (TT)/matrix product state (MPS)/matrix product operator (MPO) representations.

A tensor M is a collection of complex numbers organized by n indices,

$$M \in \mathbb{C}^{d_1 \times \cdots \times d_n}. \tag{8.1}$$

The number n is also called the number of *modes*. For simplicity, we set $d := d_1 = \cdots = d_n$ in this introduction, such that the tensor M has exactly d^n components. As mentioned in Section 2.2, the density matrix describing the quantum state of n quantum systems of Hilbert space dimension d_H may be represented as a tensor with $d = (d_H)^2$. Noting that the set $\mathbb{C}^{d_1 \times \cdots \times d_n}$ is easily equipped with the properties of a vector space, consider a linear function

$$\mathcal{F} \colon \mathbb{C}^{d_1 \times \cdots \times d_n} \to \mathbb{C}^m. \tag{8.2}$$

We shall refer to $\mathcal{F}$ as *linear measurement* on the tensor M. Now, reconstructing the tensor M from the vector $\mathcal{F}(M) \in \mathbb{C}^m$ is not possible unless both the number of components m and the rank of $\mathcal{F}$ are at least d^n (otherwise, the linear system $\mathcal{F}(M) = y$ has more than one solution). In the case $\mathrm{rk}(\mathcal{F}) < d^n$, we say that the value of $\mathcal{F}(M)$ does not provide complete information on the tensor M because it is insufficient to infer M. However, M may well be inferred from $\mathcal{F}(M)$ if additional constraints are known. As an example, M may be known to contain at most ten non-zero entries. In this case, $\mathcal{F}(M)$ is sufficient to infer M if it contains ten non-zero entries of M, independently of the rank of $\mathcal{F}$.

---

[1] Section 8.1 reproduces parts of the original publication Holzäpfel et al. 2018 with the permission of AIP Publishing.

[2] Caiafa and Cichocki 2010; Oseledets and Tyrtyshnikov 2010; Baumgratz et al. 2013a; Ballani et al. 2013; Caiafa and Cichocki 2015; Holzäpfel et al. 2018.

In order to explain the constraints we use for tensor reconstruction, we need to introduce the concept of a *tensor unfolding* or *matricization*. In the special case $n = 5$, the $\{1, 3\}$-unfolding of M is the matrix

$$\mathcal{M}_{\{1,3\}}(M) \in \mathbb{C}^{d_1 d_3 \times d_2 d_4 d_5} \tag{8.3}$$

with the same entries as M. This definition is easily generalized to a linear map $\mathcal{M}_s$ for arbitrary n and $s \subset \Lambda := \{1, 2, \ldots, n\}$ (see Section 8.2.1). For reconstruction, we demand constraints of the form

$$\mathrm{rk}(\mathcal{M}_s(M)) = \mathrm{rk}\left(L_s[\mathcal{M}_s(M)]R_s\right) \tag{8.4}$$

for certain subsets $s \subset \Lambda$ where L_s and R_s are matrices which can be chosen arbitrarily. The matrices L_s and R_s shall be called *measurement matrices* or just *measurements*. Our reconstruction procedure then takes linear measurements of the form

$$\left[\mathcal{F}(M)\right]_{sij} = \left[L_s[\mathcal{M}_s(M)]R_s\right]_{ij} \tag{8.5}$$

as well as other, similarly constructed linear functions of M as input (see Section 8.2.2). As a special case, the matrices L_s, R_s and the linear map $\mathcal{F}$ can be chosen such that $\mathcal{F}(M)$ contains only selected entries of M. It is always possible to choose L_s and R_s such that the condition (8.4) is satisfied if the number of rows in L_s and the number of columns in R_s are not smaller than the rank of $\mathcal{M}_s(M)$. Suppose that the rank of $\mathcal{M}_s(M)$ is much smaller than its maximally possible value for all relevant subsets $s \subset \Lambda$ and that measurements are chosen accordingly. In this case, our reconstruction procedure enables reconstruction of a tensor M from a vector $\mathcal{F}(M) \in \mathbb{C}^m$ where $\mathcal{F}$ is a suitably defined linear map and where m scales e.g. only linearly in n. Hence, m can be much smaller than the total number of components d^n of M.

Our tensor reconstruction method allows for different choices of the relevant subsets $s \subset \Lambda$ used in (8.4). It will become clear that the representation of the reconstructed tensor depends on this choice and that the MPS/TT, Tucker and hierarchical Tucker representations are possible options. For example, the mentioned linear dependence of m on n can be achieved in the MPS/TT representation, extending e.g. previous results by Baumgratz et al. (2013a). Indeed, our approach to tensor reconstruction has been pursued several times before and we relate our results to a selection of previous works[3] in Section 8.2.6.

A reconstruction method for low-rank matrices provides the basis of our tensor reconstruction method and is introduced in Section 8.1. The tensor reconstruction method itself is introduced, analyzed and compared to previous work in Section 8.2.

8.1. Reconstruction of low-rank matrices

In this section, we present a method which allows to recover a low-rank matrix $M \in \mathbb{C}^{m \times n}$ from less than mn entries or from less than mn linear measurements in the sense of (8.2).[1] The rank of a matrix is considered as low if it is much smaller than its maximally possible value $\min\{m, n\}$. Section 8.1.1 discusses the reconstruction method for low-rank matrices while Section 8.1.2 considers matrices which are approximately low rank. A matrix is approximately low rank if it differs from a low-rank matrix only by a small amount (as quantified by a given norm).

8.1.1. Reconstruction

Let $M \in \mathbb{C}^{m \times n}$ be a matrix. Its matrix rank is equal to r if and only if it can be written as

$$M = \sum_{i=1}^{r} u_i v_i^*, \quad u_i \in \mathbb{C}^m, \quad v_i \in \mathbb{C}^n \tag{8.6}$$

where each of the sets $\{u_1, \dots, u_r\}$ and $\{v_1, \dots, v_r\}$ is linearly independent.[4] Suppose that M has small rank in the sense that r is much smaller than m and n, i.e. $r \ll m, n$. In this case, the last equation illustrates that the low-rank matrix M, which contains mn complex numbers, can be specified in terms of only $(m+n)r \ll mn$ complex numbers. What is more, it is well known that all entries of M can reconstructed from a suitable selection of $(m+n)r$ of its entries.[5] A possible proof uses the following Proposition 8.1: If L and R are submatrices of permutation matrices, the matrix products LM and MR are given by certain rows and columns of M. Choosing $r = \operatorname{rk}(M)$ linearly independent rows and columns of M, the proposition reconstructs M from said rows and columns. The resulting decomposition of M is known as pseudoskeleton decomposition or CUR decomposition.[6]

---

[3] Oseledets et al. 2008; Caiafa and Cichocki 2010, 2015; Oseledets and Tyrtyshnikov 2010; Baumgratz et al. 2013a; Ballani et al. 2013; Savostyanov and Oseledets 2011; Espig et al. 2012; Holzäpfel et al. 2018.

[4] E.g. Horn and Johnson 1991a, 0.4.5(c).

Proposition 8.1 *Let* $L \in \mathbb{C}^{r\times m}$*,* $M \in \mathbb{C}^{m\times n}$ *and* $R \in \mathbb{C}^{n\times s}$ *be matrices. Then*

$$\operatorname{rk}(LMR) = \operatorname{rk}(M) \quad \Leftrightarrow \quad \exists\, X \in \mathbb{C}^{s\times r} \colon M = MR\,X\,LM. \tag{8.7}$$

If the condition is satisfied, $M = MR\,X\,LM$ *holds for any matrix* X *with* $CXC = C$*,* $C = LMR$*. The Moore–Penrose pseudoinverse* $X = (LMR)^+$ *has the required property* $CXC = C$*.*

Furthermore, $\operatorname{rk}(LM) = \operatorname{rk}(M)$ *implies* $\operatorname{rk}(LMR) = \operatorname{rk}(MR)$*.*

Proof "$\Rightarrow$" of Equation (8.7): Assume that $\operatorname{rk}(LMR) = \operatorname{rk}(M)$ holds. The property $CXC = C$, $C = LMR$ implies that $LMRXu = u$ holds for all $u \in \operatorname{im}(LMR)$. Let $q := \operatorname{rk}(M) = \operatorname{rk}(LMR)$. Let $u_1, \dots, u_q$ be a basis of $\operatorname{im}(LMR)$ and set $v_i := Xu_i$, $w_i := MRv_i$ (here and in the following, $i \in \{1, \dots, q\}$). The v_i are linearly independent because $LMRv_i = LMRXu_i = u_i$. The w_i are linearly independent because $Lw_i = LMRv_i = u_i$. The w_i are a linearly independent sequence of length $q = \operatorname{rk}(M)$ and they satisfy $w_i \in \operatorname{im}(M)$, i.e. they are a basis of $\operatorname{im}(M)$. Now observe

$$MR\,X\,Lw_i = MR\,Xu_i = MR\,v_i = w_i. \tag{8.8}$$

As a consequence, $MR\,X\,L$ maps any vector from $\operatorname{im}(M)$ to itself. Accordingly, $(MR\,X\,L)M = M$ holds.

$\operatorname{rk}(LM) = \operatorname{rk}(M)$ implies $\operatorname{rk}(LMR) = \operatorname{rk}(MR)$: The equality $\operatorname{rk}(LM) = \operatorname{rk}(M)$ implies $M = M(LM)^+LM$ (use the "$\Rightarrow$" direction of Equation (8.7) for $R = \mathbb{1}$). As a consequence, $MR = M(LM)^+LMR$ and $\operatorname{rk}(MR) \le \operatorname{rk}(LMR)$ hold. The converse inequality $\operatorname{rk}(LMR) \le \operatorname{rk}(MR)$ always holds and we arrive at $\operatorname{rk}(LMR) = \operatorname{rk}(MR)$.

"$\Leftarrow$" of Equation (8.7): Assume that $M = MR\,X\,LM$ holds for some matrix X. The equality $M = MR\,X\,LM$ implies $\operatorname{rk}(M) \le \operatorname{rk}(MR)$ and $\operatorname{rk}(M) \le \operatorname{rk}(LM)$. The converse inequalities $\operatorname{rk}(MR) \le \operatorname{rk}(M)$ and $\operatorname{rk}(ML) \le \operatorname{rk}(M)$ always hold. As a consequence, we have $\operatorname{rk}(LM) = \operatorname{rk}(M)$ and $\operatorname{rk}(MR) = \operatorname{rk}(M)$. Above, we saw that the former equality implies $\operatorname{rk}(LMR) = \operatorname{rk}(MR)$ which, together with the latter equality $\operatorname{rk}(MR) = \operatorname{rk}(M)$, proves the theorem. ■

---

[5] E.g. Goreinov et al. 1997.

[6] Goreinov et al. 1997; Mahoney et al. 2008.

Remark 8.2 A violation of the rank condition $\operatorname{rk}(LMR) = \operatorname{rk}(M)$ does not in general imply that there is no method to obtain M from LM and MR. As a trivial example, consider $L = \mathbb{1}$ and $R = 0$. Then, the rank condition is violated for all $M \neq 0$, but M is obtained trivially from $LM = M$. □

Remark 8.3 Several aspects of Proposition 8.1 have been covered in prior work.[7] This is discussed in more detail by Holzäpfel et al. (2018, Remark 5) and in Section 8.2.6 below. □

8.1.2. Stability of low-rank matrix reconstruction

Suppose that

$$\operatorname{rk}(LSR) = \operatorname{rk}(S) \tag{8.9}$$

holds for some matrix S, allowing for reconstruction of S from LS and SR via Proposition 8.1. This rank condition is not robust in the following sense: Consider the perturbed matrix

$$M = S + E, \quad \epsilon = \frac{\|E\|_{(\infty)}}{\|S\|_{(\infty)}} \tag{8.10}$$

where ϵ quantifies the relative magnitude of the perturbation E in the operator norm. The rank condition (8.9) may not hold for M even if ϵ is arbitrarily close to zero, such that and Proposition 8.1 does not allow for reconstruction of M. The following Theorem 8.4 provides a reconstruction of M with bounded error if ϵ is small enough, extending prior work by Caiafa and Cichocki (2015). A bound on the distance between the reconstruction $\check{M}_\tau$ and M is provided by

$$\|\check{M}_\tau - M\|_{(\infty)} \leq \|\check{M}_\tau - S\|_{(\infty)} + \|S - M\|_{(\infty)} \leq \|\check{M}_\tau - S\|_{(\infty)} + \epsilon\|S\|_{(\infty)} \tag{8.11}$$

and Theorem 8.4 provides a bound on $\|\check{M}_\tau - S\|_{(\infty)}$.

In the remainder of this section, the operator norm (largest singular value) is denoted by $\|\cdot\| := \|\cdot\|_{(\infty)}$. Recall that given a matrix M, we define $M_\tau^+ := (M_\tau)^+$ and M_τ is given by M with singular values smaller than or equal to τ replaced by zero.

---

[7] Goreinov et al. 1997; Baumgratz et al. 2013a; Caiafa and Cichocki 2015.

Theorem 8.4 *Let $M = S + E$, $\mathrm{rk}(S) = \mathrm{rk}(LSR)$, $\eta = \|L\|\|S\|\|R\|$, $\gamma = \sigma_{min}(LSR)/\eta$, $\epsilon = \|E\|/\|S\|$. Let $\gamma > 2\epsilon$ and $\epsilon \le \tau < \gamma - \epsilon$. Then*

$$\|\check{M}_\tau - S\| \le \frac{1}{\gamma - \epsilon}\left(\frac{4\epsilon}{\gamma} + 2\epsilon + \epsilon^2\right)\|S\| \le \frac{7\epsilon}{\gamma(\gamma - \epsilon)}\|S\|,$$

where $\check{M}_\tau = MR(LMR)^+_{\eta\tau}LM$.

Proof We prove the proposition for $\|L\| = \|S\| = \|R\| = 1$ (without loss of generality as explained in Remark 8.7). We have $M = S + E$, $\epsilon = \|E\|$, $\gamma = \sigma_{\min}(LSR)$ and $LMR = LSR+LER$ with $\|LER\| \le \epsilon$. We insert $S = SR(LSR)^+LS$ and use Lemma 8.6 (provided at the end of this subsection):

$$\begin{aligned}
&\|MR(LMR)^+_\tau LM - S\| && (8.12)\\
&\le \|(LMR)^+_\tau - (LSR)^+\| + 2\|(LMR)^+_\tau\|\epsilon + \|(LMR)^+_\tau\|\epsilon^2 && (8.13)\\
&\le \frac{4\epsilon}{\gamma(\gamma - \epsilon)} + \frac{2\epsilon}{\gamma - \epsilon} + \frac{\epsilon^2}{\gamma - \epsilon} \le \frac{7\epsilon}{\gamma(\gamma - \epsilon)}. && (8.14)
\end{aligned}$$

Note that by premise, we have $\epsilon < \gamma \le 1$. As a consequence, $1 \le \frac{1}{\gamma}$ and $\epsilon \le 1$ (which were used in the last equation) hold. This proves the theorem. ■

Remark 8.5 For the interpretation of the theorem, it is convenient to use the case with $\|L\| = \|S\| = \|R\| = 1$ and $\eta = 1$. Theorem 8.4 shows that the reconstruction $\check{M}_\tau$ reconstructs the low-rank component S of $M = S + E$ up to a small error if the smallest singular value γ of the low-rank component LSR is much larger than the norm ϵ of the noise component. In addition, the threshold τ must be chosen larger than the noise norm ϵ but smaller than $\gamma - \epsilon$. The appendix of Holzäpfel et al. (2018) discusses examples which show that the bound from Theorem 8.4 is optimal up to constants and that the reconstruction error can diverge as ϵ approaches zero if small singular values in LMR are not truncated. □

The following Lemma was used in the proof of Theorem 8.4:

Lemma 8.6 *Let $A, B, F \in \mathbb{C}^{m\times n}$. Let $\gamma = \sigma_{min}(A)$, $B = A + F$ with $\|F\| \le \epsilon$. Let $\gamma > 2\epsilon$ and choose τ such that $\epsilon \le \tau < \gamma - \epsilon$. Then $\sigma_{min}(B_\tau) \ge \gamma - \epsilon$, $\|B^+_\tau\| \le 1/(\gamma - \epsilon)$, $B_\tau = B_\epsilon$ and $\|B - B_\tau\| \le \epsilon$. In addition,*

$$\|B^+_\tau - A^+\| \le \frac{4\epsilon}{\gamma(\gamma - \epsilon)}. \tag{8.15}$$

Proof The singular values of $B = A + F$ satisfy[8]

$$|\sigma_i(A+F) - \sigma_i(A)| \leq \sigma_1(F) \leq \epsilon$$

and therefore, with $r = \mathrm{rk}(A)$, we obtain

$$\sigma_1(B) \geq \ldots \geq \sigma_r(B) \geq \gamma - \epsilon > \tau \geq \epsilon \geq \sigma_{r+1}(B) \geq \ldots \geq 0.$$

This shows already everything except the inequality in Equation (8.15). To show the latter one, we use[9] $\|X^+ - A^+\| \leq 2\|X^+\|\|A^+\|\|X - A\|$. Inserting $\|B_\tau - A\| \leq \|B_\tau - B\| + \|B - A\|$ shows the desired inequality. ■

Remark 8.7 In this Remark, we provide an argument which extends the proof of Theorem 8.4 from matrices S with unit operator norm to matrices S with arbitrary operator norm. Suppose that the matrix M is the sum of a signal S and a noise contribution E, $M = S + E$. The signal satisfies $\mathrm{rk}(S) = \mathrm{rk}(LSR)$, but we only know the strength $\|E\|$ of the noise. Suppose that for $\|S\| = \|L\| = \|R\| = 1$, we obtain some error bound of the form

$$\|MR(LMR)^+_\tau LM - S\| \leq f(\epsilon, \gamma, \tau), \quad \epsilon = \|E\|, \quad \gamma = \sigma_{\min}(LMR). \tag{8.16}$$

We can obtain an error bound for $M' = S' + E'$ where S', L', and R' have arbitrary norms as follows: Set $M = M'/\|S'\|$, $S = S/\|S'\|$, $E = E'/\|S'\|$ $L = L'/\|L'\|$, $R = R'/\|R'\|$. With these definitions, we have

$$\|MR(LMR)^+_\tau LM - S\| = \frac{\|M'R'(L'M'R')^+_{\tau'} L'M' - S'\|}{\|S'\|}, \tag{8.17}$$

where $\tau' = \|L\|\|R\|\|S\|\tau$. Therefore, the bound from the last but one equation implies

$$\|M'R'(L'M'R')^+_{\tau'} L'M' - S'\| \leq \|S'\| f(\epsilon, \gamma, \tau), \tag{8.18}$$

$$\epsilon = \frac{\|E'\|}{\|S'\|}, \quad \gamma = \frac{\sigma_{\min}(L'M'R')}{\|L'\|\|R'\|\|S'\|}, \quad \tau = \frac{\tau'}{\|L'\|\|R'\|\|S'\|}. \tag{8.19}$$

This enables us to assume $\|S\| = \|L\| = \|R\| = 1$ and $\epsilon = \|E\|$ in the above proof. □

8.2. Tensor reconstruction

In the following, we present a method for tensor reconstruction which is advantageous especially for low-rank tensors. As in the case of matrices, a tensor is considered as being low-rank if its ranks (defined below, cf. also (8.4)) are much smaller than their respective maximal values. The reconstruction method relies on notation and basic facts presented in Section 8.2.1. After that, we discuss the reconstruction method and its extension to recursively defined measurements in Sections 8.2.2 and 8.2.3, where the term *measurement* is used in the sense of Equations (8.2) and (8.5). Recursively defined measurements appear to relax the rank conditions necessary for reconstruction, but our analysis in Section 8.2.4 shows that this is not actually the case. Sections 8.2.5 and 8.2.6 conclude by considering further types of recursively defined measurements and by relating the results from this section to previous work.

---

[8] See e.g. Stewart 1977.

[9] Wedin 1973; Stewart 1977.

8.2.1. Notation and basic facts

A tensor is basically a matrix with more than two indices. For a tensor M with n indices, we write

$$M \in \mathbb{C}^{d_1 \times \dots \times d_n}, \qquad M_{i_1 \dots i_n} \in \mathbb{C}, \qquad i_k \in \{1 \dots d_k\}, \qquad k \in \{1 \dots n\}, \tag{8.20}$$

where d_k is the size of the k-th index i_k.[10] In quantum mechanics, the tensor M could contain e.g. the components of a pure state vector of n quantum systems of dimensions $d_1, \dots, d_n$: $M_{i_1 \dots i_n} = \langle i_1 \dots i_n | \psi \rangle$. Note that in related literature,[11] the same notation is used, except for d and n being exchanged:

$$M' \in \mathbb{C}^{n_1 \times \dots \times n_d}, \qquad M'_{i_1 \dots i_d} \in \mathbb{C}, \qquad i_k \in \{1 \dots n_k\}, \qquad k \in \{1 \dots d\}. \tag{8.21}$$

We shall stick to the notation from (8.20) where the tensor M has n indices, which is also used throughout the rest of this work.

We refer to the n indices of a tensor as n *modes* and the set of all modes is denoted by

$$\Lambda := [n] := \{1, 2, \dots, n\}. \tag{8.22}$$

The complement of a subset $t \subset \Lambda$ is denoted by t^c and we define the sizes

$$d_t := \prod_{k \in t} d_k, \qquad d'_t := \prod_{k \in t^c} d_k, \qquad t^c := \Lambda \setminus t. \tag{8.23}$$

---

[10] The following nomenclature is also used: n is also called the dimension of M or the number of modes. d_k ($k \in \{1 \dots n\}$) is called the size of the k-th index, the size of the k-th mode or the dimension of the k-th mode. See also the beginning of Chapter 2.

