Master's thesis

Master's Programme in Particle Physics and Astrophysical Sciences

Elementary particle physics and cosmology

Quantum State Tomography with Observable Commutation Graphs

Otto Veltheim

March, 2022

Supervisor: Esko Keski-Vakkuri
Examiners: Esko Keski-Vakkuri, Kimmo Tuominen

University of Helsinki
Faculty of Science

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI
MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
MATEMATISK-NATURVETENSKAPLIGA FAKULTETEN
FACULTY OF SCIENCE

Tiedekunta – Fakultet – Faculty	Koulutusohjelma – Utbildningsprogram – Degree programme
Faculty of Science	Master's Programme in Particle Physics and Astrophysical Sciences

Opintosuunta – Studieinrikting – Study track
Elementary particle physics and cosmology

Tekijä – Författare – Author
Otto Veltheim

Työn nimi – Arbetets titel – Title
Quantum State Tomography with Observable Commutation Graphs

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Master's degree	March 2022	56

Tiivistelmä – Referat – Abstract

The measurement of quantum states has been a widely studied problem ever since the discovery of quantum mechanics. In general, we can only measure a quantum state once as the measurement itself alters the state and, consequently, we lose information about the original state of the system in the process. Furthermore, this single measurement cannot uncover every detail about the system's state and thus, we get only a limited description of the system.

However, there are physical processes, e.g., a quantum circuit, which can be expected to create the same state over and over again. This allows us to measure multiple identical copies of the same system in order to gain a fuller characterization of the state. This process of diagnosing a quantum state through measurements is known as quantum state tomography. However, even if we are able to create identical copies of the same system, it is often preferable to keep the number of needed copies as low as possible. In this thesis, we will propose a method of optimising the measurements in this regard.

The full description of the state requires determining multiple different observables of the system. These observables can be measured from the same copy of the system only if they commute with each other. As the commutation relation is not transitive, it is often quite complicated to find the best way to match the observables with each other according to these commutation relations. This can be quite handily illustrated with graphs. Moreover, the best way to divide the observables into commuting sets can then be reduced to a well-known graph theoretical problem called graph colouring.

Measuring the observables with acceptable accuracy also requires measuring each observable multiple times. This information can also be included in the graph colouring approach by using a generalisation called multicolouring. Our results show that this multicolouring approach can offer significant improvements in the number of needed copies when compared to some other known methods.

Avainsanat – Nyckelord – Keywords
quantum tomography, quantum state tomography, quantum measurement, quantum observable, quantum information, graph coloring, multicoloring

Säilytyspaikka – Förvaringställe – Where deposited
Kumpula Campus Library

Contents

Acknowledgements

I would like to thank my supervisor Esko Keski-Vakkuri for suggesting looking into quantum tomography as a thesis topic and for all of the discussions, comments and encouragements that were extremely valuable during the writing of this thesis. I would also like to thank Guillermo García-Pérez for sharing his vast knowledge of quantum tomography and helping to understand the larger context with his comments and reference suggestions. Also, huge thanks to Marcel Niedermeier for providing countless helpful comments which improved the thesis. Last but not least, I'm extremely grateful to my family and friends for their tremendous support throughout the writing process!

1 Introduction

The discovery of quantum mechanics in the turn of the twentieth century came coupled with a completely new way of describing the states of particles. In classical physics, we are used to describing any physical state with some well-defined parameters such as a set of coordinates. However, quantum mechanics has completely changed this kind of thinking by showing that the parameters, which we thought were well-defined, are actually not. In the past century our understanding of this has increased tremendously but there is still a lot more to discover.

The complex behaviour of quantum systems might seem to complicate our understanding of the world, but it also enables many things that classical physics would not. As an example, even though the development of quantum computers is still in its infancy, there have already been discovered multiple significant algorithms for quantum computers, such as Shor's algorithm for prime factorization [1], which can efficiently solve problems for which there are no currently known efficient algorithms for classical computers. Also, already we have seen claims of current quantum computers having managed to solve problems which would take countless of years to solve on a classical computer [2]. These achievements emphasize the importance that we can understand, describe and verify the underlying quantum states which allow these new accomplishments. However, the importance of being able to process this quantum information is not only limited to quantum computing but is also crucial in the study of any system, large or small, which exhibits quantum behaviour.

While we cannot describe the states of a quantum system in the same way as we could with a classical state, it does not mean that the states cannot be described at all. On contrary, quantum mechanics offers a much richer set of possible physical states than classical physics. However, the biggest difference is that we cannot observe the state of a single physical system without also affecting it and, consequently, losing some information about the original state. Therefore, we need to come up with solutions for measuring the relevant information most efficiently.

Quantum tomography means the identification of an unknown quantum state through measurements. We cannot gain all of the information about a general quantum state from just one measurement and, since any measurement alters the state, we need, in general, multiple copies of the same state in order to identify the state. This, of course, presents some difficulties. Due to the no-cloning theorem, we cannot clone an unknown quantum state to a separate system without destroying the original state in the process [3]. Therefore, since we need multiple copies of the same quantum system for our measurements, in general, we can only perform quantum tomography for systems where we reproduce the same system multiple times using the same process or know that some kind of signal transmitter sends us a constant stream of similar states.

In order to make the identification of the state through quantum tomography slightly

easier, there have been developed multiple different approaches for different situations. Performing full tomography, i.e. completely identifying the state, for a general unknown state is the most difficult scenario, which requires a huge number of copies of the system and many different measurements. However, if we know that the unknown quantum state must belong to some restricted set of states, we can come up with more efficient methods. For example, [4] and [5] consider states which can be represented as matrix product states [6], and find very efficient ways of identifying those states. Even more extreme example would be the problem of trying to identify a state from only two possible states which can be represented as quantum hypothesis testing, see, e.g., [7] for a brief review.

Another way of approaching the problem is not to limit the set of possible states but instead limit the amount of information that we need to extract from the states. I.e., we select in advance some set of observables which we wish to measure from the state, but it does not have to include every possible observable if we are only interested in some subset of them. There has been some study for this approach for either some predefined set of interesting observables, such as every k-local observable in [8], or also for some more general set of observables, e.g., in [9].

In this thesis, we also focus on trying to find an efficient way to measure some freely chosen subset of the observables since this approach generalises all of the other cases. Full tomography is the special case where our subset if the full set of observables, and distinguishing a state from some restricted set of states just means that we can select our to-be-measured observables to coincide with those observables which can distinguish the states. We show that the most efficient ways to perform the measurements can be found by converting the problem into a mathematical problem called graph colouring. This graph colouring approach has also be used in the context of variational quantum eigensolvers, e.g., in [10]-[14]. In this thesis we will show, however, that by also implementing a generalisation of graph colouring called multicolouring, this approach produces the optimal measurement procedure in the context of quantum tomography.

There are some clear potential applications for efficient quantum tomography, e.g., in quantum chemistry, but the most obvious is naturally quantum computing. With the continuing development of even more powerful quantum computers, one needs to analyse and verify the correct behaviour of each individual component of the computers. To do this, we need to assess how each of the components affect the state of a qubit by measuring the states after letting the qubit interact with the component, i.e., we wish to verify with quantum tomography that the state of the qubit is what we would expect after interacting with the component. However, our future goals also include finding a way to perform quantum tomography in a wave packet basis in conformal field theory and using the graph colouring approach to find optimal measurement procedures in that setting. This could, in theory, be used to study black holes via their radiation, although the requirement of identical copies of the measured system might make this concept purely theoretical.

2 Quantum state tomography

The aim of *quantum state tomography* is to identify the unknown state of a given quantum system through measurements. In this thesis we will consider full tomography, i.e., determining the exact state of the system, only as a special case of trying to measure a given set of functions that we are interested in. The exact state can often be less relevant as we are more interested in some specific set of properties, i.e., *observables*, of the system.

2.1 Quantum states

Throughout this thesis we will assume, unless otherwise noted, that we are dealing with a system composed of n qubits. Other kinds of bases, particularly generalisations to qudits, are considered in Appendix A.

We will denote an unknown quantum state by its density matrix, ρ, which can be any positive operator with unit trace. For more information on quantum states in the context of quantum information theory, see, e.g., [15].

2.1.1 Pauli basis for linear operators

Let V be the 2^n-dimensional Hilbert space and let L_V be the space of linear operators on V. We can equip L_V with the Hilbert-Schmidt inner product.

Definition 2.1 (Hilbert-Schmidt inner product)**.** The *Hilbert-Schmidt inner product* of two linear operators A, $B \in L_V$ is defined as

$$(A,B) = \mathrm{tr}(A^\dagger B). \tag{2.1}$$

Moreover, since the familiar Pauli operators

$$\sigma_0 = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \sigma_1 = X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \sigma_2 = Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad \sigma_3 = Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{2.2}$$

span the set of linear operators for a 2-dimensional vector space, and

$$\mathrm{tr}([\sigma_{a_1} \otimes \sigma_{a_2} \otimes \cdots \otimes \sigma_{a_n}][\sigma_{b_1} \otimes \sigma_{b_2} \otimes \cdots \otimes \sigma_{b_n}]) = \prod_i 2\delta_{a_i b_i}, \tag{2.3}$$

we notice that the Pauli basis consisting of elements of the form $\sigma_{a_1} \otimes \sigma_{a_2} \otimes \cdots \otimes \sigma_{a_n}$ gives an orthogonal basis for L_V.

This also means that any linear operator A may be written in the form

$$A = \frac{1}{2^n} \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} \mathrm{tr}(PA)P. \tag{2.4}$$

Figure 2.1: ***(a)*** *A quantum state ρ which is acted on with the unitary gate U becomes the state $U\rho U^\dagger$.* ***(b)*** *An example circuit with two qubits which are acted on by two unitary operations, a single-qubit operation U_1 acting on the first qubit and a two-qubit operation U_2 acting on both of the qubits.*

The elements $P \in \{I,X,Y,Z\}^{\otimes n}$ are called *Pauli strings* and the number of non-I terms in the tensor product is the *length* of the Pauli string.

As is usual, we will use short-hand notation for the Pauli strings so that, e.g., X_i means $I^{\otimes i-1} \otimes X \otimes I^{\otimes n-i}$ giving, e.g., the following Pauli string of length 3

$$X_1Y_2X_5 = X \otimes Y \otimes I \otimes I \otimes X \text{ (for } n = 5). \tag{2.5}$$

2.1.2 Quantum circuits

In this thesis, we will generally think of the quantum measurements occurring in quantum circuits and, consequently, we will speak of the necessary quantum gates to construct the circuit performing the measurement. In the quantum circuit model, we think of quantum states travelling through some wires and being acted on by single-qubit or multi-qubit gates. Usually, unless otherwise noted, each wire is used to represent just a single qubit. Generally, we expect that the circuit represents a closed system where the states can only evolve through unitary transformations and these unitary transformations are represented by the gates.

Figure 2.1a shows how this unitary transformation is represented in the circuit: a quantum state ρ which goes through a unitary gate U becomes the state $U\rho U^\dagger$. Figure 2.1b shows an example involving also two qubit-gates. In this example, the first unitary transformation could equally well also be thought as the two-qubit operation $U_1 \otimes I$.

Naturally, we could replace the whole circuit with just a single unitary gate acting on all of the qubits, but, instead of using arbitrary unitary operations, we want to be able to construct the circuits with just some limited set of gates. The quantum gates that we use are listed in Table 2.1. The first three are not strictly necessary but they are easy to implement and since they serve also as our Hilbert-Schmidt basis, they are listed for reference.

We will consider three different sets of basic gates. The first and simplest one, which is used for what we will call *Pauli measurements*, consists of only single qubit operations. Furthermore, these operations all have the property that they will transform Pauli matrices into Pauli matrices under conjugation, at most including some phase factor, e.g., $HXH = Z$

Table 2.1: Quantum gates needed for each of the three considered measurement schemes.

Gate	Matrix representation	Pauli measurements	Clifford measurements	Arbitrary measurements
X	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	(x)	(x)	(x)
Y	$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$	(x)	(x)	(x)
Z	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	(x)	(x)	(x)
H	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$	x	x	x
S	$\begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$	x	x	x
$CNOT$	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$		x	x
T	$\begin{pmatrix} 1 & 0 \\ 0 & \frac{1+i}{\sqrt{2}} \end{pmatrix}$			x

and $ZXZ = -X$. Using products of these gates for each of the qubits, we can transform any Pauli string to any other Pauli string with the only constraint being that the identity operators have to be in the exactly same locations in the string, e.g., $X \otimes I \otimes Y$ could become $Z \otimes I \otimes X$ or $X \otimes I \otimes X$ but not $X \otimes X \otimes X$.

The second set of measurement gates, used for what we will call *Clifford measurements*, includes also the two-qubit $CNOT$, or controlled-NOT gate in addition to the previous ones. This gate also transforms n-qubit Pauli operators into other n-qubit Pauli operators, so, like before, Pauli operators still cannot be transformed to anything more general. However, with this set we can transform any n-qubit Pauli operator to any other n-qubit Pauli operator (not including $I^{\otimes n}$). Furthermore, as we will see later, if we are not considering just single Pauli operators, this gate set also offers more choices in how a set of Pauli operators transforms compared to each other with the same circuit. This fact will be important when constructing the measurement circuits.

The final set of measurement gates adds also the T gate and is used for what we will call *arbitrary measurements*[1]. This is a so-called *universal* set of quantum gates, meaning

[1]Even though we call these arbitrary measurements, it should be noted that we are still dealing with only projective measurements instead of more general measurements such as POVM measurements. The word arbitrary means here only that we can measure any arbitrary operator in contrast to the other two discussed measurement types. For more information on POVM measurements, see, e.g., [15].

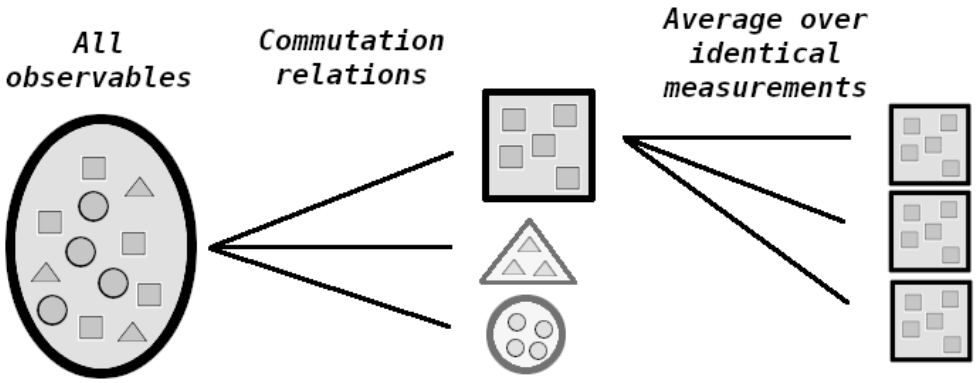


Figure 2.2: Factors affecting the number of needed measurements. Not all observables can be measured at the same time, so we must first identify which observables can by their commutation relations. In addition, we must measure each of these subsets multiple times, since we get the correct value of an observable only in expectation.

