Purifying Quantum States with Repeated Measurements

by

Timo Gils
Student number: 5869447

to obtain the degree of
Bachelor of Science
in Applied Mathematics and Applied Physics
at the Delft University of Technology,
to be defended publicly on Monday July 7, 2025 at 13:30

Supervisor:
Dr. Bas Janssens
Dr. Viatcheslav Dobrovitski

Thesis Committee:
Dr. Bas Janssens
Dr. Eliška Greplová
Dr. Barbara Terhal
Dr. Michael Wimmer

TU Delft Delft University of Technology

1 Abstract

This paper covers the perfect indirect quantum measurement, specifically in the context of repeated measurement. The indirect measurement is useful as it allows information to be obtained from quantum systems without inflicting much disturbance on them. We restrict ourselves to cases with no evolution of the measured system between measurements and to perfect measurements, that is, measurements from which no outgoing information is missed and no extra information is added. In this case we can make use of the work by M. A. Nielsen (2005). It says that the expected amount of information following a perfect indirect measurement is larger than the information before the measurement. We make use of this result to show that the repeated indirect perfect measurement of a quantum state has two mutually exclusive outcomes. The first outcome is that the measured state becomes a pure state almost surely. The second is that the measurement eventually stops resulting in information being revealed. In the latter case, further measurements on the system result in the state switching through spaces with the same dimension, and thus it does not become a pure state. This paper builds on the work by Maassen and Kümmerer from 2005, which already proved this, by expanding their proofs and adding additional theorems and proofs to create a more self-contained result. Further studies might look at the rate at which states become pure, and what might influence this rate.

Contents

2 Introduction

The indirect quantum measurement is an important concept describing how information can be obtained about a quantum system. It provides a method to get details on the state of the quantum system without disrupting the system much [1]. This is in contrast to the direct (projective) measurement in which, after the measurement, the system is set to the pure state corresponding to the given outcome. Although you now have complete information regarding the current system, the system was also greatly disrupted. As a trade-off, the indirect measurement generally returns less information. After an indirect measurement, the state is altered but does not necessarily collapse. This means that indirect measurements can be repeated on a previously measured system. But does repeating the indirect measurement necessarily lead to the complete information of a pure state? Thereby showing us the path the system took to a pure state? M. A. Nielsen has shown that the perfect indirect quantum measurement will on average lead to a gain of information [2]. We make use of this result to explore the effect on repeated indirect measurements. We restrict ourselves to cases with no evolution of the measured system between measurements and to perfect measurements, such that the measurement captures all outgoing information and no extra information is going into the system. With this restriction, we show that repeated measurements will either almost surely converge to a pure state, or the purification process will stop before that point and some information regarding the state will not be revealed. In the latter case, the state will eventually cycle through 'dark subspaces', named after the inability to extract information from them.

The concept of purification of quantum trajectories has already been covered by H. Maassen and B. Kümmerer in their article 'Purification of Quantum Trajectories' [3], and our work is based on this article. We want to create a stand-alone and expanded version of this result.

If the repeated indirect measurements do lead to a pure state almost surely or not depends on the measurement (Kraus) operators. A special case to this would be in 2 dimensions, as mentioned in the work of Maassen and Kümmerer [3][Corollary 2]. If all the measurement operators correspond to scaled unitaries, then the dark subspace is the entire space. Whilst if the measurement operators are not all scaled unitaries, then the repeated measurements converge to a pure state with probability 1.

In Section 3 we give a brief introduction to the quantum mechanics used and explain the perfect indirect measurement. Then in Section 4 we introduce the trajectory of repeated measurements on a state and show that it is a stochastic process. In Section 5 we step away from the quantum trajectory for a moment to give a self-contained proof of a crucial inequality based on the work by Nielsen, for which we need to delve into the majorization relation of vectors. Lastly, we state and prove the purification of the quantum trajectories in Section 6, followed by clarifications and an extension to this theorem.

3 Prerequisite knowledge

3.1 States

To understand the purification of quantum trajectories, it is crucial to have a basic understanding of pure and mixed states, which are descriptions of quantum systems. Pure states contain all possible information of the system, while mixed states are a probabilistic mixture of pure states, thus giving a far less clear image of the state. We describe these states using density matrices. A good description of these topics can be found in the *"Introduction to Quantum Mechanics"* by D. J. Griffiths and D. F. Schroeter [4][p. 575].

A quantum system can be modelled by a Hilbert space $\mathcal{H}$. There are two possible types of state that describe the system, the first being the pure state. A pure state is represented by a normalised vector $|\psi\rangle \in \mathcal{H}$, this state can also be represented as a square matrix, the density matrix, $\theta = |\psi\rangle \langle\psi|$. This notation is useful for defining the mixed state. This state is a probabilistic mixture of pure states with a density matrix.

$$\theta = \sum_k p_k |\psi_k\rangle \langle\psi_k| , \tag{1}$$

where the $|\psi_k\rangle$ are orthogonal pure states, and the p_k are the 'probabilities', with $0 < p_k < 1$ and $\sum_k p_k = 1$. The p_k are also the eigenvalues of the density matrix.

As the pure state is defined using a normalised vector, the trace of the density matrix of a pure state must be one. And as the probabilities in a mixed state must sum to one, the trace of a mixed state is also one. For a pure state, we have that the state vector is the eigenvector of the density matrix representation, with eigenvalue one. All other eigenvectors are orthogonal to the state vector and thus have an eigenvalue of zero. This makes the density matrix positive semidefinite. A mixed state is an affine combination of pure states, thus also positive semidefinite. Thus θ is a density matrix only if it has trace one and is positive semidefinite. Furthermore, it can be shown that if a matrix has trace one and is positive semidefinite, it must be a density matrix; see *"Quantum Computation and Quantum Information"* by M. A. Nielsen and I. L. Chuang [5], page 101.

What we can see from equation 1 due to the orthogonality of the vectors is that for every density matrix

$$\theta^n = \big(\sum_k p_k |\psi_k\rangle \langle\psi_k| \big)^n = \sum_k p_k^n |\psi_k\rangle \langle\psi_k| ,$$

with $0 < p_k \leq 1$ and $\sum_k p_k = 1$. From this it follows that

$$\mathrm{tr}(\theta^n) = \sum_k p_k^n. \tag{2}$$

Following Maassen and Kümmerer [3] we call $\mathrm{tr}(\theta^n)$ the n^{th} moment of the density matrix θ. Now we conclude the following observation, which we use later to identify pure states.

Observation 1. *Let θ be a density matrix, then θ is a pure state if and only if for all $n \in \mathbb{N}$, $\mathrm{tr}(\theta^n) = 1$.*

Proof. If θ is a pure state, then we have that $\theta^n = (|\psi\rangle \langle\psi|)^n = |\psi\rangle \langle\psi| = \theta$, as the vector is normalised. It follows that $\mathrm{tr}(\theta^n) = \mathrm{tr}(\theta) = 1$.
If for all $n \in \mathbb{N}$, $\mathrm{tr}(\theta^n) = 1$, then from equation 2 it follows that $\mathrm{tr}(\theta^n) = \sum_k p_k^n = 1$ for all $n \in \mathbb{N}$, with $0 < p_k \leq 1$ and $\sum_k p_k = 1$. This means we must have one p_k equal to one and the rest equal to zero, otherwise with increasing n the trace would decrease. But one p_k equal to one and the rest equal to zero implies that θ is a pure state. □

Remark 1. *In fact, for mixed states θ we have that $\mathrm{tr}(\theta^n) < 1$ for $n \geq 2$. This is because $p_k^n < p_k$ for $n \geq 2$, as p_k is strictly less than one. So that $\mathrm{tr}(\theta^n) = \sum_k p_k^n < \sum_k p_k = 1$. So for a density matrix θ it suffices to have that $\mathrm{tr}(\theta^2) = 1$ to show that it is a pure state.*

The observables are the hermitian matrices $A \in \mathcal{A}$, where $\mathcal{A}$ is a *-subalgebra of $\mathcal{L}(\mathcal{H})$, the algebra of bounded operators on $\mathcal{H}$. The observables correspond to measurable quantities, whose possible outcomes are given by the eigenvalues of the corresponding observable. The eigenvectors of an observable form a basis of the space $\mathcal{H}$. The expectation of an observable A for a density matrix θ is given by $\mathrm{tr}(\theta A)$.