[11] E.g. Ballani et al. 2013.

When we work with a matrix such as $A \in \mathbb{C}^{a_1 a_2 a_3 \times c}$, we assume that we can also treat it as a four-mode tensor from $\mathbb{C}^{a_1 \times a_2 \times a_3 \times c}$:

Definition 8.8 (Matrices with tensor structure) For matrices whose dimensions are given by products of integers, a decomposition of column and row indices into corresponding multi-indices shall be available. For example, components of a matrix $A \in \mathbb{C}^{a_1 a_2 a_3 \times b}$ may be denoted as $A_{i_1 i_2 i_3, c}$ where $i_k \in [a_k]$, $c \in [b]$ and $k \in [3]$. The order of rows in A is arbitrary but has to remain fixed. □

This definition enables a commutative notation for the tensor product $A \otimes B$:

Definition 8.9 (Non-standard tensor product) Matrix products of tensor products and matrices are evaluated according to the variable names that occur in the shape specification of the involved matrices. Example: A product involving $A \in \mathbb{C}^{a_1 a_2 a_3 \times b}$ and $X^{(k)} \in \mathbb{C}^{x_k \times a_k}$ ($k \in [3]$) shall be defined as

$$\left[\left(X^{(3)} \otimes X^{(1)} \otimes X^{(2)}\right) A\right]_{y_3 y_1 y_2, c} = \sum_{i_1, i_2, i_3} X^{(1)}_{y_1, i_1} X^{(2)}_{y_2, i_2} X^{(3)}_{y_3, i_3} A_{i_1 i_2 i_3, c}. \tag{8.24}$$

where $y_k \in [x_k]$, $i_k \in [a_k]$ and $c \in [b]$. This differs from the standard definition of the tensor product.[12] □

Definition 8.10 (Matricization, tensorization) A *tensor unfolding* or *matricization* of a tensor M is the matrix $\mathcal{M}_t(M) \in \mathbb{C}^{d_t \times d'_t}$ with the same entries as M, where $t \subset \Lambda$.[13] We formalize this definition using suitable multi-indices I_t and I'_t:

$$[\mathcal{M}_t(M)]_{I_t, I'_t} := M_{i_1, \dots, i_n}, \qquad I_t := (i_k)_{k \in t}, \qquad I'_t := (i_k)_{k \in t^c}. \tag{8.25}$$

We can choose to order the entries of $\mathcal{M}_t(M)$ lexicographically according to the multi-indices I_t and I'_t but any other defined order is equally permissible. Elsewhere, matricizations are often denoted as $M^{(t)} := \mathcal{M}_t(M)$ or $M_{(t)} := \mathcal{M}_t(M)$. Below, we also use the notation $M_t := \mathcal{M}_t(M)$.

The *tensorization* of a matrix $A \in \mathbb{C}^{d_t \times d'_t}$ is the tensor $\mathcal{T}_t(A) \in \mathbb{C}^{d_1 \times \dots \times d_n}$ with the same components, i.e.

$$[\mathcal{T}_t(A)]_{i_1, \dots, i_n} := A_{I_t, I'_t}. \tag{8.26}$$

---

[12] The *tensor product, Kronecker product* or *direct product* $A \otimes B$ of two matrices $A \in \mathbb{C}^{a_1 \times a_2}$ and $B \in \mathbb{C}^{b_1 \times b_2}$ is elsewhere defined as the matrix $A \otimes B \in \mathbb{C}^{a_1 b_1 \times a_2 b_2}$ with the entries $A_{i_1, i_2} B_{j_1, j_2} =: (A \otimes B)_{b_1(i_1 - 1) + j_1, b_2(i_2 - 1) + j_2}$ where $i_1 \in [a_1]$, $i_2 \in [a_2]$, $j_1 \in [b_1]$, $j_2 \in [b_2]$. Cf. e.g. Horn and Johnson 1991b, Definition 4.2.1.

[13] E.g. Grasedyck 2010, Definition 2.2.

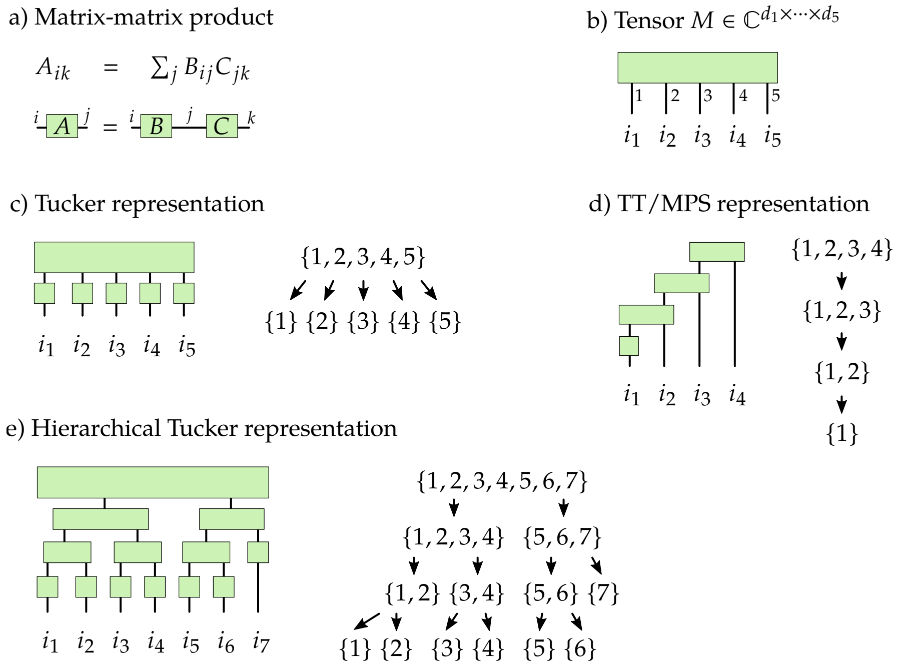

a) Matrix-matrix product
$A_{ik} = \sum_j B_{ij} C_{jk}$
i A j = i B j C k
b) Tensor $M \in \mathbb{C}^{d_1 \times \cdots \times d_5}$
1 2 3 4 5
i_1 i_2 i_3 i_4 i_5
c) Tucker representation
i_1 i_2 i_3 i_4 i_5
{1, 2, 3, 4, 5}
{1} {2} {3} {4} {5}
d) TT/MPS representation
i_1 i_2 i_3 i_4
{1, 2, 3, 4}
{1, 2, 3}
{1, 2}
{1}
e) Hierarchical Tucker representation
i_1 i_2 i_3 i_4 i_5 i_6 i_7
{1, 2, 3, 4, 5, 6, 7}
{1, 2, 3, 4} {5, 6, 7}
{1, 2} {3, 4} {5, 6} {7}
{1} {2} {3} {4} {5} {6}

Figure 8.1: (a) Graphical representation of a matrix-matrix product, which constitutes the basic building block of the graphical representations of tensor networks. (b) Graphical representations of a tensor with five modes (indices). (c–e) Dimension trees and corresponding tensor representations.

Tensorization is the inverse operation of matricization. Strictly speaking, we should write $\mathcal{M}_t^{d_1 \times \cdots \times d_n}(M)$ and $\mathcal{T}_t^{d_1 \times \cdots \times d_n}(A)$ because both operations depend on the mode sizes $(d_1, \ldots, d_n)$. To enhance readability and because the mode sizes can usually be inferred from the context, we use the shortened notation $\mathcal{M}_t(M)$ and $\mathcal{T}_t(A)$. □

Example 8.11 As an example, consider a matrix $A \in \mathbb{C}^{m_{12} m_3 d_4 \times m'_{1234}}$ where d_4 is the size of mode 4 and m_{12}, m_3 and m'_{1234} are certain integers (this example is taken from the proof of Lemma 8.26 below). This matrix can be tensorized and unfolded into the matrix $\mathcal{M}_{m_{12}} \mathcal{T}_{m_{12} m_3 d_4} A$, where we have already abused notation by referencing modes not by their label (1, 2, etc.) but by the variable which determines their size (d_1, d_2, m_3, etc.). For the tensorization we shall always introduce one tensor mode for each integer, obtaining tensors and matrices of the following shapes:

$$\mathcal{T}_{m_{12} m_3 d_4} A \in \mathbb{C}^{m_{12} \times m_3 \times d_4 \times m'_{1234}}, \qquad \mathcal{M}_{m_{12}} \mathcal{T}_{m_{12} m_3 d_4} A \in \mathbb{C}^{m_{12} \times m_3 d_4 m'_{1234}}. \tag{8.27}$$

Definition 8.12 (Tensor-matrix-product) The product of a matrix $A \in \mathbb{C}^{m \times d_k}$ with mode k of a tensor $\mathbb{C}^{d_1 \times \cdots \times d_n}$ is denoted by $M \times_k A$ and its components are given by[14]

$$(M \times_k A)_{i_1,\ldots,i_n} := \sum_j M_{i_1 \ldots i_{k-1} j i_{k+1} \ldots i_n} A_{i_k j}. \tag{8.28}$$

As was mentioned above, the number of elements of a tensor with n modes increases exponentially with n. To reduce the storage cost where possible and to make handling tensors with many modes practical, tensors are represented by a collection of tensors with fewer modes which, when matrix-matrix products are evaluated between certain modes, return the original tensor. This approach is illustrated in Figure 8.1. For more details on the MPS/TT representation and the graphical notation shown in the figure, refer to Sections 2.2 and 2.2.2, respectively. In order to describe different tensor representations formally, we define a dimension tree as follows:

Definition 8.13 (Dimension tree) Given a finite set of modes Λ, a dimension tree T is a finite set with the following properties:[15]

1. The root of the tree is $\Lambda \in T$.
2. Each element $t \in T$ is a non-empty subset $t \subset \Lambda$.
3. There is a successor function $S\colon T \to T \cup \{\emptyset\}$ with the following properties:

   For all $t \in T$ and $s \in S(t)$, $s \subsetneq t$ holds. In addition, the intersection of any two elements of $S(t)$ is empty.
4. For each $s \in T$, either $s = \Lambda$ or there is a $t \in \Lambda$ such that $s \in S(t)$. (All elements $s \in T$ can be reached from the root Λ via the successor function). □

Example 8.14 (Tensor representations) In Figure 8.1, every dimension tree element $t \in T$ is associated to one tensor in a tensor network representation of the original tensor $M \in \mathbb{C}^{d_1 \times \cdots \times d_n}$. The sizes of open modes are fixed to $(d_1, \ldots, d_n)$ while the sizes of modes connecting two tensors are free parameters which determines the complexity of the representation. In the MPS/TT dimension tree, for example, the size of modes connecting two tensors are the MPS bond dimensions or TT representation ranks. If the MPS bond dimensions are bounded by a constant, the storage cost of the representation is only linear in n (instead of exponential). Figure 8.1 shows the following tensor representations:

---

[14] E.g. Lathauwer et al. 2000, Definition 8.

[15] Cf. Ballani et al. (2013). They impose an additional restriction on dimension trees, requiring either (i) $S(t) = \emptyset$ or (ii) $S(t) = \{t_1, t_2\}$ with $t = t_1 \dot{\cup} t_2$ (Ballani et al. 2013, Definition 3).

- Matrix product state (MPS)/tensor train (TT):[16]
  $S(\{1,\dots,k\}) = \{\{1,\dots,k-1\}\}$ for $k \in \{2,\dots,n\}$ and $S(\{1\}) = \emptyset$.
- Tucker representation:[17] $S(\Lambda) = \{\{1\},\{2\},\dots,\{n\}\}$ and $S(\{i\}) = \emptyset$, $i \in \Lambda$.
- Hierarchical Tucker representation:[18] A binary tree with half of each node $t \in T$ as successors and with single-mode leaves. □

Below, we use the following Lemma, whose basic idea is not new and which is closely related to the Tucker representation and the related, so-called higher-order singular value decomposition (HO-SVD):[19]

Lemma 8.15 *Let $M \in \mathbb{C}^{d_1 \times d_2 \times d_3}$ be a tensor with three modes and denote unfoldings by $M^{(t)} := \mathcal{M}_t M$, $t \subset \{1,2,3\}$. Then*

$$M^{(12)} = \big[[M^{(1)}(M^{(1)})^+] \otimes [M^{(2)}(M^{(2)})^+]\big] M^{(12)}, \tag{8.29}$$

or, in other words, $\operatorname{im}(M^{(12)}) \subset \operatorname{im}(M^{(1)}) \otimes \operatorname{im}(M^{(2)})$.

Proof We take a part of the right-hand side and write it as mode-1 product:

$$\begin{aligned}\mathcal{T}_{12}\Big([[M^{(1)}(M^{(1)})^+] \otimes \mathbb{1}_{d_2}]M^{(12)}\Big) &= M \times_1 [M^{(1)}(M^{(1)})^+] = \mathcal{T}_1\Big(M^{(1)}(M^{(1)})^+ M^{(1)}\Big) \\ &= \mathcal{T}_1(M^{(1)}) = M.\end{aligned} \tag{8.30}$$

Taking the mode-12 unfolding of the left and right hand sides, we obtain[20]

$$[[M^{(1)}(M^{(1)})^+] \otimes \mathbb{1}_{d_2}]M^{(12)} = M^{(12)}. \tag{8.31}$$

In the same way, we obtain $[\mathbb{1}_1 \otimes [M^{(2)}(M^{(2)})^+]]M^{(12)} = M^{(12)}$. This completes the proof. ■

---

[16] E.g. Schollwöck 2011; Oseledets 2011. See also Section 2.2.

[17] Tucker 1966. Lathauwer et al. (2000) further cite the earlier publication Tucker 1964.

[18] Hackbusch and Kühn 2009; Grasedyck 2010.

[19] The transformation in Equation (8.30) was used e.g. by Tucker (1966, Eqs. (17)–(19)) and Lathauwer et al. (2000, Eqs. (14)–(15)).

[20] Written in components, the argument runs as follows: By i, j, k we denote indices referring to the first, second and third modes, respectively. We have $(M^{(12)})_{ij,k} = M_{ijk} = (M^{(1)})_{i,jk}$, where the comma separates row and column indices of a matrix. Note that $(M^{(1)})_{i,jk} = [M^{(1)}(M^{(1)})^+ M^{(1)}]_{i,jk} = \sum_{i'} [M^{(1)}(M^{(1)})^+]_{i,i'} (M^{(1)})_{i',jk}$ holds. Inserting $(M^{(12)})_{ij,k} = (M^{(1)})_{i,jk}$ on both sides yields the desired $M^{(12)} = [[M^{(1)}(M^{(1)})^+)] \otimes \mathbb{1}_{d_2}]M^{(12)}$.

8.2.2. Tensor reconstruction

Our main result on tensor reconstruction, Theorem 8.17, uses the following Definition 8.16, which is illustrated in Figure 8.2.

Definition 8.16 (Reconstruction setting) Let $M \in \mathbb{C}^{d_1 \times \cdots \times d_n}$ be a tensor with unfoldings denoted by $M_t = \mathcal{M}_t M \in \mathbb{C}^{d_t \times d'_t}$, $t \subset \Lambda$. Let T be a dimension tree with successor function S.

Set $m'_\Lambda := 1$ and $R_\Lambda := 1$. Let $t \in T$ and $s \in S(t)$. Choose positive integers m_s, m'_s and matrices

$$L_s \in \mathbb{C}^{m_s \times d_s}, \qquad\qquad R_s \in \mathbb{C}^{d'_s \times m'_s}. \tag{8.32}$$

For the successors of t in some arbitrary, fixed order, we write $\{s_1, \dots, s_h\} := S(t)$ where $h := |S(t)|$. The subset of t which is not covered by any successor is given by $x := t \setminus \bigcup_{s \in S(t)} s$. Keep in mind that both h and x depend on t. We define matrices

$$A_t \in \mathbb{C}^{m_{s_1} \dots m_{s_h} d_x \times m'_t}, \qquad B_s \in \mathbb{C}^{m_s \times m'_s}, \qquad C_t \in \mathbb{C}^{d_t \times m'_t} \tag{8.33}$$

as[21]

$$A_t := \left[\left(\bigotimes\nolimits_{s \in S(t)} L_s \right) \otimes \mathbb{1}_{d_x} \right] M_t R_t, \qquad\qquad B_s := L_s M_s R_s \tag{8.34}$$

and (see Figure 8.2)

$$C_t := \left[\left(\bigotimes\nolimits_{s \in S(t)} C_s (B_s)^+ \right) \otimes \mathbb{1}_{d_x} \right] A_t. \tag{8.35}$$

The recursion in C_t stops at leaves t, for which $S(t) = \emptyset$ and $C_t = A_t = M_t R_t$ hold.□

Theorem 8.17 *Assume Definition 8.16. If the rank conditions*

$$\operatorname{rk}(M_s) = \operatorname{rk}(L_s M_s R_s), \quad t \in T, s \in S(t) \tag{8.36}$$

hold, then $C_t = M_t R_t$ ($t \in T$) and $C_\Lambda = M_\Lambda$ hold, i.e. M can be reconstructed from A_t and B_s ($t \in T$, $s \in S(t)$).

Remark 8.18 As mentioned in the introduction, we call the matrices L_s and R_s *measurement matrices* or just *measurements*. The measurement L_s acts on the mode set $s \subset \Lambda$ to the right and its output size is given by the integer m_s. Conversely, the measurement R_s acts on the mode set $s^c = \Lambda \setminus s$ to the left, and its output size is given by the integer m'_s. The matrices A_t and B_s are linear functions of M and they can both be represented by a linear measurement function $\mathcal{F}\colon \mathbb{C}^{d_1\times\cdots\times d_n} \to \mathbb{C}^m$ as defined in (8.2) in the introduction. Theorem 8.17 enables us to reconstruct M using the tensorization $M = \mathcal{T}_\Lambda(C_\Lambda)$, where C_Λ can be constructed using only A_t and B_s ($t \in T$, $s \in S(t)$). Furthermore, a tensor representation in accordance with the dimension tree T can be directly constructed from A_t and B_s. The representation ranks (or "bond dimensions") can be chosen as m_s or m'_s, as illustrated by Figure 8.2c and d. If the tensor ranks $\operatorname{rk}(M_s)$ and the measurement output sizes are sufficiently small, this representation can be much more efficient than a full representation of M as multi-dimensional array. By using the corresponding dimension tree, the reconstructed tensor can be represented e.g. as MPS/TT and an example of this case is presented below in Example 8.20. □

---

[21] Below, we may sometimes omit the identity matrix $\mathbb{1}_{d_x}$.