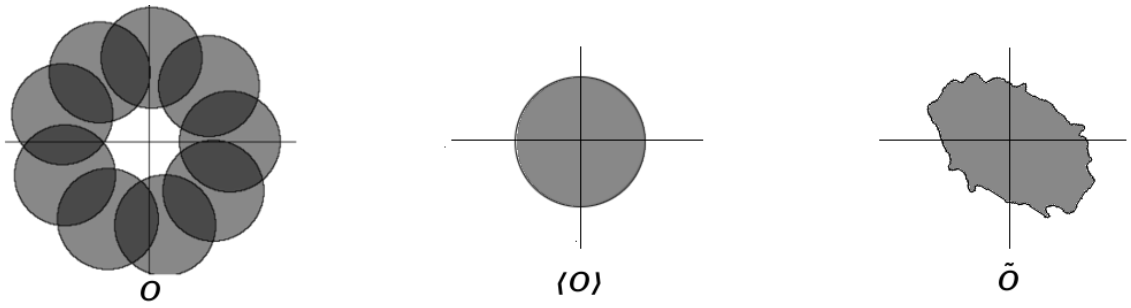
that any unitary transformation can be written as a combination of these gates [16]. That means that for any Hermitian operator we can construct a circuit which diagonalises the operator, and, as we will discuss in the next section, this is exactly the requirement for also measuring an operator.

2.2 Quantum measurements

The density matrix ρ is a linear operator in the space L_V, so the easiest and most common way to do quantum state tomography is to try to find the values $\mathrm{tr}(P\rho)$ as in Equation 2.4. However, we are not always interested in finding the full description of the unknown density matrix, but might instead wish to measure some limited set of observables. The latter situation is more common and hence we will consider full tomography only as a special case of measuring some predefined set of observables.

Our objective in this thesis is to study how we can minimize the quantum computational workload in tomography by minimizing the number of needed quantum measurements. Therefore, we first need to identify the factors which affect that number. Figure 2.2 illustrates how the measurements build up. First of all, we cannot measure all of the observables at the same time but instead must divide the set of observables into subsets which can be measured at the same time. This will be discussed in more detail in Sections 2.2.6 and 2.3. Secondly, we do not get the correct value of even a single observable from a single measurement but instead need to measure it multiple times and then average over the measurement results. This will be discussed further in Sections 2.2.1 and 2.2.7.

Besides the aspects mentioned above, the number of measurements are naturally also affected by all factors which contribute to the error of the measurements, such as inaccurate quantum gates in a measurement circuit, but we will not be considering those factors here



(a) Observable *(b) Expectation value* *(c) Measurement estimate*

Figure 2.3: The quantum state that we are measuring is only rarely in an eigenstate of a given observable. Instead, it is most likely in a superposition of several eigenstates, and the value that we are interested in is the expectation value. By measuring the observable several times from multiple copies of the system, we get an estimate of the expectation value. Note that these images do not represent any real situation but are just meant to illustrate the concept and introduce the notation used in this thesis.

but instead assume that our measurement instruments are perfectly accurate. The error caused by measurement devices would, in general, only introduce a constant factor to the number of measurements.

We have already alluded to measuring multiple observables *at the same time*. What we mean by these *simultaneous* or *parallel measurements* is that they can be performed with the same copy of the quantum system without the order of the measurements affecting the results. An example of this can be seen in Figure 2.4. The two meter symbols at the end of both wires denote the measurements of both qubits of the system. Since the order at which these two measurements are performed is irrelevant, we can just as well consider them being simultaneous but, strictly speaking, there are still two separate measurements that we are performing. However, in this thesis we will count such a measurement as a single measurement since it can be performed on a single copy of the system and, ultimately, we are interested in minimising the number of copies that we need.

2.2.1 Expectation value of an observable

First, we must determine what it means to measure an observable. As is common, we will restrict to considering only Hermitian operators. While it would be entirely possible to consider also complex-valued observables, their usefulness in real applications is questionable while at the same time making it more difficult to consider inaccuracies in measurements. However, when considering qudits, i.e., quantum states similar to qubits but with a d-dimensional Hilbert space, it is easier to consider complex-valued measurements as we will see in Appendix A.

In addition, we will in general assume that the observables are traceless. This assumption

is reasonable because we can always replace an observable O with a traceless observable $O - \mathrm{tr}(O)I/2^n$, where 2^n is the dimension of the total Hilbert space. The only difference between these two observables is that every eigenvalue is shifted by an amount $-\mathrm{tr}(O)/2^n$, so, in the end, we can shift our measurement results back with the corresponding amount.

However we wish to limit the definition of an observable, the expectation value of an observable O is in any case

$$\langle O \rangle = \mathrm{tr}(O\rho), \tag{2.6}$$

i.e., the Hilbert-Schmidt inner product of O and ρ as defined in Definition 2.1. Therefore this is also the value we would like to get in expectation from measuring the observable. We will denote the estimate of $\langle O \rangle$, which we get from the measurements, by $\tilde{O}$. These notations are illustrated in Figure 2.3.

2.2.2 Computational basis

In the quantum circuit model, the actual measurement step usually means measuring the state of the system in the so-called computational basis. Whenever we use matrix representations, the matrices will also be with respect to this computational basis unless otherwise specified. In ket-notation, we are measuring for each qubit whether they are in the state $|0\rangle$ or $|1\rangle$.

Measurement in the computational basis corresponds to measuring the Z-operator for each of the qubits, giving the value $\mathrm{tr}(Z^{\otimes n}\rho)$ when multiplied together. In addition, at the same time we get all the values $\mathrm{tr}(P\rho)$, where $P \in \{I,Z\}^{\otimes n}$, by just ignoring the measurement results on those qubits which would correspond to the identity operator. E.g., to get the measurement value for $Z \otimes Z \otimes I \otimes Z$, we just ignore the third qubit and multiply the results from the other three.

In the matrix representation these tensor products of Z and I matrices and their linear combinations correspond exactly to the diagonal elements of the matrix. This means that by measuring all of the qubits in a circuit (and replicating the measurement on a number of copies of the system), we get estimates of all the diagonal elements of the density matrix as it was at the measurement point but only the diagonal elements. In order to measure other elements of the matrix, first we need to diagonalise the desired values.

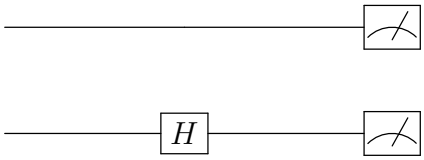


Figure 2.4: Example of a very simple circuit measuring the observables Z_1, X_2 and/or Z_1X_2 of a two qubit system. Note that in order to do these parallel measurements, we need to measure both qubits individually denoted by the meter symbol at the end of both wires. However, since the order at which we perform these measurements does not matter, we can consider them happening simultaneously.

2.2.3 Diagonalising an operator

Since we are considering only Hermitian operators, we can diagonalise our desired observable O with some unitary operation U, i.e.,

$$UOU^\dagger = \begin{bmatrix} \lambda_1 & 0 & 0 & \\ 0 & \lambda_2 & 0 & \cdots \\ 0 & 0 & \lambda_3 & \\ & \vdots & & \ddots \end{bmatrix}, \tag{2.7}$$

where λ_i are the eigenvalues of O. Now, since

$$\langle O \rangle = \mathrm{tr}(O\rho) = \mathrm{tr}(UOU^\dagger U\rho U^\dagger), \tag{2.8}$$

we can instead first perform the unitary operation U to the unknown state ρ, making our desired observable diagonal. In the quantum circuit model, this reduces the problem into measuring in the computational basis. Finally, we just need to assign the corresponding eigenvalue λ_i to the measurement result and this way we gain the value $\langle O \rangle$ in expectation just like we wanted.

2.2.4 Measuring Pauli operators

In order to measure other Pauli operators than just tensor products of the Z matrices, we can first operate on the state with some unitary operators, i.e., the quantum gates discussed in Section 2.1.2. With the conjugation relations, we can transform the Z measurement into X or Y measurements using just the gates in the Pauli measurement set in Table 2.1.

E.g., since $HZH = X$, by using the H gate on the second qubit before the measurement of a two-qubit system, we will be measuring the operators Z_1X_2, Z_1 and/or X_2 of the original state. This example is illustrated in Figure 2.4.

2.2.5 Arbitrary measurements

Like we have discussed in the previous sections, in order to measure an arbitrary observable, it is enough that we are able to diagonalise it. If U is an unitary operation which

diagonalizes our observable O, then we wish to operate on the unknown state ρ with some combination of gates which approximates U. The full set of gates listed in Table 2.1 forms a universal set, meaning that any unitary operation can be approximated with those gates with arbitrary accuracy [16].

There are multiple caveats with this compared to the simpler Pauli measurements. First of all, diagonalising the observable O might be very difficult. Considering that the size of the matrix increases exponentially with the number of qubits, it is often not realistic to attempt diagonalising an arbitrary observable. However, finding the correct U to diagonalise O can be done entirely classically without requiring any quantum computation so it is often a trade-off between decreasing the need for quantum operations at an exponential cost of classical operations.

Another caveat is the length of the circuit needed to approximate U. If we are truly working with only the limited set of gates from Table 2.1, in the worst case, approximating an arbitrary U would take an exponential number of gates with respect to the number of qubits [17].

One final thing to consider is the use of the $CNOT$ gate. While the Pauli measurements need only single-qubit gates, both the Clifford measurements as well as the arbitrary measurements need also multi-qubit operations using the $CNOT$. Single-qubit gates are fairly simple to implement fault-tolerantly, but as the $CNOT$ gate requires interactions between two qubits, it is much harder to do in practice. For this reason it is common to consider only Pauli measurements.

2.2.6 Commuting observables

As we already noted in earlier sections, it is possible to measure multiple observables with the same measurement (or, to be more exact, with the same copy of the state). E.g., measuring each qubit in the computational basis gave us measurements of each $P \in \{I,Z\}^{\otimes n}$. The reason we are able to measure all of these observables at the same time is because they all commute with each other.

Two observables A and B are simultaneously diagonalisable, i.e., can be measured at the same time, if and only if they commute with each other, i.e. $[A,B] = AB - BA = 0$. However, herein lies the biggest difference between the Pauli measurements and Clifford measurements. As we have discussed, both of these measurement types are able to measure only tensor products of Pauli matrices. However, with Pauli measurements, we are able to do unitary transformations only on single qubits. This also means that we cannot measure two observables at the same time unless they commute at every single-qubit level. On the other hand, with Clifford measurements, we are able to measure any two commuting Pauli observables.

This is best illustrated with an example. Let $A = X_1X_2$ and $B = Z_1Z_2$. These two observables commute and can be reduced to computational basis measurements with the unitary transformation $U = H_1C_{12}$, where H_1 is the Hadamard gate acting on the first qubit and C_{12} is the $CNOT$ gate with the first qubit as the control qubit and second qubit as the target qubit. Namely,

$$UAU^\dagger = Z_1, \qquad UBU^\dagger = Z_2. \tag{2.9}$$

It should be clear, however, that with only single-qubit operations, we are not able to reduce both of the observables into tensor products of Z matrices at the same time, since, e.g., the measurement on the first qubit will never match between the two. Therefore, Pauli measurements cannot measure these two observables at the same time.

Thus, while the observables that can be measured with Pauli measurements and Clifford measurements are exactly the same, i.e., any n-qubit Pauli operators, when considering parallel measurements, Clifford measurements can measure any commuting Pauli operators at the same time while Pauli measurements can do it only if the observables commute at every single-qubit level.

2.2.7 Hoeffding's inequality

So far we have only explained how we get the correct value of $\langle O\rangle = \mathrm{tr}(O\rho)$ for an observable O in expectation. However, from any single measurement we get only a single eigenvalue of the observable, with no chance to predict the actual expectation value. However, by performing multiple measurements, we can estimate the real value by averaging over our measurement results.

Let us denote the measurement results of observable O from m different copies of the system by $o_1, \ldots, o_m$. We can treat the o_i as independent, identically distributed (i.i.d.) random variables. We wish to somehow bound the probability that our measured average of $\tilde{O} = \frac{\sum_i o_i}{m}$ differs from the actual expectation value $\langle O\rangle$ by more than some error tolerance ε:

$$P(|\tilde{O} - \langle O\rangle| \geq \varepsilon) \leq \delta. \tag{2.10}$$

To find a suitable bound, we can use Hoeffding's inequality [18].

Lemma 2.1 (Hoeffding's inequality)**.** *Let X_1,X_2,…,X_N be i.i.d. random variables bounded by the interval $[\lambda_{min},\lambda_{max}]$ and let $\bar{X}$ be their empirical mean. Then*

$$P\left(\left|\bar{X} - E\left[\bar{X}\right]\right| \geq \varepsilon\right) \leq 2\exp\left(-\frac{2N\varepsilon^2}{(\lambda_{max} - \lambda_{min})^2}\right). \tag{2.11}$$

Thus, by setting the bound on error probability and solving for N, we get the following corollary for our current situation

Corollary 2.1.1. *Let O be a quantum observable with largest eigenvalue λ_{max} and smallest eigenvalue λ_{min}. In order to measure $\langle O \rangle$ with accuracy ε with probability at least $1-\delta$, it is enough to measure the observable*

$$N = \frac{\log\left(\frac{2}{\delta}\right)}{2\varepsilon^2}(\lambda_{max} - \lambda_{min})^2 \tag{2.12}$$

times.

For Pauli strings, the eigenvalues are just ± 1, so the number of needed measurements can be bounded by

$$N = \frac{2\log\left(\frac{2}{\delta}\right)}{\varepsilon^2}. \tag{2.13}$$

For arbitrary operators, we must consider the difference between the largest and smallest eigenvalue as in Corollary 2.1.1.

Since we are dealing with traceless Hermitian operators, the value $\lambda_{\max} - \lambda_{\min}$ actually defines a norm.

Proposition 2.2. *Let O be a traceless Hermitian operator with largest eigenvalue λ_{max} and smallest eigenvalue λ_{min}. The value $\lambda_{max} - \lambda_{min}$ defines a norm in the space of traceless Hermitian operators.*

Proof. A norm must satisfy homogeneity, triangle inequality and the implication $||x|| = 0 \Rightarrow x = 0$.

As the eigenvalues scale with the scaling of the operator, $\lambda_{\max} - \lambda_{\min}$ is clearly a homogeneous function of the operator. Also

$$\lambda_{\max} - \lambda_{\min} = 0 \tag{2.14}$$

$$\Rightarrow \lambda_i = \lambda \quad \forall i \tag{2.15}$$

and since we are dealing with traceless operators, this would imply that all the eigenvalues are 0 and hence $O = 0$.