As it suffices for the paper, we consider only the finite-dimensional Hilbert spaces $\mathcal{H} \cong \mathbb{C}^n$ for simplicity. Thus also $\mathcal{L}(\mathcal{H}) \cong M_n(\mathbb{C}^n)$. For every finite-dimensional Hilbert space $\mathcal{H}$ we have corresponding state space denoted as $\mathcal{D}(\mathcal{H})$, the space containing all density matrices.

3.2 Evolution of states

Now we are interested in the evolution of these quantum states, specifically in the context of measurements. In the finite-dimensional case this process can be described using maps between state spaces. Specifically, given two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, by maps $\epsilon : \mathcal{D}(\mathcal{H}_1) \to \mathcal{D}(\mathcal{H}_2)$. Restricting ourselves to linear maps, according to Nielsen and Chuang [5], these maps $\epsilon : \mathcal{D}(\mathcal{H}_1) \to \mathcal{D}(\mathcal{H}_2)$, must be completely positive. A positive map implies that $\epsilon(\theta)$ is a positive semidefinite matrix if θ is a positive semidefinite matrix, such that density matrices are sent to density matrices. Completely positive is the added demand that the operator remains positive in combined systems. Thus $(\epsilon \otimes \mathrm{Id}_n)(\theta)$ must be positive semidefinite for any positive semidefinite matrix θ and any n, where Id_n represents the identity map on $M_n(\mathbb{C}^n)$. The meaning of positivity here is that the operation is valid on systems in isolation, meaning that when the system interacts with no other system, the operation sends states to states. Complete positivity adds to this by demanding that the operation remains valid

even when applied to a system part of a larger entangled system. This means that, even if the system is entangled with some ancillary system, applying the operation locally, leaving the ancillary system untouched, still results in the entire entangled system remaining a valid state.

Now to look at measurements in more detail, we make use of the work of Maassen and Kümmerer [3] to describe a measurement as a linear completely positive map. When we measure a prior state θ, we get information and possibly change the state; say a measurement has outcomes $i \in \{1, \ldots, k\}$, with corresponding new states θ'_i. All these outcomes have a certain probability to occur, given by p_i. As these are all possible outcomes, we have that $\sum_{i=1}^{k} p_i = 1$. We define the map $T_i : \theta \to p_i \theta'_i$, which says that there is a probability $p_i = \mathrm{tr}(T_i(\theta))$ to get the result i, and thus the posterior state $\frac{T_i(\theta)}{\mathrm{tr}(T_i(\theta))}$. By definition, the map T_i is a completely positive map. In the event that the measurement process takes place but no outcome is measured, so we do not condition on the outcome, then the posterior state is given by the trace-preserving completely positive map $T : \theta \to \sum_{i=1}^{k} p_i \theta'_i = \sum_{i=1}^{k} T_i(\theta)$. Meaning we average over all possible outcomes of the measurement.

To get a more detailed description of the operation T_i we work out the Kraus representation of the measurement, following the approach used by Maassen and Kümmerer [3] to explain this. The Kraus measurement makes use of the von Neumann measurement, so we start there. The von Neumann measurement is a projective measurement. Suppose we have the observable $A \in \mathcal{A}$, whose eigenspaces are the ranges of P_i, which are mutually orthogonal projections in $\mathcal{A}$ adding up to $\mathbf{1}$. A measurement of A is described by operation $T_i(\theta) = P_i \theta P_i$. The probability is given by $\mathrm{tr}(P_i \theta P_i) = \mathrm{tr}(\theta P_i)$, due to the cyclic nature of trace and projection property $P_i^2 = P_i$.

For an indirect (Kraus) measurement of our system $\mathcal{A}$ in the state θ, we make use of a second system $\mathcal{B}$, the 'ancilla', in the state β. We bring these two systems into contact such that they interact under Schrödinger's evolution, described by a unitary $u \in \mathcal{B} \otimes \mathcal{A}$. We then uncouple the ancilla and perform a projective measurement on the ancilla. As the two systems have interacted, the information gathered from this measurement will also contain information regarding the original system.

To deduce the posterior state of our original system in this case, we make use of an observable projection $q \in \mathcal{A}$ in our original system. Following our formalism, where θ'_i denotes the posterior state after decoupling the ancilla and measuring i, we have that the probability of then measuring q is

$$\mathbb{P}(q|P_i) = \mathrm{tr}(\theta'_i q) = \mathrm{tr}\left(\frac{T_i \theta}{\mathrm{tr}(T_i \theta)} q\right) = \frac{\mathrm{tr}((T_i \theta) q)}{\mathrm{tr}(T_i \theta)}$$

The projections $(P_i \otimes \mathbf{1})$ and $(\mathbf{1} \otimes q)$ commute, as they apply to other systems. So again using von Neumann, we can write the probability of both happening as:

$$\mathbb{P}(\text{measurement } i \text{ and event } q) = tr \otimes tr(u(\beta \otimes \theta)u^*(P_i \otimes q))$$

From which we can gather the conditional probability as well.

$$\mathbb{P}(\text{event } q|\text{measurement } i) = \frac{tr \otimes tr(u(\beta \otimes \theta)u^*(P_i \otimes q))}{tr \otimes tr(u(\beta \otimes \theta)u^*(P_i \otimes \mathbf{1}))}$$

Thus we have that $T_i\theta = tr \otimes id(u(\beta \otimes \theta)u^*(P_i \otimes \mathbf{1}))$. Where id is the identity map. This notation is unwieldy, so we first denote a simplification.

Proposition 1. *Let $T_i\theta = tr \otimes id(u(\beta \otimes \theta)u^*(P_i \otimes \mathbf{1}))$. If we have that:*

1. *$\mathcal{B}$ consists of all $k \times k$ matrices*
2. *P_i is one-dimensional ($P_i = |g_i\rangle \langle g_i|$ for some orthonormal basis $\{|g_i\rangle\}_i$)*
3. *β is a pure state (with state vector $|\beta\rangle = (\beta_1, \dots \beta_k)^T$).*

Then we can denote $T_i\theta = a_i\theta a_i^*$. Here $a_i := \sum_{j=1}^k \beta_j w_{i,j}$, where $w_{i,j}$ are $d \times d$ matrices, the blocks of a unitary $k \times k$ block matrix w.