Proof (of Theorem 8.17) As a consequence of Lemma 8.15,

$$M_t = \Big[\bigotimes\nolimits_{s\in S(t)} M_s (M_s)^+\Big] M_t, \quad t \in T. \tag{8.37}$$

Due to the rank conditions, Proposition 8.1 for $L = L_s$, $M = M_s$ and $R = R_s$ implies[22]

$$M_s = M_s R_s \; (L_s M_s R_s)^+ \; L_s M_s, \qquad t \in T, s \in S(t). \tag{8.38}$$

Inserting (8.38) into (8.37) provides

$$\begin{aligned} M_t &= \Big[\bigotimes\nolimits_{s\in S(t)} M_s R_s \; (L_s M_s R_s)^+ \; L_s \; M_s (M_s)^+\Big] \; M_t \\ &= \Big[\bigotimes\nolimits_{s\in S(t)} M_s R_s \; (L_s M_s R_s)^+ \; L_s\Big] \; M_t, \end{aligned} \tag{8.39}$$

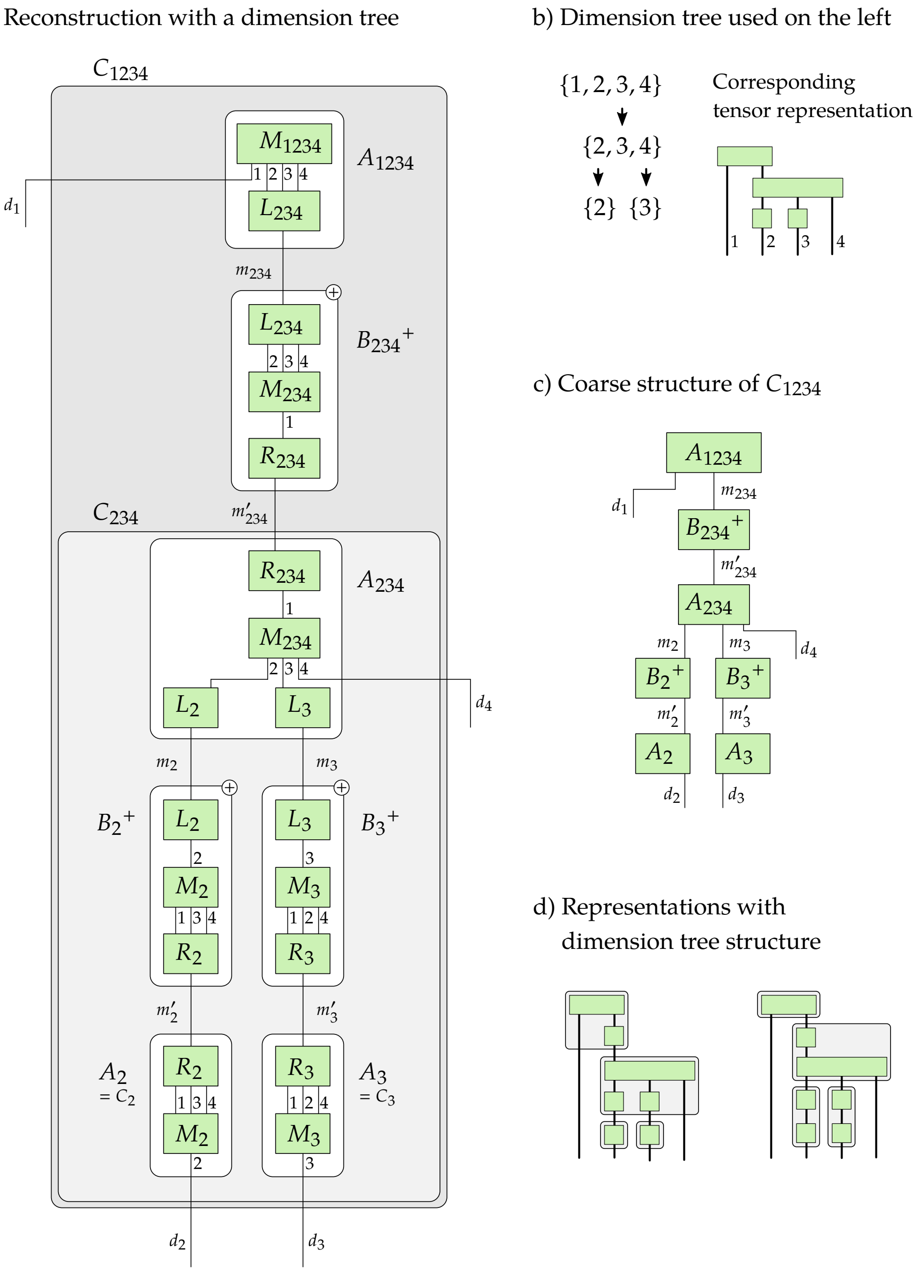
where the second line has been obtained by using (8.37) again. Multiplying with R_t provides

$$\begin{aligned} M_t R_t &= \Big[\bigotimes\nolimits_{s\in S(t)} M_s R_s \; (L_s M_s R_s)^+ \; L_s\Big] \; M_t R_t \\ &= \Big[\bigotimes\nolimits_{s\in S(t)} M_s R_s \; (B_s)^+ \;\Big] \; A_t. \end{aligned} \tag{8.40}$$

Recall that $C_t = M_t R_t$ holds if t is a leaf, i.e. if $S(t) = \emptyset$. We show that $C_t = M_t R_t$ holds for all $t \in T$ with induction from the leaves to the root of the tree: Let $t \in T$ and suppose that $C_s = M_s R_s$ holds for all $s \in S(t)$. Then, (8.40) shows that $C_t = M_t R_t$ holds. This completes the proof. ■

---

[22] Spaces in formulas were added to improve readability.



a) Reconstruction with a dimension tree

b) Dimension tree used on the left

c) Coarse structure of C_{1234}

d) Representations with dimension tree structure

C_{1234}
M_{1234}
A_{1234}
1 2 3 4
d_1
L_{234}
m_{234}
L_{234}
$B_{234}{}^+$
2 3 4
M_{234}
1
R_{234}
C_{234}
m'_{234}
R_{234}
A_{234}
1
M_{234}
2 3 4
L_2
L_3
d_4
m_2
m_3
$B_2{}^+$
L_2
L_3
$B_3{}^+$
2
3
M_2
M_3
1 3 4
1 2 4
R_2
R_3
m'_2
m'_3
A_2
$= C_2$
R_2
R_3
A_3
$= C_3$
1 3 4
1 2 4
M_2
M_3
2
3
d_2
d_3
$\{1,2,3,4\}$
$\{2,3,4\}$
$\{2\}$ $\{3\}$
Corresponding tensor representation
1 2 3 4
A_{1234}
d_1
m_{234}
$B_{234}{}^+$
m'_{234}
A_{234}
m_2
m_3
d_4
$B_2{}^+$
$B_3{}^+$
m'_2
m'_3
A_2
A_3
d_2
d_3

Figure 8.2: (a) Structure of the tensor reconstruction C_{1234} from Theorem 8.17 for the exemplary dimension tree shown in (b). (c) Coarse structure of C_{1234}. (d) Tensors can be contracted in two different ways to obtain a tensor representation which adheres to the structure from (b). The figure uses the shortened notation $C_{\{1,2,3,4\}} =: C_{1234}$, $L_{\{2,3,4\}} =: L_{234}$, etc.

Remark 8.19 Theorem 8.17 and $C_t = M_t R_t$ still hold if we set

$$C_t = \Big[\Big(\bigotimes\nolimits_{s \in S(t)} C_s (L_s C_s)^+ \Big) \otimes \mathbb{1}_{d_x} \Big] A_t, \tag{8.41}$$

eliminating the need for B_s. This also applies to Theorem 8.23 below. It is not clear which ansatz may yield a better reconstruction if the entries of A_t and B_s are perturbed by noise and this should be investigated in practical applications. □

Example 8.20 In order to illustrate Theorem 8.17, we use it to derive the matrix product operator (MPO) reconstruction scheme proposed by Baumgratz et al. (2013a). Let $\rho \in \mathcal{B}(\mathcal{H}_{1,\dots,n})$ be a linear operator on a tensor product $\mathcal{H}_{1,\dots,n} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ of n finite-dimensional Hilbert spaces $\mathcal{H}_k$, $k \in [n]$. For example, ρ could be a density matrix describing the state of n quantum systems. For notational simplicity, we assume that all Hilbert spaces have the same dimension d_H. Set $d := (d_H)^2$ and let $\{F_1^{(k)}, \dots, F_d^{(k)}\}$ be an operator basis of $\mathcal{B}(\mathcal{H}_k)$. Further, let M be the tensor which contains ρ's coefficients in the tensor product of these operator bases, i.e.

$$\rho = \sum_{i_1=1}^{d} \cdots \sum_{i_n=1}^{d} M_{i_1 \dots i_n} F_{i_1}^{(1)} \otimes \cdots \otimes F_{i_n}^{(n)}. \tag{8.42}$$

The linear operator ρ is thus represented by the tensor M with n modes of sizes $d_1 = \cdots = d_n = d = (d_H)^2$. In order to apply Theorem 8.17, we make the following definitions. We choose integers $l, r \geq 1$ such that $l + r \leq n - 2$. The set of modes is $\Lambda := [n]$ and we use the following successor function:

$$S(\Lambda) := \{1 \dots n - r - 1\}, \tag{8.43a}$$

$$S([k]) := [k - 1] \qquad (k \in \{l + 2 \dots n - r - 1\}), \tag{8.43b}$$

$$S([l + 1]) := \emptyset. \tag{8.43c}$$

This is the MPS/TT dimension tree introduced in Example 8.14 with the first $l + 1$ and last $r + 1$ sites treated as a single site. In the following, let $k \in \{l + 1, \dots, n - r - 1\}$. We choose the matrices $L_{[k]}$ and $R_{[k]}$ by means of the following linear maps:[23]

$$\tilde{L}_{[k]}\colon \quad \mathcal{B}(\mathcal{H}_{1\dots k}) \to \mathcal{B}(\mathcal{H}_{k-l+1\dots k}), \quad \sigma \mapsto \operatorname{Tr}_{1\dots k-l}(\sigma). \tag{8.44a}$$

$$\tilde{R}'_{[k]}\colon \quad \mathcal{B}(\mathcal{H}_{k+1\dots n}) \to \mathcal{B}(\mathcal{H}_{k+1\dots k+r}), \quad \sigma \mapsto \operatorname{Tr}_{k+r+1\dots n}(\sigma). \tag{8.44b}$$

---

[23] The partial trace, reduced density matrices and reduced linear operators were defined in Section 1.3.

Let $L_{[k]}$ and $R'_{[k]}$ be the matrix representations of $\tilde{L}_{[k]}$ and $\tilde{R}'_{[k]}$ in the operator bases $\{F^{(k)}_{i_k}\}$ from above. This defines $L_{[k]}$ and we set $R_{[k]} = (R'_{[k]})^\top$. This result in matrices

$$L_{[k]} \in \mathbb{C}^{d^l \times d^k}, \qquad R_{[k]} \in \mathbb{C}^{d^{n-k} \times d^r}. \tag{8.45}$$

Correspondingly, we have to set $m_{[k]} := d^l = (d_H)^{2l}$ and $m'_k := d^r = (d_H)^{2r}$. Applying Theorem 8.17 results in matrices of the sizes

$$A_{[k]} \in \mathbb{C}^{d^{l+1} \times d^r}, \qquad B_{[k]} \in \mathbb{C}^{d^l \times d^r}, \qquad C_{[k]} \in \mathbb{C}^{d^k \times d^r} \tag{8.46}$$

as well as

$$A_\Lambda \in \mathbb{C}^{d^l d^{r+1} \times 1}, \qquad C_\Lambda \in \mathbb{C}^{d^n \times 1}. \tag{8.47}$$

Clearly, $A_{[k]}$, $B_{[k]}$ and A_Λ are representations of the reduced density matrices (or linear operators)[23] $\rho_{k-l\dots k+r}$, $\rho_{k-l+1\dots k+r}$ and $\rho_{n-l-r\dots n}$. The rank condition (8.36) can be rewritten in terms of the operator Schmidt rank (OSR, see Section 1.3) as

$$\mathrm{OSR}(1\dots k : k+1\dots n)_\rho = \mathrm{OSR}(k-l+1\dots k : k+1\dots k+r)_\rho \tag{8.48}$$

where, as before, $k \in \{l+1 \dots n-r-1\}$.

If the rank conditions are satisfied, the equality $M_\Lambda = C_\Lambda$ is provided by Theorem 8.17. The sought linear operator ρ can thus be obtained from its representation C_Λ, which can be constructed from the matrices A_t and B_s ($t \in T$, $s \in S(t)$). The named matrices A_t and B_s correspond the reduced density matrices mentioned in the last paragraph. In conclusion, ρ can be reconstructed from reduced density matrices on at most $l + r + 1$ neighbouring sites if the rank conditions (8.48) hold. Furthermore, the structure of C_Λ is such that an MPS/TT representation of M, which corresponds to an MPO representation of ρ, can be constructed directly and the bond dimension of the representation can be chosen as $m_{[k]} = (d_H)^{2l}$ or $m'_{[k]} = (d_H)^{2r}$. This provides an alternate proof of the result by Baumgratz et al. (2013a, Theorem 1).[24] □

8.2.3. Reconstruction with recursive measurements

In this section, we augment the results from the previous section with a result which exploits a recursive structure in the measurement matrices R_s. The two results are compared in the following Section 8.2.4.

---

[24] For the resulting bond dimension or representation rank, see Remark 8.18 and Figure 8.2. To reproduce the cited result exactly, we have to adapt our definitions to $k \in \{l+1, \dots, n-r-1\}$ and $S([l]) := \emptyset$ in (8.43) and to $k \in \{l, \dots, n-r-1\}$ in the remainder of this example.

Definition 8.21 (Extended reconstruction setting) Assume the reconstruction setting from Definition 8.16 and use the following procedure to choose the matrices R_s: Set $E_\Lambda := 1$ and $\tilde{m}_\Lambda := 1$. Let $t \in T$ and $s \in S(t)$. Choose integers $\tilde{m}_s$ and set $y := s^c \setminus t^c = t \setminus s$ (whose dependence on s and t shall be kept in mind). Recalling d_t and d'_t from (8.23), choose matrices

$$F_s \in \mathbb{C}^{\tilde{m}_t d_y \times \tilde{m}_s}, \qquad G_s \in \mathbb{C}^{\tilde{m}_s \times m'_s} \tag{8.49}$$

and define

$$E_s \in \mathbb{C}^{d'_s \times \tilde{m}_s}, \qquad R_s \in \mathbb{C}^{d'_s \times m'_s} \tag{8.50}$$

via

$$E_s := (E_t \otimes \mathbb{1}_{d_y})F_s, \qquad R_s := E_s G_s. \tag{8.51}$$

Proof For $t \in T$ and $s \in S(t)$, $s \subset t$ and $t^c \subset s^c$ hold. Therefore, $d'_t d_y = d_{t^c} d_y = d_{s^c} = d'_s$ holds for $y = s^c \setminus t^c$. Induction starting at the root $\Lambda \in T$ shows that the definition of E_s yields matrices of the claimed sizes. ■

Remark 8.22 The main idea of Definition 8.21 is as follows. For $\tilde{m}_s = m'_s$ and $G_s = \mathbb{1}$, we have $R_s = E_s = (R_t \otimes \mathbb{1}_{d_y})F_s$. This means that the transformation R_s which "acts on" s^c consists of first applying R_t, which "acts on" the subset $t^c \subsetneq s^c$, followed by the additional transformation F_s. The measurements R_s are therefore defined recursively.

The factor G_s has been included in R_s because it allows to recover Definition 8.16 as a possible special case of Definition 8.21 as follows. Set $\tilde{m}_s = d'_s$, which implies $\tilde{m}_t d_y = d_{t^c} d_{s^c \setminus t^c} = d'_s = \tilde{m}_s$, turning F_s into a square matrix. Setting $F_s = \mathbb{1}$ then implies $E_s = \mathbb{1}$ and $R_s = G_s$, which recovers Definition 8.16. □

Theorem 8.23 *Assume Definition 8.21 (recalling $M_s := \mathcal{M}_s(M)$). If the rank conditions*

$$\operatorname{rk}(M_s(E_t \otimes \mathbb{1}_{d_y})) = \operatorname{rk}(L_s M_s R_s), \quad t \in T, s \in S(t) \tag{8.52}$$

hold, then $C_t = M_t R_t$ ($t \in T$) and $C_\Lambda = M_\Lambda$ hold, i.e. M can be reconstructed from A_t and B_s ($t \in T$, $s \in S(t)$).

Proof For brevity, we omit the identity matrix $\mathbb{1}_{d_y}$ in this proof. Let $t \in T$ and $S(t) \neq \emptyset$. Consider a tensor N with mode sizes $\tilde{m}_t \times (\times_{i \in t} d_i)$ whose t-unfolding is given by $N_t := \mathcal{M}_t(N) = M_t E_t$. Let $s \in S(t)$, i.e. $s \subset t$. Then the s-unfolding of N is given by $N_s = M_s(E_t \otimes \mathbb{1}_{d_y})$ where $y = s^c \setminus t^c$ (from now on, we omit $\mathbb{1}_{d_y}$ again). As a consequence of Lemma 8.15,

$$M_t E_t = N_t = \Big[\bigotimes_{s \in S(t)} N_s (N_s)^+\Big] N_t = \Big[\bigotimes_{s \in S(t)} M_s E_t (M_s E_t)^+\Big] M_t E_t. \tag{8.53}$$

Inserting $R_s = E_t F_s G_s$ into the rank condition provides[25]

$$\operatorname{rk}(M_s E_t) = \operatorname{rk}(L_s \; M_s E_t \; F_s G_s), \quad t \in T, \; s \in S(t). \tag{8.54}$$

Proposition 8.1 for $L = L_s$, $M = M_s E_t$ and $R = F_s G_s$ implies

$$\begin{aligned} M_s E_t &= M_s E_t \; F_s G_s (L_s \; M_s E_t \; F_s G_s)^+ L_s \; M_s E_t \\ &= M_s R_s \; (L_s M_s R_s)^+ \; L_s \; M_s E_t \,, \qquad\qquad t \in T, \; s \in S(t). \end{aligned} \tag{8.55}$$

Inserting (8.55) into (8.53) provides

$$M_t E_t = \Big[\bigotimes_{s \in S(t)} M_s R_s \; (L_s M_s R_s)^+ \; L_s \; M_s E_t \; (M_s E_t)^+\Big] \; M_t E_t. \tag{8.56}$$

Using (8.53) again, we obtain

$$M_t E_t = \Big[\bigotimes_{s \in S(t)} M_s R_s \; (L_s M_s R_s)^+ \; L_s\Big] \; M_t E_t \tag{8.57}$$

and, multiplying with $F_s G_s$,

$$\begin{aligned} M_t R_t = M_t E_t F_s G_s &= \Big[\bigotimes_{s \in S(t)} M_s R_s \; (L_s M_s R_s)^+ \; L_s\Big] \; M_t R_t \\ &= \Big[\bigotimes_{s \in S(t)} M_s R_s \; (B_s)^+ \;\Big] \; A_t. \end{aligned} \tag{8.58}$$

Recall that $C_t = M_t R_t$ holds if t is a leaf, i.e. if $S(t) = \emptyset$. We show that $C_t = M_t R_t$ holds for all $t \in T$ with induction from the leaves of the tree to its root: Let $t \in T$ and suppose that $C_s = M_s R_s$ holds for all $s \in S(t)$. Then, (8.58) shows that $C_t = M_t R_t$ holds. This completes the proof. ■

Example 8.24 To illustrate Theorem 8.23 and to compare it with Theorem 8.17, we take the same example as before, the MPO reconstruction scheme proposed by Baumgratz et al. (2013a). We use the same definitions as in Example 8.20 but provide a recursive construction to define the same matrices $R_{[k]}$ as before. Let $k \in \{l+1 \dots n-r-1\}$. Set $\tilde{m}_{[k]} = m'_{[k]} = d^r$, $G_{[k]} = \mathbb{1}_{m'_{[k]}}$ and

$$\tilde{F}'_{[k]}\colon \quad \mathcal{B}(\mathcal{H}_{k+1\dots k+r+1}) \to \mathcal{B}(\mathcal{H}_{k+1\dots k+r}), \quad \sigma \mapsto \operatorname{Tr}_{k+r+1}(\sigma). \tag{8.59}$$

---

[25] Spaces in formulas were added to improve readability.

Similarly as before, let $F'_{[k]}$ be the matrix representation of $\tilde{F}'_{[k]}$ in the operator bases $\{F^{(k)}_{i_k}\}$ and set $F_{[k]} = (F'_{[k]})^\top$. This results in matrices

$$F_{[k]} \in \mathbb{C}^{d^{r+1} \times d^r}. \tag{8.60}$$

It remains to show that $E_s = R_s$ with E_s from Definition 8.21 and R_s from Example 8.20 holds for $s = [k]$, where $k \in \{l + 1 \dots n - r - 1\}$. For $k = n - r - 1$, we have $y = \{n - r \dots n\}$, $d_y = d^{r+1}$ and $E_{[k]} = F_{[k]} = R_{[k]}$ clearly holds. Suppose that $E_{[k+1]} = R_{[k+1]}$ holds for some $k \in \{l + 1 \dots n - r - 2\}$. We have

$$E_{[k]} = (E_{[k+1]} \otimes \mathbb{1}_{d_y}) F_{[k]} \tag{8.61}$$

where $y = \{k+1\}$ and $d_y = d$. In addition, $E_{[k+1]} = R_{[k+1]}$ is the matrix representation of

$$\tilde{R}'_{[k+1]}: \quad \mathcal{B}(\{k+2 \dots n\}) \to \mathcal{B}(\{k+2 \dots k+r+1\}), \quad \sigma \mapsto \mathrm{Tr}_{k+r+2 \dots n}(\sigma). \tag{8.62}$$

Taking a closer look reveals that $E_{[k]}$ is the matrix representation of $\tilde{R}'_{[k]}$ and that $E_{[k]} = R_{[k]}$ holds. In other words, the matrices $R_{[k]}$ constructed recursively in this example equal the matrices $R_{[k]}$ from Example 8.20. However, the rank conditions (8.52) correspond to ($k \in \{l + 1 \dots n - r - 1\}$)

$$\mathrm{OSR}(1 \dots k : k+1 \dots k+r+1)_\rho = \mathrm{OSR}(k-l+1 \dots k : k+1 \dots k+r)_\rho, \tag{8.63}$$

which differ from the conditions (8.48) obtained in the previous example. Apart from that, the density matrix can be reconstructed in the same way as in the mentioned example if either set of conditions holds.

The conditions (8.48) were assumed by the original proposal of MPO reconstruction,[26] whereas the conditions (8.63) were shown sufficient in a more recent work.[27] One notices that for each $k \in \{l + 1 \dots n - r - 2\}$, the condition in (8.63) is weaker than the corresponding condition in (8.48).[28] However, we set out to show that the two condition sets are indeed equivalent in the next subsection. □

---

[26] Baumgratz et al. 2013a, Theorem 1.

[27] Holzäpfel et al. 2018, Theorem 17.

[28] For each value of k, (8.48) implies (8.63): $\mathrm{OSR}(k-l+1 \dots k : k+1 \dots k+r)_\rho \le \mathrm{OSR}(1 \dots k : k+1 \dots k+r+1) \le \mathrm{OSR}(1 \dots k : k+1 \dots n) = \mathrm{OSR}(k-l+1 \dots k : k+1 \dots k+r)$, where the inequalities follow from the data processing inequality (DPI) for the operator Schmidt rank (Holzäpfel et al. 2018, Corollary 11). See also Lemma 8.25.