The only thing left to prove is that the triangle inequality applies. We will start with a useful lemma.

Lemma 2.3. *Let A and B be two Hermitian operators with largest eigenvalues a_{max} and b_{max} respectively. Let $C = A + B$ and c_{max} be the largest eigenvalue of C. Then*

$$c_{max} \leq a_{max} + b_{max}. \tag{2.16}$$

Proof.

$$\begin{aligned} c_{\max} &= \max_{|\psi\rangle}(\langle\psi|C|\psi\rangle) = \max_{|\psi\rangle}(\langle\psi|A|\psi\rangle + \langle\psi|B|\psi\rangle) \\ &\leq \max_{|\psi\rangle}(\langle\psi|A|\psi\rangle) + \max_{|\varphi\rangle}(\langle\varphi|B|\varphi\rangle) = a_{\max} + b_{\max} \end{aligned} \tag{2.17}$$

□

Proof of Proposition 2.2 continued. Now, let A and B be two traceless Hermitian operators with largest and smallest eigenvalues $a_{\max}$, $a_{\min}$, $b_{\max}$ and $b_{\min}$. Let $c_{\max}$ and $c_{\min}$ be the largest and smallest eigenvalues of $C = A + B$. Then, Lemma 2.3 tells us that

$$c_{\max} \leq a_{\max} + b_{\max} \tag{2.18}$$

but also, using the same lemma for $-C = -A + (-B)$, we get

$$-c_{\min} \leq -a_{\min} - b_{\min}. \tag{2.19}$$

Therefore

$$c_{\max} - c_{\min} \leq (a_{\max} - a_{\min}) + (b_{\max} - b_{\min}) \tag{2.20}$$

which proves the triangle inequality. □

The fact that we can define a norm this way is useful for a couple of reasons. First of all, just for a matter of convenience, since we will have to use the eigenvalue difference for the operators quite frequently, this will allow us to define an easier shorthand notation for it. Secondly, in Section 2.3.3 we will consider a geometrical interpretation of the space of observables and this new norm provides a useful measure of distance in that vector space.

Definition 2.2 (λ-norm)**.** We define the *λ-norm* of a traceless Hermitian operator O as the difference between its largest and smallest eigenvalue $\lambda_{\max} - \lambda_{\min}$ and we will denote it with $||O||_\lambda$.

With this definition we can rewrite Corollary 2.1.1.

Theorem 2.4 (Rewrite of Corollary 2.1.1)**.** *Let O be a traceless Hermitian operator. In order to measure $\langle O \rangle$ with accuracy ε with probability at least $1 - \delta$, it is enough to measure the observable*

$$N = \frac{\log\left(\frac{2}{\delta}\right)}{2\varepsilon^2} ||O||_\lambda^2 \tag{2.21}$$

times.

The Hilbert-Schmidt inner product from Definition 2.1 also defines a norm, which is useful as well.

Definition 2.3 (Hilbert-Schmidt norm)**.** The *Hilbert-Schmidt norm* of an operator O is defined as

$$||O|| = \sqrt{\mathrm{tr}(O^\dagger O)}. \tag{2.22}$$

We will denote it with $||O||_{HS}$ in order to distinguish it from the λ-norm.

Proposition 2.5. *The λ-norm and the H-S norm satisfy the following inequalities for any traceless Hermitian operator O*

$$\frac{||O||_\lambda}{\sqrt{2}} \leq ||O||_{HS} \leq 2^{n/2-1} ||O||_\lambda, \tag{2.23}$$

where 2^n is the dimension of the Hilbert space.

Proof. Let $\lambda_{\max}$ and $\lambda_{\min}$ be the largest and smallest eigenvalue of O and let λ_i be the ith eigenvalue. Then

$$\begin{aligned} ||O||_\lambda = \lambda_{\max} - \lambda_{\min} &= \sqrt{\lambda_{\max}^2 + \lambda_{\min}^2 - 2\lambda_{\max}\lambda_{\min}} \leq \sqrt{2(\lambda_{\max}^2 + \lambda_{\min}^2)} \\ &\leq \sqrt{2\sum_i \lambda_i^2} = \sqrt{2}||O||_{HS}, \end{aligned} \tag{2.24}$$

which proves the first inequality.

To prove the second, we will start by noting that

$$0 \leq (\lambda_{\max} - \lambda_i)(\lambda_i - \lambda_{\min}) = -\lambda_i^2 + \lambda_i(\lambda_{\max} + \lambda_{\min}) - \lambda_{\max}\lambda_{\min}. \tag{2.25}$$

Therefore,

$$\begin{aligned} ||O||_{HS} = \sqrt{\sum_i \lambda_i^2} &\leq \sqrt{\sum_i [\lambda_i(\lambda_{\max} + \lambda_{\min}) - \lambda_{\max}\lambda_{\min}]} \\ &= \sqrt{2^n|\lambda_{\max}\lambda_{\min}| + (\lambda_{\max} + \lambda_{\min})\sum_i \lambda_i} = 2^{n/2}\sqrt{|\lambda_{\max}\lambda_{\min}|} \\ &\leq 2^{n/2}\sqrt{\frac{(|\lambda_{\max}| + |\lambda_{\min}|)^2}{4}} = 2^{n/2-1}||O||_\lambda. \end{aligned} \tag{2.26}$$

□

Both of these bounds are also tight. The first equality is attained by any operator with only two non-zero eigenvalues, $\pm\lambda$ (with multiplicities 1) and the rest of the eigenvalues being zero. The second equality is attained by any operator with exactly half of the eigenvalues being λ and the second half $-\lambda$ such as any Pauli string.

The first inequality of Proposition 2.5 shows that we can replace the λ-norm in Theorem 2.4 by the more well-known Hilbert-Schmidt norm. However, the second inequality shows that this could scale extremely badly with increasing system size and, e.g., the most commonly used Pauli strings would need an exponential number of measurements compared to the constant number with the λ-norm.

2.3 Measuring a set of observables

Since we are most likely interested in measuring multiple observables, we also need to take into account how the error accumulates. We will assume that the accuracy ε is constant for all the observables, since we could just rescale the observables to make that the case. In other words, if one of the observables, O, has error tolerance ε', then we could instead

aim to measure $\frac{\varepsilon}{\varepsilon'}\langle O\rangle$ with accuracy ε.

We wish to measure all of the observables within accuracy ε with some probability $1-\delta$. The observables all have their respective error probabilities δ_i so we need to bound the total error δ with respect to these individual errors.

Lemma 2.6. *Let $O_1, O_2, \ldots, O_m$ be a set of observables which will be measured accurately with probabilities $1-\delta_i$ for every O_i. Then we can bound the probability of error δ, that at least one of them is inaccurate by*

$$\delta \leq \sum_i \delta_i \tag{2.27}$$

Proof. Let A and B be two events with probabilities $P(A)$ and $P(B)$. Then the probability of at least one of the events occurring is

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) \leq P(A) + P(B) \tag{2.28}$$

where the probability of both events occurring $P(A \cup B)$ is non-negative and hence the lemma follows. □

2.3.1 Commuting observables

Lemma 2.6 tells us that we don't need to worry about any correlation of the observable errors. If we just set, e.g., $\delta_i = \frac{\delta}{m}$ for all observables $O_1, \ldots, O_m$, then the total probability of error will be less than our threshold. However, setting the individual error probabilities equal for all the observables is not necessarily optimal when using the bound from Lemma 2.6. Instead, in the case of commuting observables, since we can measure all of the observables at the same time, we would rather want to make the number of measurements for each observable equal.

Proposition 2.7. *Let $O_1, \ldots, O_m$ be a set of commuting observables which we wish to measure all within accuracy ε and the probability of one or more of them being inaccurate is δ. A sufficient number of measurements can be found as the solution N for the equation*

$$\sum_{i=1}^{m} \left[\exp\left(-\frac{2\varepsilon^2}{||O_i||_\lambda^2} \right) \right]^N = \frac{\delta}{2}. \tag{2.29}$$

Proof. This follows straight from Lemma 2.1,

$$\delta_i \leq 2 \exp\left(-\frac{2\varepsilon^2 N}{||O_i||_\lambda^2} \right) \tag{2.30}$$

and Lemma 2.6

$$\delta \leq \sum_i \delta_i. \tag{2.31}$$

□

If we are only interested in measuring Pauli strings, i.e., tensor products of Pauli matrices, each with the same accuracy, then the number of measurements becomes

$$N = \frac{2\log\left(\frac{2m}{\delta}\right)}{\varepsilon^2}. \tag{2.32}$$

Even for other kinds of observables, we usually get a good approximation by just setting $\delta_i = \delta/m$, giving

$$N = \frac{\log\left(\frac{2m}{\delta}\right)}{2\varepsilon^2} \max_i(||O_i||_\lambda^2). \tag{2.33}$$

We inferred earlier that we can ignore any correlation between the observable measurements by bounding the total error with the sum of the individual errors. However, taking also the correlations into account can be highly beneficial. Let us consider a couple of examples.

Example 2.1. Let $O_1 = X_1$, $O_2 = X_2$ and $O_3 = (X_1 + X_2)/2$ be three observables which we wish to measure. All of these have the same λ-norm, $||O_i||_\lambda = 2$, so by our earlier considerations, we would assign all of them an equal chance of error, $\delta/3$. Then, the number of measurements would be

$$N = \frac{2\log\left(\frac{6}{\delta}\right)}{\varepsilon^2}. \tag{2.34}$$

However, we can easily see that $\tilde{O}_3$ can be inaccurate only if at least one of $\tilde{O}_1$ or $\tilde{O}_2$ is inaccurate as well. Therefore, it would be enough to make sure that we measure $\langle O_1\rangle$ and $\langle O_2\rangle$ accurately, so

$$N = \frac{2\log\left(\frac{4}{\delta}\right)}{\varepsilon^2}. \tag{2.35}$$

is sufficient.

Example 2.2. Let $O_1 = X_1$ and $O_2 = \frac{\varepsilon-2\alpha}{\varepsilon}X_1 + \alpha X_2$, where $\alpha \ll \varepsilon$. Unlike in the previous example, these two observables are linearly independent. However, again $\tilde{O}_2$ can be inaccurate only if $\tilde{O}_1$ is inaccurate as well, so it is enough to make sure that we measure $\langle O_1\rangle$ accurately.

While it would be possible to take the improvements from these examples into account, their effect would in most cases be negligible. Since the number of measurements in Equation 2.21 is logarithmic in $1/\delta$, we would need to be able to ignore a large majority of the observables in order to see any significant improvement in the number of measurements needed. In the first example, if we assume $\delta \leq 0.1$, losing one of the observables decreased the number of measurements by less than 10%. For that reason, unless we are dealing with a very specific situation, it makes more sense to consider each observable independently, ignoring the correlations. Then, we will have approximately logarithmic dependence in the number of measurements with respect to the number of observables like in Equation 2.33.

2.3.2 Non-commuting observables

It is highly unlikely that we would be dealing with a case where all of the observables we wish to measure would commute with each other. Instead, we will need to divide our set of observables into subsets of commuting observables, which we will then measure each separately.

It would seem that we want to include each observable in one of the commuting subsets and minimize the number of these subsets in order to minimize the number of needed measurements. Then, we could rewrite Proposition 2.7 as

Proposition 2.8. *Let* $\{O_{11}, O_{12}, \ldots, O_{1m_1}\}, \{O_{21}, O_{22}, \ldots, O_{2m_2}\}, \ldots, \{O_{l1}, O_{l2}, \ldots, O_{lm_l}\}$ *be disjoint sets of commuting observables which we wish to measure all within accuracy* ε *and the probability of one or more of them being inaccurate is* δ*. A sufficient number of measurements can be found as the minimum of the sum*

$$N = \sum_{i=1}^{l} N_i \tag{2.36}$$

while the summands N_i *have to satisfy the condition*

$$\sum_{i=1}^{l} \sum_{j=1}^{m_j} \left[\exp\left(-\frac{2\varepsilon^2}{||O_{ij}||_\lambda^2} \right) \right]^{N_j} = \frac{\delta}{2}. \tag{2.37}$$

This could offer some improvement when considering observables with different λ-norms compared to the easier approach of assigning each observable the same error chance:

$$N_j = \frac{\log\left(\frac{2\sum m_i}{\delta}\right)}{2\varepsilon^2} \max_i(||O_{ij}||_\lambda^2). \tag{2.38}$$

However, if we assume different λ-norms, we are more likely to get a significant improvement if we also allow the same observable to belong into different subsets of commuting observables. This way, we can try to divide the workload of the observables with the largest λ-norms into as many subsets as possible.

Example 2.3. Let $O_1 = X_1$, $O_2 = Z_1$ and $O_3 = \sqrt{2}X_2$ be three observables which we wish to measure. We need to measure O_3 approximately twice as many times as the other two but, also, we can measure it at the same time as either of the other two. Therefore, we should choose to have two different measurement sets: one where we measure $\{O_1, O_3\}$ and another where we measure $\{O_2, O_3\}$.

Furthermore, as we will see in later sections, even if the observables each share the same λ-norm, for optimal results we might have to allow allocating the same observables into multiple subsets. We will demonstrate in Section 4.1 how this division into different subsets

can be done efficiently. While it would, in theory, be possible to combine this idea with trying to distribute the error probabilities unequally like in Propositions 2.7 and 2.8, in practice, finding the optimal distribution for both the observable subsets and error distribution could be incredibly difficult. Therefore, we will focus on the more significant of these two: trying to find the optimal division into subsets.

As long as we are only interested in measuring Pauli strings, there is not much freedom of choice in the way the measurements are carried out besides choosing which observables are measured at the same time. However, if we want to measure more general observables and possibly even with arbitrary measurements, then there are other significant considerations to take into account. The examples from Section 2.3.1, which didn't have much effect for commuting observables, have potentially more significant counterparts when applied to non-commuting observables.

Example 2.4. Let $O_1 = X_1$, $O_2 = Z_1$ and $O_3 = (X_1 + Z_1)/2$ be three observables which we wish to measure. None of these commute, so if we assign each of them an equal chance of error, $\delta/3$, and measure them separately, the total number of measurements would be

$$N = 3 \times \frac{2\log\left(\frac{6}{\delta}\right)}{\varepsilon^2}. \tag{2.39}$$

However, $\tilde{O}_3$ can be inaccurate only if at least one of $\tilde{O}_1$ or $\tilde{O}_2$ is inaccurate as well so it is enough to measure $\langle O_1 \rangle$ and $\langle O_2 \rangle$ accurately, so

$$N = 2 \times \frac{2\log\left(\frac{4}{\delta}\right)}{\varepsilon^2}. \tag{2.40}$$

is sufficient.