Proof. First we wish to remove the trace operator. Using the fact that $\mathrm{tr} \otimes \mathrm{id}(X \otimes Y) = \sum_i (\langle g_i| \otimes \mathbf{1})(X \otimes Y)(|g_i\rangle \otimes \mathbf{1})$, for any orthonormal basis. Picking the orthonormal basis of the projection for the trace, the sum is removed due to the orthogonality property, and we write

$$\begin{aligned} T_i\theta &= \mathrm{tr} \otimes \mathrm{id}(u(|\beta\rangle \otimes \mathbf{1})(\mathbf{1} \otimes \theta)(\langle\beta| \otimes \mathbf{1})u^*(|g_i\rangle \langle g_i| \otimes \mathbf{1})) \\ &= ((\langle g_i| \otimes \mathbf{1})u(|\beta\rangle \otimes \mathbf{1})\theta(\langle\beta| \otimes \mathbf{1})u^*(|g_i\rangle \otimes \mathbf{1})). \end{aligned} \tag{3}$$

Following point 2, we have a unitary matrix v with the orthonormal basis vectors $|g_i\rangle$ as its columns such that $|g_i\rangle = v|e_i\rangle$, and $\langle g_i| = \langle e_i| v^*$, with $\{|e_i\rangle\}_i$ representing the standard basis. We then have a new unitary matrix $w := (v^* \otimes I)u$. Applying this on the element left of θ in equation 3 we see that

$$(\langle g_i| \otimes \mathbf{1})u(|\beta\rangle \otimes \mathbf{1}) = (\langle e_i| v^* \otimes \mathbf{1})u(|\beta\rangle \otimes \mathbf{1}) = (\langle e_i| \otimes \mathbf{1})w(|\beta\rangle \otimes \mathbf{1}). \tag{4}$$

Due to point 1, u can be written as a $k \times k$ matrix $(u_{i,j})$ of $d \times d$ matrices. As explained above, we then have a new unitary matrix $w = (v^* \otimes I)u$, which can also be written as a $k \times k$ matrix $(w_{i,j})$ of $d \times d$ matrices. Thus $w = \sum_{i=1}^k \sum_{j=1}^k |e_i\rangle \langle e_j| \otimes w_{i,j}$. Similarly we can write a vector as a sum of the standard basis, $|\beta\rangle = \sum_{l=1}^k \beta_l |e_l\rangle$. Substituting both of these in equation 4 we get

$$(\langle g_i| \otimes \mathbf{1})u(|\beta\rangle \otimes \mathbf{1}) = (\langle e_i| \otimes \mathbf{1})(\sum_{i=1}^k \sum_{j=1}^k |e_i\rangle \langle e_j| \otimes w_{i,j})(\sum_{l=1}^k \beta_l |e_l\rangle \otimes \mathbf{1}) = \sum_{j=1}^k \beta_j w_{i,j}, \tag{5}$$

by the orthogonality of the standard basis. By a similar argument for the element to the right of θ we find that,

$$(\langle\beta| \otimes \mathbf{1})u^*(|g_i\rangle \otimes \mathbf{1}) = \sum_{j=1}^{k} \bar{\beta}_j (w_{i,j})^* \tag{6}$$

Substituting equations 5 and 6 in equation 3, we get that

$$T_i\theta = (\sum_{j=1}^{k} \beta_j w_{i,j})\theta(\sum_{j=1}^{k} \bar{\beta}_j (w_{i,j})^*) = a_i \theta a_i^*,$$

with $a_i := \sum_{j=1}^{k} \beta_j w_{i,j}$. This completes our simplification.

□

We now show that this simplified notation comes with the completeness property of the a_i. Meaning that $\sum_{i=1}^{k} a_i^* a_i = \mathbf{1}_d$, where $\mathbf{1}_d$ denotes the $d \times d$ identity matrix. Starting with

$$\sum_{i=1}^{k} a_i^* a_i = \sum_{i=1}^{k} ([\sum_{j=1}^{k} \bar{\beta}_j (w_{i,j})^*][\sum_{j'=1}^{k} \beta_{j'} w_{ij'}]) = \sum_{j,j'=1}^{k} \sum_{i=1}^{k} \bar{\beta}_j \beta_{j'} (w_{i,j})^* w_{ij'}. \tag{7}$$

As w is unitary, we have that $ww^* = w^*w = \mathbf{1}_{dk}$, the $dk \times dk$ identity matrix. Using the block matrix representation, this means that $\sum_{j=1}^{k} w_{i,j}(w_{i'j})^* = \mathbf{1}_d \delta_{i,i'}$ and $\sum_{i=1}^{k} (w_{i,j})^* w_{ij'} = \mathbf{1}_d \delta_{j,j'}$ respectively. Substituting this in equation 7 we get

$$\sum_{i=1}^{k} a_i^* a_i = \sum_{j,j'=1}^{k} \bar{\beta}_j \beta_{j'} \mathbf{1}_d \delta_{j,j'} = \sum_{j=1}^{k} |\beta_j|^2 \mathbf{1}_d = \mathbf{1}_d,$$

where in the last step we used that as β is a pure state, the vector $|\beta\rangle$ is normalised. The Kraus (or indirect) measurement that we just described is a perfect measurement by Maassen and Kümmerer, as they define it as follows:

'By a perfect measurement on $\mathcal{A}$ we mean a k-tuple $(T_1, ..., T_k)$ of operations on $\mathcal{S}$, where $T_i\theta$ is of the form $a_i \theta a_i^*$ with $\sum_{i=1}^{k} a_i^* a_i = \mathbf{1}$'[3].

4 Repeated measurement

We consider repeated measurement of a system, potentially up to infinite measurements. Using the notation of the previous section, in the case that we first obtain outcome i and then, after a second measurement, outcome j, the posterior state after these two measurements is:

$$\theta'_{i,j} = \frac{(T_j \circ T_i)(\theta)}{\mathrm{tr}((T_j \circ T_i)(\theta))}.$$

This procedure can be repeated for any finite number of measurements to determine the posterior state. However, to make statements about future measurements, it is helpful to view this process as a stochastic process. Based on the explanation given in the book 'A First Course in Stochastic Processes' by S. Karlin and H. M. Taylor [6] a stochastic process is a collection of random variables, say X_t, where $t \in T$ is the index set. All individual random variables X_t are defined on the same probability space and take values in the same state space.

To this end we now create a probability space for the repeated measurements, following the approach of Maassen and Kümmerer [3]. Starting with the set of infinite outcome sequences (sample space) $\Omega = \{\omega = (\omega_1, \omega_2, \dots)|\omega_i \in \{1, \dots, k\}\}$. To fix the first m outcomes for each $m \in \mathbb{N}$ we make use of the cylinder sets $\Lambda_{i_1,\dots i_m} = \{\omega|\omega_1 = i_1, \dots, \omega_m = i_m\}$. For our event space we could take the σ-algebra generated by the cylinder sets for all $m \in \mathbb{N}$, but later we see that it is helpful to make the distinction, such that we first take the finite σ-algebra Σ_m generated by all cylinder sets which fix the first m outcomes and use these Σ_m to generate the σ-algebra Σ over all m.

4.1 Creating a measure

With this in place we need a measure on this probability space (Ω, Σ) with respect to an arbitrary starting state θ_0. To this end we apply Kolmogorov's extension theorem from *"Probability: Theory and Examples"* by R. Durrett [7] appendix A.3. The extension theorem applies to the following set $\mathbf{R}^{\mathbb{N}} = \{\omega = (\omega_1, \omega_2, \dots)|\omega_i \in \mathbf{R}\}$ with product σ-algebra $\mathcal{R}^{\mathbb{N}}$ generated by sets of the form $\{\omega|\omega_i \in (a_i, b_i], i = 1, \dots, m\}$, for all $m \in \mathbb{N}$. This is a more general real-valued version of our integer-based infinite outcome sequence Ω with product σ-algebra Σ generated by cylinder sets $\Lambda_{i_1,\dots,i_m}$. The theorem then states that given probability measures μ_m on $(\mathbf{R}^m, \mathcal{R}^m)$ for all $m \in \mathbb{N}$ that are consistent, meaning that

$$\mu_{m+1}((a_1, b_1] \times \cdots \times (a_m, b_m] \times \mathbf{R}) = \mu_m((a_1, b_1] \times \cdots \times (a_m, b_m]),$$

then there exists a unique probability measure μ on $(\mathbf{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}})$ with

$$\mu(\omega|\omega_i \in (a_i, b_i], i = 1, \dots, m) = \mu_m((a_1, b_1] \times \cdots \times (a_m, b_m]).$$

In our case, we work with $\omega_i \in \{1, \dots, k\}$, and we create the following measures on the Σ_m algebras,

$$P_m(\bigcup_{i_1 \in J_1, \dots, i_m \in J_m} \Lambda_{i_1,\dots i_m}) = P_m(J_1 \times \cdots \times J_m) = \sum_{i_1 \in J_1} \cdots \sum_{i_m \in J_m} tr[(T_{i_m} \circ \cdots \circ T_{i_1})(\theta_0)],$$

where $J_i \subseteq \{1, \dots, k\}$ for all i, and we used the fact that a measure on the length m cylinder sets solely depends on the m subsets J_i to write it as a measure on $\{1, \dots, k\}^m$ to simplify the notation.