8.2.4. Recursive versus non-recursive measurements

In the following we compare the different conditions for successful reconstruction used in Theorems 8.17 and 8.23. Specifically, we can ask the following question: Assume a fixed set of measurement matrices L_s and R_s where R_s has been constructed in accordance with the extended reconstruction setting from Definition 8.21. Is the set of tensors M for which (8.36) holds strictly smaller than the set for which (8.52) holds? The following Lemma 8.25 implies that the first set is indeed a subset or equal to the second set:

Lemma 8.25 *Assume Definition 8.21. Given a single, fixed $t \in T$ and a single, fixed $s \in S(t)$, the condition (8.36),*

$$\mathrm{rk}(M_s) = \mathrm{rk}(L_s M_s R_s), \tag{8.64}$$

implies (8.52),

$$\mathrm{rk}(M_s(E_t \otimes \mathbb{1}_{d_y})) = \mathrm{rk}(L_s M_s R_s). \tag{8.65}$$

Proof Observe

$$\mathrm{rk}(L_s M_s R_s) = \mathrm{rk}(L_s M_s (E_t \otimes \mathbb{1}_{d_y}) F_s G_s) \le \mathrm{rk}(M_s(E_t \otimes \mathbb{1}_{d_y})) \le \mathrm{rk}(M_s),$$

where we have used $\mathrm{rk}(ABC) \le \mathrm{rk}(B)$ twice. ∎

A counterexample for the converse of Lemma 8.25 is easy to construct, showing that (8.64) and (8.65) are not equivalent. However, it turns out that there is an equivalence between the two sets of conditions containing (8.64) or (8.65) for all $t \in T$ and $s \in S(t)$. This is shown in Theorem 8.28, whose proof uses the following Lemmata 8.26 and 8.27.

Lemma 8.26 *Assume Definition 8.16, $t \in T$ and $s \in S(t)$. Let $s' \subseteq s$, recalling that $s \subsetneq t$ holds by definition. Define $D_s := C_s(B_s)^+$. The following rank inequality holds:*

$$\mathrm{rk}\left(\mathcal{M}_{s'} \mathcal{T}_t C_t\right) \le \mathrm{rk}\left(\mathcal{M}_{s'} \mathcal{T}_s D_s\right) \le \mathrm{rk}\left(\mathcal{M}_{s'} \mathcal{T}_s C_s\right). \tag{8.66}$$

Proof (using graphical representations) Without loss of generality,[29] we set

$$s' := \{1\}, \qquad s := \{1,2\}, \qquad S(t) := \{\{1,2\},\{3\}\}, \qquad t := \{1,2,3,4\} \tag{8.67}$$

and $\Lambda := \{1,2,3,4,5\}$, where Λ is the dimension tree's root. Recall from Definition 8.16 that C_t and D_s are defined as

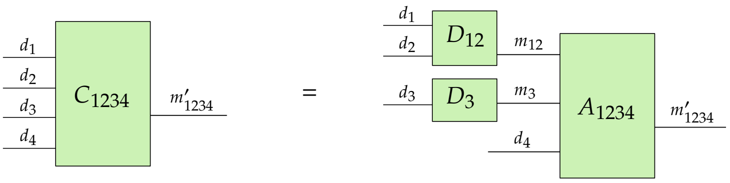
$$C_{1234} = \big(D_{12} \otimes D_3 \otimes \mathbb{1}_{d_4}\big) A_{1234}, \qquad\qquad D_{12} = C_{12}(B_{12})^{+}, \tag{8.68}$$

where C_{12} corresponds to C_s. For matrices, the rank inequality $\operatorname{rk}(AB) \leq \operatorname{rk}(B)$ holds. The simple idea behind the proof is that this rank inequality still applies if we tensorize and unfold as specified above in Equation (8.66). In order to see that this is indeed the case, we recall the shapes of all the involved matrices and use the graphical notation introduced in Section 2.2.2. The matrices have the following shapes:

$$A_t \in \mathbb{C}^{m_{s_1} \dots m_{s_h} d_y \times m'_t} \qquad B_s \in \mathbb{C}^{m_s \times m'_s} \qquad (B_s)^{+} \in \mathbb{C}^{m'_s \times m_s} \tag{8.69a}$$

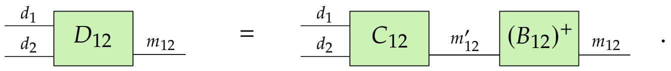
$$C_t \in \mathbb{C}^{d_t \times m'_t} \qquad D_s \in \mathbb{C}^{d_s \times m_s} \tag{8.69b}$$

The graphical representation of (8.68) is thus given by



(8.70a)

and

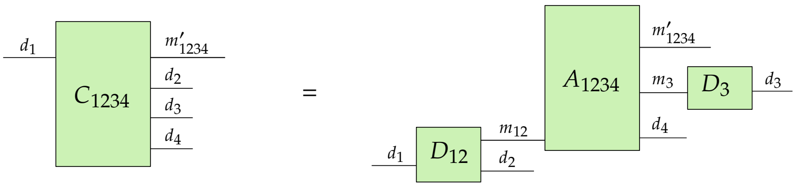


(8.70b)

---

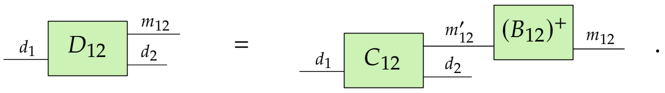
[29] The following sequence of observations shows that generality is not lost: s' is a potentially non-empty set of modes, therefore we assign $s' := \{1\}$ (grouping several modes into one mode if necessary). s is a possibly strictly larger superset of s', therefore we assign $s' := \{1,2\}$ where mode 2 has unit dimension if s equals s'. s must be a successor of t and t can have other successors than s, such that we set $S(t) := \{\{1,2\},\{3\}\}$ where mode 3 results from grouping all of t's successors different from s into one mode. t may contain modes which are not successors of t and we consider this via $t := \{1,2,3,4\}$. Modes which are not part of t are grouped into mode 5, such that the complete set of modes is $\Lambda := \{1,2,3,4,5\}$.

These are easily rearranged into



(8.71a)

and



(8.71b)

The last two equations in turn provide

$$\mathcal{M}_1\mathcal{T}_{1234}C_{1234} = (\mathcal{M}_1\mathcal{T}_{12}D_{12})\big[(\mathcal{M}_{m_{12}}\mathcal{T}_{m_{12}m_3d_4}A_{1234}) \otimes \mathbb{1}_{d_2}\big](\mathbb{1}_{m'_{1234}} \otimes D_3^\top \otimes \mathbb{1}_{d_4d_2})$$

and

$$\mathcal{M}_1\mathcal{T}_{12}D_{12} = (\mathcal{M}_1\mathcal{T}_{12}C_{12})[(B_{12})^+ \otimes \mathbb{1}_{d_2}]. \tag{8.72}$$

Using the matrix rank inequality $\mathrm{rk}(AB) \leq \mathrm{rk}(A)$, we obtain

$$\mathrm{rk}(\mathcal{M}_1\mathcal{T}_{1234}C_{1234}) \leq \mathrm{rk}(\mathcal{M}_1\mathcal{T}_{12}D_{12}) \leq \mathrm{rk}(\mathcal{M}_1\mathcal{T}_{12}C_{12}), \tag{8.73}$$

which proves the lemma. ■

Proof (using algebra) The preceding proof of Lemma 8.26 can also be expressed without graphical representations, which we shall do in the following. We start from Equation (8.69) and introduce an index for each dimension involved in the given shapes:

$$i_1 \in \{1 \dots d_1\}, \qquad j_1 \in \{1 \dots m_1\}, \qquad j'_1 \in \{1 \dots m'_1\}. \tag{8.74}$$

We shall further use the notation $m_{12} := m_{\{1,2\}}$ and corresponding indices such as $j_{12} \in \{1 \dots m_{12}\}$. We define $X \in \mathbb{C}^{m_{12} \times d_3 d_4 j'_{123}}$ via

$$\Big[X\Big]_{j_{12},i_3i_4j'_{123}} := \sum_{j_3} \Big[D_3\Big]_{i_3,j_3} \Big[A_{1234}\Big]_{j_{12}j_3i_4,j'_{123}} \tag{8.75}$$

With this definition, simple rearrangements provide

$$
\begin{aligned}
&\Big[(\mathcal{M}_1\mathcal{T}_{12}D_{12})\big(\mathbb{1}_{d_2}\otimes X\big)\Big]_{i_1,i_2i_3i_4j'_{123}} && (8.76a)\\
&=\sum_{j_{12}}\Big[\mathcal{M}_1\mathcal{T}_{12}D_{12}\Big]_{i_1,i_2j_{12}}\Big[X\Big]_{j_{12},i_3i_4j'_{123}} && (8.76b)\\
&=\sum_{j_{12}}\sum_{j_3}\Big[D_{12}\Big]_{i_1i_2,j_{12}}\Big[D_3\Big]_{i_3,j_3}\Big[A_{1234}\Big]_{j_{12}j_3i_4,j'_{123}} && (8.76c)\\
&=\Big[\big(D_{12}\otimes D_3\otimes\mathbb{1}_{d_4}\big)A_{1234}\Big]_{i_1i_2i_3i_4,j'_{123}} && (8.76d)\\
&=\Big[\mathcal{M}_1\mathcal{T}_{1234}C_{1234}\Big]_{i_1,i_2i_3i_4j'_{123}}\,. && (8.76e)
\end{aligned}
$$

The last Equation implies

$$\mathcal{M}_1\mathcal{T}_{1234}C_{1234}=(\mathcal{M}_1\mathcal{T}_{12}D_{12})\big(\mathbb{1}_{d_2}\otimes X\big) \tag{8.77}$$

and (use $\operatorname{rk}(AB)\le\operatorname{rk}(A)$ for arbitrary matrices A, B)

$$\operatorname{rk}(\mathcal{M}_1\mathcal{T}_{1234}C_{1234})\le\operatorname{rk}(\mathcal{M}_1\mathcal{T}_{12}D_{12}). \tag{8.78}$$

Slightly simpler rearrangements provide

$$
\begin{aligned}
&\Big[(\mathcal{M}_1\mathcal{T}_{12}C_{12})\big(\mathbb{1}_{d_2}\otimes(B_{12})^+\big)\Big]_{i_1,i_2j_{12}}=\sum_{j'_{12}}\Big[\mathcal{M}_1\mathcal{T}_{12}C_{12}\Big]_{i_1,i_2j'_{12}}\Big[(B_{12})^+\Big]_{j'_{12},j_{12}} && (8.79a)\\
&=\sum_{j'_{12}}\Big[C_{12}\Big]_{i_1i_2,j'_{12}}\Big[(B_{12})^+\Big]_{j'_{12},j_{12}}=\Big[C_{12}(B_{12})^+\Big]_{i_1i_2,j_{12}}=\Big[D_{12}\Big]_{i_1i_2,j_{12}} && (8.79b)\\
&=\Big[\mathcal{M}_1\mathcal{T}_{12}D_{12}\Big]_{i_1,i_2j_{12}}\,. && (8.79c)
\end{aligned}
$$

This in turn implies

$$\mathcal{M}_1\mathcal{T}_{12}D_{12}=(\mathcal{M}_1\mathcal{T}_{12}C_{12})\big(\mathbb{1}_{d_2}\otimes(B_{12})^+\big) \tag{8.80}$$

and

$$\operatorname{rk}(\mathcal{M}_1\mathcal{T}_{12}D_{12})\le\operatorname{rk}(\mathcal{M}_1\mathcal{T}_{12}C_{12}) \tag{8.81}$$

Combining the last equation with (8.78) completes the proof. ■

The next lemma constitutes the essential step in the proof of the subsequent Theorem 8.28:

Lemma 8.27 *Assume Definition 8.16. If* $C_\Lambda = M_\Lambda$ *holds, it implies that*

$$\operatorname{rk}(M_s) = \operatorname{rk}(L_s M_s R_s) \tag{8.82}$$

holds for all $s \in T$, $s \neq \Lambda$.

Proof Let $s \in T$, $s \neq \Lambda$. Set $t_0 := s$ and choose $t_{l+1} \in T$ such that $t_l \in S(t_{l+1})$ until the root Λ is reached after $k \geq 1$ steps, i.e. $l \in \{0 \dots k-1\}$ and $t_k = \Lambda$. Applying Lemma 8.26 k times yields

$$\operatorname{rk}\left(\mathcal{M}_{t_0}\mathcal{T}_{t_k}C_{t_k}\right) \leq \operatorname{rk}\left(\mathcal{M}_{t_0}\mathcal{T}_{t_{k-1}}C_{t_{k-1}}\right) \leq \cdots \leq \operatorname{rk}\left(\mathcal{M}_{t_0}\mathcal{T}_{t_1}C_{t_1}\right) \tag{8.83a}$$

$$\leq \operatorname{rk}\left(\mathcal{M}_{t_0}\mathcal{T}_{t_0}D_{t_0}\right) = \operatorname{rk}\left(C_{t_0}\left(B_{t_0}\right)^{+}\right) \leq \operatorname{rk}\left(B_{t_0}\right). \tag{8.83b}$$

Inserting $C_{t_k} = C_\Lambda = M_\Lambda$ provides $\mathcal{T}_{t_k}C_{t_k} = \mathcal{T}_\Lambda M_\Lambda = M$. In addition, we insert $B_{t_0} = B_s = L_s M_s R_s$ and obtain

$$\operatorname{rk}(M_s) \leq \operatorname{rk}(L_s M_s R_s). \tag{8.84}$$

Since the converse inequality always holds, the proof is complete. ■

Theorem 8.28 *Assume Definition 8.21. The following conditions are equivalent:*

1. *The rank condition (8.36),*

   $$\operatorname{rk}(M_s) = \operatorname{rk}(L_s M_s R_s), \tag{8.85}$$

   holds for all $s \in T$, $s \neq \Lambda$.

2. *The rank condition (8.52),*

   $$\operatorname{rk}(M_s(E_t \otimes \mathbb{1}_{d_y})) = \operatorname{rk}(L_s M_s R_s), \tag{8.86}$$

   holds for all $t \in T$ *and* $s \in S(t)$.

3. $C_t = M_t R_t$ *holds for all* $t \in T$.

4. $C_\Lambda = M_\Lambda$, *i.e. the tensor* M *can be reconstructed from the set of matrices* $\{A_t, B_s \colon t \in T, s \in S(t)\}$.

If Definition 8.16 is assumed, Items 1, 3 and 4 are equivalent.

Proof 1 ⇒ 2: Lemma 8.25. 2 ⇒ 3: Theorem 8.23. 3 ⇒ 4 is evident because $R_\Lambda = 1$. 4 ⇒ 1: Lemma 8.27.[30] ■

---

[30] In addition, 1 ⇒ 3 is shown by Theorem 8.17.

Theorem 8.28 states that any tensor M which can be reconstructed with the construction from Definition 8.16 satisfies the two sets of rank conditions required by Theorems 8.17 and 8.23. In other words, exactly the same set of tensors is reconstructed successfully by the two Theorems for a given set of recursively defined measurements conforming to Definition 8.21. Other types of recursively defined measurements can be conceived, but a suitably improved theorem such as 8.23 reconstructs exactly the same set of tensors successfully if C_Λ from Definition 8.16 is used. In order to prepare the comparison of Theorem 8.17 to related work, we briefly discuss other types of recursively defined measurements in the following section.

8.2.5. Further types of recursive measurements

Before we commence the discussion of prior related work, it is helpful to consider further possibilities for recursive construction of the measurement matrices L_s and R_s employed above. Definition 8.21 above introduced a scheme where the matrices R_s are constructed recursively. The essential idea of this scheme is, as was mentioned in Remark 8.22, to use the following definition:

$$R_s = (R_t \otimes \mathbb{1}_{d_y}) F_s \in \mathbb{C}^{d'_s \times m'_s}, \qquad F_s \in \mathbb{C}^{m'_t d_y \times m'_s}. \tag{8.87}$$

Here, $t \in T$ is a node in the dimension tree, $s \in S(t)$ is a successor of t, $R_\Lambda := 1$ and $y := s^c \setminus t^c$. The matrices F_s can be chosen freely and determine the matrices R_s. The measurement matrix acts on the modes $s^c = \Lambda \setminus s$ and it is constructed by first applying R_t, which acts on the $t^c \subsetneq s^c$, followed by applying F_s. F_s acts on the output of R_t and on the remaining modes y.

A similar recursive construction is easily devised for the measurement matrices $L_s \in \mathbb{C}^{m_s \times d_s}$. For this purpose, we demand $t \neq \Lambda$, and set $h := |S(t)|$, $\{s_1, \dots, s_h\} := S(t)$ in some arbitrary, fixed order and define

$$L_t = K_t \left(L_{s_1} \otimes \cdots \otimes L_{s_h} \otimes \mathbb{1}_{d_x} \right) \in \mathbb{C}^{m_t \times d_t}, \qquad K_t \in \mathbb{C}^{m_t \times m_{s_1} \dots m_{s_h} d_x}. \tag{8.88}$$

Here, the set of modes $x := t \setminus \bigcup_{s \in S(t)} s$ covers those modes from t which are not covered by any successor. The recursion stops at leaves t, for which $S(t) = \emptyset$ and $L_t = K_t$ holds. Here, the measurement matrix L_t is given by applying L_s on $s \subsetneq t$ for each successor $s \in S(t)$. The matrix K_t takes input from the L_s and it also acts on the remaining modes $x \subseteq t$.

Last but not least, the matrices L_s can be used in constructing a given R_s. To this end, note that the matrix F_s in (8.87) acted on all the modes in $y = s^c \setminus t^c = t \setminus s$

(where $s \in S(t)$ and $t \in T$; $t = \Lambda$ is permissible). If t has more than one successor, the matrices $L_{s'}$ for all the other successors $s' \neq s$ can be applied next to R_t in (8.87), which becomes apparent in the next two equations. The set of modes s^c can be decomposed as[31]

$$s^c \quad = \quad t^c \mathbin{\dot\cup} (t \setminus s) \quad = \quad t^c \mathbin{\dot\cup} s_2 \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} s_h \mathbin{\dot\cup} x \tag{8.89}$$

where $s_1 = s$ was assumed without loss of generality. This decomposition shows that we are able to define

$$R_s := \left(R_t \otimes L_{s_2}^\top \otimes \cdots \otimes L_{s_h}^\top \otimes \mathbb{1}_{d_x}\right) F_s \in \mathbb{C}^{d'_s \times m'_s}, \qquad F_s \in \mathbb{C}^{m'_t m_{s_2} \ldots m_{s_h} d_x \times m'_s}. \tag{8.90}$$

As already mentioned, F_s acts on the mode set x and takes the output from R_t and L_{s_i}, which, together with the identity on x, cover the full set s^c on which R_s is supposed to act.

Recursively defined measurement matrices may arise naturally if measurement matrices correspond to physical measurement devices. In addition, they have be proven advantageous in numerical algorithms[32] which determine tensor entries allowing for the reconstruction of a given tensor with an iterative search. All the above definitions for recursively constructed measurement matrices allow for attempted reconstruction of a given tensor via Theorem 8.17. If the reconstruction is successful, Theorem 8.28 shows that all conditions in the set $\{\mathrm{rk}(M_s) = \mathrm{rk}(L_s M_s R_s) \colon s \in S(t), t \in T\}$ hold. Therefore, a seemingly weaker condition set such as $\{\mathrm{rk}(M_s R_t) = \mathrm{rk}(L_s M_s R_s)\}$, which is sufficient under (8.87) and Theorem 8.23, is indeed unable to reconstruct a strictly larger set of tensors.

8.2.6. Related work and conclusion

In the following, we compare Theorem 8.17 and Theorem 8.23 with previous works which use a similar approach to reconstruct a tensor. Table 8.1 describes the special cases of Theorem 8.17 included in different contributions.[33] In each column, the most general case is highlighted in green. The comparison is limited to explicitly derived, mathematical results and it deliberately neglects the great extent of empirical studies reported by the cited works.

---

[31] To see that this decomposition is valid, consider $s_1 \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} s_h \subseteq t$, $t = s_1 \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} s_h \mathbin{\dot\cup} x$, $\Lambda = t \mathbin{\dot\cup} t^c = t^c \mathbin{\dot\cup} s_1 \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} s_h \mathbin{\dot\cup} x$ and $\Lambda = s_1 \mathbin{\dot\cup} s_1^c$.

[32] Cf. Section 8.2.6.

[33] The first column denotes the following publications by author's last names and years: Goreinov et al. 1997; Caiafa and Cichocki 2010; Oseledets et al. 2008; Caiafa and Cichocki 2015; Oseledets and Tyrtyshnikov 2010; Baumgratz et al. 2013a; Holzäpfel et al. 2018; Ballani et al. 2013.

Contributions are compared by means of the following properties. The tensor to be reconstructed in Theorem 8.17 is $M \in \mathbb{C}^{d_1 \times \cdots \times d_n}$ and its modes are $\Lambda = \{1, 2, \dots, n\}$. Reconstruction proceeds along a dimension tree T with successor function $S(\,\cdot\,)$. For each $t \in T$ and $s \in S(t)$, the matrices

$$B_s := L_s M_s R_s \in \mathbb{C}^{m_s \times m_s'} \tag{8.91}$$

and their ranks $r_s := \operatorname{rk}(L_s M_s R_s)$ are central to the reconstruction. Reconstruction uses either the regular inverse or the Moore–Penrose pseudoinverse of B_s. The matrix $M_s := \mathcal{M}_s(M)$ is an unfolding of M (Definition 8.10). The matrices $L_s \in \mathbb{C}^{m_s \times d_s}$ and $R_s \in \mathbb{C}^{d_s' \times m_s'}$ are called *measurement matrices* or just *measurements (on M)*. The measurement L_s acts on the modes in $s \subset \Lambda$ (to the right) and the measurement R_s acts on the modes in $s^c = \Lambda \backslash s$ (to the left). The measurement's output sizes are m_s and m_s'; their shapes are $m_s \times d_s$ and $d_s' \times m_s'$. As explored in Section 8.2.5, different recursive structures can be used to define the measurements. If L_s and R_s are submatrices of permutation matrices,[34] they indeed select certain entries of M. Many previous contributions focus on determining suitable entry-selecting measurements efficiently (column "Measurements / Adaptive" in Table 8.1); such a search typically benefits from measurements with recursive structure. Theorems 8.17 and 8.28 are more limited in that they assume fixed measurements L_s, R_s and restrict themselves to the question whether reconstruction is possible for a given tensor M.