Example 2.5. Let $O_1 = X_1$ and $O_2 = \frac{\varepsilon - 2\alpha}{\varepsilon} X_1 + \alpha Z_1$, where $\alpha \ll \varepsilon$. $\tilde{O}_2$ can be inaccurate only if $\tilde{O}_1$ is inaccurate as well, so it is enough to make sure that we measure $\langle O_1 \rangle$ accurately.

While the improvements in the corresponding commuting cases were negligible, here we have already decreased the number of measurements by more than 33% in the first example and by more than half in the second example. Obviously, in a real scenario, it would be unlikely that we would be able to reduce such significant portion of the observables into other observables. However, for optimal performance, we should try to identify such "irrelevant" observables before devising the measurement scheme.

2.3.3 Geometrical interpretation

We can represent the previous examples also geometrically in the vector space formed by traceless Hermitian operators and equipped with the λ-norm using the Pauli operators as a

(a) (b)

Figure 2.5: ***(a)*** *The darker shaded circle illustrates observables within radius 2ε (λ-norm distance) from the origin. We can always estimate these observables as the average of their largest and smallest eigenvalue without ever having to measure them. Additionally, when we measure, say, the Pauli operator X, we get for free any observables that are within the cone from that point to the circle around the origin as illustrated by the lighter shaded cone.* ***(b)*** *When we measure two observables, e.g., X and Z, we also get free estimates for any observable within the rhombus with corners at $\pm X$ and $\pm Z$.*

basis. Figure 2.5a illustrates Example 2.5. Within radius 2ε from the origin, the expectation value of any observable O must be within

$$\frac{\lambda_{\max} + \lambda_{\min}}{2} - \varepsilon \leq \lambda_{\min} \leq \langle O \rangle \leq \lambda_{\max} \leq \frac{\lambda_{\max} + \lambda_{\min}}{2} + \varepsilon. \tag{2.41}$$

Therefore, we get these "for free" by just estimating $\tilde{O} = \frac{\lambda_{\max}+\lambda_{\min}}{2}$ since it will always be within our error tolerance. Moreover, by measuring X within the desired accuracy, we now get also everything within the shaded cone for free as well. The circle of radius 2ε is defined by the λ-norm. This means that in the case of two anti-commuting Pauli observables such as in Figure 2.5a the metric resembles the familiar 2-norm. In the case of two commuting Pauli observables, however, it would resemble the 1-norm. In the general case, the "hyperspheres" of constant radius around the origin would have more complex shapes.

Figure 2.5b illustrates Example 2.4. Here, we measure the observables X and Z and consequently get also estimates for any observables within the rhombus with corners at $\pm X$ and $\pm Z$. This can be generalised for higher dimensions and more observables by determining the convex hull of the observables.

Definition 2.4 (Convex hull)**.** The convex hull of a set of points is the intersection of every convex set containing those points.

Lemma 2.9. *Let $\{O_1, O_2, \ldots, O_m\}$ be a set of observables and let C be the convex hull of $\{\pm O_1, \pm O_2, \ldots, \pm O_m\}$. Then, $A \in C$ iff there exist real coefficients $a_1, \ldots, a_m \in \mathbb{R}$*

such that

$$A = \sum_{i=1}^{m} a_i O_i \tag{2.42}$$

and

$$\sum_{i=1}^{m} |a_i| \leq 1. \tag{2.43}$$

Proof. Since $\pm O_i \in C$, also

$$a_i O_i = \frac{1+a_i}{2} O_i + \frac{1-a_i}{2}(-O_i) \in C \quad \forall a_i \in \mathbb{R}, |a_i| \leq 1. \tag{2.44}$$

By induction it is also easy to see that

$$\sum_{i=1}^{m} |a_i| \leq 1 \quad \Rightarrow \quad \sum_{i=1}^{m} a_i O_i \in C. \tag{2.45}$$

Since the set of observables defined this way is clearly convex, no other observables can belong to the convex hull. □

Theorem 2.10. *Let $\{O_1,O_2,\ldots,O_m\}$ be a set of observables which we have measured within accuracy ε and let C be the convex hull of $\{\pm O_1, \pm O_2, \ldots, \pm O_m\}$. Any observable within C can be estimated within accuracy ε just from our estimates of $\{O_1,O_2,\ldots,O_m\}$.*

Proof. Let $A = \sum_{i=1}^{m} a_i O_i \in C$, $\sum_{i=1}^{m} |a_i| \leq 1$ and let us define

$$\tilde{A} := \sum_{i=1}^{m} a_i \tilde{O}_i. \tag{2.46}$$

Then

$$P\left([|\tilde{A} - \langle A \rangle| \geq \varepsilon] \wedge \bigwedge_{i=1}^{m} [\tilde{O}_i - \langle O_i \rangle \leq \varepsilon] \right) = 0 \tag{2.47}$$

so $\tilde{A}$ is guaranteed to be a good estimate of A as long as $\tilde{O}_i$ are good estimates. □

Corollary 2.10.1. *Let $S = \{O_1,O_2,\ldots,O_m\}$ be a set of observables which we wish to measure within accuracy ε and let C be the convex hull of $\{\pm O_1, \pm O_2, \ldots, \pm O_m\}$. Let $\tilde{S} = \{O_{a_1},O_{a_2},\ldots,O_{a_l}\} \subseteq S$ be the smallest subset of S which results in the same convex hull. Measuring $\tilde{S}$ gives us all the necessary information.*

The convex hull of a set of observables is illustrated for the two-dimensional case in Figure 2.6. We can now ignore any observable which is not in the corners of the convex hull and this could ideally decrease the number of observables significantly. However, this is far from optimal. Instead of having to measure every corner of this convex hull, it would

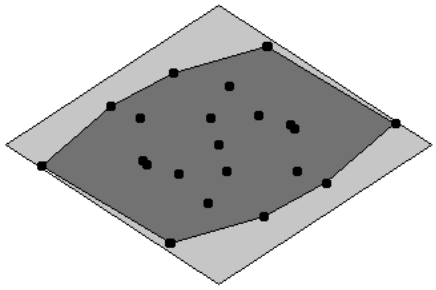


Figure 2.6: The convex hull of a set of observables is illustrated with the darker shaded region. However, it might be easier to measure the larger, lighter shaded region which has less corners, i.e., less observables to measure.

more likely be preferable to decrease the number of these corners by increasing the size of the convex set. While it would require us to measure the remaining corners with higher accuracy, the decrease in the number of observables would more than likely make up for it. This is illustrated as the larger shaded region in Figure 2.6.

However, with higher dimensions this becomes extremely complicated. In addition to having to find a convex set to cover all of the observables, we need to take into account how to best get the actual to-be-measured observables to commute with each other. For this reason, even if we are able to do arbitrary measurements, it might be easier to reduce the corners of the convex set to unitary transformations of Pauli observables since finding a set of observables with as favourable commutation relations would be difficult. Thus, the problem is just to find a suitable unitary transformation and the optimal coefficients for our transformed Pauli observables so that their convex hull covers all of the observables. Also, by choosing not to transform the Pauli basis (i.e., picking I as the unitary transformation), it is enough to use either Pauli or Clifford measurements to complete the measurements.

Problem statement 1. Let $\{O_1,O_2,\ldots,O_m\}$ be a set of n-qubit observables. We wish to find

- Unitary transformation on n qubits, U (for Pauli or Clifford measurements $U = I$)
- Real, positive coefficients $T_{\alpha_1\alpha_2\ldots\alpha_n} \in \mathbb{R}_+$, $\alpha_i \in \{0,1,2,3\}$

  which satisfy the constraints

$$\frac{1}{2^n}\sum_{\alpha_1\ldots\alpha_n} |\mathrm{tr}(UT_{\alpha_1\ldots\alpha_n}\sigma_{\alpha_1}\otimes\cdots\otimes\sigma_{\alpha_n}U^\dagger O_i)| \leq 1 \quad \forall i. \tag{2.48}$$

Theorem 2.11. *Let U and $T_{\alpha_1\ldots\alpha_n}$ be the solutions found for Problem statement 1. Then, we get estimates of the observables $\{O_1,\ldots,O_m\}$ by measuring the observables*

$$\frac{1}{T_{\alpha_1\ldots\alpha_n}} U\sigma_{\alpha_1}\otimes\cdots\otimes\sigma_{\alpha_n}U^\dagger \tag{2.49}$$

with the desired accuracy.

Proof. We can write O_i as

$$\begin{aligned} O_i &= \frac{1}{2^n} \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} \mathrm{tr}(UPU^\dagger O_i) UPU^\dagger \\ &= \frac{1}{2^n} \sum_{\alpha_1 \ldots \alpha_n} \mathrm{tr}(UT_{\alpha_1 \ldots \alpha_n} \sigma_{\alpha_1} \otimes \cdots \otimes \sigma_{\alpha_n} U^\dagger O_i) \times \frac{1}{T_{\alpha_1 \ldots \alpha_n}} U\sigma_{\alpha_1} \otimes \cdots \otimes \sigma_{\alpha_n} U^\dagger. \end{aligned} \tag{2.50}$$

If we make an estimation of $\langle O_i \rangle$ by taking a linear combination of our measurement results, the corresponding error $\tilde{\varepsilon}$ can be bounded by

$$\tilde{\varepsilon} \leq \frac{1}{2^n} \sum_{\alpha_1 \ldots \alpha_n} |\mathrm{tr}(UT_{\alpha_1 \ldots \alpha_n} \sigma_{\alpha_1} \otimes \cdots \otimes \sigma_{\alpha_n} U^\dagger O_i)| \varepsilon \leq \varepsilon. \tag{2.51}$$

□

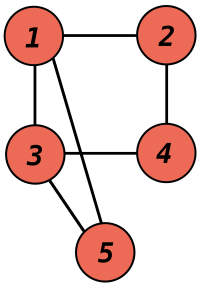
There are obviously an infinite number of solutions to the above problem since we can increase the size of the convex set (i.e., decrease the coefficients $T_{\alpha_1 \ldots \alpha_n}$) as much as we want. However, we clearly want to find a solution which requires the least number of measurements. If the set of desired observables consisted of only (scaled) Pauli strings to begin with, then the job is easy, as they already define the optimal set to measure by themselves. However, for more general cases, we should compute the number of measurements for each possible solution satisfying the constraints and then pick the optimal among those. This optimisation is not applied for the methods developed in the rest of this thesis but it remains as a further possible path of generalisation.

3 Graph theory

Graph theory is a very active field of mathematics which has applications very widely to different study areas. The general applicability of it can be easily understood as many more complex situations can be reduced to a graph representation which is easy to understand but at the same time also offers the tools to study difficult systems. The basic idea behind graphs is simple: we have elements, called *vertices*, and interactions between the elements, called *edges*.

Definition 3.1 (Graph)**.** A graph G is a pair $G = (V,E)$, where V is a set of *vertices* $V = \{v_1, \ldots, v_m\}$, and E is a set of vertex pairs called *edges*. The number of vertices $|V|$ is called the *order* of the graph.

In this thesis, we will be dealing only with simple, undirected graphs meaning that there can be at most one edge between each vertex pair and that the vertex pairs forming an edge are unordered. We will also not be considering edges from a vertex to itself, that is,



(a)

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

(b)

Figure 3.1: ***(a)*** *An example of a graph. The numbers denote some arbitrary indexing of the vertices.* ***(b)*** *An adjacency matrix corresponding to the graph in (a).*

$\{v_i,v_i\} \notin E$ for all i.

We say that two vertices, v_i and v_j, are *adjacent* or *neighbours* if E contains an edge $\{v_i,v_j\}$.

The same graph can be represented in multiple different ways. Most commonly graphs are depicted as a picture where each vertex is represented by a dot and each vertex is represented as a line between the dots. An example of this can be seen in Figure 3.1a. Another way to represent a graph is by its *adjacency matrix* A which has the same number of rows and columns as there are vertices in the graph. In this form, the elements of the matrix A_{ij} are 1 if there exists a corresponding edge in the graph between vertices v_i and v_j and 0 if there is no edge. I.e.,

$$A_{ij} = \begin{cases} 1, & \text{if } \{v_i,v_j\} \in E \\ 0, & \text{if } \{v_i,v_j\} \notin E \end{cases} \tag{3.1}$$

An example of an adjacency matrix can be seen in Figure 3.1b.

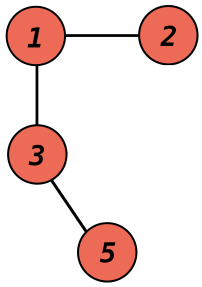
Two other concepts we need are those of a subgraph and a complete graph

Definition 3.2 (Subgraph)**.** $\tilde{G} = (\tilde{V},\tilde{E})$ is called a subgraph of graph $G = (V,E)$ iff

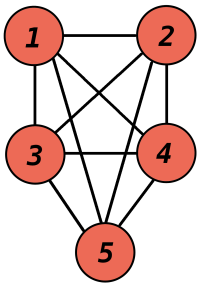
- $\tilde{V} \subseteq V$,
- $\tilde{E} \subseteq E$ and
- $\{v_i,v_j\} \in \tilde{E} \Rightarrow \{v_i,v_j\} \subseteq \tilde{V}$

Definition 3.3 (Complete graph)**.** A complete graph is a graph which contains all possible edges for a given vertex set. The complete graph of order m is denoted by K_m. I.e.,

$$K_m = (V,E) \qquad V = \{v_1,\ldots,v_m\} \qquad \{v_i,v_j\} \in E \ \forall i,j \in \{1,\ldots,m\},\ i \neq j. \tag{3.2}$$



(a)



(b)

Figure 3.2: ***(a)*** *This graph is a subgraph of the graphs in both Figure 3.1a and Figure 3.2b.* ***(b)*** *A complete graph of order five,* K_5.

Figure 3.2a shows an example of a subgraph and Figure 3.2b shows an example of a complete graph.

3.1 Graph colouring

Graph colouring is a problem where we try to assign each vertex of a graph a colour with the only rule being that two neighbouring vertices cannot be coloured the same. More formally

Definition 3.4 (Graph colouring)**.** Graph colouring for a graph $G = (V,E)$ is a mapping $f : V \to C$ from the set of vertices V to a set of colours C such that

$$\{v_i,v_j\} \in E \Rightarrow f(v_i) \neq f(v_j). \tag{3.3}$$

Examples of valid graph colourings are shown in Figure 3.3.

It is easy to find valid graph colourings; after all, we can just colour every vertex with a different colour. However, usually we try to minimize the number of colours that we need for a given graph. The minimum number of colours that is needed to colour a graph is called the graph's *chromatic number* and is denoted by χ.

Definition 3.5 (Chromatic number)**.** The chromatic number χ of a graph $G = (V,E)$, is the smallest number of colours $|C|$ for which it is possible to find a valid graph colouring $f : V \to C$.