P_m is a measure only accepting values from $\{1,\dots,k\}$, for a measure accepting real numbers, we create $P'_m(A) = P_m(A \cap \{1,\dots,k\}^m)$, and we require that $\{1,\dots,k\} \subseteq \mathbf{R}$. Now for the proof of consistency of P'_m we denote $W := \{1,\dots,k\}$,

$$
\begin{aligned}
&P'_{m+1}((a_1,b_1]\times\cdots\times(a_m,b_m]\times\mathbf{R}) = \\
&P_{m+1}([(a_1,b_1]\cap W]\times\cdots\times[(a_m,b_m]\cap W]\times W) = \\
&\sum_{i_1\in((a_1,b_1]\cap W)}\cdots\sum_{i_m\in((a_m,b_m]\cap W)}\sum_{i_{m+1}=1}^{k} tr[(T_{i_{m+1}}\circ T_{i_m}\circ\cdots\circ T_{i_1})(\theta_0)] = \\
&\sum_{i_1\in((a_1,b_1]\cap W)}\cdots\sum_{i_m\in((a_m,b_m]\cap W)} tr[(T\circ T_{i_m}\circ\cdots\circ T_{i_1})(\theta_0)] = \\
&\sum_{i_1\in((a_1,b_1]\cap W)}\cdots\sum_{i_m\in((a_m,b_m]\cap W)} tr[(T_{i_m}\circ\cdots\circ T_{i_1})(\theta_0)] = \\
&P_{m+1}([(a_1,b_1]\cap W]\times\cdots\times[(a_m,b_m]\cap W]) = \\
&P'_{m+1}((a_1,b_1]\times\cdots\times(a_m,b_m]),
\end{aligned}
$$

where we used the fact that T is trace preserving. Thus we have a measure P' on $(\mathbf{R}^{\mathbb{N}},\mathcal{R}^{\mathbb{N}})$ and by the same relation between P_m and P'_m we also have a measure P on (Ω,Σ). Until now we have neglected specifying the starting state θ_0, but it should be noted that this measure works for any valid starting state. Thus now we have the probability space $(\Omega,\Sigma,P_{\theta_0})$, on which we can create the random variable

$$\Theta_n = \frac{(T_{\omega_n}\circ\cdots\circ T_{\omega_1})(\theta_0)}{\mathrm{tr}((T_{\omega_n}\circ\cdots\circ T_{\omega_1}))}.$$

Which denotes the posterior state, given that we started with state θ_0, and we measured outcomes $\omega_1,\dots,\omega_n$, in that order after n measurements. For every $n\in\mathbb{N}$ this random variable is Σ_n measurable. As for every density matrix, the pre-image is a union of cylinder sets.

We call $(\Theta_n)_{n\in\mathbb{N}}$ the quantum trajectory, and this stochastic process is a Markov process, as any posterior state depends solely on the prior state.

With the trajectory in place, we can start looking ahead towards purification. By purification we mean that the state in a trajectory converges to a pure state, formally a quantum trajectory $(\Theta_n(\omega))_{n\in\mathbb{N}}$ purifies when

$$\forall_{m\in\mathbb{N}} : \lim_{n\to\infty}\mathrm{tr}(\Theta_n(\omega)^m) = 1.$$

We consider this purification, as only pure states have that the trace of all powers of their density matrix is equal to 1, as shown in observation 1. Thus, a quantum trajectory purifies when the state eventually becomes pure. With that out of the way, we can finally state the theorem on purification.

Theorem 1 (From [3]). *Let $(\Theta_n)_{n\in\mathbb{N}}$ be a quantum trajectory with an initial state θ_0 and process as described above. Then we have either of the following options occur:*

- *(i) The quantum trajectory $(\Theta_n)_{n\in\mathbb{N}}$ purifies almost surely.*
- *(ii) There exists a projection $p \in \mathcal{A}$ with a dimension of at least 2, such that*

$$\forall_{i\in\{1,\dots,k\}}\exists_{\lambda_i\geq 0} : pa_i^* a_i p = \lambda_i p. \tag{8}$$

The proof of Theorem 1 is also discussed in the work of Maassen and Kümmerer [3], we have added detail to this proof. But before we can start proving this theorem we need more information on the relation between the m^{th} moment of subsequent density matrices in a trajectory.

5 Nielsen's Inequality

We derive an inequality first shown by Nielsen [2], that gives a relation between the moments of density matrices before and after an indirect measurement. This inequality allows us to describe the moments as submartingales, and the submartingale properties are the main components of the proof of Theorem 1.

This inequality is derived from a majorization relation, so we first need to define the majorization relation $\prec$. For real vectors x and y, $x \prec y$ means that y majorizes x, thus that the vectors with their components in non-increasing order, denoted as $x^\downarrow = (x_1^\downarrow, \dots, x_n^\downarrow)$, with $x_k^\downarrow \geq x_l^\downarrow$ for $k \leq l$ have the following relation:

$$\sum_{i=1}^{k} x_i^\downarrow \leq \sum_{i=1}^{k} y_i^\downarrow,$$

where $k \in \{1, \dots, n\}$, with n the number of components of the vectors. We also require equality for $k = n$, so

$$\sum_{i=1}^{n} x_i^\downarrow = \sum_{i=1}^{n} y_i^\downarrow.$$

5.1 Horn's lemma

Horn's lemma serves as the foundation of Nielsen's inequality, as it allows us to relate specific equalities between vectors to a majorization relation. Leading us to eventually relate our measurement process to a majorization relation.

For the description of Horn's lemma, we make use of the following article by Nielsen [8].

Lemma 1 (Horn's Lemma)**.** *Let x and y be two vectors of length n then $x \prec y$ if and only if there exists a unitary matrix $u = (u_{i,j})$ such that for the matrix $D_{i,j} = |u_{i,j}|^2$ we have $x = Dy$.*

This means that $x_i = \sum_{j=1}^{n} |u_{i,j}|^2 y_j$. Furthermore, D is called unitary-stochastic, and due to the unitary matrix, it has the property that

$$\sum_{j=1}^{n} |u_{i,j}|^2 = \sum_{i=1}^{n} |u_{i,j}|^2 = 1, \tag{9}$$

so that every row and column sum to one. If the matrix u were orthogonal, then D is called orthostochastic.

In our argument we only need the reverse implication of Horn's lemma, but for completeness, we also provide a proof of the forward implication. Starting with the latter, the proof of the forward implication is based on the same work by Nielsen [8], and will be done for orthogonal matrices. We make use of the matrices called T-transforms, which are identical to the identity matrix except in two dimensions, where they have the following structure:

$$\begin{bmatrix} t & 1-t \\ 1-t & t \end{bmatrix}$$

with $0 \leq t \leq 1$. Keep in mind that the given structure may occur on any set of coordinates $i \neq j$ in the T-transform, not necessarily on consecutive coordinates. The matrix has t on the diagonal entries $A_{i,i}$ and $A_{j,j}$, $1-t$ on the entries $A_{i,j}$ and $A_{j,i}$, and equals the identity matrix in all other entries. We first state and prove a supporting lemma before moving on to the proof of Horn's lemma.

Lemma 2. *Let x and y be vectors with the relation $x \prec y$, then there exists a finite number of T-transforms T_i such that $x = T_1 \dots T_m y$.*

Proof. For ease of notation in the entire proof we set the components of the vectors in non-increasing order, which is allowed as this can be done through a finite series of T-transforms with $t = 1$. We prove this lemma using induction.