Our tensor reconstruction method is based on a matrix reconstruction scheme known as skeleton decomposition or CUR decomposition if the matrix is reconstructed from some of its entries.[35] The skeleton decomposition relies on the regular matrix inverse and $m_s = m_s' = r_s$ and it was generalized to $m_s = m_s' \geq r_s$ using the Moore–Penrose pseudoinverse.[36] A first step towards tensor reconstruction in the spirit of Theorem 8.17 was achieved by Oseledets et al. (2008). For the Tucker dimension tree with three modes, they showed the existence of a tensor decomposition in terms of selected tensor entries. The structure of their decomposition is similar to our reconstruction formula. Furthermore, their proof of existence can be turned into a constructive formula with the restrictions $R_2 = (L_1^\top \otimes \mathbb{1}_{d_2}) F_2$ and $R_3 = (L_1^\top \otimes L_2^\top) F_3$; R_1, L_k and F_k are arbitrary submatrices of permutation

---

[34] A permutation matrix has exactly one non-zero, unit entry in each row and each column (e.g. Horn and Johnson 1991a, Sec. 0.9.5). A submatrix of a permutation matrix has at most one non-zero, unit entry in each row and each column; therefore, it is a so-called partial permutation matrix (Horn and Johnson 1991b, Definition 3.2.5).

[35] Goreinov et al. 1997; Gantmacher 1986.

[36] Caiafa and Cichocki 2010.

Source	Dim. tree	Condition	Inv.	Measurements			
				Shape	Kind	Structure	Adaptive
E.g. GTZ97, Eq. (1.3)	Matrix	$\mathrm{rk}(M_s) = r_s$	Reg.	$m_s = m'_s = r_s$	Entries	–	Yes
CC10, Thm. 1	Matrix	$\mathrm{rk}(M_s) = r_s$	MP	$m_s = m'_s$	Entries	–	No
OST08, Thm. 3.1	Tucker ($n = 3$)	$\mathrm{rk}(M_s) \leq r_s$	Reg.	$m_s = m'_s = r_s$	Ex./entr.	Restr.	Yes
CC10, Thm. 3, FSTD-1	Tucker	$\mathrm{rk}(M_s) \leq r_s$	MP	Arb.	Entries	$R_s = \bigotimes L^\top_{s'}$	Yes
CC10, Thm. 5, FSTD-2	Tucker	$\mathrm{rk}(M_s) \leq r_s$	Reg.	$m_s = m'_s = r_s$	Entries	Arb.	No
CC15, Thm. p. 784	Tucker	$\mathrm{rk}(M_s) = r_s$	MP	$m_s = r_s$	Arb.	$R_s = \bigotimes L^\top_{s'}$	No
OT10, Thm. 3.1	MPS/TT	$\mathrm{rk}(M_s) = r_s$	Reg./ other	$m_s = m'_s = r_s$	Entries	$R_s = R_t F_s$ $L_t = K_t L_s$ (opt.)	Yes
BGCP13, Thm. 1	MPS/TT	$\mathrm{rk}(M_s) = r_s$	MP	Indep. of r_s	Fixed	Fixed	No
HCDP18, Thm. 27	MPS/TT	$\mathrm{rk}(M_s R_t) = r_s$	MP	Arb.	Arb.	$R_s = R_t F_s$	No
BGK13, Eqs. (7)–(9)	H-Tucker	$\mathrm{rk}(M_s) = r_s$	Reg.	$m_s = m'_s = r_s$	Entries	$R_s = R_t F_s$	Yes
This work	Arb.	$\mathrm{rk}(M_s E_t) = r_s$	MP	Arb.	Arb.	Arb.	No

Table 8.1: Comparison of different proposals for tensor reconstruction.[33] The table describes which special case of Theorem 8.17 has been included in each proposal. In each column, the most general property is highlighted in green.

Formulas assume $t \in T$, $s \in S(t)$ and $r_s := \mathrm{rk}(L_s M_s R_s)$. The two conditions $\mathrm{rk}(M_s) = r_s$ and $\mathrm{rk}(M_s) \leq r_s$ are equivalent because $r_s \leq \mathrm{rk}(M_s)$ always holds. Tensor products run over $s' \in S(t)$, $s' \neq s$. Identity matrices in formulas were dropped. See Section 8.2.5 for details regarding the (recursive) structure of measurements.

Abbreviations used are dimension tree (Dim. tree), matrix inverse (Inv.), arbitrary/no restriction (Arb.), (regular) matrix inverse (Reg.), Moore–Penrose pseudoinverse (MP), independent (Indep.), optional (opt.). OST08 showed the existence of a related decomposition based on tensor entries (Ex./entr.) with restricted structure (Restr.) described in the main text. The fixed measurements used in BGCP13 were described in Examples 8.20 and 8.24.[38]

matrices of predetermined shapes. Further results for the Tucker dimension tree were presented by Caiafa and Cichocki (2010, 2015). They reconstruct a tensor from selected entries with restrictions on either the shape or the recursive structure of measurements. In addition, they showed that the restriction of measurement matrices to submatrices of permutation matrices can be lifted under restrictions on both the shape and the recursive structure of measurements.

Oseledets and Tyrtyshnikov (2010) showed that reconstruction from tensor entries is possible with the MPS/TT dimension tree under restrictions on the shape and recursive structure of measurements.[37] Baumgratz et al. (2013a) showed that reconstruction with the MPS/TT dimension tree is possible without limiting the measurement output sizes to the ranks r_s; however, they restricted to certain, fixed measurements which correspond to selecting fixed tensor entries.[38] Holzäpfel et al. (2018) improved the previous two results by showing that arbitrary, recursive measurements can be employed and by introducing a modified rank condition. Reconstruction from tensor entries for the H-Tucker dimension tree was demonstrated by Ballani et al. (2013) with restrictions on the shape and recursive structure of measurements. Another related reconstruction result by Espig et al. (2012) for tensor chains[39] is not mentioned in Table 8.1 because the cyclical structure of tensor chains cannot be represented with a dimension tree as defined above.

---

[37] They lift the restriction of measurement's output size by replacing the regular matrix inverse with a

Theorem 8.17 provides a common basis for most of the cited approaches to tensor reconstruction. It is based on a generalized H-Tucker dimension tree which was modified to allow faithful reproduction of previous results for the MPS/TT dimension tree. It requires the set of conditions $\{\mathrm{rk}(M_s E_t) = r_s\}$, which was shown to be equivalent to the set $\{\mathrm{rk}(M_s) = r_s\}$ but which may in some cases be easier to verify since each individual condition $\mathrm{rk}(M_s E_t) = r_s$ is necessary but not sufficient[40] for $\mathrm{rk}(M_s) = r_s$. Theorem 8.17 allows for entry-selecting or arbitrary measurements with or without recursive structure, each of which may be advantageous for particular applications: Entry-selecting measurements are a natural choice if individual tensor entries can be computed[41] while more general measurement matrices are relevant if the task at hand is e.g. quantum state tomography.[42] General measurement matrices also allow B_s to have the same singular values as M_s, which can improve reconstruction stability[43] and which is usually not possible with entry-selecting measurements. Such a preservation of singular values can be observed e.g. using the measurements constructed in an MPS reconstruction scheme by Cramer et al. (2010).[44] Measurements with recursive

---

procedure which is outside the scope of Theorem 8.17 and this discussion because it uses submatrices of quasi-maximal volume (e.g. Goreinov et al. 1997).

[38] These fixed measurements correspond to selecting fixed tensor entries if the tensor is represented in an operator basis which is orthogonal in the Hilbert–Schmidt inner product and which includes the (normalized) identity matrix as one of its elements. See also Examples 8.20 and 8.24.

[39] Tensor chains are also known as MPS with *periodic boundary conditions*, see Section 2.2.1.

[40] This holds for $s \in T$, $s \neq \Lambda$ but not for $s = \Lambda$. See Lemma 8.25 and Theorem 8.28.

[41] E.g. Oseledets and Tyrtyshnikov 2010.

[42] E.g. Cramer et al. 2010; Holzäpfel et al. 2018.

[43] Cf. Theorem 8.4 and Caiafa and Cichocki 2015.

[44] Cramer et al. 2010 describe a method involving unitary quantum operations on few neighbouring qubits. Preservation of singular values is observed if Theorem 8.17 is used to reconstruct an MPO representation of a pure density matrix ρ and if measurements are constructed from the named

structure are beneficial when determining optimal measurements iteratively[45] but future applications might benefit from measurements with a non-recursive structure. In Theorem 8.17, the output size of measurements is not restricted to the corresponding tensor rank which allows measurements to be chosen independently from the ranks of the tensor. In addition, measurements with output size strictly larger than the corresponding tensor rank are advantageous because they have a higher chance to capture the full support of the unfolding M_s (or its full null space complement) e.g. if measurements are chosen randomly.

If the tensor to be reconstructed is only approximately low rank, i.e. if it is the sum of a large low-rank component and a tiny high-rank component (e.g. noise), the reconstruction from Theorem 8.17 will become unstable due to the matrix pseudoinverse. This instability can be avoided by restricting the output size of measurements to the rank of B_s (i.e. $m_s = m'_s = r_s$) and by additionally choosing measurements in an optimal or pseudo-optimal way.[46] Another, potentially simpler means to avoid the instability is the introduction of a soft or hard threshold for singular values in the matrices B_s.[47] The conditions under which a hard threshold guarantees an improved reconstruction were discussed in Theorem 8.4.

In conclusion, Theorem 8.17 establishes a unified and concise description of several previous results and exhibits several desirable theoretical properties. Potential advantages in practical applications remain to be explored in future work.

---

unitary operations. The relevant preserved singular values of $\rho = |\psi\rangle\langle\psi|$ are then given by all products of two Schmidt coefficients of $|\psi\rangle$.

[45] E.g. Oseledets and Tyrtyshnikov 2010; Ballani et al. 2013.

[46] Goreinov et al. 1997; Oseledets et al. 2008; Oseledets and Tyrtyshnikov 2010. Note that an iterative method which optimizes measurements and estimates tensor ranks simultaneously has been proposed by Savostyanov and Oseledets (2011), building upon concepts from the DMRG. See also Footnote 37.

[47] Goreinov et al. 1997; Espig et al. 2012; Baumgratz et al. 2013a; Caiafa and Cichocki 2015.

Chapter 9.

Conclusions and outlook

Two novel applications of existent efficient quantum state estimation methods were demonstrated in this work. We demonstrated successful efficient estimation of a mixed quantum state of up to six infinite-dimensional continuous-variable quantum systems from simulated measurement data of quadrature amplitudes. This shows that a state of a continuous-variable system can be estimated by the MPS-MLE method if it is supported on a finite-dimensional subspace; furthermore, one can expect that states approximately supported on finite-dimensional subspaces can be estimated with MPS-MLE in the same way. An open problem concerns the convergence speed of the MPS-MLE algorithm. Both the original $R\rho R$ algorithm for maximum likelihood estimation of quantum states and its MPS-based variants were observed to suffer from relatively slow convergence. Recent work demonstrated improvements for the former case on the basis of advanced gradient descent methods and one can hope that similar improvements can be achieved for MPS-MLE and its variants. Another open question concerns the usefulness of MPS-SVT for estimating mixed low-rank states or certain low-rank tensors. Developments in this direction could benefit from recently developed tensor train algorithms. In addition, MPS-MLE and MPS-SVT are iterative algorithms which can benefit from carefully chosen step sizes. In principle, the necessary resources for MPS-MLE and MPS-SVT increase only polynomially with the number of subsystems because matrix product state (MPS) representations are used. However, the number of iterations required to obtain a good estimate can depend on the number of subsystems as well. Existing results suggest that the number of iterations does not increase too strongly with the number of subsystems but the precise scaling could be investigated in future work.

We constructed efficient representations of quantum processes on the basis of existing representations of quantum states. Numerical simulations involving unitary processes were used to demonstrate that the same construction allows for efficient quantum process tomography on the basis of existing quantum state tomography algorithms. The underlying theoretical description relies on augmenting each subsystem with an additional subsystem of the same dimension but we showed that the additional systems are not required in an experiment.

The results from the ion trap quantum simulator experiment showed that the MPS-SVT and MPS-MLE estimators as well as the parent Hamiltonian certificate are well able to estimate and verify the state of a real-world quantum system. However, the experiment also highlighted some shortcomings in our approach for experimental states which possibly exhibit a moderate amount of mixedness. These shortcomings might be alleviated if a mixed state estimator such as MPO-MLE or PMPS-MLE and/or a different or improved certificate were used. Alternatively, the performance of state estimation and verification might be improved by considering longer-ranged observables. For example, incorporating information on all pairs or triplets of sites, independently of their distance, should still allow for efficient estimation. Another avenue for future improvements concerns reducing the estimation error by using the same measurement data for state estimation and state verification. This most likely results in a biased estimator for the fidelity of the estimated state and the unknown state in the experiment, but the resulting estimator may exhibit a smaller mean squared error (MSE) than the estimator which is currently used. The MSE of the resulting, biased estimator could be estimated with a parametric or non-parametric bootstrap simulation or other numerical simulations, providing insight on the usefulness of this estimator as a replacement for the current method.

The complexity of simulating and verifying a time evolution under a local Hamiltonian on a lattice of arbitrary spatial dimension was analyzed theoretically. It was revealed that the necessary effort increases exponentially with time but only quasi-polynomially with the number of subsystems and for one-dimensional systems, the quasi-polynomial scaling reduces to a polynomial one. It would be interesting to compare the predicted worst-case behaviour of the theoretically constructed certificate with the certificate based on the empirical parent Hamiltonians used in evaluating the ion trap experiment. Regarding simulation, our decomposition of time evolution on an arbitrary non-cubic lattice is non-optimal because it evolves local observables into observables which act on a region whose diameter grows polynomially instead of linearly with time. This property of the decomposition is not optimal because the Lieb–Robinson bound used in the decomposition's proof already states that the diameter should grow at most linearly with time; our decomposition for hypercubic lattices shows exactly this behaviour. The latter decomposition could allow computing expectation values of a number of time-evolved tensor product observables if PEPS algorithms for approximate computation of expectation values are used[1] and in some cases this might go beyond what can be achieved with the Lieb–Robinson bound alone. In any case, our results provide an upper bound on the resources required by numerical, PEPS-based algorithms for simulating quantum time evolution. A numerical algorithm, on the other hand, can require fewer resources for particular Hamiltonians and initial states.

Our result for the verification of an initial, pure product state evolved in time under a local Hamiltonian also allows for the verification of the corresponding unitary quantum operation in the following way. As before, each lattice site is augmented by an ancilla lattice site of the same dimension. A time evolution starting from a state with maximal entanglement between each site and corresponding ancilla site then contains complete information on the unitary process induced by the local Hamiltonian. Moreover, this state can be verified with our method because the ancilla sites can be incorporated without changing the lattice by squaring the dimensions of local Hilbert spaces. Determining a local Hamiltonian under an assumption on its maximal interaction range with existing methods[2] and verifying the result with the method just described therefore enables the assumption-free reconstruction of local Hamiltonians. This is an important result because local Hamiltonians were shown to be sufficient to perform universal, adiabatic quantum computation.[3]

Recently, the matrix product state representation has started receiving additional attention under the alias *tensor train (TT) representation*. Matrix product states and tensor trains are identical concepts and newly developed numerical algorithms formulated for tensor trains, the related hierarchical Tucker representation or other related tensor representations may prove beneficial for the simulation and analysis of quantum many-body systems. In this direction, we compared a method for reconstructing a matrix product operator representation of a density matrix with several effectively related methods for tensor reconstruction. As a result, we obtained a tensor reconstruction method with improved properties which can also be used to reconstruct quantum states. Furthermore, many of the named tensor reconstruction proposals go beyond our discussion based on fixed measurements by proposing algorithms for determining suitable measurements adaptively and efficiently, providing new possibilities for adaptive and efficient quantum state estimation. In light of the multitude of proposed numerical algorithms for tensors and quantum many-body systems, it would be desirable to develop a software platform which provides implementations of a preferably large number of numerical algorithms. Such a platform would both simplify using tensor algorithms in practical applications and it would facilitate the development of new or improved algorithms. Open source libraries such as `ttpy`[4] or `mpnum`[5] can provide a starting point for the MPS/TT representation, but it would of course be desirable to also incorporate other, related tensor decompositions.

---

[1] Verstraete and Cirac 2004.

[2] da Silva et al. 2011; Holzäpfel et al. 2015. See also Chapter 5.

[3] Aharonov et al. 2007.

---

[4] Oseledets 2013.

[5] Suess and Holzäpfel 2017.

Acknowledgements

This work would not have been possible without the help from numerous people. I would like to thank my supervisor Prof. Dr. Martin B. Plenio for giving me the opportunity to work on the presented topics and for help, guidance and discussions in the process. During my first two years in Ulm, I could also greatly benefit from help and supervision by Dr. Marcus Cramer. I would like to thank Ben P. Lanyon, Christine Maier, Daniel Suess, Ish Dhand, Nicolai Friis, Nilanjana Datta, Oliver Marty and Tillmann Baumgratz for productive and stimulating collaborations. They as well as Alexander Nüßeler, Dario Egloff, Jan Haase, Joachim Rosskopf and Robert Rosenbach helped with many insightful discussions and valuable feedback during my work on this project. Finally, I would particularly like to thank my friends and my family who kept me company and who supported me.

Appendix A.

A.1. Interaction picture

The Lieb–Robinson bounds discussed in Section 6.2 show that information propagates with a maximal velocity v, the Lieb–Robinson velocity, if the Hamiltonian $H(t) = \sum_{Z \subset \Lambda} h_Z(t)$ which governs the dynamics satisfies certain conditions. The Lieb–Robinson velocity is given by $v = J\mathcal{Z} \exp(1)$ where $J = 2 \sup_{t, Z \subset \Lambda} \|h_Z(t)\|_{(\infty)}$ is twice the maximal norm of a local term of the Hamiltonian (Equation (6.4)). Adding a term h_x acting only on a single lattice site $x \in \Lambda$ to the Hamiltonian can increase the Lieb–Robinson velocity arbitrarily but one would not expect that it affects how information propagates in the system because it acts only on a single site. In infinite-dimensional systems, Lieb–Robinson bounds unaffected even by unbounded single-site terms have been proven.[1] In the following, we provide a simple way to use Theorem 6.1 without single-site terms influencing the Lieb–Robinson velocity. This is achieved by switching to a suitable interaction or Dirac picture before applying the theorem. Lemma A.1 introduces the interaction picture we use and Corollary A.2 applies it to the Lieb–Robinson bound from Theorem 6.1.

The interaction or Dirac picture is introduced in most quantum mechanics textbooks and we present our version in the following lemma. A Hamiltonian H is split into two parts, $H = F + G$. Observables $A_{\mathrm{D}}(t)$ evolve according to F and states $|\psi_{\mathrm{D}}(t)\rangle$ evolve such that the correct expectation values arise.

Lemma A.1 *Fix a time $r \in \mathbb{R}$ and let the two times $s, t \in \mathbb{R}$ be arbitrary. Let $H(t) = F(t) + G(t)$ be a Hamiltonian and let $A(t)$ be an observable. We have $|\psi(t)\rangle = U^H_{tr}|\psi(r)\rangle$ and set*

$$A_D(t) := U^F_{rt}\, A(t)\, U^F_{tr}\,, \qquad |\psi_D(t)\rangle := U^F_{rt}\, U^H_{tr}\, |\psi(r)\rangle. \tag{A.1}$$

---

[1] Nachtergaele et al. 2009, 2010; Nachtergaele and Sims 2014.

Expectation values are given by

$$\langle \psi_D(t)| A_D(t) |\psi_D(t)\rangle = \langle \psi(t)| A(t) |\psi(t)\rangle \tag{A.2}$$

Set

$$U^{(D)}_{ts} := U^F_{rt} U^H_{ts} U^F_{sr} \,. \tag{A.3}$$

This operator propagates the states $|\psi_D(t)\rangle$ *via*

$$|\psi_D(t)\rangle = U^{(D)}_{ts} |\psi_D(s)\rangle. \tag{A.4}$$

and it is the solution of the differential equation

$$\partial_t U^{(D)}_{ts} = -\mathrm{i}\tilde{G}(t) U^{(D)}_{ts} \tag{A.5}$$

where $\tilde{G}(t) := U^F_{rt} G(t) U^F_{tr}$ *and* $U^{(D)}_{ss} = \mathbb{1}$*, i.e.* $U^{(D)}_{ts} = U^{\tilde{G}}_{ts}$*.*

Proof Equations (A.2) and (A.4) follow directly from the definitions. Equation (A.5) is shown by

$$\partial_t U^{(\mathrm{D})}_{ts} = +\mathrm{i}U^F_{rt}[F(t) - H(t)]U^H_{ts}U^F_{sr} = -\mathrm{i}U^F_{rt}G(t)U^F_{tr}U^F_{rt}U^H_{ts}U^F_{sr} = -\mathrm{i}\tilde{G}(t)U^{(\mathrm{D})}_{ts} \,,$$

which completes the proof. ■

Corollary A.2 *In the setting from Section 6.2, let* $H(t) = \sum_{Z\subset\Lambda} h_Z(t)$*,* $F(t) = \sum_{x\in\Lambda} h_{\{x\}}(t)$ *and* $G(t) = \sum_{Z\subset\Lambda, |Z|\geq 2} h_Z(t)$*. Assume that* $G(t) \neq 0$ *for some* t *and consider the parameters defined in Equations (6.4) to (6.8). The maximal range* a *of* H*,* G *and* $\tilde{G}$ *is the same. The maximal norm* J *satisfies* $J(\tilde{G}) = J(G) \leq J(H)$ *and the same holds for the parameters* $\mathcal{Z}$*,* M*,* κ *and* $\mathcal{Y}$*.*

Let $Y \subset R \subset \Lambda$ *and let* A *act on* Y*. Theorem 6.1 provides a bound*

$$\left\| \tau^H_{ts}(A) - \tau^{H_{\bar{R}}}_{ts}(A) \right\|_{(\infty)} \leq \epsilon(H) \tag{A.6}$$

where $\epsilon(H)$ *depends on the parameters of* H *just mentioned, as specified in Theorem 6.1. Equation (A.6) still holds if* $\epsilon(H)$ *is replaced by the smaller* $\epsilon(\tilde{G}) = \epsilon(G)$*.*

Proof $\tilde{G}$ and G have the same value of J because the operator norm is unitarily invariant. $\tilde{G}$ and G have the same value of a, $\mathcal{Z}$, M, κ and $\mathcal{Y}$ because the tensor product U^F_{ts} does not change the set of sites on which a local term acts non-trivially. Inspection of Equations (6.4) to (6.8) yields claimed inequalities between parameters of H and parameters of G.