It can be easily seen that, e.g., the chromatic number of a complete graph K_m is $\chi = m$ and that the chromatic number of a subgraph is at most the chromatic number of the original graph. However, figuring out generally whether a given graph can be coloured with some given number of colours is an NP-complete problem (if the number of colours is greater

(a) *(b)*

Figure 3.3: Examples of valid graph colourings. The graph in (a) is the same as in Figure 3.1a but the colouring here is not optimal as the graph could also be coloured with only three colours. The colouring for (b) is optimal.

than 2) [19], meaning that there is no known algorithm which could solve it in a polynomial time with respect to the number of vertices. Finding the actual chromatic number or an optimal colouring for a graph are naturally even harder problems and so they belong to the category of NP-hard problems.

3.2 Examples of graph colouring algorithms

Finding the best possible colouring can be extremely hard and, for that reason, it is often more reasonable to look for some "good" colouring with algorithms which can find almost the best colouring for most of the time. We will review here just a few known algorithms. For more information on different algorithms see, e.g., Ref. [20].

3.2.1 Brute-force search and other exact methods

If we don't have to care about the computational cost of finding the graph colouring solution, the simplest and most foolproof method is to brute-force through every possible colouring starting from the smallest number of colours until we find one that works. This way we are guaranteed to find the optimal solution but the simplest way of implementing it will take $\mathcal{O}(\chi^m)$ time to compute where m is the number of vertices since we are going through every possibility of assigning each of the m vertices each of the χ colours.

There are some obvious and some less-obvious ways to improve the algorithm for finding the exact solution. For example, instead of going through every possible combination of

colours, we could backtrack whenever we notice that the current colouring is either not legal or is going to result in a colouring with more colours than the best known result. Plenty of more sophisticated improvements exist as well [20] but, since we are dealing with an NP-hard problem, even the best of those currently known methods take exponential time to compute. For that reason, when dealing with graphs with at least thousands of vertices like we do in this thesis, it is not realistic to try to find the optimal solution.

3.2.2 Greedy algorithm

Perhaps the most intuitive, non-exact method is the greedy algorithm, where we go through the vertices one-by-one in some order and colour the next vertex always with the first colour available so that the neighbouring conditions are satisfied. This will result in the optimal colouring if, by chance, we happen to have the vertices in an optimal order, but can also result in a very bad colouring otherwise.

Greedy algorithm can be implemented in $\mathcal{O}(n+m)$ time, where n is the number of vertices and m is the number of edges [20]. The linear complexity is of course a huge improvement compared to the exponential complexity of the exact methods, but, as a trade-off, we will rarely get even close to the optimal solution. Since the ordering of the vertices is very crucial for this method, one might want to first reorder the vertices in some more preferable order which could increase the computational complexity. However, in our case, we will only use either a random initial order for the vertices or reorder them based on some known solution.

3.2.3 Hybrid evolutionary algorithms

While there are vast numbers of different complex algorithms for solving the graph colouring problem, out of the more sophisticated methods, we will consider here only the hybrid evolutionary algorithm (HEA) proposed by Hao and Galinier [21] as implemented by Lewis [20]. In evolutionary algorithms, the first step is to create a population of candidate solutions by some easy method. Then we pick random pairs of the population to produce *offspring* starting as some kind of combination of the two parent solutions. This offspring then replaces the worse of the two parent solutions regardless whether it is better than the replaced solution or not. After some fixed number of steps, we pick the best solution that the algorithm has produced so far.

3.3 Multicolouring problem

There exists a generalisation of the graph colouring problem for weighted graphs called *multicolouring* problem [20]. Here, each vertex of the graph is assigned an integer value called *weight* and we want to assign each vertex as many distinct colours as indicated by the weight. Each vertex having weight 1 then just defines the regular graph colouring problem.

Definition 3.6 (Multicolouring)**.** Let $G = (V,E)$ be a graph and let $w : V \to \mathbb{Z}_+$ be a weight function from the vertices of the graph to positive integers. Let C be a set of colours. *Multicolouring* of the weighted graph (V,E,w) is a mapping $f : V \to 2^C$ from the vertices to the power set of colours such that

$$\{v_i,v_j\} \in E \Rightarrow f(v_i) \cap f(v_j) = \emptyset \tag{3.4}$$

and

$$|f(v_i)| = w(v_i). \tag{3.5}$$

The multicolouring problem can quite easily be reduced to the basic graph colouring problem.

Theorem 3.1. *Multicolouring can be reduced to regular graph colouring by replacing each vertex v_i with a complete graph of order $w(v_i)$ in such a way that every vertex of the complete graph is adjacent to every neighbour of v_i. The minimum number of colours is the same for both representations of the problem.*

Proof. Let $f : V \to 2^C$ be a solution to the multicolouring problem for the weighted graph (V,E,w) and let us assign labels to the colours so that

$$c_{ij} \in f(v_i),\ c_{ij} \neq c_{ik} \quad \forall j,k \in \{1,\dots,w(v_i)\}, j \neq k\ \forall v_i \in V \tag{3.6}$$

so that c_{ij} corresponds to the jth colour of the ith vertex. This of course means that two different labels could also correspond to the same colour since two non-adjacent vertices can share a colour.

Now let us replace each vertex v_i with a complete graph of order $w(v_i)$ and label the vertices in this graph as $v_{ij} \in \tilde{V}$. Then, let us define a map $\tilde{f} : \tilde{V} \to C$, $\tilde{f}(v_{ij}) = c_{ij}$. Now, if v_{ij} is adjacent to v_{kl}, then either $i = k, j \neq l$ or $\{v_i,v_k\} \in E$. In the first case, as defined above, $c_{ij} \neq c_{kl}$. In the second case,

$$c_{ij} \in f(v_i) \qquad\qquad c_{kl} \in f(v_k) \tag{3.7}$$

and the definition of multicolouring tells us that $f(v_i) \cap f(v_k) = \emptyset$ so $c_{ij} \neq c_{kl}$. Therefore $\tilde{f}$ is a valid colouring of the new graph.

Tracing the same steps backwards, we can also show that if $\tilde{f}$ is a valid colouring of the transformed graph, then f is a valid multicolouring of the weighted graph. Since we can transform the problem both ways, this guarantees that the minimum number of colours stays the same. □

The important part of the above theorem is that the number of colours stays the same whether we are considering the multicolouring problem or transforming it to the regular graph colouring, so the chromatic number is the same in both cases. Since we already know methods to solve the graph colouring problem, those same methods can then be applied to the multicolouring problem as well. The transformation of the graph from a weighted graph to an unweighted graph solvable by graph colouring is illustrated in Figure 3.4.

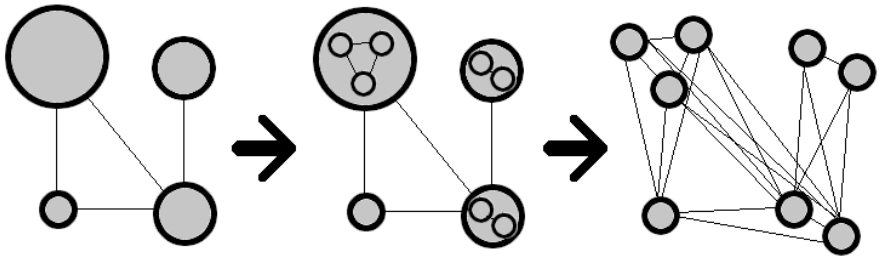


Figure 3.4: Transformation of the multicolouring problem into regular graph colouring. If we have a graph with different weights assigned on the vertices, we can instead replace those weighted vertices with complete subgraphs in order to transform the problem to the more familiar graph colouring.

4 Observable commutation graphs

As we have seen, the observable commutation relations determine how many of the observables can be measured at the same time. One easy way to describe these relations is by a binary matrix M where the element M_{ij} is 1 if the elements do not commute and 0 if they do commute[2]. Consequently, this seems a lot like an adjacency matrix for a graph and, indeed, we can illustrate these commutation relations also as a graph.

What's more, the problem of finding ways to divide the observables into simultaneously measurable subsets is now just the graph colouring problem.

Theorem 4.1. *Let $\{O_1, \ldots, O_m\}$ be a set of observables which we wish to measure and let A be a binary matrix whose elements are chosen so that*

$$A_{ij} = \begin{cases} 1, & \text{if } O_i \text{ and } O_j \text{ do not commute.} \\ 0, & \text{if } O_i \text{ and } O_j \text{ commute.} \end{cases} \tag{4.1}$$

Let $G = (V,E)$ be a graph defined by the adjacency matrix A with the vertex set being the set of observables $V = \{O_1, \ldots, O_m\}$ and let f be a colouring of G. Then those observables which are mapped to the same colour by f can be measured at the same time.

[2]One might wonder why we wouldn't choose to represent commuting pairs with 1 and non-commuting pairs with 0. This would indeed be a possibility but instead of leading to the graph colouring problem, it would lead to a problem called *minimum clique cover* problem. While some previous studies on variational quantum eigensolvers have chosen this approach [10][11][12][13][14], we have chosen the graph colouring approach as it is perhaps the better-known problem of the two. The two problems are however easily converted into each other so there is no real reason why one should be considered better than the other.

(a) Pauli measurements *(b) Clifford measurements*

Figure 4.1: The number of needed measurements depends on what types of measurements we are able to do. In (a) we need three different measurement sets for Pauli measurements while in (b) the same can be done with only two measurement sets using Clifford measurements.

Proof. If f maps some subset of V to the same colour, then there are no edges between any of the vertices in that subset. This is equivalent to saying that all of the corresponding observables commute and hence can be measured at the same time. □

One important thing to note is that we can construct these commutation graphs regardless of whether we are equipped with just the Pauli measurements or if we are able to do more complicated measurements. The types of measurement only affect how we construct the adjacency matrix in Equation 4.1: for Pauli measurements, the operators need to commute on every single-qubit level, while for more complicated measurements, they just need to commute as a whole.

Example 4.1. Let us consider a Hamiltonian

$$H = Y_1 + Y_2 + X_1X_2 + Z_1Z_2. \tag{4.2}$$

If we wish to measure this Hamiltonian with Pauli measurements, only the Y_1 and Y_2 operators commute at a single-qubit level. The corresponding commutation graph is shown in Figure 4.1a. As a result, we will need at least three sets of measurements, one for the pair $\{Y_1, Y_2\}$, one for X_1X_2 and one for Z_1Z_2.

If, instead, we measure the Hamiltonian using Clifford measurements, then also X_1X_2 and Z_1Z_2 commute with each other. The corresponding graph is shown in Figure 4.1b. Consequently, we now need only two sets of measurements.

Finally, if we assume that we are able to diagonalise the Hamiltonian with arbitrary measurements, then we clearly need only one set of measurements since we are trying to measure only one observable.

However, like we have already noticed earlier in Section 2.3, merely finding out whether observables commute or not is not enough to determine the optimal way to measure them in all cases. Consequently, this current graph colouring method cannot possibly cover everything. We can improve on this by taking into account the number of measurements each individual observable needs in order to achieve the desired accuracy for that observable.

(a) *(b)*

Figure 4.2: ***(a)*** *The observables from Example 4.2 can be coloured with two colours. If X_1 requires two measurements while the other two need only one, this colouring would lead to three measurements.* ***(b)*** *If we assign a weight two for the X_1 vertex, we can find a multicolouring with two colours. This way, only two measurements are needed.*

4.1 Weighted commutation graphs

As we noted in Equation 2.33, the number of measurements needed for a given commuting set of observables is proportional to the largest λ-norm of those observables squared. The λ-norms of the different observables can be initially different, or, in the case that we needed different accuracies for different observables, we decided to rescale the observables in order for the accuracies to match. We can include all of this necessary information in the graph representation by assigning weights to each vertex, each weight indicating how many measurements that corresponding observable would need by itself.

We also found out that when dividing these observables into commuting subsets, it might make sense to put some observables with large λ-norms into multiple sets if possible in order to divide the workload between those sets. All of this suggests that we should generalise the graph colouring problem and, indeed, the multicolouring problem discussed in Section 3.3 gives us a way to represent this situation.

Theorem 4.2. *Let $V = \{O_1, \dots O_m\}$ be a set of observables which we wish to measure and let $G = (V,E)$ be the corresponding commutation graph. Let $w : V \to \mathbb{Z}$ describe the number of measurements each observable needs and let f be a multicolouring of the weighted graph (V,E,w). We can measure the observables by letting each colour of f represent one measurement.*

The proof of this theorem is essentially the same as before, now we just have multiple colours for each observable. What is important though is that, as we have shown in Theorem 3.1, we can transform this problem back to regular graph colouring by replacing the vertices with complete graphs of corresponding order.

Example 4.2. Assume that we want to measure the observable X_1 twice and observables X_2 and Z_2 once. These observables form a graph which can be coloured with two

(a) *(b)*

Figure 4.3: ***(a)*** *A cycle graph of five vertices can be coloured with three colours.* ***(b)*** *Giving each vertex from (a) weight 2 for multicolouring, we can find a solution with only five colours instead of six.*

colours as shown in Figure 4.2a. Since one of the colours needs two measurements and the other needs one measurement, we need a total of three measurements.

If, instead, we give the vertex X_1 weight two for multicolouring, we can still find a solution with two colours as shown in Figure 4.2b. Now, we have a solution using only two measurements.

Example 4.3. Let $\{X_1, X_2, Z_1Z_2, Z_1X_3, Z_2Z_3\}$ be a set of observables which we wish to measure twice each. These observables form a *cycle graph* of five vertices, which can be coloured with three colours as shown in Figure 4.3a. This means that the total number of measurements becomes 6.

However, if we multicolour the same graph, giving each vertex weight 2, the graph can be coloured with five colours as shown in Figure 4.3b. This way, a total number of 5 measurements is enough.

In order to find the optimal way of measuring the state, the weights of the vertices should be chosen to be the number of times we need to measure the corresponding observables. However, as we have already noted, finding the optimal colouring for a graph becomes extremely difficult very quickly when increasing the size of the graph and, here, using the number of needed measurements for each observable could easily increase the system size by several magnitudes. Therefore, it might actually make more sense to pick some constant, large enough integer and divide the measurement count for each observable by that integer (rounding up). This would limit the increase in the graph size while still solving our problem since we could just repeat every measurement at the measurement step by the chosen constant number of times.

Example 4.4. Let us assume that we wish to get estimates of $\langle X_1 \rangle$ and $\langle 2Z_1 \rangle$ with accuracies $\varepsilon \leq 0.1$ and total chance of error $\delta \leq 0.05$. This means that we need to measure X_1 877 times and Z_1 3506 times. If we ignore the fact that, for this trivial example, we wouldn't really need the utilization of graphs, in this case it would seem that we would be increasing the graph size already over 1000-fold by giving the first observable 877 vertices and the second one 3506 vertices.