For vectors x and y with two components and $x \prec y$, we have that $x_1 \leq y_1$ and $x_1 + x_2 = y_1 + y_2$, thus also $y_2 \leq x_2$. By the non-increasing order of the vectors $x_2 \leq x_1$. Together we have that $y_2 \leq x_2 \leq x_1 \leq y_1$. Thus there exists a $t \in [0, 1]$ such that $x_1 = ty_1 + (1-t)y_2$ and as $x_1 + x_2 = y_1 + y_2$, we also conclude that $x_2 = (1-t)y_1 + ty_2$. Proving the lemma for vectors with two components.

For induction we assume that for vectors with n components, if $x \prec y$, there exists a finite number of T-transforms such that $x = T_1 \dots T_m y$. To finish the induction step, we prove that the same holds for vectors with $n+1$ components. For this we first show that there exists a T-transform which sets the first component of of both vectors equal, before applying the induction assumption. Suppose we have two vectors x and y with $n+1$ components such that $x \prec y$, then $x_1 \leq y_1$. Furthermore, for $x \prec y$ with $n+1$ components, $\sum_{i=1}^{n} x_i \leq \sum_{i=1}^{n} y_i$, but then for the sum over one more component, we have

equality, $\sum_{i=1}^{n+1} x_i = \sum_{i=1}^{n+1} y_i$. This means that $x_{n+1} \geq y_{n+1}$, and by the non-increasing order of the vectors $x_1 \geq x_{n+1}$, so that $x_1 \geq x_{n+1} \geq y_{n+1}$. As we now have that $y_1 \geq x_1 \geq y_{n+1}$, we can find the component y_k such that $y_k \leq x_1 \leq y_{k-1}$. This means that there exists a $t \in [0,1]$ such that $x_1 = ty_1 + (1-t)y_k$. Applying the corresponding T-transform matrix to vector y, we can write $Ty = (x_1, y')$, with $y' = (y_2, \dots y_{k-1}, (1-t)y_1 + ty_k, y_{k+1}, \dots, y_{n+1})$ a vector of n components, we note that this vector y' is not necessarily in non-increasing order. Similarly we define $x' = (x_2, \dots, x_{n+1})$, also as a vector of n components, which is still ordered.

Now we have shown that there exists a T-transform matrix that sets the first component of y equal to the first component of x. From here, to prove the lemma, we need to show that there exists a finite number of T-transforms to set the other n components equal. For this we make use of the induction assumption so that we only need to show that $x' \prec y'$, as these are vectors with n components. For the new component $(1-t)y_1 + ty_k$ of y', we have that $(1-t)y_1 + ty_k \geq y_k \geq y_{k+1}$, so in the ordered vector $y'^{\downarrow}$ we know that this new component $(1-t)y_1 + ty_k$ has position $k-1$ or lower. position $k-1$ and not k as from y to y' we removed the largest component, so all vectors rise one position when ordered. The position could be lower, as this new component could be larger than y_{k-1} etc. Call the position of this new component k', with $k' \leq k-1$. As $[(1-t)y_1 + ty_k] + [ty_1 + (1-t)y_k] = y_1 + y_k$ and $x_1 = ty_1 + (1-t)y_k$, we have that this new component $(1-t)y_1 + ty_k = y_1 + y_k - x_1$.

To show the majorization relation $x' \prec y'$, we start by comparing summations over the ordered components before we include this new component. As x' is still ordered, we have that $\sum_{i=1}^{k'-1} x_i'^{\downarrow} = \sum_{i=2}^{k'} x_i$, and as x_1 is the largest component, we have $\sum_{i=2}^{k'} x_i \leq (k'-1)x_1$. The vector y' is also ordered non-increasingly except for the new component, which is placed too high. So, for sums over the ordered components of y' that end before the index k', so before the new component is added to the sum, we have a similar notation to x' in that $\sum_{i=1}^{k'-1} y_i'^{\downarrow} = \sum_{i=2}^{k'} y_i$. Then by our previous assumption that $y_k \leq x_1 \leq y_{k-1}$, combined with $k' \leq k-1$, we see that

$$\sum_{i=1}^{k'-1} x_i'^{\downarrow} = \sum_{i=2}^{k'} x_i \leq (k'-1)x_1 \leq \sum_{i=2}^{k'} y_i = \sum_{i=1}^{k'-1} y_i'^{\downarrow}.$$

By the same argument, we have inequality for sums over fewer components, say $l \leq k'-1$.

$$\sum_{i=1}^{l} x_i'^{\downarrow} = \sum_{i=2}^{l+1} x_i \leq lx_1 \leq \sum_{i=2}^{l+1} y_i = \sum_{i=1}^{l} y_i'^{\downarrow}.$$

When we do include this new component, say a sum over $m \geq k'$, we see that the sum over the first m components of the $y'^{\downarrow}$ vector is

$$\sum_{i=1}^{m} y_i'^{\downarrow} = \sum_{i=2}^{m+1} y_i - y_k + (y_1 + y_k - x_1) = \sum_{i=1}^{m+1} y_i - x_1, \tag{10}$$

where the term $-y_k + (y_1 + y_k - x_1)$ was used to replace the y_k in the sum with the new component $(1-t)y_1 + ty_k = y_1 + y_k - x_1$. Given that $x \prec y$, we know that $\sum_{i=1}^{m+1} y_i \geq \sum_{i=1}^{m+1} x_i$, subtracting x_1 from both sides, $\sum_{i=1}^{m+1} y_i - x_1 \geq \sum_{i=2}^{m+1} x_i$, meaning that

$$\sum_{i=1}^{m} y_i'^{\downarrow} = \sum_{i=1}^{m+1} y_i - x_1 \geq \sum_{i=2}^{m+1} x_i = \sum_{i=1}^{m} x_i'^{\downarrow}. \tag{11}$$

For equality when summing over all n components, it suffices to note that as $x \prec y$, we have $\sum_{i=1}^{n+1} y_i = \sum_{i=1}^{n+1} x_i$. Subtracting x_1 from both sides we conclude

$$\sum_{i=1}^{n} y_i'^{\downarrow} = \sum_{i=1}^{n+1} y_i - x_1 = \sum_{i=2}^{n+1} x_i = \sum_{i=1}^{n} x_i'^{\downarrow}. \tag{12}$$

Thus proving that $x' \prec y'$.

As x' and y' are vectors with only n components by our induction assumption, there exists a finite number of T-transforms such that $x' = T_1 \dots T_r y'$. These T-transforms can be extended to work on vectors with $n+1$ components, working as the identity on the first element, and thus we can write $x = T_1 \dots T_r T y$, proving what we wanted. □

With this lemma in place we briefly cover an observation that aids the proof of Horn's lemma before moving on to the proof of Horn's lemma.

Observation 2. *Let D be a unitary-stochastic or orthostochastic matrix, so $D = |u_{i,j}|^2$. Then for permutation matrices P and Q, we have that PDQ is also a unitary-stochastic or orthostochastic matrix, respectively.*

Proof. As permutation matrices are orthogonal (thus also unitary), we can define the unitary or orthogonal matrix $u' := PuQ$. As the relation between u and D is that D has the absolute square elements of u, as $D = |u_{i,j}|^2$, permutations of D can be seen as permutations of u. This means that $PDQ = \left|u'_{i,j}\right|^2$, which is unitary-stochastic or orthostochastic, respectively. □

As the proof of Horn's lemma is rather long I will split the lemma into two different lemmas, a forward and reverse implication and proof both separately. Note that for the forward implication we prove a stronger case with orthostochastic matrices.