Applying Theorem 6.1 to $\tilde{G}$ and A' acting on Y provides

$$\left\| \tau_{ts}^{\tilde{G}}(A') - \tau_{ts}^{\tilde{G}_{\bar{R}}}(A') \right\|_{(\infty)} \leq \epsilon'(\tilde{G}). \tag{A.7}$$

As $\mathcal{Z}$ appears in the denominator of ϵ, the claimed $\epsilon'(\tilde{G}) \leq \epsilon(H)$ might fail to hold if $\mathcal{Z}(\tilde{G}) < \mathcal{Z}(H)$. $\mathcal{Z}(\tilde{G}) = \mathcal{Z}(H)$ can be ensured by keeping arbitrarily small single-site terms in G instead of removing them completely. If we similarly set $\kappa(G) = \kappa(H)$ and $M(G) = M(H)$, Equation (6.7) is satisfied for G. Inserting $A' = \tau_{rs}^{F}(A) = \tau_{rs}^{F_Y}(A)$, which acts only on Y, into (A.7) and using the unitary invariance of the operator norm provides

$$\left\| \tau_{ts}^{H}(A) - \left(\tau_{tr}^{F_{\bar{R}}} \tau_{ts}^{\tilde{G}_{\bar{R}}} \tau_{rs}^{F_Y} \right)(A) \right\|_{(\infty)} = \left\| \tau_{tr}^{F} \left(\tau_{ts}^{\tilde{G}} \left(\tau_{rs}^{F}(A) \right) - \tau_{ts}^{\tilde{G}_{\bar{R}}} \left(\tau_{rs}^{F_Y}(A) \right) \right) \right\|_{(\infty)} \leq \epsilon'(\tilde{G}).$$

Here, we used $U_{ts}^{H} = U_{tr}^{F} U_{ts}^{\tilde{G}} U_{rs}^{F}$ (Equation (A.3)). Note that $\left(\tau_{tr}^{F_{\bar{R}}} \tau_{ts}^{\tilde{G}_{\bar{R}}} \tau_{rs}^{F_Y}\right)(A) = V_{ts} A V_{ts}^{*}$ where $V_{ts} = U_{tr}^{F_{\bar{R}}} U_{ts}^{\tilde{G}_{\bar{R}}} U_{rs}^{F_{\bar{R}}}$. Applying Lemma A.1 to $H_{\bar{R}} = F_{\bar{R}} + G_{\bar{R}}$, where $H_{\bar{R}}$ was split in the same way as H, provides $U_{ts}^{\tilde{G}_{\bar{R}}} = U_{rt}^{F_{\bar{R}}} U_{ts}^{H_{\bar{R}}} U_{sr}^{F_{\bar{R}}}$, i.e. $U_{ts}^{H_{\bar{R}}} = U_{tr}^{F_{\bar{R}}} U_{ts}^{\tilde{G}_{\bar{R}}} U_{rs}^{F_{\bar{R}}} = V_{ts}$, which completes the proof. ■

Remark A.3 Before applying Corollary A.2, it can be worthwhile to minimize the norm of h_Z with $|Z| \geq 2$ by subtracting single-site terms from it. These single-site terms can reduce the norm of h_Z (i.e. J and v) and they are added to the Hamiltonian as single-site terms in order to leave the total Hamiltonian unchanged. □

A.2. Selected inequalities

The following inequalities are used in Chapter 6 and elsewhere.

Lemma A.4 *Let $\|\cdot\|$ be a unitarily invariant norm and let U_2, V_1 be unitary, $i \in \{1, 2\}$. Let A be an arbitrary matrix. Then $\|U_1 A U_2 - V_1 A V_2\| \leq \|(U_1 - V_1)A\| + \|A(U_2 - V_2)\|$.*

Proof

$$\begin{aligned} \|U_1 A U_2 - V_1 A V_2\| &= \|U_1 A U_2 - V_1 A U_2 \quad + \quad V_1 A U_2 - V_1 A V_2\| \\ &\leq \|U_1 A U_2 - V_1 A U_2\| + \|V_1 A U_2 - V_1 A V_2\| \\ &= \|(U_1 - V_1)A\| \qquad + \|A(U_2 - V_2)\| \end{aligned}$$

where the triangle inequality and unitary invariance have each been used once. ■

Lemma A.5 *Let* $n \geq 0$, $a > 0$ *and* $x \geq \max\{0, \frac{2n}{a}\ln(\frac{n}{a})\}$. *Then* $x^n \exp(-ax) \leq 1$.

Proof For $n = 0$ or $x = 0$, the Lemma holds. Let $n > 0$ and $x > 0$. Let $z = \frac{a}{n}x$ and $c = \ln(\frac{n}{a})$. The inequalities $z \geq 2c$ (implied by the premise) and $\ln(z) \leq \frac{z}{2}$ (see Lemma A.6) imply $\ln(z) + c \leq \frac{z}{2} + c \leq z$. We have

$$\ln(z) + c \leq z \quad \Leftrightarrow \quad \ln(x) \leq \frac{ax}{n} \quad \Leftrightarrow \quad n\ln(x) - ax \leq 0 \quad \Leftrightarrow \quad x^n \mathrm{e}^{-ax} \leq 1.$$

This completes the proof because the inequality on the very left is implied by the premise. ■

Lemma A.6 $\ln(x) \leq \frac{x}{2} - (1 - \ln 2) < \frac{x}{2}$ *for* $x \in [0, \infty)$ *with equality if and only if* $x = 2$.

Proof Let $f(x) = \frac{x}{2} - \ln(x) - (1 - \ln 2)$. The derivative satisfies

$$f'(x) = \frac{1}{2} - \frac{1}{x} \qquad \begin{cases} > 0, & \text{if } x > 2, \\ = 0, & \text{if } x = 2, \\ < 0, & \text{if } x < 2. \end{cases} \tag{A.8}$$

In addition, $f(2) = 0$. This shows the claim. ■

Lemma A.7 *(i) Let* $\|\cdot\|_{(1)}$ *denote the trace norm,* $\psi = |\psi\rangle\langle\psi|$ *and* $\psi' = |\psi'\rangle\langle\psi'|$. *If* $\||\psi\rangle - |\psi'\rangle\| \leq \epsilon \leq \sqrt{2}$, *then* $\|\psi - \psi'\|_{(1)} \leq 2\epsilon$.

(ii) Let $1 - |\langle\psi|\psi'\rangle| = \epsilon$. *Then* $\min_{\alpha\in[0,2\pi]} \||\psi\rangle - \mathrm{e}^{\mathrm{i}\alpha}|\psi'\rangle\| = \sqrt{2\epsilon}$. *Let in addition* $\epsilon \leq 1$, *then* $\|\psi - \psi'\|_{(1)} \leq 2\sqrt{2\epsilon}$.

Proof (i) Assume that $\||\psi\rangle - |\psi'\rangle\| \leq \epsilon$ holds. This gives us

$$\epsilon^2 \geq \||\psi\rangle - |\psi'\rangle\|^2 = 2(1 - \mathrm{Re}(\langle\psi|\psi'\rangle)) \geq 2(1 - \sqrt{F}) \tag{A.9}$$

where $F = |\langle\psi|\psi'\rangle|^2 = F(|\psi\rangle, |\psi'\rangle)$. This gives $\sqrt{F} \geq 1 - \epsilon^2/2$ and $1 - F \leq 1 - (1 - \epsilon^2/2)^2 = \epsilon^2 - \epsilon^4/4 \leq \epsilon^2$. The equality $\|\psi - \psi'\|_{(1)} = 2\sqrt{1 - F}$ completes the proof.[2]

(ii) Choose $\alpha \in \mathbb{R}$ such that, with $|\psi''\rangle = \mathrm{e}^{\mathrm{i}\alpha}|\psi'\rangle$, the equalities $|\langle\psi|\psi'\rangle| = \langle\psi|\psi''\rangle = \mathrm{Re}(\langle\psi|\psi''\rangle)$ hold. In this case, we have

$$\min_{\alpha\in[0,2\pi]} \||\psi\rangle - \mathrm{e}^{\mathrm{i}\alpha}|\psi'\rangle\| \leq \||\psi\rangle - |\psi''\rangle\|^2 = 2[1 - \mathrm{Re}(\langle\psi|\psi''\rangle)] = 2\epsilon \tag{A.10}$$

and it is clear that for all other values of $\alpha \in \mathbb{R}$, the value of $1 - \mathrm{Re}(\langle\psi|\psi''\rangle)$ will be larger. Part (i) proofs the remaining part of (ii). ■

---

[2] Nielsen and Chuang 2007, Eqs. 9.11, 9.60, 9.99.

A.3. Metric spaces

Remark A.8 Given two sets A and B, the expression $A \subset B$ is used to refer to the implication $x \in A \Rightarrow x \in B$. □

Definition A.9 Let Λ be a set. A function $d\colon \Lambda \times \Lambda \to \mathbb{R}$ is called a metric if, for all $x, y, z \in \Lambda$, $d(x,y) \geq 0$, $d(x,y) = 0$ if and only if $x = y$, $d(x,y) = d(y,x)$ and $d(x,z) \leq d(x,y) + d(y,z)$ (triangle inequality). The pair (Λ, d) is called a metric space and a finite metric space is a metric space where Λ has finitely many elements. Statements in this section for infinite metric spaces should be treated with caution (they are not used in the main text).

Distances between sets are given by $d(A,B) = \inf_{a \in A, b \in B} d(a,b)$ and the infimum turns into a minimum if both sets are finite. Accordingly, we have

$$\exists\, a_0 \in A, b_0 \in B\colon d(a_0,b_0) < r \quad \Rightarrow \quad d(A,B) \leq d(a_0,b_0) < r, \tag{A.11a}$$

$$\forall\, a \in A, b \in B\colon d(a,b) > r \quad \Rightarrow \quad d(A,B) > r. \tag{A.11b}$$

Strict inequalities can be replaced by equalities in both equations. If the metric space is infinite, the strict inequality in the second equation turns into an inequality.

The diameter of a subset $Y \subset \Lambda$ is given by $\operatorname{diam}(Y) = \sup_{x,y \in Y} d(x,y)$ and the supremum turns into a maximum for a finite set Y. Let $\mathcal{M}$ be a set of subsets of Λ with $a = \sup_{Z \in \mathcal{M}} \operatorname{diam}(Z) < \infty$. Define the extension of $R \subset \Lambda$ via $\bar{R} = \bigcup_{Z \in \mathcal{M}, Z \cap R \neq \emptyset} Z$.

The open and closed ball around $Y \subset \Lambda$ are defined by

$$B_r^o(Y) = \{x \in \Lambda\colon d(x,Y) < r\}, \tag{A.12a}$$

$$B_r^c(Y) = \{x \in \Lambda\colon d(x,Y) \leq r\}. \tag{A.12b}$$

Lemma A.10 *The following hold ($Y \subset \Lambda$, $r, s \geq 0$):*

$$d(Y, \Lambda \setminus B_r^o(Y)) \geq r \tag{A.13a}$$

$$d(Y, \Lambda \setminus B_r^c(Y)) > r \tag{A.13b}$$

$$d(B_s^o(Y), \Lambda \setminus R) > d(Y, \Lambda \setminus R) - s \tag{A.13c}$$

$$d(B_s^c(Y), \Lambda \setminus R) \geq d(Y, \Lambda \setminus R) - s \tag{A.13d}$$

$$B_r^o(Y) \subset R \quad \textit{where} \quad r = d(Y, \Lambda \setminus R) \tag{A.13e}$$

$$[B_r^c(B_s^o(Y)) \cup B_r^o(B_s^c(Y)) \cup B_r^o(B_s^o(Y))] \;\subset\; B_{r+s}^o(Y) \tag{A.13f}$$

$$\bar{R} \subset B_a^c(R) \tag{A.13g}$$

$$\operatorname{diam}(B_r^o(Y)) < 2r + \operatorname{diam}(Y). \tag{A.13h}$$

Strict inequalities turn into non-strict inequalities for infinite metric spaces.

Let $x, y \in \Lambda$*. Then* $d(x, y) \geq r + s$ *implies* $B_r^o(\{x\}) \cap B_s^c(\{y\}) = \emptyset$*.*

Proof For all $y \in Y$ and $z \in \Lambda \setminus B_r^o(Y)$, it is true that $z \notin B_r^o(Y)$ and thus $d(y, z) \geq r$. (A.11b) thus implies (A.13a).

For all $y \in Y$ and $z \in \Lambda \setminus B_r^c(Y)$, it is true that $z \notin B_r^c(Y)$ and thus $d(y, z) > r$. (A.11b) thus implies (A.13b).

Let $z \in \Lambda \setminus R$ and $x \in B_s^o(Y)$. There is a $y \in Y$ such that $d(x, y) < s$. Note that $d(y, z) \geq d(Y, \Lambda \setminus R)$ holds. This implies that $d(z, x) \geq d(z, y) - d(y, x) > d(Y, \Lambda \setminus R) - s$. (A.11b) implies (A.13c).

Let $z \in \Lambda \setminus R$ and $x \in B_s^c(Y)$. There is a $y \in Y$ such that $d(x, y) \leq s$. Note that $d(y, z) \geq d(Y, \Lambda \setminus R)$ holds. This implies that $d(z, x) \geq d(z, y) - d(y, x) \geq d(Y, \Lambda \setminus R) - s$. (A.11b) implies (A.13d).

Let $x \in B_r^o(Y)$, then there is a $y \in Y$ such that $d(x, y) < r$. If $x \in \Lambda \setminus R$ was true, it would imply $d(Y, \Lambda \setminus R) \leq d(x, y) < r$ (see (A.11a)), which is a contradiction. Therefore, we infer $x \notin \Lambda \setminus R$ and thus $x \in R$. This shows (A.13e).

Let $x \in B_r^c(B_s^o(Y))$. Then there are $z \in B_s^o(Y)$ and $y \in Y$ such that $d(x, z) \leq r$ and $d(z, y) < s$. This implies $d(x, y) < r + s$ and thus $x \in B_{r+s}^o(Y)$. The remaining parts of (A.13f) are shown in the same way.

Let $x \in \bar{R}$. If $x \in R$, then $x \in B_a^c(R)$ holds. Let $x \in \bar{R} \setminus R$. Then there is a $Z \subset \Lambda$ such that $\mathrm{diam}(Z) \leq a$ and $x \in Z$ and $Z \cap R \neq \emptyset$. Let $y \in Z \cap R$, then $d(x, y) \leq \mathrm{diam}(Z) \leq a$. Because $y \in R$, we can conclude $x \in B_a^c(R)$. This shows (A.13g).

Let $x, y \in B_r^o(Y)$. Then there are $x', y' \in Y$ such that $d(x, x') < r$ and $d(y, y') < r$. This implies $d(x, y) \leq d(x, x') + d(x', y') + d(y', y) < 2r + \mathrm{diam}(Y)$. This shows (A.13h).

Assume that $z \in B_r^o(\{x\}) \cap B_s^c(\{y\})$ exists. Then $d(x, y) \leq d(x, z) + d(z, y) < r + s$ contradicts the assumption. ∎

The following lemmata use definitions from Section 6.4.3.

Lemma A.11 *Let* d *be a metric with property (6.58). Let* $x, y \in \Lambda$*. For any* $r \geq 0$*, we have*

$$B_r^c(C(x, y)) \subset C_s(C(x, y)), \quad s = \lfloor r \rfloor. \tag{A.14}$$

Proof Let $z \in B_r^c(C(x, y))$, then there is a $b \in C(x, y)$ such that $d(z, b) \leq r$; this implies $|z_i - b_i| \leq r$ for all $i \in [1 : \eta]$. In addition, $b \in C(x, y)$ implies $x_i \leq b_i \leq y_i$. Combining both yields $x_i - r \leq z_i \leq y_i + r$ and this shows that $z \in C(x - su, y + su)$ where $s = \lfloor r \rfloor$ and $u := (1, 1, \ldots, 1) \in \mathbb{Z}^\eta$. ∎

Lemma A.12 *Let* $Y, Z \subset \Lambda$ *with* $Y \cap Z \neq \emptyset$*. If* $r \geq \operatorname{diam}(Z)$ *then* $Z \subset B_r^c(Y)$.

Proof Let $z \in Z$ and $y \in Z \cap Y$. Then $d(z, y) \leq \operatorname{diam}(Z, Y) \leq r$, i.e. $z \in B_r^c(Y)$. ■

Lemma A.13 *Let* $a, b, c, d \in \Lambda$*. Then,* $C(a, b) \cap C(c, d) = C(x, y)$ *where* $x_i = \max\{a_i, c_i\}$ *and* $y_i = \min\{b_i, d_i\}$.

Proof $C(a, b) \cap C(c, d) = \bigtimes_{i=1}^{\eta}[a_i : b_i] \cap [c_i : d_i] = \bigtimes_{i=1}^{\eta}[x_i : y_i] = C(x, y)$. ■

Bibliography

Aaronson, S. (2007). “The learnability of quantum states”. Proc. R. Soc. London, Ser. A 463(2088), 3089–3114. DOI: 10.1098/rspa.2007.0113. arXiv: quant-ph/0608142.

Aharonov, D., W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev (2007). “Adiabatic quantum computation is equivalent to standard quantum computation”. SIAM J. Comput. 37(1), 166–194. DOI: 10.1137/S0097539705447323. arXiv: quant-ph/0405098 [quant-ph].

Altepeter, J. B., D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. O’Brien, M. A. Nielsen, and A. G. White (2003). “Ancilla-assisted quantum process tomography”. Phys. Rev. Lett. 90(19), 193601. DOI: 10.1103/physrevlett.90.193601. arXiv: quant-ph/0303038.

Aspelmeyer, M., T. J. Kippenberg, and F. Marquardt (2014). “Cavity optomechanics”. Rev. Mod. Phys. 86(4), 1391–1452. DOI: 10.1103/RevModPhys.86.1391. arXiv: 1303.0733.

Babichev, S. A., J. Appel, and A. I. Lvovsky (2004). “Homodyne tomography characterization and nonlocality of a dual-mode optical qubit”. Phys. Rev. Lett. 92(19), 193601. DOI: 10.1103/PhysRevLett.92.193601. arXiv: quant-ph/0312135.

Bai, J., J. Li, and F. Xu (2017). “Accelerated method for optimization over density matrices in quantum state estimation”. Linear and Multilinear Algebra 66(5), 869–880. DOI: 10.1080/03081087.2017.1330864.

Baldwin, C. H., I. H. Deutsch, and A. Kalev (2016). “Strictly-complete measurements for bounded-rank quantum-state tomography”. Phys. Rev. A 93(5), 052105. DOI: 10.1103/PhysRevA.93.052105. arXiv: 1605.02109 [quant-ph].

Baldwin, C. H., A. Kalev, and I. H. Deutsch (2014). “Quantum process tomography of unitary and near-unitary maps”. Phys. Rev. A 90(1), 012110. DOI: 10.1103/physreva.90.012110. arXiv: 1404.2877.

Ballani, J., L. Grasedyck, and M. Kluge (2013). “Black box approximation of tensors in hierarchical Tucker format”. Linear Algebra Appl. 438(2), 639–657. DOI: 10.1016/j.laa.2011.08.010.

Banaszek, K., G. M. D’Ariano, M. G. Paris, and M. F. Sacchi (1999). “Maximum-likelihood estimation of the density matrix”. Phys. Rev. A 61, 010304. DOI: 10.1103/PhysRevA.61.010304. arXiv: quant-ph/9909052 [quant-ph].

Bañuls, M. C., D. Pérez-García, M. M. Wolf, F. Verstraete, and J. I. Cirac (2008). “Sequentially generated states for the study of two-dimensional systems”. Phys. Rev. A 77(5), 052306. DOI: 10.1103/PhysRevA.77.052306. arXiv: 0802.2472 [quant-ph].

Barthel, T. and M. Kliesch (2012). “Quasilocality and efficient simulation of Markovian quantum dynamics”. Phys. Rev. Lett. 108(23), 230504. DOI: 10.1103/PhysRevLett.108.230504. arXiv: 1111.4210 [quant-ph].

Baumann, T. (2016). “Scalable maximum likelihood quantum state tomography with local purification form matrix product states”. Bachelor thesis. Ulm University.