However, a much easier way would be to give observable X_1 just one vertex and observable Z_1 $\left\lceil \frac{3506}{877} \right\rceil = 4$ vertices. Then, at the end, we would just consider every vertex to represent 877 measurements each. This way, we managed to keep our graph size small while still solving the problem efficiently. The difference in the needed number of measurements at the last step would be just 4385 vs 4383, i.e., an increase of less than 0.05% while the decrease in the graph size is almost 99.9%.

However, as we will notice in Section 5, we should still make sure that the weights of the observables scale reasonably with the system size in order to find measurement schemes which scale optimally as well.

4.2 Optimality for Pauli observables

With the graph colouring, we are able to solve the correct measurement scheme classically first before having to apply quantum methods for the actual measurements. The algorithm is described in Algorithm 1 and is divided into classical and quantum parts.

As long as we are dealing with only (scaled) Pauli observables, it can be shown that the quantum measurement scheme that this algorithm provides is optimal in the number of needed measurements as long as we are distributing the error probability equally among the observables.

Theorem 4.3. *Given a set of scaled Pauli observables to measure, the graph colouring method described in Algorithm 1 finds an optimal measurement scheme in terms of the number of needed quantum measurements by choosing $c = 1$ and finding an optimal colouring for the graph in the classical part.*

Proof. Let us have an optimal measurement scheme consisting of measurements M_1,M_2,...,M_N, where each individual measurement M_i describes the set of observables that is measured at that step, $M_i = \{O_{i1}, O_{i2}, \dots\}$.

We can now construct a graph so that we have a vertex v_{ij} corresponding to each observable O_{ij}. A pair of vertices should have an edge between them if either (1) the corresponding observables do not commute or (2) the two observables are equal. This graph can be coloured with N colours since we can colour all the observables in $M_i = \{O_{i1}, O_{i2}, \dots\}$ with the same colour.

Algorithm 1 Measurement with commutation graphs - Classical part

Inputs: (1) Set of observables $\{O_1, O_2, \ldots, O_m\}$, (2) Desired accuracy ε, (3) Error probability δ, (4) Constant integer c (For optimality, pick $c = 1$)

Output: Graph with colouring

Create a graph with a vertex v_i corresponding to each observable O_i.
for $i = 1 : m$ **do**
    for $j = i + 1 : m$ **do**
        if O_i and O_j do not commute **then**
            Add an edge between v_i and v_j
        end if ▷ (For Pauli measurements, commutation should be checked on single-qubit level)
    end for
end for
for $i = 1 : m$ **do**
    for $j = 2 : \left\lceil \frac{\log\left(\frac{2m}{\delta}\right)||O_i||_\lambda^2}{2c\varepsilon^2} \right\rceil$ **do**
        Add a copy of v_i with all of the same edges, and add an edge from the new copy to all of the previous copies.
    end for
end for
Find an (optimal) colouring of the graph (e.g., by brute forcing)

Algorithm 1 Measurement with commutation graphs - Quantum part

Inputs: (1) Quantum state ρ, (2) All the inputs and outputs from the classical part

Output: Estimates of $\langle O_i \rangle$ for all of the observables.

for $i = 1 :$ number of colours **do**
    $S = \emptyset$
    for $j = 1 :$ number of vertices **do**
        if vertex j is coloured with colour i **then**
            Add the observable corresponding to vertex j into set S
        end if
    end for
    for $j = 1 : c$ **do**
        Measure all of the observables in set S at the same time and adjust their estimates accordingly
    end for
end for

We can also construct another graph using Algorithm 1. If we look at the vertices in each graph corresponding to a given observable O_i, the measurement graph should have at least $\left\lceil \frac{\log\left(\frac{2m}{\delta}\right)||O_i||_\lambda^2}{2\varepsilon^2} \right\rceil$ such vertices in order to have measured that observable with good enough accuracy. On the other hand, the graph from Algorithm 1 has exactly $\left\lceil \frac{\log\left(\frac{2m}{\delta}\right)||O_i||_\lambda^2}{2\varepsilon^2} \right\rceil$ such vertices.

Furthermore, the edges follow the same rules for both graphs, so we can see that the graph from Algorithm 1 must be a subgraph of the measurement graph. Therefore, the chromatic number of the former must be at most N, and, consequently, the number of measurements needed in Algorithm 1 is at most N.

□

As a caveat, Algorithm 1 assumes that we should distribute the error probability equally which is not fully optimal in all cases as noted in Section 2.3. However, if we wished to use a more optimal distribution, this would only affect the weights given for the observables, and would consequently only affect the number of created vertices on line 10 of the algorithm's classical part. By replacing that with the more optimal distribution, the algorithm would once again output the optimal measurement scheme.

However, in the case of more general observables, Algorithm 1 is not yet optimal and we would have to take into account all of the considerations from Section 2.3. Especially, we should remove from the graph those observables that do not need to be measured. This can be done by trying to optimize within the set of graphs which satisfy the constraints of Theorem 2.11. However, in order to do this efficiently, it would require further generalisation of the graph colouring problem, which will not be studied here.

5 Computational experiments

In this section, we will consider the problem of finding every k-qubit reduced density matrix (RDM) of an n-qubit system using Pauli measurements. This means that we want to measure every Pauli string of length at most k. We will call this the (n,k) *reduced density matrix (RDM) problem.*

5.1 Comparison methods

We will compare our graph colouring approach with two other methods. *Overlapping tomography* [8] is a tomography method specifically designed to measure the (n,k) RDM problem. In overlapping tomography one tries to find efficient ways of partitioning the set of Pauli observables into commuting subsets by using *families of perfect hash functions.* The idea is that we try to find a set of strings of length n using alphabet of size k such that if we pick any k positions in the string, there exists at least one string where each

of those k letters are different. Then, by letting the alphabet go through every possible combination of Pauli bases for each string, we can make sure that each k-qubit subset is measured in every possible Pauli combination. For more details, see the original paper by Cotler and Wilczek [8].

For pairwise tomography, i.e., the case $k = 2$, a slightly better implementation still using the same general idea is developed in [22]. In this pairwise tomography case there are efficient ways to find the families of hash functions. However, there are no general methods for finding them in the case $k > 2$ but since these have been studied extensively, there exist plenty of prior results which give the relevant families of hash functions for a wide range of (n,k) pairs. However, a bigger problem is, as we shall see, that this method scales very much suboptimally.

Another method which we will use as comparison is *random Pauli measurements*, which is considered, e.g., in [9]. In this method, we will just measure each of the qubits in a random Pauli basis (X, Y or Z). Perhaps surprisingly, this method scales much better than overlapping tomography. However, this better scaling can be explained qualitatively with the fact that, while we need to measure every observable multiple times, the random measurements allow us to measure each k-qubit observable with many different observable subsets while with overlapping tomography, we restrict each observable to one particular subset.

5.2 Single-colouring the (n,k) RDM problem

Much like in overlapping tomography, we will use graph colouring to find valid partitions of the observables. In this case, we will only use unit weights for each observable in the graph. As we will discuss later, using unit weights instead of the number of needed measurements and multicolouring will produce significantly worse scaling with the problem size. However, the benefit of this approach is that the number of different measurement settings stays smaller so we don't need to configure the measurement circuit again between every measurement. Additionally, even with this worse scaling we will obtain clearly better results than overlapping tomography.

We will consider both Pauli and Clifford measurements for the graph colouring approach. However, as we will see, for this particular problem these two will give generally the same results. The reason for this is that the benefit of the Clifford measurements is more significant when our observables are more global but when we are only interested in k-local observables there is actually not that much difference between the two measurement protocols. This can be most easily seen by considering the number of neighbours for each observable in the graph: the relative difference in the number of neighbours for each observable is of the order $\mathcal{O}(k/n)$ which approaches zero as n grows large.

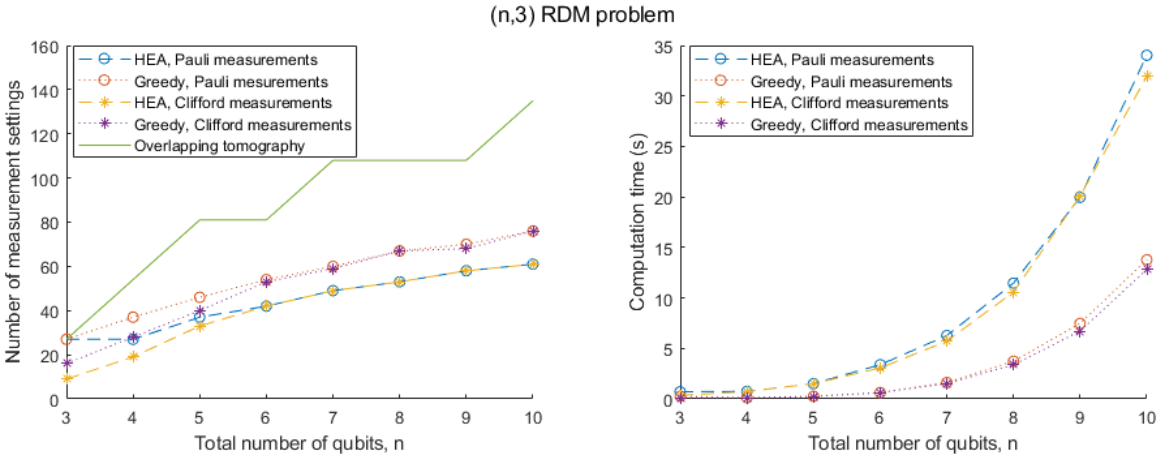


Figure 5.1: $(n,3)$ RDM problem for system sizes up to 10 qubits. As shown by the graph in the left, there are significant improvements in the number of needed measurement configurations found by both greedy and hybrid evolutionary graph colouring algorithms compared to the overlapping tomography method in [8]. For the total number of measurements, these numbers still need to be multiplied by the number of measurements that each observable needs, which grows logarithmically in the number of qubits. The graph on the right shows the computational time of these graph colouring solutions. For these system sizes, these non-optimal solutions can still be found relatively fast. However trying to find or prove the optimal solutions for the graph colouring problems would take very unreasonable amounts of time.

We will also consider two different colouring algorithms for the commutation graphs. First of all, we will use the very fast and simple greedy algorithm (see Section 3.2.2). One would not expect it to produce very competitive results as it is the simplest method available. Secondly, we will use the more complex hybrid evolutionary algorithm (see Section 3.2.3). This will naturally be slower than the greedy algorithm but should also provide better results.

Figure 5.1 shows our results for the 3-qubit reduced density matrices for up to 10-qubit systems and Figure 5.2 shows similarly results for 2-qubit RDMs for up to 40-qubit systems. Perhaps slightly surprisingly, we found in our computations that both of the used algorithms seemed to generally find better results for the Pauli measurements than for the Clifford measurements. Since any solution of the Pauli measurements is also a valid solution for the Clifford measurements, this must only mean that the algorithms find the solutions more easily for the Pauli measurements for some reason. However, this is probably less surprising when noticing that, in general, increasing the number of edges while keeping the chromatic number the same should make it easier for the algorithms to find the optimal solution. However, for this reason we decided to augment the algorithm for the Clifford measurements so that they start with the solution for the Pauli measurements and then try to improve from there. This way, we could guarantee that the Clifford measurements performed at least as good as the Pauli measurements, but as the graphs show, there is

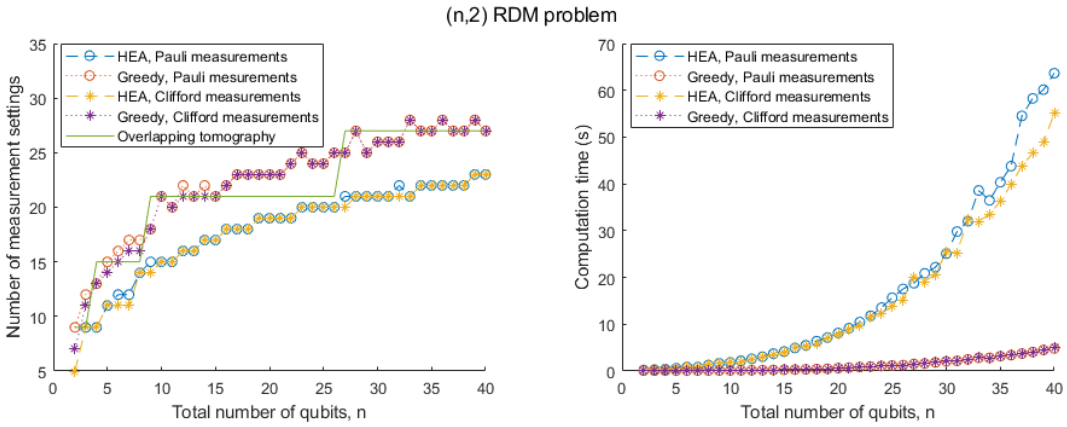


Figure 5.2: $(n,2)$ RDM problem for system sizes up to 40 qubits. The reference values for the overlapping tomography line are computed using the method from [22]. The greedy algorithm performs only barely as well as the reference value but the hybrid evolutionary algorithm shows some clear improvement.

not much difference.

However, a much more significant difference can be seen between the algorithms themselves and between the algorithms and the reference value provided by overlapping tomography. For the $(n,3)$ case in Figure 5.1, we can see that both of the graph colouring algorithms perform a lot better than the overlapping tomography approach. Even the greedy algorithm, while not as good as the HEA, seems to find much more efficient ways to perform the measurements than overlapping tomography. The graph on the right shows the computational time used to find these graph colouring solutions. As can be seen, the computational time still stays relatively small for these system sizes (35s for the HEA and 15s for greedy to solve the 10-qubit case). However, the limiting case for our computations was not so much the computational time in this case but the memory requirements. The $(n,3)$ RDM problem produces already for a system size of 10 qubits a graph with more than 3000 vertices and more than 2×10^6 edges. However, for the computational time, when comparing with overlapping tomography, one should also consider that there is no easy way to find the families of perfect hash functions needed for overlapping tomography either.

For the $(n,2)$ RDM case in Figure 5.2, we have used the slightly improved version for pairwise overlapping tomography provided in [22] for the reference value. In this case, the greedy algorithm performs only as well as overlapping tomography, but with HEA we can already see an improvement to both.

5.3 Limits of the single-colouring approach

We can show that the graphs in Figures 5.1 and 5.2 will always grow at least logarithmically. Let us consider a graph which consists of only the observables X_iY_j or Y_iX_j for all pairs (i,j). Then the colouring of the easiest of these graphs corresponds to the $(n,2,2)$ perfect hash family which grows logarithmically. Since this is a subgraph of any of the (n,k) RDM problems, also the colourings of those graphs must grow logarithmically.