Lemma 3 (Forward implication of Horn's lemma)**.** *Let x and y be two vectors of length n, then $x \prec y$ implies that there exists a orthostochastic matrix D, such that $x = Dy$.*

Proof. Without loss of generalization we may assume that the vectors x and y are in non-increasing order. Indeed, any vector can be created from an ordered vector and a permutation matrix. If we have that $x^{\downarrow} = Dy^{\downarrow}$, for orthostochastic matrix D. Then for any two permutation matrices P and Q such that $Px = x^{\downarrow}$ and $Qy = y^{\downarrow}$, we have $x = P^{-1}DQy$. As P^{-1} is also a permutation matrix, by observation 2, $P^{-1}DQ$ is also orthostochastic.

Now we only have to prove the lemma for vectors with components in non-increasing order. The proof of Lemma 2 directly gives us a method to generate the T-transforms required for these ordered vectors such that for $n+1$ components $x = T_1 \dots T_n y$. In this product of T-transforms, T_{n+1-k} acts on the k^{th} and some d_k^{th} component with $k < d_k$.

Using induction, we prove that this specific sequence of T-transforms can be written as an orthostochastic matrix. In the case that $n = 2$, we shape the T transform with the orthogonal matrix

$$U = \begin{bmatrix} \sqrt{t} & -\sqrt{1-t} \\ \sqrt{1-t} & \sqrt{t} \end{bmatrix}.$$

The matrix U is orthogonal as $UU^* = U^*U = I$ and has the property that $T_1 = (U_{i,j}^2)$.

Now for induction we assume that the product of $n-1$ T-transforms for vectors with n components is equal to some $n \times n$ orthostochastic matrix D_n such that $T_1' \dots T_{n-1}' = D_n$. The accent denotes that these are T-transforms on n dimensions. Then for the case of vectors with $n+1$ components, we extend $T_1' \dots T_{n-1}'$ to act on $n+1$ components by letting it act as the identity on the first element, just like in the previous proof. Let us denote these extended T-transform without the accent, then

$$T_1 \dots T_{n-1} = \begin{bmatrix} 1 & 0 \\ 0 & T_1' \dots T_{n-1}' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & D_n \end{bmatrix}.$$

Besides $T_1 \dots T_{n-1}$, we also need a new T-transform T_n, which influences the first and d_1^{th} component of the vector. For convenience we introduce the permutation matrix P which switches this d_1^{th} row with the second row, such that

$$PT_nP = \begin{bmatrix} t & 1-t & 0 \\ 1-t & t & 0 \\ 0 & 0 & I_{n-1} \end{bmatrix},$$

where I_{n-1} is the $n-1 \times n-1$ identity matrix. We also want to apply the permutation to $T_1 \dots T_{n-1}$. As the permutation matrix P switches the d_1^{th} row with the second row, we can write P as

$$P = \begin{bmatrix} 1 & 0 \\ 0 & P' \end{bmatrix},$$

with P' being a $n \times n$ permutation matrix, the switching first and $(d_1 - 1)^{\text{th}}$ row. This makes it easier to apply the permutation to $T_1 \dots T_{n-1}$, as we now have that

$$PT_1 \dots T_{n-1}P = \begin{bmatrix} 1 & 0 \\ 0 & D_n' \end{bmatrix},$$

where $D_n' := P'D_nP'$. As P' is a permutation matrix and D_n is orthostochastic, by observation 2 D_n' is also an orthostochastic matrix. For the following step it helps to split the first column of D_n' such that $D_n' = [\delta \quad \tilde{D}_n']$, with δ the first column and $\tilde{D}_n'$ being D_n' without the first column. Then we can write

$$PT_1 \dots T_{n-1}P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \delta & \tilde{D}'_n \end{bmatrix}.$$

Now to multiply $PT_1 \dots T_{n-1}P$ by PT_nP, as $P^2 = I$, we see that

$$PT_1 \dots T_{n-1}PPT_nP = PT_1 \dots T_nP = \begin{bmatrix} t & 1-t & 0 \\ (1-t)\delta & t\delta & \tilde{D}'_n \end{bmatrix}.$$

This is an orthostochastic matrix, as if we take the orthogonal matrix V corresponding to D'_n, and have v as the first column and $\tilde{V}$ as the matrix without the first column, we can write a new matrix

$$W = \begin{bmatrix} \sqrt{t} & -\sqrt{1-t} & 0 \\ \sqrt{(1-t)}v & \sqrt{t}v & \tilde{V} \end{bmatrix}.$$

W is orthogonal, as the columns have unit length. For the first column we have $\sqrt{t + (1-t)|v|^2} = 1$ as column v is a normalised vector. The argument is similar for the second column, and all other columns are from the orthogonal matrix V, so they must have unit length. The columns are also orthogonal, as for the first and second column we see that the dot product $\sqrt{t(1-t)} - \sqrt{t(1-t)}|v|^2 = 0$. For the other columns, the first element is zero, and the remainder are orthonormal vectors, including the split column v, so all other columns are orthogonal. So we have now proven that $PT_1 \dots T_nP = (W^2_{i,j})$ for an orthogonal matrix W. So $T_1 \dots T_n = P^{-1}(W^2_{i,j})P^{-1}$, and as P^{-1} is also a permutation matrix, we apply observation 2, showing that $T_1 \dots T_n$ is orthostochastic. Thus, we have proven the forward implication. □

Lemma 4 (Reverse implication of Horn's lemma)**.** *Let x and y be two vectors of length n and let D be a unitary-stochastic matrix, then $x = Dy$ implies that $x \prec y$.*

Proof. If we have that $x = Dy$, for some unitary-stochastic D. Then we also have that $x^{\downarrow} = D'y$ for some other unitary-stochastic D'. As $x^{\downarrow}$, is the vector x ordered non-increasingly, so there exists a permutation matrix P such that $Px = x^{\downarrow} = PDy$. Then by observation 2, PD is also unitary-stochastic for some unitary $u' := Pu$, so we define $D' := PD = (\left|u'_{i,j}\right|^2)$. It follows that $x^{\downarrow} = D'y$, such that $x^{\downarrow}_i = \sum_{j=1}^{n} \left|u'_{i,j}\right|^2 y_j$. Now to show that to show that $x \prec y$. Summing over the first k components of $x^{\downarrow}$, we get

$$\sum_{i=1}^{k} x^{\downarrow}_i = \sum_{i=1}^{k}\sum_{j=1}^{n} \left|u'_{i,j}\right|^2 y_j = \sum_{j=1}^{n} \Big(\sum_{i=1}^{k} \left|u'_{i,j}\right|^2\Big) y_j.$$

Now let us denote $\alpha_j = \sum_{i=1}^{k} \left|u'_{i,j}\right|^2$ as the coefficient for the y_j. Using the unitary-stochastic matrix property 9, we see that

$$0 \leq \alpha_j = \sum_{i=1}^{k} \left|u'_{i,j}\right|^2 \leq \sum_{i=1}^{n} \left|u'_{i,j}\right|^2 = 1. \tag{13}$$

And looking at the sum over all α_j, we have that

$$\sum_{j=1}^{n} \alpha_j = \sum_{j=1}^{n} \sum_{i=1}^{k} \left|u'_{i,j}\right|^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} \left|u'_{i,j}\right|^2 = \sum_{i=1}^{k} 1 = k. \tag{14}$$

Putting this all together, we have that

$$\sum_{i=1}^{k} x_i^{\downarrow} = \sum_{j=1}^{n} \alpha_j y_j, \tag{15}$$

with $0 \leq \alpha_j \leq 1$ and $\sum_{j=1}^{n} \alpha_j = k$ for any $k \leq n$. Now to prove the majorization relation, we need to show that $\sum_{j=1}^{n} \alpha_j y_j \leq \sum_{i=1}^{k} y_i^{\downarrow}$, with equality for $k = n$. We do this by showing that

$$\max\{\sum_{j=1}^{n} \alpha_j y_j | 0 \leq \alpha_j \leq 1, \sum_{j=1}^{n} \alpha_j = k\} = \sum_{i=1}^{k} y_i^{\downarrow}.$$