Baumgratz, T., D. Gross, M. Cramer, and M. B. Plenio (2013a). "Scalable reconstruction of density matrices". Phys. Rev. Lett. 111(2), 020401. DOI: `10.1103/physrevlett.111.020401`. arXiv: `1207.0358`.

Baumgratz, T., A. Nüßeler, M. Cramer, and M. B. Plenio (2013b). "A scalable maximum likelihood method for quantum state tomography". New J. Phys. 15(12), 125004. DOI: `10.1088/1367-2630/15/12/125004`. arXiv: `1308.2395`.

Baumgratz, T. (2014). "Efficient system identification and characterization for quantum many-body systems". PhD thesis. Ulm University. DOI: `10.18725/OPARU-3293`.

Bhatia, R. (1997). "Matrix analysis". New York; Heidelberg: Springer. ISBN: 0-387-94846-5.

Bolduc, E., G. C. Knee, E. M. Gauger, and J. Leach (2017). "Projected gradient descent algorithms for quantum state tomography". npj Quantum Inf. 3(1), 44. DOI: `10.1038/s41534-017-0043-1`. arXiv: `1612.09531`.

Bondy, J. A. and U. S. R. Murty (2008). "Graph theory". Graduate texts in mathematics. New York: Springer. ISBN: 978-1-84628-969-9.

Boyd, S. P. and L. Vandenberghe (2009). "Convex optimization". 7th ed. Cambridge: Cambridge University Press. ISBN: 978-0-521-83378-3.

Bravyi, S., M. B. Hastings, and F. Verstraete (2006). "Lieb-Robinson bounds and the generation of correlations and topological quantum order". Phys. Rev. Lett. 97(5), 050401. DOI: `10.1103/physrevlett.97.050401`. arXiv: `quant-ph/0603121`.

Busch, P. (1991). "Informationally complete sets of physical quantities". Int. J. Theor. Phys. 30(9), 1217–1227. DOI: `10.1007/BF00671008`.

Buyskikh, A. S. (2017). "Dynamics of quantum many-body systems with long-range interactions". PhD thesis. University of Strathclyde. `http://suprimo.lib.strath.ac.uk/SUVU01:LSCOP_SU:SUDIGI28798`.

Cai, J.-F., E. J. Candès, and Z. Shen (2010). "A singular value thresholding algorithm for matrix completion". SIAM J. Optim. 20(4), 1956–1982. DOI: `10.1137/080738970`. arXiv: `0810.3286 [math.OC]`.

Caiafa, C. F. and A. Cichocki (2010). "Generalizing the column–row matrix decomposition to multi-way arrays". Linear Algebra Appl. 433(3), 557–573. DOI: `10.1016/j.laa.2010.03.020`.

– (2015). "Stable, robust, and super fast reconstruction of tensors using multi-way projections". IEEE Trans. Signal Process. 63(3), 780–793. DOI: `10.1109/TSP.2014.2385040`.

Candès, E. J. and B. Recht (2008). "Exact low-rank matrix completion via convex optimization". In: 2008 46th Annu. Allerton Conf. Comm. Control Comput. 806–812. DOI: `10.1109/ALLERTON.2008.4797640`.

Candès, E. J. and B. Recht (2009). "Exact matrix completion via convex optimization". Found. Comput. Math. 9(6), 717. DOI: `10.1007/s10208-009-9045-5`. arXiv: `0805.4471 [cs.IT]`.

Cerf, N. J. and C. Adami (1997). "Negative entropy and information in quantum mechanics". Phys. Rev. Lett. 79, 5194–5197. DOI: `10.1103/PhysRevLett.79.5194`. arXiv: `quant-ph/9512022 [quant-ph]`.

Choi, M.-D. (1975). "Completely positive linear maps on complex matrices". Linear Algebra Appl. 10(3), 285–290. DOI: `10.1016/0024-3795(75)90075-0`.

Coppersmith, D. (1994). "An approximate Fourier transform useful in quantum factoring". IBM Research Report No. RC 19642. arXiv: `quant-ph/0201067`.

Cramer, M., M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu (2010). "Efficient quantum state tomography". Nat. Commun. 1(9), 149. DOI: `10.1038/ncomms1147`. arXiv: `1101.4366`.

da Silva, M. P., O. Landon-Cardinal, and D. Poulin (2011). "Practical characterization of quantum devices without tomography". Phys. Rev. Lett. 107(21), 210404. DOI: `10.1103/physrevlett.107.210404`. arXiv: `1104.3835`.

Dalcin, L., R. Bradshaw, K. Smith, C. Citro, S. Behnel, and D. S. Seljebotn (2010). "Cython: The best of both worlds". Comput. Sci. Eng. 13, 31–39. DOI: `10.1109/MCSE.2010.118`.

D'Ariano, G. M. and P. Lo Presti (2001). "Quantum tomography for measuring experimentally the matrix elements of an arbitrary quantum operation". Phys. Rev. Lett. 86(19), 4195–4198. DOI: `10.1103/PhysRevLett.86.4195`. arXiv: `quant-ph/0012071`.

Datta, A. and G. Vidal (2007). "Role of entanglement and correlations in mixed-state quantum computation". Phys. Rev. A 75(4), 042310. DOI: `10.1103/PhysRevA.75.042310`. arXiv: `quant-ph/0611157`.

De las Cuevas, G., N. Schuch, D. Pérez-García, and J. I. Cirac (2013). "Purifications of multipartite states: Limitations and constructive methods". New J. Phys. 15(12), 123021. DOI: `10.1088/1367-2630/15/12/123021`. arXiv: `1308.1914 [quant-ph]`.

De Raedt, H. (1987). "Product formula algorithms for solving the time dependent Schrödinger equation". Comput. Phys. Rep. 7(1), 1–72. DOI: `10.1016/0167-7977(87)90002-5`.

Dolgov, S. V. and D. V. Savostyanov (2014). "Alternating minimal energy methods for linear systems in higher dimensions". SIAM J. Sci. Comput. 36(5), A2248–A2271. DOI: `10.1137/140953289`. arXiv: `1304.1222 [math.NA]`.

Dollard, J. D. and C. N. Friedman (1979a). "Product integration of measures and applications". J. Differ. Equat. 31(3), 418–464. DOI: `10.1016/S0022-0396(79)80009-1`.

– (1979b). "Product integration with applications to differential equations". Encycl. Math. Appl. Reading, Massachusetts: Addison-Wesley. ISBN: 0-201-13509-4.

Dunn, T. J., I. A. Walmsley, and S. Mukamel (1995). "Experimental determination of the quantum-mechanical state of a molecular vibrational mode using fluorescence tomography". Phys. Rev. Lett. 74 (6), 884–887. DOI: `10.1103/PhysRevLett.74.884`.

Dür, W. and J. I. Cirac (2001). "Nonlocal operations: Purification, storage, compression, tomography, and probabilistic implementation". Phys. Rev. A 64(1), 012317. DOI: `10.1103/PhysRevA.64.012317`. arXiv: `quant-ph/0012148`.

Eckart, C. and G. Young (1936). "The approximation of one matrix by another of lower rank". Psychometrika 1(3), 211–218. DOI: `10.1007/BF02288367`.

Efron, B. and R. Tibshirani (1993). "An introduction to the bootstrap". Vol. 57. Monogr. Statistics Appl. Probab. New York: Chapman & Hall. ISBN: 0-412-04231-2.

Eisert, J., M. Cramer, and M. B. Plenio (2010). "Colloquium: Area laws for the entanglement entropy". Rev. Mod. Phys. 82(1), 277–306. DOI: `10.1103/RevModPhys.82.277`. arXiv: `0808.3773`.

Eisert, J. and T. J. Osborne (2006). "General entanglement scaling laws from time evolution". Phys. Rev. Lett. 97(15), 150404. DOI: `10.1103/physrevlett.97.150404`. arXiv: `quant-ph/0603114`.

Espig, M., K. K. Naraparaju, and J. Schneider (2012). "A note on tensor chain approximation". Comput. Vis. Sci. 15(6), 331–344. DOI: `10.1007/s00791-014-0218-7`.

Fannes, M., B. Nachtergaele, and R. F. Werner (1992). "Finitely correlated states on quantum spin chains". Commun. Math. Phys. 144(3), 443–490. DOI: `10.1007/bf02099178`.

Finkelstein, J. (2004). "Pure-state informationally complete and 'really' complete measurements". Phys. Rev. A 70(5), 052107. DOI: `10.1103/PhysRevA.70.052107`. arXiv: `quant-ph/0407078`.

Fiurášek, J. (2001). "Maximum-likelihood estimation of quantum measurement". Phys. Rev. A 64, 024102. DOI: `10.1103/PhysRevA.64.024102`. arXiv: `quant-ph/0101027 [quant-ph]`.

Flammia, S. T., D. Gross, S. D. Bartlett, and R. Somma (2010). "Heralded polynomial-time quantum state tomography". arXiv: `1002.3839 [quant-ph]`.

Flammia, S. T., A. Silberfarb, and C. M. Caves (2005). "Minimal informationally complete measurements for pure states". Found. Phys. 35, 1985–2006. DOI: `10.1007/s10701-005-8658-z`. arXiv: `quant-ph/0404137`.

Flammia, S. T., D. Gross, Y.-K. Liu, and J. Eisert (2012). "Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators". New J. Phys. 14(9), 095022. DOI: `10.1088/1367-2630/14/9/095022`. arXiv: `1205.2300`.

Flammia, S. T. and Y.-K. Liu (2011). "Direct fidelity estimation from few Pauli measurements". Phys. Rev. Lett. 106(23), 230501. DOI: `10.1103/physrevlett.106.230501`. arXiv: `1104.4695`.

Foreman-Mackey, D., D. W. Hogg, D. Lang, and J. Goodman (2013). "emcee: The MCMC hammer". Publ. Astron. Soc. Pac. 125, 306–312. DOI: `10.1086/670067`. arXiv: `1202.3665 [astro-ph.IM]`.

Friis, N., O. Marty, C. Maier, C. Hempel, M. Holzäpfel, P. Jurcevic, M. B. Plenio, M. Huber, C. F. Roos, R. Blatt, and B. P. Lanyon (2018). "Observation of entangled states of a fully-controlled 20 qubit system". Phys. Rev. X 8 (2), 021012. DOI: `10.1103/PhysRevX.8.021012`. arXiv: `1711.11092 [quant-ph]`.

Gantmacher, F. R. (1986). "Matrizentheorie". Berlin, Heidelberg: Springer. ISBN: 978-3-642-71244-9.

Gilchrist, A., N. K. Langford, and M. A. Nielsen (2005). "Distance measures to compare real and ideal quantum processes". Phys. Rev. A 71(6), 062310. DOI: `10.1103/physreva.71.062310`. arXiv: `quant-ph/0408063`.

Gonçalves, D. S., M. A. Gomes-Ruggiero, and C. Lavor (2014). "Global convergence of diluted iterations in maximum-likelihood quantum tomography". Quantum Inf. Comput. 14, 966. arXiv: `1306.3057 [math-ph]`. `http://www.rintonpress.com/journals/qiconline.html#v14n1112`.

Gonçalves, D., M. Gomes-Ruggiero, and C. Lavor (2016). "A projected gradient method for optimization over density matrices". Optim. Method. Softw. 31(2), 328–341. DOI: `10.1080/10556788.2015.1082105`.

Goreinov, S. A., E. E. Tyrtyshnikov, and N. L. Zamarashkin (1997). "A theory of pseudoskeleton approximations". Linear Algebra Appl. 261(1), 1–21. DOI: `10.1016/S0024-3795(96)00301-1`.

Granade, C., J. Combes, and D. G. Cory (2016). "Practical Bayesian tomography". New J. Phys. 18(3), 033024. DOI: `10.1088/1367-2630/18/3/033024`. arXiv: `1509.03770 [quant-ph]`.

Grasedyck, L. (2010). "Hierarchical singular value decomposition of tensors". SIAM J. Matrix Anal. Appl. 31(4), 2029–2054. DOI: `10.1137/090764189`.

Grasedyck, L., D. Kressner, and C. Tobler (2013). "A literature survey of low-rank tensor approximation techniques". GAMM-Mitt. 36(1), 53–78. DOI: `10.1002/gamm.201310004`.

Greville, T. N. E. (1966). “Note on the generalized inverse of a matrix product”. SIAM Rev. 8(4), 518–521. DOI: `10.1137/1008107`.

Groisman, B., S. Popescu, and A. Winter (2005). “Quantum, classical, and total amount of correlations in a quantum state”. Phys. Rev. A 72, 032317. DOI: `10.1103/PhysRevA.72.032317`. arXiv: `quant-ph/0410091 [quant-ph]`.

Gross, D. (2011). “Recovering low-rank matrices from few coefficients in any basis”. IEEE Trans. Inf. Theory 57(3), 1548–1566. DOI: `10.1109/TIT.2011.2104999`. arXiv: `0910.1879 [cs.IT]`.

Gross, J. L., J. Yellen, and P. Zhang, eds. (2014). “Handbook of graph theory”. 2nd ed. Discrete Math. Appl. Boca Raton, Fla.: CRC Press. ISBN: 978-143-988-018-0.

Hackbusch, W. and S. Kühn (2009). “A new scheme for the tensor representation”. J. Fourier Analysis Appl. 15(5), 706–722. DOI: `10.1007/s00041-009-9094-9`.

Håstad, J. (1990). “Tensor rank is np-complete”. J. Algorithms 11(4), 644–654. DOI: `10.1016/0196-6774(90)90014-6`.

Hastings, M. B. and T. Koma (2006). “Spectral gap and exponential decay of correlations”. Commun. Math. Phys. 265, 781–804. DOI: `10.1007/s00220-006-0030-4`. arXiv: `math-ph/0507008`.

Hayden, P., R. Jozsa, D. Petz, and A. Winter (2004). “Structure of states which satisfy strong subadditivity of quantum entropy with equality”. Commun. Math. Phys. 246, 359–374. DOI: `10.1007/s00220-004-1049-z`. arXiv: `quant-ph/0304007`.

Heinosaari, T., L. Mazzarella, and M. M. Wolf (2013). “Quantum tomography under prior information”. Commun. Math. Phys. 318, 355–374. DOI: `10.1007/s00220-013-1671-8`. arXiv: `1109.5478 [quant-ph]`.

Henderson, L. and V. Vedral (2001). “Classical, quantum and total correlations”. J. Phys. A: Math. Gen. 34, 6899–6905. DOI: `10.1088/0305-4470/34/35/315`. arXiv: `quant-ph/0105028`.

Holtz, S., T. Rohwedder, and R. Schneider (2012). “The alternating linear scheme for tensor optimization in the tensor train format”. SIAM J. Sci. Comput. 34(2), A683–A713. DOI: `10.1137/100818893`.

Holzäpfel, M., T. Baumgratz, M. Cramer, and M. B. Plenio (2015). “Scalable reconstruction of unitary processes and Hamiltonians”. Phys. Rev. A 91(4), 042129. DOI: `10.1103/PhysRevA.91.042129`. arXiv: `1411.6379 [quant-ph]`.

Holzäpfel, M. and M. B. Plenio (2017). “Efficient certification and simulation of local quantum many-body Hamiltonians”. arXiv: `1712.04396 [quant-ph]`.

Holzäpfel, M., M. Cramer, N. Datta, and M. B. Plenio (2018). “Petz recovery versus matrix reconstruction”. J. Math. Phys. 59(4), 042201. DOI: `10.1063/1.5009658`. arXiv: `1709.04538 [quant-ph]`.

Horn, R. A. and C. R. Johnson (1991a). “Matrix analysis”. Cambridge: Cambridge University Press. ISBN: 0-521-30586-1.

– (1991b). “Topics in matrix analysis”. 1st ed. Cambridge: Cambridge University Press. ISBN: 0-521-30587-X.

Hradil, Z. (1997). “Quantum-state estimation”. Phys. Rev. A 55, R1561–R1564. DOI: `10.1103/PhysRevA.55.R1561`. arXiv: `quant-ph/9609012`.

Hradil, Z., J. Řeháček, J. Fiurášek, and M. Ježek (2004). “Maximum-likelihood methods in quantum mechanics”. In: Quantum State Estimation. Ed. by M. G. A. Paris and J. Řeháček. Vol. 649. Lect. Notes Phys. Berlin: Springer, 59–112. DOI: `10.1007/b98673`.

Hübener, R., V. Nebendahl, and W. Dür (2010). “Concatenated tensor network states”. New J. Phys. 12(2), 025004. DOI: `10.1088/1367-2630/12/2/025004`. arXiv: `0904.1925 [quant-ph]`.

Jamiołkowski, A. (1972). “Linear transformations which preserve trace and positive semidefiniteness of operators”. Rep. Math. Phys. 3(4), 275–287. DOI: `10.1016/0034-4877(72)90011-0`.

Jaynes, E. T. (2003). “Probability theory: The logic of science”. Ed. by G. L. Bretthorst. Cambridge: Cambridge University Press. ISBN: 978-0-521-59271-0.

Jones, E., T. Oliphant, P. Peterson, et al. (2001). Scipy: Open source scientific tools for Python. `http://www.scipy.org/`.

Jozsa, R. (2006). “On the simulation of quantum circuits”. arXiv: `quant-ph/0603163`.

Jurcevic, P., P. Hauke, C. Maier, C. Hempel, B. P. Lanyon, R. Blatt, and C. F. Roos (2015). “Spectroscopy of interacting quasiparticles in trapped ions”. Phys. Rev. Lett. 115(10), 100501. DOI: `10.1103/PhysRevLett.115.100501`. arXiv: `1505.02066 [quant-ph]`.

Kim, I. H. (2014). “On the informational completeness of local observables”. arXiv: `1405.0137 [quant-ph]`.

Kliesch, M., D. Gross, and J. Eisert (2014a). “Matrix-product operators and states: NP-hardness and undecidability”. Phys. Rev. Lett. 113(16), 160503. DOI: `10.1103/PhysRevLett.113.160503`. arXiv: `1404.4466 [quant-ph]`.

Kliesch, M., C. Gogolin, and J. Eisert (2014b). “Lieb-Robinson bounds and the simulation of time-evolution of local observables in lattice systems”. In: Many-Electron Approaches Phys. Chem. Math.: A Multidiscip. View. Ed. by V. Bach and L. Delle Site. Cham: Springer, 301–318. ISBN: 978-3-319-06379-9. DOI: `10.1007/978-3-319-06379-9_17`. arXiv: `1306.0716 [quant-ph]`.

Knee, G. C., E. Bolduc, J. Leach, and E. M. Gauger (2018). “Quantum process tomography via completely positive and trace-preserving projection”. Phys. Rev. A 98 (6), 062336. DOI: `10.1103/PhysRevA.98.062336`. arXiv: `1803.10062`.

Knill, E. (1995). “Approximation by quantum circuits”. arXiv: `quant-ph/9508006`.

Kolda, T. and B. Bader (2009). “Tensor decompositions and applications”. SIAM Rev. 51(3), 455–500. DOI: `10.1137/07070111X`.

Kressner, D., M. Steinlechner, and A. Uschmajew (2014). “Low-rank tensor methods with subspace correction for symmetric eigenvalue problems”. SIAM J. Sci. Comput. 36(5), A2346–A2368. DOI: `10.1137/130949919`.

Lam, S. K., A. Pitrou, and S. Seibert (2015). “Numba: A LLVM-based Python JIT compiler”. In: Proc. Second Workshop LLVM Compiler Infrastructure HPC. LLVM ’15. Austin, Texas: ACM, 7:1–7:6. ISBN: 978-1-4503-4005-2. DOI: `10.1145/2833157.2833162`.

Lanyon, B. P., C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos (2017). “Efficient tomography of a quantum many-body system”. Nat. Phys. 13, 1158–1162. DOI: `10.1038/nphys4244`. arXiv: `1612.08000 [quant-ph]`.

Lathauwer, L. D., B. D. Moor, and J. Vandewalle (2000). “A multilinear singular value decomposition”. SIAM J. Matrix Anal. Appl. 21(4), 1253–1278. DOI: `10.1137/S0895479896305696`.

Lee, N. and A. Cichocki (2015). “Estimating a few extreme singular values and vectors for large-scale matrices in tensor train format”. SIAM J. Matrix Anal. Appl. 36(3), 994–1014. DOI: `10.1137/140983410`.

– (2016). “Regularized computation of approximate pseudoinverse of large matrices using low-rank tensor train decompositions”. SIAM J. Matrix Anal. Appl. 37(2), 598–623. DOI: 10.1137/15M1028479. arXiv: 1506.01959 [math.NA].

Lehmann, E. L. and G. Casella (2003). “Theory of point estimation”. 2nd ed. New York: Springer. ISBN: 0-387-98502-6.

Leonhardt, U. and M. Munroe (1996). “Number of phases required to determine a quantum state in optical homodyne tomography”. Phys. Rev. A 54, 3682–3684. DOI: 10.1103/PhysRevA.54.3682.

Lieb, E. H. and D. W. Robinson (1972). “The finite group velocity of quantum spin systems”. Comm. Math. Phys. 28(3), 251–257. DOI: 10.1007/BF01645779. https://projecteuclid.org/euclid.cmp/1103858407.

Liu, Y.-K. (2011). “Universal low-rank matrix recovery from Pauli measurements”. In: NIPS 24, 1638–1646. arXiv: 1103.2816 [quant-ph]. http://papers.nips.cc/paper/4222-universal-low-rank-matrix-recovery-from-pauli-measurements.pdf.