Furthermore, to get the total number of measurements, we must multiply the number of colours with $\left\lceil \frac{2\log\left(\frac{2m}{\delta}\right)}{\varepsilon^2} \right\rceil$ as we must measure each observable that number of times. As that is also logarithmic in n (as m is polynomial in n), our total number of measurements scales as $\mathcal{O}(\log^2(n))$.

However, there are also measurement methods which scale only as $\mathcal{O}(\log n)$ using random measurements. In [9] it is shown that

$$N = \frac{68 \times 4^k \log\left(\frac{2m}{\delta}\right)}{\varepsilon^2} \tag{5.1}$$

random Pauli measurements is sufficient for the (n,k) RDM problem. In Appendix B we will further improve this bound to

$$N = \frac{8 \times 3^{k-1} \log\left(\frac{2m}{\delta}\right)}{\varepsilon^2} \tag{5.2}$$

measurements. The reason for the better scaling with these random measurements compared to our earlier $\log^2(n)$ scaling for the graph colouring is because the random measurements allow us to measure the same observable with multiple different measurement settings. This could just as well be accomplished by using the multicolouring approach so this just goes to show that for optimal results, we also need to consider the weights of the vertices. However, there are also clear benefits for the simpler approach: in addition to keeping the graphs significantly smaller, we also don't need to use new measurement setup for every measurement like with random measurements or multicolouring. Instead, for each observable we have assigned only one different measurement scheme and then the same measurement is replicated multiple times using that scheme. However, as we have seen, the number of different measurement configurations grows logarithmically either way so it might be preferable to use multicolouring nonetheless.

5.4 Multicolouring the (n,k) RDM problem

We have seen that in order to compete with the random measurement scheme, our graph colouring approach will eventually require multicolouring the vertices. In fact, the new bound that we introduced for the random measurements in Appendix B performs so well that the single-colouring approach loses already on very small system sizes with the (n,k)

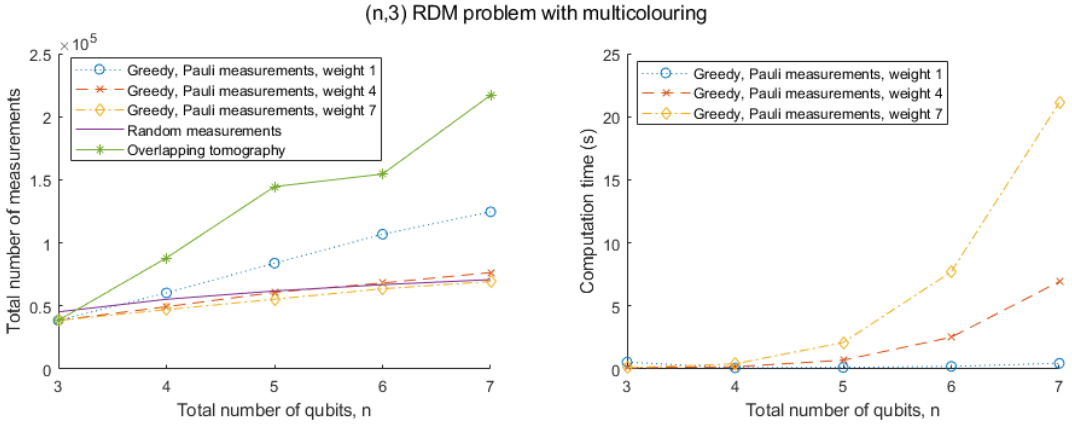


Figure 5.3: $(n,3)$ RDM problem using multicolouring with three different weights for each observable. While the single colouring approach loses very quickly to the random measurement bound, already seven colours for each observable is enough to outdo the random measurements at least for up to seven qubit systems. However, in order to keep up with the random bound, the weights of the observables have to be increased at least logarithmically with the system size. Note that unlike in Figures 5.1 and 5.2, here we are counting the total number of measurements (i.e., copies of the system) instead of the number of different measurement setups.

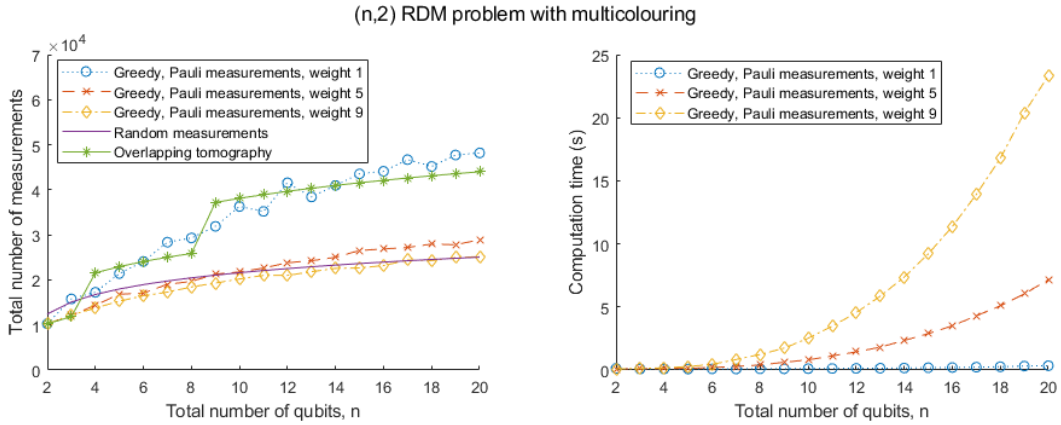


Figure 5.4: $(n,2)$ RDM problem using multicolouring with three different weights for each observable. Again, we can see that by increasing the weights of the observables enough, even greedy algorithm allows us to find better results than the comparison methods. The weights here are still moderately small and could still be increased for the largest system sizes by over 200-fold if computational constraints allowed it.

RDM problem.

Therefore, we have used multicolouring with weights 1 (single-colouring), 4 and 7 to solve the (n,3) RDM problem in Figure 5.3 with a reference value for the random measurement scheme computed according to the results in Appendix B. Here, we also need to assign values for the error tolerance and probability, so we have used the values $\delta = \varepsilon = 0.1$. In this case we used only the naive greedy algorithm but even that was able to outperform the random measurement scheme with the weight 7 multicolouring for at least up to 7-qubit systems. The weight of 7 is still quite small though, as, for optimal performance on a 7-qubit system, we should assign each observable a weight of $\left\lceil \frac{2\log\left(\frac{2m}{\delta}\right)}{\varepsilon^2} \right\rceil = 2010$. However, even with these small weights and way-less-than-optimal algorithm, we can clearly see how increasing the weights reduces the number of measurements and, with large enough weights, multicolouring can exceed any other method.

Figure 5.4 shows similarly our results for the (n,2) RDM problem with weights 1, 5 and 9. Again, for the largest system sizes of 20 qubits, optimally, we should really use weights of over 2000, but even nine colours per observable is enough to match the results of the random measurements with the greedy algorithm. For the perfect performance however, we should still maximize the weights as well as use a better-performing algorithm for the graph colouring.

6 Summary

We have outlined some general principles of quantum state tomgraphy and shown that graph colouring and especially multicolouring can be used to find more efficient measurement procedures for quantum state tomography. Graph colouring can, in theory, find the optimal measurement setups but, in practice, there are at least three restricting factors which make it more difficult.

First of all, in order to prove optimality, one would have to use exact algorithms for solving the graph colouring with minimal number of colours and, consequently, the computational time would increase exponentially with the system size. However, we have shown that even the most simple graph colouring algorithm, the greedy algorithm, can find more efficient solutions than some other known methods for quantum tomography while only requiring polynomial time with respect to the graph size to compute the solutions.

Second hindrance for the optimal solution are the vertex weights for multicolouring. For a perfect algorithm, these weights would need to be extremely large even for small system sizes, which would slow down the graph colouring algorithm even further. We saw in Section 5 for the case of the (n,k) RDM problem that while we need to increase the weights together with the system size in order for our solutions to scale optimally. However, even

with relatively small weights, we were able to find efficient solutions for small system sizes. This would indicate that even relatively large system sizes could be solved somewhat efficiently with still reasonable weights.

The third obstacle for finding the optimal solution is finding the optimal error probabilities for the individual observables as discussed in Section 2.3. For perfectly symmetric problems, such as the (n,k) RDM problem, it makes sense that the error probabilities should be equal for all of the observables like we have generally assumed in this thesis. However, in general, one might have to consider assigning the error probabilities and, consequently, the weights of the observables unevenly. Furthermore, taking correlation of the observables into account, one could find better division of the probabilities instead of just using the sum of the individual probabilites as a bound for the total error probability.

Additionally, we also considered how a set of more general observables should be optimally arranged into measurements in Section 2.3.3. While using the Pauli basis works well also in this general case, we speculate that it might be possible to find a generalisation of the proposed graph colouring method which also allows arbitrary unitary transformations of the Pauli basis. As a more general method, this would guarantee better performance for these general observables. However, as it is unlikely that there already exists an analogue of this problem in graph theory, algorithms for this problem might have to be built from scratch unlike the graph colouring algorithms considered in this thesis.

Despite all these possible impediments and directions of further improvement, the results that we have shown here are extremely promising. Even a small overhead in the classical computations can significantly lower the cost of the quantum measurements. Therefore, this approach should prove useful in various different applications requiring quantum tomography.

References

[1] P. Shor, Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer, SIAM J.Sci.Statist.Comput. 26 (1997)

[2] F. Arute et al. Quantum supremacy using a programmable superconducting processor, Nature 574, 505–510 (2019)

[3] H. Barnum et al., Noncommuting mixed states cannot be broadcast, Phys.Rev.Lett. 76, 2818-2821 (1996)

[4] M. Cramer et al., Efficient quantum state tomography, Nat. Commun., 1 (9) 149 (2010)

[5] B.P. Lanyon et al., Efficient tomography of a quantum many-body system, Nature Physics 13, 1158-1162 (2017)

[6] M. Niedermeier, From Quantum Entanglement to Interactions of Elementary Excitations in Coupled Spin Chains, `https://helda.helsinki.fi/bitstream/handle/10138/333721/Niedermeier_Marcel_thesis_2021.pdf` (2021)

[7] J. de Boer et al., Quantum hypothesis testing in many-body systems, SciPost Phys. Core 4, 019 (2021)

[8] J. Cotler, F. Wilczek, Quantum Overlapping Tomography, arXiv:1908.02754 (2019)

[9] H.-Y. Huang, R. Kueng, J. Preskill, Predicting Many Properties of a Quantum System from Very Few Measurements, arXiv:2002.08953 (2020)

[10] V. Verteletskyi, T-C. Yen, A.F. Izmaylov, Measurement Optimization in the Variational Quantum Eigensolver Using a Minimum Clique Cover, arXiv:1907.03358 (2019)

[11] A. Jena, S. Genin, M. Mosca, Pauli Partitioning with Respect to Gate Sets, arXiv:1907.07859 (2019)

[12] T-C. Yen, V. Verteletskyi, A.F. Izmaylov, Measuring all compatible operators in one series of a single-qubit measurements using unitary transformations, arXiv:1907.09386 (2019)

[13] W.J. Huggins et al., Efficient and Noise Resilient Measurements for Quantum Chemistry on Near-Term Quantum Computers, npj Quantum Inf 7, 23 (2021)

[14] P. Gokhale et al., Minimizing State Preparations in Variational Quantum Eigensolver by Partitioning into Commuting Families, arXiv:1907.13623 (2019)

[15] M.A. Nielsen, I.L. Chuang, Quantum Computation and Quantum Information, 10th anniv. ed. (Cambridge University Press, 2010)

[16] P.O. Boykin et al., On Universal and Fault-Tolerant Quantum Computing, arXiv:quant-ph/9906054 (1999)

[17] E. Knill, Approximation by Quantum Circuits, arXiv:quant-ph/9508006

[18] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. 58, 13–30 (1963)

[19] R.M. Karp, Reducibility among Combinatorial Problems, In: R.E. Miller, J.W. Thatcher, J.D. Bohlinger (eds), Complexity of Computer Computations, 85-103 (Springer, Boston, 1972)

[20] R.M.R. Lewis, Guide to Graph Colouring: Algorithms and Applications, 2nd ed. (Springer, Cham, 2021)

[21] P. Galinier, J-K. Hao, Hybrid evolutionary algorithms for graph coloring, J. Comb. Optim. 3, 379–397 (1999)

[22] G. García-Pérez et al., Pairwise tomography networks for many-body quantum systems, arXiv:1909.12814 (2019)

A Qudits and other bases

While we have decided to restrict ourselves to systems of qubits, all of the results can be generalised to other types of bases as well. In order to measure an observable, the only two things we need are

- the ability to perform unitary operations diagonalizing the observable and
- the ability to measure the system in a "computational basis".

As an example, let us assume we have a system comprised of n qudits each of which occupies a d-dimensional Hilbert space, and that we are trying to measure some observable O. Like with qubits, we need a unitary operator U which diagonalises O and we need to be able to perform that same unitary operation on the unknown quantum state. In addition, we need to be able to measure each qudit in the computational basis, meaning that we need to be able to tell apart the states $|0\rangle, \ldots, |d-1\rangle$.

However, while the Pauli strings provided us with a nice basis for the linear operators in the case of qubits, they won't work similarly here. This won't matter for measuring general observables, but if, for instance, we wanted to do full tomography, then we would need a basis with which to represent the state. One possibility would be a common generalisation of the Pauli matrices. First we need to divide the dimensions of the qudits into their prime factors and then for each prime q, we can define the "shift" and "clock" matrices:

$$S_q = \begin{pmatrix} 0 & 0 & & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \qquad C_q = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & e^{2\pi i \frac{1}{q}} & 0 & \cdots & 0 \\ 0 & 0 & e^{2\pi i \frac{2}{q}} & \cdots & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & e^{2\pi i \frac{q-1}{q}} \end{pmatrix}. \tag{A.1}$$

Now, our generalized basis for the q-dimensional Hilbert space's linear operators is formed from the matrices $P_{(q,a,b)} = S_q^a C_q^b$ for $a,b \in \{0,\ldots,q-1\}$ and for non-prime dimensions, we just take tensor products of the corresponding matrices of their prime factors.

The following lemma shows that the matrices $P_{(q,a,b)} = S_q^a C_q^b$ share some of the traits of the original Pauli strings, the main difference being that they are not Hermitian.