The argument goes as follows: to maximise $\sum_{j=1}^{n} \alpha_j y_j$, we wish to maximise the coefficients of the largest components y_j. As α_j is upper bounded by one, we set α_j equal to one for the largest components y_j. Then by the constraint $\sum_{j=1}^{n} \alpha_j = k$, only the coefficients of the k largest components y_j get set to one. We note that the vector $y^{\downarrow}$ with components $y_j^{\downarrow}$ is defined as the vector y with its components set in non-increasing order. So $\sum_{i=1}^{k} y_i^{\downarrow}$ represents the sum over the k largest components of the vector y. Now we have shown that

$$\sum_{i=1}^{k} x_i^{\downarrow} = \sum_{j=1}^{n} \alpha_j y_j \leq \sum_{i=1}^{k} y_i^{\downarrow}.$$

To show equality when $k = n$, we note that in this case $\alpha_j = 1$ for all j following from the constraints, and so

$$\sum_{i=1}^{n} x_i^{\downarrow} = \sum_{j=1}^{n} y_j = \sum_{i=1}^{n} y_i^{\downarrow}.$$

Showing that $x \prec y$. □

Combining Lemma 3 and Lemma 4 we have also proven Horn's lemma, Lemma 1.

5.2 Ky Fan's maximum principle

With Horn's lemma in place, we can show Ky Fan's maximum principle, whose description and proof are based upon the following work by Nielsen [2]. This principle allows us to relate majorization to sums of vectors.

Lemma 5 (Ky Fan's maximum principle)**.** *Let A be a hermitian matrix, then the sum of the k largest eigenvalues is the maximum value of* $\mathrm{tr}(AP)$*, where P denotes the projections onto k dimensional spaces.*

We denote $\lambda(A)$ as the vector of eigenvalues of A, with components $\lambda_j(A)$ in non-increasing order, so that the lemma claims that

$$\sum_{j=1}^{k} \lambda_j(A) = \max_P \mathrm{tr}(AP).$$

Proof. The proof of this is based upon the work by Nielsen [2]. Of course we have equality when P projects onto the eigenspaces of the k eigenvectors of A with the largest eigenvalues. Now denote $P = \sum_{j=1}^{k} |g_j\rangle \langle g_j|$ for an orthonormal basis $|g_1\rangle, \ldots, |g_n\rangle$, then

$$\mathrm{tr}(AP) = \sum_{j=1}^{k} \langle g_j| A |g_j\rangle .$$

As A is hermitian, there exists an orthonormal basis $|f_1\rangle, \ldots, |f_n\rangle$ such that $A = \sum_{i=1}^{n} \lambda_i(A) |f_i\rangle \langle f_i|$. Note that the ordering of eigenvalues is non-increasing. This means that

$$\langle g_j| A |g_j\rangle = \langle g_j| (\sum_{i=1}^{n} \lambda_i(A) |f_i\rangle \langle f_i|) |g_j\rangle = \sum_{i=1}^{n} \lambda_i(A) |\langle g_j|f_i\rangle|^2.$$

This means that for the vector $(\langle g_j| A |g_j\rangle)$ with components from $j = 1$ to $j = k$, we have that

$$(\langle g_j| A |g_j\rangle) = \lambda(A)D, \tag{16}$$

where we define the unitary-stochastic matrix $D_{i,j} := |\langle g_j|f_i\rangle|^2$. To show D is unitary-stochastic, we have to show that the matrix $U_{i,j} := (\langle g_j|f_i\rangle)$ is unitary. Say i denotes the row, then the inner product of two columns j and m is $\sum_{i=1}^{n} \langle g_j|f_i\rangle \langle f_i|g_m\rangle = \langle g_j|g_m\rangle$.

Using equation 16, we can use Horn's lemma 1 and state $(\langle j| A |j\rangle) \prec \lambda(A)$, meaning that

$$\mathrm{tr}(AP) = \sum_{j=1}^{k} \langle j| A |j\rangle \leq \sum_{j=1}^{k} \lambda_j(A),$$

which is what we wanted to prove.

□

This can now be used to show that

$$\lambda(A + B) \prec \lambda(A) + \lambda(B) \tag{17}$$

for hermitian matrices A and B, as there exists a P such that $\sum_{j=1}^{k} \lambda_j(A+B) = \mathrm{tr}([A+B]P) = \mathrm{tr}(AP) + \mathrm{tr}(BP) \leq \sum_{j=1}^{k} [\lambda_j(A) + \lambda_j(B)]$ for all integers $k \leq n$, where n denotes the size of the matrices.

5.3 Density matrices and measurement

Considering density matrices θ, as they are hermitian, there exists a matrix A such that $\theta = A^*A$, and we can split the matrix $A = [A_1 A_2]$ such that

$$\theta = A^*A = \begin{bmatrix} A_1^*A_1 & A_1^*A_2 \\ A_2^*A_1 & A_2^*A_2 \end{bmatrix}.$$

We will show that

$$\lambda(\theta) \prec \lambda(A_1^*A_1) + \lambda(A_2^*A_2). \tag{18}$$

By using that $\lambda(EF) = \lambda(FE)$ for two square matrices E and F of the same size, as shown in the book "Matrix Analysis" by R. A. Horn and C. R. Johnson [9, Theorem 1.3.22]. And noting that $AA^* = A_1A_1^* + A_2A_2^*$ combined with equation 17, we conclude that

$$\begin{aligned} &\lambda(\theta) = \lambda(A^*A) = \lambda(AA^*) = \\ &\lambda(A_1A_1^* + A_2A_2^*) \prec \lambda(A_1A_1^*) + \lambda(A_2A_2^*) = \\ &\lambda(A_1^*A_1) + \lambda(A_2^*A_2). \end{aligned}$$

For any complete set of orthogonal projections P_i, we have that the density operator $\theta = \sum_{i,j=1}^{n} P_i\theta P_j$. If we choose a basis such that each projection acts as the identity on its own subspace, we can write the density matrix as

$$\theta = \begin{bmatrix} P_1\theta P_1 & \dots & P_1\theta P_n \\ \vdots & \ddots & \\ P_n\theta P_1 & & P_n\theta P_n \end{bmatrix}.$$

Then using our previous result 18 with induction, we can deduce that

$$\lambda(\theta) \prec \sum_i \lambda(P_i\theta P_i). \tag{19}$$

Keep in mind that the vectors $\lambda(P_i\theta P_i)$ are first padded with zeroes to enable summation and majorization.

This can then be applied to our measurement process to show that

$$\lambda(\theta) \prec \sum_{i=1}^{k} p_i\lambda(\theta_i'). \tag{20}$$

For the projective measurement on the ancilla, we used a complete set of orthogonal projections $P_i = |g_i\rangle\langle g_i|$, and then $(P_i \otimes I) = (|g_i\rangle\langle g_i| \otimes I)$ is also a complete set of orthogonal projections. Furthermore, for our density matrices θ and pure state ancilla $|\beta\rangle\langle\beta|$, we have that $\lambda(\theta) = \lambda(|\beta\rangle\langle\beta| \otimes \theta)$ as we have the same eigenvectors $|\psi_k\rangle$, only now expanded to $|\beta\rangle \otimes |\psi_k\rangle$, thus with the same eigenvalues. We also applied a unitary to this system, which does not influence the eigenvalues, so that $\lambda(|\beta\rangle\langle\beta| \otimes \theta) = \lambda(u(|\beta\rangle\langle\beta| \otimes \theta)u^*)$. Lastly, we need to recall our Kraus measurement $T_i\theta = a_i\theta a_i^*$, where $a_i = \sum_{j=1}^{k} \beta_j w_{i,j} =$