Lvovsky, A. I. (2004). “Iterative maximum-likelihood reconstruction in quantum homodyne tomography”. J. Opt. B: Quantum Semiclassical Opt. 6, 556. DOI: 10.1088/1464-4266/6/6/014. arXiv: quant-ph/0311097.

Lvovsky, A. I. and M. G. Raymer (2009). “Continuous-variable optical quantum-state tomography”. Rev. Mod. Phys. 81, 299–332. DOI: 10.1103/RevModPhys.81.299. arXiv: quant-ph/0511044.

Mahoney, M. W., M. Maggioni, and P. Drineas (2008). “Tensor-CUR decompositions for tensor-based data”. SIAM J. Matrix Anal. Appl. 30(3), 957–987. DOI: 10.1137/060665336.

Molina-Terriza, G., A. Vaziri, J. Řeháček, Z. Hradil, and A. Zeilinger (2004). “Triggered qutrits for quantum communication protocols”. Phys. Rev. Lett. 92(16), 167903. DOI: 10.1103/PhysRevLett.92.167903. arXiv: quant-ph/0401183.

Molnar, A., N. Schuch, F. Verstraete, and J. I. Cirac (2015). “Approximating Gibbs states of local Hamiltonians efficiently with projected entangled pair states”. Phys. Rev. B 91(4), 045138. DOI: 10.1103/PhysRevB.91.045138. arXiv: 1406.2973 [quant-ph].

Mood, A. M., F. A. Graybill, and D. C. Boes (1974). “Introduction to the theory of statistics”. 3rd ed. Tokyo: McGraw-Hill. ISBN: 0-07-042864-6.

Moore, E. H. (1935). “General analysis”. Ed. by R. W. Barnard. Vol. 1. Philadelphia: The American Philosophical Society. https://catalog.hathitrust.org/api/volumes/oclc/10623960.html.

– (1939). “General analysis”. Ed. by R. W. Barnard. Vol. 2. Philadelphia: The American Philosophical Society. https://catalog.hathitrust.org/api/volumes/oclc/10623960.html.

Murg, V., F. Verstraete, and J. I. Cirac (2007). “Variational study of hard-core bosons in a two-dimensional optical lattice using projected entangled pair states”. Phys. Rev. A 75(3), 033605. DOI: 10.1103/PhysRevA.75.033605. arXiv: cond-mat/0611522.

Nachtergaele, B., Y. Ogata, and R. Sims (2006). “Propagation of correlations in quantum lattice systems”. J. Stat. Phys. 124, 1–13. DOI: 10.1007/s10955-006-9143-6. arXiv: math-ph/0603064.

Nachtergaele, B., H. Raz, B. Schlein, and R. Sims (2009). “Lieb-Robinson bounds for harmonic and anharmonic lattice systems”. Commun. Math. Phys. 286, 1073–1098. DOI: 10.1007/s00220-008-0630-2. arXiv: 0712.3820 [math-ph].

Nachtergaele, B., B. Schlein, R. Sims, S. Starr, and V. Zagrebnov (2010). "On the existence of the dynamics for anharmonic quantum oscillator systems". Rev. Math. Phys. 22, 207–231. DOI: `10.1142/S0129055X1000393X`. arXiv: `0909.2249 [math-ph]`.

Nachtergaele, B. and R. Sims (2006). "Lieb-Robinson bounds and the exponential clustering theorem". Commun. Math. Phys. 265, 119–130. DOI: `10.1007/s00220-006-1556-1`. arXiv: `math-ph/0506030`.

– (2014). "On the dynamics of lattice systems with unbounded on-site terms in the Hamiltonian". arXiv: `1410.8174 [math-ph]`.

Nielsen, M. A. (1999). "Conditions for a class of entanglement transformations". Phys. Rev. Lett. 83, 436–439. DOI: `10.1103/PhysRevLett.83.436`. arXiv: `quant-ph/9811053 [quant-ph]`.

Nielsen, M. A., C. M. Dawson, J. L. Dodd, A. Gilchrist, D. Mortimer, T. J. Osborne, M. J. Bremner, A. W. Harrow, and A. Hines (2003). "Quantum dynamics as a physical resource". Phys. Rev. A 67(5), 052301. DOI: `10.1103/PhysRevA.67.052301`. arXiv: `quant-ph/0208077`.

Nielsen, M. A. and I. L. Chuang (2007). "Quantum computation and quantum information". 9th ed. Cambridge: Cambridge University Press. DOI: `10.1017/CBO9780511976667`.

NIST Digital Library of Mathematical Functions (2010). Release 1.0.22 of 2019-03-15. `https://dlmf.nist.gov/`. (Online companion to Olver, Lozier, Boisvert, and Clark 2010.)

Ollivier, H. and W. H. Zurek (2002). "Quantum discord: A measure of the quantumness of correlations". Phys. Rev. Lett. 88(1), 017901. DOI: `10.1103/PhysRevLett.88.017901`. arXiv: `quant-ph/0105072`.

Olver, F. W. J., D. W. Lozier, R. F. Boisvert, and C. W. Clark, eds. (2010). "NIST handbook of mathematical functions". New York, NY: Cambridge University Press. ISBN: 9780521140638. (Print companion to *NIST Digital Library of Mathematical Functions* 2010.)

Osborne, T. J. (2006). "Efficient approximation of the dynamics of one-dimensional quantum spin systems". Phys. Rev. Lett. 97(15), 157202. DOI: `10.1103/physrevlett.97.157202`. arXiv: `quant-ph/0508031`.

Oseledets, I. V. (2011). "Tensor-train decomposition". SIAM J. Sci. Comput. 33(5), 2295–2317. DOI: `10.1137/090752286`.

Oseledets, I. V. and S. V. Dolgov (2012). "Solution of linear systems and matrix inversion in the TT-format". SIAM J. Sci. Comput. 34(5), A2718–A2739. DOI: `10.1137/110833142`.

Oseledets, I. V., D. V. Savostyanov, and E. E. Tyrtyshnikov (2008). "Tucker dimensionality reduction of three-dimensional arrays in linear time". SIAM J. Matrix Anal. Appl. 30(3), 939–956. DOI: `10.1137/060655894`.

Oseledets, I. V. (2013). Python implementation of the TT-toolbox. `https://github.com/oseledets/ttpy`.

Oseledets, I. V. and E. E. Tyrtyshnikov (2010). "TT-cross approximation for multidimensional arrays". Linear Algebra Appl. 432(1), 70–88. DOI: `10.1016/j.laa.2009.07.024`.

Östlund, S. and S. Rommer (1995). "Thermodynamic limit of density matrix renormalization". Phys. Rev. Lett. 75(19), 3537–3540. DOI: `10.1103/physrevlett.75.3537`. arXiv: `cond-mat/9503107`.

Pedersen, L. H., N. M. Møller, and K. Mølmer (2007). "Fidelity of quantum operations". Phys. Lett. A 367(1-2), 47–51. DOI: `10.1016/j.physleta.2007.02.069`. arXiv: `quant-ph/0701138`.

Penrose, R. (1955). “A generalized inverse for matrices”. Math. Proc. Cambridge Philos. Soc. 51 (03), 406–413. DOI: `10.1017/S0305004100030401`.

Perez-Garcia, D., F. Verstraete, M. M. Wolf, and J. I. Cirac (2007). “Matrix product state representations”. Quantum Inf. Comput. 7, 401. ISSN: 1533-7146. arXiv: `quant-ph/0608197`. `http://www.rintonpress.com/journals/qiconline.html#v7n56`.

Petz, D. (2003). “Monotonicity of quantum relative entropy revisited”. Rev. Math. Phys. 15, 79–91. DOI: `10.1142/S0129055X03001576`. arXiv: `quant-ph/0209053`.

Piani, M., P. Horodecki, and R. Horodecki (2008). “No-local-broadcasting theorem for multipartite quantum correlations”. Phys. Rev. Lett. 100(9), 090502. DOI: `10.1103/PhysRevLett.100.090502`. arXiv: `0707.0848 [quant-ph]`.

Poulin, D., A. Qarry, R. Somma, and F. Verstraete (2011). “Quantum simulation of time-dependent Hamiltonians and the convenient illusion of Hilbert space”. Phys. Rev. Lett. 106(17), 170501. DOI: `10.1103/PhysRevLett.106.170501`. arXiv: `1102.1360 [quant-ph]`.

Poulin, D. and M. B. Hastings (2011). “Markov entropy decomposition: A variational dual for quantum belief propagation”. Phys. Rev. Lett. 106(8), 080403. DOI: `10.1103/physrevlett.106.080403`. arXiv: `1012.2050`.

Poyatos, J. F., R. Walser, J. I. Cirac, P. Zoller, and R. Blatt (1996). “Motion tomography of a single trapped ion”. Phys. Rev. A 53, R1966–R1969. DOI: `10.1103/PhysRevA.53.R1966`. arXiv: `atom-ph/9601001 [physics.atom-ph]`.

Prugovečki, E. (1977). “Information-theoretical aspects of quantum measurement”. Int. J. Theor. Phys. 16(5), 321–331. DOI: `10.1007/BF01807146`.

Rado, R. (1956). “Note on generalized inverses of matrices”. Math. Proc. Cambridge Philos. Soc. 52 (03), 600–601. DOI: `10.1017/S0305004100031601`.

Raginsky, M. (2001). “A fidelity measure for quantum channels”. Phys. Lett. A 290(1-2), 11–18. DOI: `10.1016/s0375-9601(01)00640-5`. arXiv: `quant-ph/0107108`.

Řeháček, J., Z. Hradil, E. Knill, and A. I. Lvovsky (2007). “Diluted maximum-likelihood algorithm for quantum tomography”. Phys. Rev. A 75(4), 042108. DOI: `10.1103/PhysRevA.75.042108`. arXiv: `quant-ph/0611244`.

Renes, J. M., R. Blume-Kohout, A. J. Scott, and C. M. Caves (2004). “Symmetric informationally complete quantum measurements”. J. Math. Phys. 45(6), 2171–2180. DOI: `10.1063/1.1737053`. arXiv: `quant-ph/0310075`.

Renner, R. (2005). “Security of quantum key distribution”. PhD thesis. Swiss Federal Institute of Technology, Zurich.

Safieldeen, B. T. (2016). “Quantum state tomography using maximum likelihood estimation on purified matrix product states”. Bachelor thesis. Ulm University.

Savostyanov, D. V. and I. V. Oseledets (2011). “Fast adaptive interpolation of multi-dimensional arrays in tensor train format”. In: 2011 Int. Workshop Multidim. (nD) Syst. 1–8. DOI: `10.1109/nDS.2011.6076873`.

Schindler, P., D. Nigg, T. Monz, J. T. Barreiro, E. Martinez, S. X. Wang, S. Quint, M. F. Brandl, V. Nebendahl, C. F. Roos, M. Chwalla, M. Hennrich, and R. Blatt (2013). “A quantum information processor with trapped ions”. New J. Phys. 15(12), 123012. DOI: `10.1088/1367-2630/15/12/123012`. arXiv: `1308.3096 [quant-ph]`.

Schollwöck, U. (2011). "The density-matrix renormalization group in the age of matrix product states". Ann. Phys. 326(1), 96–192. DOI: `10.1016/j.aop.2010.09.012`. arXiv: `1008.3477`.

Schuch, N., M. M. Wolf, F. Verstraete, and J. I. Cirac (2007). "Computational complexity of projected entangled pair states". Phys. Rev. Lett. 98(14), 140506. DOI: `10.1103/PhysRevLett.98.140506`. arXiv: `quant-ph/0611050`.

Schuch, N., M. M. Wolf, F. Verstraete, and J. I. Cirac (2008). "Entropy scaling and simulability by matrix product states". Phys. Rev. Lett. 100(3), 030504. DOI: `10.1103/physrevlett.100.030504`. arXiv: `0705.0292`.

Schwemmer, C., L. Knips, D. Richart, H. Weinfurter, T. Moroder, M. Kleinmann, and O. Gühne (2015). "Systematic errors in current quantum state tomography tools". Phys. Rev. Lett. 114(8), 080403. DOI: `10.1103/PhysRevLett.114.080403`. arXiv: `1310.8465 [quant-ph]`.

Shang, J., H. Khoon Ng, and B.-G. Englert (2014). "Quantum state tomography: Mean squared error matters, bias does not". arXiv: `1405.5350 [quant-ph]`.

Shang, J., Z. Zhang, and H. Khoon Ng (2017). "Superfast maximum-likelihood reconstruction for quantum tomography". Phys. Rev. A 95 (6), 062336. DOI: `10.1103/PhysRevA.95.062336`. arXiv: `1609.07881 [quant-ph]`.

Siah Teo, Y. (2013). "Numerical estimation schemes for quantum tomography". PhD thesis. National University of Singapore. arXiv: `1302.3399 [quant-ph]`. `http://scholarbank.nus.edu.sg/handle/10635/34694`.

Smithey, D. T., M. Beck, M. G. Raymer, and A. Faridani (1993). "Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography – application to squeezed states and the vacuum". Phys. Rev. Lett. 70, 1244–1247. DOI: `10.1103/PhysRevLett.70.1244`.

Stewart, G. W. (1977). "On the perturbation of pseudo-inverses, projections and linear least squares problems". SIAM Rev. 19(4), 634–662. DOI: `10.1137/1019104`.

Suess, D. and M. Holzäpfel (2017). "mpnum: A matrix product representation library for Python". J. Open Source Softw. 2(20), 465. DOI: `10.21105/joss.00465`. (Source code and documentation available at `https://github.com/dseuss/mpnum` and `https://mpnum.readthedocs.io/`.)

Suzuki, M. (1985). "Decomposition formulas of exponential operators and Lie exponentials with some applications to quantum mechanics and statistical physics". J. Math. Phys. 26(4), 601–612. DOI: `10.1063/1.526596`.

Terhal, B. M. and P. Horodecki (2000). "Schmidt number for density matrices". Phys. Rev. A 61, 040301. DOI: `10.1103/PhysRevA.61.040301`. arXiv: `quant-ph/9911117 [quant-ph]`.

Top500 supercomputer list (2018). November 2018. `https://www.top500.org/`.

Tucker, L. R. (1964). "The extension of factor analysis to three-dimensional matrices". In: Contributions to mathematical psychology. Ed. by H. Gulliksen and N. Frederiksen. New York: Holt, Rinehart and Winston, 109–127.

Tucker, L. R. (1966). "Some mathematical notes on three-mode factor analysis". Psychometrika 31(3), 279–311. DOI: `10.1007/BF02289464`.

Vanner, M. R., J. Hofer, G. D. Cole, and M. Aspelmeyer (2013). "Cooling-by-measurement and mechanical state tomography via pulsed optomechanics". Nat. Commun. 4, 2295. DOI: `10.1038/ncomms3295`. arXiv: `1211.7036 [quant-ph]`.

Verstraete, F. and J. I. Cirac (2004). “Renormalization algorithms for quantum-many body systems in two and higher dimensions”. arXiv: `cond-mat/0407066`.

Verstraete, F., J. J. García-Ripoll, and J. I. Cirac (2004). “Matrix product density operators: Simulation of finite-temperature and dissipative systems”. Phys. Rev. Lett. 93(20), 207204. DOI: `10.1103/PhysRevLett.93.207204`. arXiv: `cond-mat/0406426`.

Verstraete, F., V. Murg, and J. I. Cirac (2008). “Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems”. Adv. Phys. 57, 143–224. DOI: `10.1080/14789940801912366`. arXiv: `0907.2796 [quant-ph]`.

Vidal, G. (2003). “Efficient classical simulation of slightly entangled quantum computations”. Phys. Rev. Lett. 91(14), 147902. DOI: `10.1103/PhysRevLett.91.147902`. arXiv: `quant-ph/0301063`.

– (2004). “Efficient simulation of one-dimensional quantum many-body systems”. Phys. Rev. Lett. 93(4), 040502. DOI: `10.1103/physrevlett.93.040502`. arXiv: `quant-ph/0310089`.

Vidyasagar, M. (2011). “The complete realization problem for hidden Markov models: A survey and some new results”. Math. Control Signals Syst. 23(1), 1–65. DOI: `10.1007/s00498-011-0066-7`.

Vogel, K. and H. Risken (1989). “Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase”. Phys. Rev. A 40 (5), 2847–2849. DOI: `10.1103/PhysRevA.40.2847`.

Wedin, P.-Å. (1973). “Perturbation theory for pseudo-inverses”. BIT Numer. Math. 13(2), 217–232. DOI: `10.1007/BF01933494`.

Werner, R. F. (1989). “Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model”. Phys. Rev. A 40 (8), 4277–4281. DOI: `10.1103/PhysRevA.40.4277`.

Yoran, N. and A. J. Short (2007). “Efficient classical simulation of the approximate quantum Fourier transform”. Phys. Rev. A 76(4), 042321. DOI: `10.1103/physreva.76.042321`. arXiv: `quant-ph/0611241`.

Zhao, Y.-Y., Z. Hou, G.-Y. Xiang, Y.-J. Han, C.-F. Li, and G.-C. Guo (2017). “Experimental demonstration of efficient quantum state tomography of matrix product states”. Opt. Express 25, 9010. DOI: `10.1364/OE.25.009010`.

Zwolak, M. and G. Vidal (2004). “Mixed-state dynamics in one-dimensional quantum lattice systems: A time-dependent superoperator renormalization algorithm”. Phys. Rev. Lett. 93(20), 207205. DOI: `10.1103/PhysRevLett.93.207205`. arXiv: `cond-mat/0406440`.

177 entries

Lists and indices

List of figures

2.1. Operator product of projected entangled pair operators (PEPOs) on an arbitrary graph 33

4.1. Estimation error of efficient MLE with a finite number of measurements 64
4.2. Estimation of a three-mode state from three-mode measurements . 69
4.3. Estimation of a six-mode state from three-mode measurements . . 70
4.4. Estimation of a six-mode state from three-mode measurements (part 2) 71

5.1. Basic principle of scalable ancilla-assisted process tomography (AAPT) 76
5.2. Reconstruction error of unitary circuits 82
5.3. Reconstruction quality of local Hamiltonians 85
5.4. Two-time reconstruction error of local Hamiltonians 87

6.1. Certifying time-evolved states (summary) 90
6.2. Time evolution of a local observable under a local Hamiltonian . . 94
6.3. Operator norm error of the first order Trotter decomposition as a function of the number of sites 106
6.4. Decomposition of an η-dimensional hypercube 114

7.1. Bipartite entanglement of time-evolved states and estimates. 130
7.2. Certified tomography of 8 qubits. 131
7.3. Certified tomography of 14 qubits. 132

8.1. Different tensor representations and their dimension trees 144
8.2. Structure of the tensor reconstruction from Theorem 8.17 149

List of tables

1.1. Projective, generalized and positive operator-valued measure (POVM) measurements 14

2.1. Tensor notation and terminology 18

7.1. Estimators of energy and its standard deviation 129

8.1. Comparison of tensor reconstruction proposals 163

Acronyms

AAPT ancilla-assisted process tomography. iii, 73, 74, 76, 79

BHD balanced homodyne detection. 66

BME Bayesian mean estimate. 47

BSGS block sequentially generated state. 105

CCD charge-coupled device. 122, 123

CP completely positive. 73, 74

CPTP completely positive and trace preserving. 74, 80

CW continuous wave. 122

DFE direct fidelity estimation. 55, 132

DFS depth-first search. 34

DMRG density matrix renormalization group. 30, 54, 64, 80, 83, 165

DPI data processing inequality. 154

FCS finitely correlated state. 22

GHZ Greenberger–Horne–Zeilinger. 30, 81

HMM hidden Markov model. 22

HO-SVD higher-order singular value decomposition. 146

i.i.d. independent and indentically distributed. 5, 6, 128

IC informationally complete. 38, 41, 43, 44, 49, 67, 100–102, 124, 127

LOCC local operations and classical communication. 12

MBQC measurement-based quantum computation. 104

MLE maximum likelihood estimation. iii, 43, 45, 47–52, 63–65, 67–70, 79, 125, 133, 167, 168

MPDO matrix product density operator. 21

MPO matrix product operator. 21–24, 28–33, 49, 50, 52–54, 60, 75, 77, 78, 80, 82, 83, 107, 133, 135, 150, 151, 153, 154, 164, 168, 169

MPS matrix product state. iii, iv, 17–21, 23–26, 28–32, 34, 45, 48–50, 52–55, 57–59, 63–65, 73, 75, 78–80, 91, 105, 125, 126, 135, 137, 145, 146, 148, 150, 151, 163, 164, 167–170

MSE mean squared error. 5, 168

OBC open boundary conditions. 21

OSR operator Schmidt rank. 12, 13, 23, 37, 54, 151

PBC periodic boundary conditions. 21, 164

PDF probability density function. 3, 4, 6

PEPO projected entangled pair operator. 32–35, 82, 91, 120

PEPS projected entangled pair state. iii, 17, 19, 32–34, 89, 91, 104, 105, 107, 110, 111, 113, 118–120, 169

PMPS locally purified matrix product state. 21–24, 31, 49–52, 63, 70, 71, 77, 133, 168

POVM positive operator-valued measure. 14, 15, 41–45, 47, 49, 51, 52, 65, 67–70, 72, 79, 80, 100–102, 123–128, 130

QST quantum state tomography. 37–39, 45, 49, 53, 73, 74, 76, 121

RAM random access memory. 17

SVD singular value decomposition. 9, 10, 19, 28–30, 53

SVT singular value thresholding. 49, 52–54, 65, 125, 167, 168

TNS tensor network state. 104

TT tensor train. iii, 17–21, 24–26, 28, 30, 31, 34, 135, 137, 145, 146, 148, 150, 151, 163, 164, 169, 170