Lemma A.1. *The matrices $P_{(q,a,b)} = S_q^a C_q^b$ satisfy the following properties:*

1. *$P_{(q,a,b)}$ is unitary.*
2. *$P_{(q,a,b)}$ is traceless, except for $I = P_{(q,0,0)}$.*
3. *$P_{(q,a,b)}$ and $P_{(q,a',b')} \neq P_{(q,a,b)}$ are orthogonal with respect to the Hilbert-Schmidt inner product.*

4. $P_{(q,a,b)}P_{(q,a',b')} = e^{2\pi i\frac{a'b}{q}}P_{(q,a+a',b+b')} = e^{2\pi i\frac{a'b-b'a}{q}}P_{(q,a',b')}P_{(q,a,b)}$

5. *The eigenvalues of* $P_{(q,a,b)} \neq I$ *are every* q*th root of unity.*

Proof. First, we can write the matrices in the outer product representation as

$$\begin{aligned} P_{(q,a,b)} = S_q^a C_q^b &= \left(\sum_{j=0}^{q-1} |j+1\rangle\langle j|\right)^a \left(\sum_{k=0}^{q-1} e^{2\pi i\frac{k}{q}}|k\rangle\langle k|\right)^b \\ &= \sum_{j=0}^{q-1} e^{2\pi i\frac{jb}{q}}|j+a\rangle\langle j|, \end{aligned} \tag{A.2}$$

where the sums inside the kets are to be taken modulo q.

1. Clearly S_q and C_q are unitary so likewise is any product of them.

2.
$$\operatorname{tr}(P_{(q,a,b)}) = \sum_{j=0}^{q-1} e^{2\pi i\frac{jb}{q}}\langle j|j+a\rangle = \delta_{a,0}\delta_{b,0} \tag{A.3}$$

3.
$$\begin{aligned} (P_{(q,a,b)},P_{(q,a',b')}) &= \operatorname{tr}\left(\sum_{j,j'=0}^{q-1} e^{2\pi i\frac{j'b'-jb}{q}}|j\rangle\langle j+a|j'+a'\rangle\langle j'|\right) \\ &= \sum_{j=0}^{q-1} e^{2\pi i\frac{j(b'-b)}{q}}\langle j+a|j+a'\rangle \\ &= \delta_{b,b'}\delta_{a,a'} \end{aligned} \tag{A.4}$$

4.
$$\begin{aligned} P_{(q,a,b)}P_{(q,a',b')} &= \sum_{j,j'=0}^{q-1} e^{2\pi i\frac{jb+j'b'}{q}}|j+a\rangle\langle j|j'+a'\rangle\langle j'| \\ &= \sum_{j=0}^{q-1} e^{2\pi i\frac{jb+j'b'+a'b}{q}}|j+a+a'\rangle\langle j| \\ &= e^{2\pi i\frac{a'b}{q}}P_{(q,a+a',b+b')} \\ &= e^{2\pi i\frac{a'b-b'a}{q}}P_{(q,a',b')}P_{(q,a,b)} \end{aligned} \tag{A.5}$$

As we might expect, $P_{(q,a,b)}$ and $P_{(q,a',b')}$ commute exactly when $a'b = b'a \mod (q)$, since that also quarantees that they can be expressed as powers of each other with some additional phase factors.

5. First of all, using the previous result, we know that

$$P_{(q,a,b)}{}^q = e^{2\pi i \frac{q(q-1)ab}{q}} P_{(q,qa,qb)} = P_{(q,0,0)} = I, \tag{A.6}$$

so any eigenvalue of $P_{(q,a,b)}$ must be a qth root of unity. Let $|\varphi\rangle$ be an eigenvector of $P_{(q,a,b)}$ with eigenvalue λ. Now, let us choose a',$b' \in \{0,\ldots,q-1\}$ such that

$$a'b - b'a = 1 \mod q, \tag{A.7}$$

which is always possible unless $a = b = 0$. Then,

$$P_{(q,a,b)} P_{(q,a',b')} |\varphi\rangle = e^{2\pi i \frac{1}{q}} \lambda P_{(q,a',b')} |\varphi\rangle, \tag{A.8}$$

so $e^{2\pi i \frac{1}{q}} \lambda$ is also an eigenvalue. Therefore, every qth root of unity must be an eigenvalue of $P_{(q,a,b)}$.

□

There is nothing inherently different about measuring a complex-valued observable. E.g., the measurements still produce the correct value $\mathrm{tr}(O\rho)$ in expectation. However, we need to redefine how we interpret the accuracy of the measurement, since Lemma 2.1 only considers real-valued random variables.

The obvious way to define the error tolerance of the result is with the absolute value of the difference from the real value

$$|\tilde{O} - \langle O \rangle| \leq \varepsilon. \tag{A.9}$$

While Lemma 2.1 does not give us the probability of success directly, one easy way to work around this is to consider the real and imaginary parts of the measurements as separate random values.

Lemma A.2. *Let X_1,X_2,$\ldots$,X_N be i.i.d. complex-valued random variables such that their absolute value $|X_i| \leq 1$ and let $\bar{X}$ be their empirical mean. Then*

$$P\left(\left|\bar{X} - E\left[\bar{X}\right]\right| \geq \varepsilon\right) \leq 4 \exp\left(-\frac{N\varepsilon^2}{4}\right). \tag{A.10}$$

Proof. Using Lemma 2.1, we find

$$\begin{aligned} &P\left(\left|\bar{X} - E\left[\bar{X}\right]\right| \geq \varepsilon\right) \\ \leq &P\left(\left|\mathrm{Re}\left(\bar{X}\right) - E\left[\mathrm{Re}\left(\bar{X}\right)\right]\right| \geq \frac{\varepsilon}{\sqrt{2}}\right) + P\left(\left|\mathrm{Im}\left(\bar{X}\right) - E\left[\mathrm{Im}\left(\bar{X}\right)\right]\right| \geq \frac{\varepsilon}{\sqrt{2}}\right) \\ \leq &4 \exp\left(-\frac{N\varepsilon^2}{4}\right). \end{aligned} \tag{A.11}$$

□

So now we can determine a sufficient number of measurements for the observables

Corollary A.2.1. *Let O be a complex-valued quantum observable with eigenvalues bounded by $|\lambda_i| \leq 1$. In order to measure $\langle O \rangle$ with probability at least $1-\delta$, it is enough to measure the observable*

$$N = \frac{4 \log\left(\frac{4}{\delta}\right)}{\varepsilon^2} \tag{A.12}$$

times.

With this, we are able to measure the generalised Pauli matrices and can use it, e.g., for performing full tomography on qudits as well. One important thing to consider is, though, how these generalised Pauli matrices commute, since that determines how many of them we can measure at the same time. It is clear that, once again, the Pauli matrices acting on different qudits commute with each other. However, in this generalised case, also Pauli matrices acting on the same qudit can commute.

Proposition A.3. *Ignoring I, the generalised Pauli matrices $P_{(q,a,b)}$ can be divided into $q+1$ commuting subsets of size $q-1$ each. $P_{(q,a,b)}$ and $P_{(q,a',b')}$ belong to the same subset iff*

$$a'b = b'a \mod q \tag{A.13}$$

Proof. It should be clear that both $P_{(q,a,0)} = S_q^a$ and $P_{(q,0,b)} = C_q^b$ form commuting subsets of size $q-1$, so we will assume that $a,b \neq 0$. We know from Lemma A.1 that $P_{(q,a,b)}$ and $P_{(q,a',b')}$ commute iff

$$a'b = b'a \mod q. \tag{A.14}$$

Furthermore, if

$$a'b = b'a \mod q \qquad \text{and} \qquad a''b = b''a \mod q \tag{A.15}$$

then

$$a''b'ab = b''a'ab \qquad \Rightarrow \qquad a''b' = b''a', \tag{A.16}$$

so the commutation relation is transitive in this set (if we included I, it would not be). Now, since q is prime, for each $a' \in \{1, \ldots, q-1\}$ we can find $b' \in \{1, \ldots, q-1\}$ such that

$$a'b = b'a \mod q \tag{A.17}$$

so there must be $q-1$ commuting operators. Since the total number of operators is q^2-1, there must be $q+1$ of these commuting subsets. □

With this result, we know that we can reduce the number of different measurement sets for each qudit to $q+1$. Now, it would make sense to ask, how would the Pauli and Clifford measurements compare between qubits and qudits in efficiency?

Theorem A.4. *With generalised Pauli measurements, we can perform full tomography on a d-dimensional Hilbert space divided into q-dimensional qudits by dividing the (generalised) Pauli operators into*

$$f(q) = d^{\frac{\log(q+1)}{\log q}} \tag{A.18}$$

commuting subsets. For a constant d, f is a strictly decreasing function of q for $q \geq 2$.

Proof. It is enough for us to make sure that we measure every generalised Pauli string where none of the single-qudit operators equal I, since this produces at the same time also those Pauli strings where some of the single-qudit operators are replaced by I. Also, like we noticed in Proposition A.3, the single-qudit operators can be divided into $q+1$ commuting subsets. This means that we just need to measure each possible combination of these subsets on n qudits and there are

$$f(q) = (q+1)^n = (q+1)^{\frac{\log d}{\log q}} = d^{\frac{\log(q+1)}{\log q}} \tag{A.19}$$

such combinations.

Now, the derivative of the exponent is

$$D\left(\frac{\log(q+1)}{\log q}\right) = \frac{q\log q - (q+1)\log(q+1)}{q(q+1)(\log q)^2} < 0 \tag{A.20}$$

so the exponent is strictly decreasing and therefore f is also strictly decreasing for all $q \geq 2$. □

This theorem tells us that if we are constrained to single-qudit measurements, then qubits give us the worst performance. The reason for this should be clear: with qubits we assume the least amount of freedom in choosing unitary operators to transform between bases. Also, by choosing the upper limit $q = d$, giving the lower limit of $d+1$ subsets, we can show that this performance cannot be beaten with any kind of projective measurements regardless of the dimension of the qudits.

Theorem A.5. *Full tomography with projective measurements will always require at least $d+1$ different commuting subsets.*

Proof. To show that we cannot achieve better performance, let us consider some particular projective measurement and let us represent the state ρ in the basis which diagonalises that measurement. In this basis, the measurement can only give us information about the diagonal elements of ρ, so we can see that any projective measurement can only provide information about at most $d-1$ linearly independent variables. Since full description of the state requires knowing d^2-1 variables, we see that it is not possible to do full tomography with less than $d+1$ separate measurement setups.

□

We would also like to point out that while the scaling $d + 1$ might at first glance look quite promising, d still scales exponentially with respect to the number of qudits n. Therefore, increasing the number of qubits or qudits still increases the number of measurement settings exponentially, making full tomography extremely difficult for larger system sizes even in this best case scenario. For this reason, it is usually preferable to try to limit the number of observables we are interested in.

One final thing to note is that there is no reason why we could not also form an orthogonal Hermitian basis just for the space of Hermitian operators. Since ρ is also Hermitian, these would also enable full tomography and we could do the measurements with only real-valued results.

One such Hermitian basis, written using the outer product representation, could be

$$A_j = |j\rangle\langle j| \qquad B_{j,k} = |j\rangle\langle k| + |k\rangle\langle j| \qquad C_{j,k} = i|j\rangle\langle k| - i|k\rangle\langle j| \tag{A.21}$$

for $j,k \in \{0,\ldots,d-1\}$, $j < k$. These operators are all Hermitian and orthogonal with each other and they span the whole space of Hermitian operators for any dimension d. We might notice that A_j are not traceless. While this doesn't really matter, we could also replace all but one of them by traceless operators by first removing the trace from them and then applying the Gram-Schmidt procedure. One potential problem with this basis is, though, that since almost all of the eigenvalues of the operators are zero, the expectation values are likely to be quite small, and, with these small expectation values, even small errors could produce a significant difference. Furthermore, unlike with the generalised Pauli matrices, we don't have a simple way of dividing these operators into as few commuting subsets so any advantage that we get from the fact that they are Hermitian is likely to be lost with the number of separate subsets we would need to consider.

B Improved upper bound for the k-local problem with randomised measurements

The methods in [9] can be used to determine an upper bound on the needed measurements when measuring a set of k-local observables with randomised Pauli measurements. A special case of this k-local problem would be the (n,k) RDM problem considered in Section 5 and in [8]. The upper bound for the random measurements from [9] would be

$$N = \frac{68 \times 4^k \log\left(\frac{2m}{\delta}\right)}{\varepsilon^2}. \tag{B.1}$$

However, this bound can be improved upon significantly.

Theorem B.1. *Let $\{O_1, \ldots, O_m\}$ be a set of Pauli strings, each of same length k, which we wish to measure with random Pauli measurements. If we require that all of the measurement estimates are within accuracy ε of the correct expectation values with probability $1 - \delta$, we can bound the number of needed measurements by*

$$N = \frac{8 \times 3^{k-1} \log\left(\frac{2m}{\delta}\right)}{\varepsilon^2}. \tag{B.2}$$

Proof. We will denote with $P_{N,R}()$ the probability of something occurring with $N \in \mathbb{N}$ random measurements, and with $P_j()$ something occurring with $j \in \mathbb{N}$ measurements measuring explicitly the given observable.

Let p be the probability that a random measurement measures O and, consequently, with probability $1 - p$ we get no information about O. Then, the probability that, after N random measurements, our estimate $\tilde{O}$ is inaccurate by more than ε can be bounded by

$$\begin{aligned} P_{N,R}(|\tilde{O} - \langle O \rangle| > \varepsilon) &= \sum_{j=1}^{N} \binom{N}{j} p^j (1-p)^{N-j} P_j(|\tilde{O} - \langle O \rangle| > \varepsilon) \\ &\leq 2 \sum_{j=1}^{N} \binom{N}{j} p^j (1-p)^{N-j} \exp\left(-\frac{j\varepsilon^2}{2}\right) \\ &= 2\left[1 - p + p\exp(-\varepsilon^2/2)\right]^N. \end{aligned} \tag{B.3}$$

Since we want to bound the probability of error by δ/m, we want to find N such that

$$2\left[1 - p + p\exp(-\varepsilon^2/2)\right]^N = \frac{\delta}{m}. \tag{B.4}$$

This is given by

$$\begin{aligned} N &= \frac{\log\left(\frac{\delta}{2m}\right)}{\log[1-p+p\exp(-\varepsilon^2/2)]} \\ &\leq \frac{\log\left(\frac{2m}{\delta}\right)}{p[1-\exp(-\varepsilon^2/2)]} \\ &\leq \frac{\log\left(\frac{2m}{\delta}\right)}{p(\varepsilon^2/2-\varepsilon^4/8)} \\ &\leq \frac{8\log\left(\frac{2m}{\delta}\right)}{3p\varepsilon^2}. \end{aligned} \tag{B.5}$$

The probability that a Pauli string of length k is measured with some random Pauli measurement is $p = \frac{1}{3^k}$ so the number of measurements can finally be bounded by

$$N = \frac{8 \times 3^{k-1} \log\left(\frac{2m}{\delta}\right)}{\varepsilon^2}. \tag{B.6}$$

□

Corollary B.1.1. *We can measure every k-qubit reduced density matrix with*

$$N = \frac{8 \times 3^{k-1} \log\left(\frac{2m}{\delta}\right)}{\varepsilon^2} \tag{B.7}$$

random measurements.