$(\langle g_i| \otimes \mathbf{1})u(|\beta\rangle \otimes \mathbf{1})$. Putting this all together with our previous statement 19, we see that

$$\begin{aligned}\lambda(\theta) &= \lambda(u(|\beta\rangle \langle\beta| \otimes \theta)u^*) \\ &\prec \sum_{i=1}^{k} \lambda((|g_i\rangle \langle g_i| \otimes \mathbf{1})u(|\beta\rangle \langle\beta| \otimes \theta)u^*(|g_i\rangle \langle g_i| \otimes \mathbf{1})) \\ &= \sum_{i=1}^{k} \lambda(|g_i\rangle \langle g_i| \otimes a_i\theta a_i^*).\end{aligned}$$

Now we can apply the same argument as with $\lambda(\theta) = \lambda(|\beta\rangle \langle\beta| \otimes \theta)$ to state $\lambda(|g_i\rangle \langle g_i| \otimes a_i\theta a_i^*) = \lambda(a_i\theta a_i^*)$. Furthermore, we can substitute $a_i\theta a_i^* = p_i\theta_i'$, where θ_i' represents the new density state, and $p_i = \mathrm{tr}(a_i\theta a_i^*)$ is the probability of measuring outcome i. This gives us that

$$\sum_{i=1}^{k} \lambda(|g_i\rangle \langle g_i| \otimes a_i\theta a_i^*) = \sum_{i=1}^{k} \lambda(a_i\theta a_i^*) = \sum_{i=1}^{k} p_i\lambda(\theta_i'),$$

Proving what we wanted to show.

5.4 Deriving the inequality

To obtain an inequality from this, we make use of Schur-convex functions, as explained in the article by Nielsen [2].

Definition 1. *a Schur-convex function $f()$ is a function mapping a vector to a real number. It has the property that when applied to a majorization $x \prec y$ it preserves the relation such that $f(x) \leq f(y)$.*

In that same article, they mention that the function $f(x) = \sum_{i=1}^{n} x_i^m$ is Schur convex for any $m \geq 1$, where n denotes the length of vector x with components x_i.

Applying this function to the constraint 20 above, we get that for all $m \in \mathbb{N}$ and states θ,

$$\mathrm{tr}(\theta^m) \leq \sum_{i=1}^{k} p_i \mathrm{tr}((\theta_i')^m). \tag{21}$$

We show this by first working out the left hand side

$$f(\lambda(\theta)) = \sum_{j=1}^{n} \lambda_j(\theta)^m = \mathrm{tr}(\theta^m). \tag{22}$$

Doing the same to the right-hand side, we get

$$f(\sum_{i=1}^{k} p_i\lambda(\theta_i')) = \sum_{j=1}^{n}[\sum_{i=1}^{k} p_i\lambda_j(\theta_i')]^m \leq \sum_{j=1}^{n}\sum_{i=1}^{k} p_i[\lambda_j(\theta_i')]^m,$$

where we used that the function x^m is convex for values in $[0,1]$. Keep in mind that the eigenvalues $\lambda_j(\theta'_i)$ are from density matrices and thus represent a probability, so that $0 \leq \lambda_j(\theta'_i) \leq 1$. This is the reason that the convex sum is in $[0,1]$. Switching the sums, we see that

$$\sum_{j=1}^{n}\sum_{i=1}^{k} p_i[\lambda_j(\theta'_i)]^m = \sum_{i=1}^{k}\sum_{j=1}^{n} p_i[\lambda_j(\theta'_i)]^m = \sum_{i=1}^{k} p_i \mathrm{tr}((\theta'_i)^m). \tag{23}$$

Now that we have worked out both sides, equation 22 and 23, applying the Schur-convex relation to equation 20 gives us the stated result 21. This inequality states that the expectation of any moment of the posterior state is larger or equal to the same moment of the prior state θ. As this inequality holds for all states θ, it also works for our random variable Θ_n for all $n \in \mathbb{N}$, as these random variables output a state. Flipping the inequality, this results in the following:

$$\mathbb{E}(\mathrm{tr}(\Theta_{n+1}^m)|\Sigma_n) \geq \mathrm{tr}(\Theta_n^m), \tag{24}$$

where on the left-hand side we specified in the expectation that we are only considering the expectation of the $(n+1)^{\text{th}}$ measurement, as the first n measurements are known.

5.5 Submartingales and their convergence theorem

Equation 24 is a very important result for the proof of purification, but to effectively use it in the proof, we first cover the context of the notation. This is because equation 24 allows us to describe the stochastic processes $\mathrm{tr}(\Theta_n^m)$ as submartingales. We now cover what submartingales are and give a convergence theorem, to then show that this applies to $\mathrm{tr}(\Theta_n^m)$. This section is based upon the book 'Probability with Martingales' by David Williams [10], with chapter 10 covering martingales and chapter 11 the convergence theorem.

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we define the filtration $\{\mathcal{F}_n : n \geq 0\}$ as an increasing family of sub-σ-algebras of $\mathcal{F}$, $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$. This can be interpreted such that $\mathcal{F}_n$ contains all possible outcome sequences, where we know the first n outcomes. Then a stochastic process $(X_n)_{n\in\mathbb{N}}$ is a martingale relative to the filtration $\mathcal{F}_n$ and measure $\mathbb{P}$ when:

- (i) X is adapted, meaning that X_n is $\mathcal{F}_n$-measurable for all n
- (ii) $\mathbb{E}(|X_n|) < \infty, \forall n$
- (iii) $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n, \text{a.s.} \forall n.$

We subsequently define super- and submartingales similarly, but with (iii) changed to $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \leq X_n,$ a.s. $\forall n$ and $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n,$ a.s. $\forall n$ respectively.

Now for Doob's 'Forward' Convergence Theorem: Let X be a supermartingale bounded in $\mathcal{L}^1$: $\sup \mathbb{E}(|X_n|) < \infty$. Then $lim_{n\to\infty} X_n$ exists and is finite.

For the proof of this theorem we refer back to the book 'Probability with Martingales' by David Williams [10] page 109. We note that by definition if X is a supermartingale, then $-X$ is a submartingale, and every martingale is both a super- and a submartingale. This means that the previous theorem also holds for submartingales and martingales.

Looking back at the stochastic process Θ_n^m, we see that the moments, which we note as $M_n^{(m)} := \mathrm{tr}(\Theta_n^m)$, are submartingales. The filtration in this case is given by the sets Σ_n, and $\mathrm{tr}(\Theta_n^m)$ is Σ_n measurable for all $n, m \in \mathbb{N}$. For point two, these moments take values between $[0, 1]$, and for point three we refer to the inequality we gathered from the works of Nielsen, equation 24.

This then directly implies that we can use the convergence theorem on the submartingales, so the moments $M_n^{(m)}$ converge to some finite random variable $M_\infty^{(m)}$.

6 Purification

Now we have all the tools necessary to prove Theorem 1. Following the approach taken by Maassen and Kümmerer [3] we first state and prove a related lemma before showing how this lemma proves the purification theorem.

Lemma 6 (From [3])**.** *Let $(\Theta_n)_{n\in\mathbb{N}}$ be a quantum trajectory with a process as described in Section 4.1 and an initial state θ_0. Then we have either of the following options occur:*

- *(i) The quantum trajectory $(\Theta_n)_{n\in\mathbb{N}}$ purifies almost surely.*
- *(ii) There exists a mixed state density matrix $\rho \in \mathcal{S}$ such that*

$$\forall_{i=1,\dots,k} \exists_{\lambda_i \geq 0} : a_i \rho a_i^* \sim \lambda_i \rho$$

For the proof of this lemma, we make use of the following function for each $m \in \mathbb{N}$

$$\delta_m : \mathcal{S} \to [0, \infty) : \theta \to \sum_{i=1}^{k} \mathrm{tr}(a_i \theta a_i^*)[\mathrm{tr}([\frac{a_i \theta a_i^*}{\mathrm{tr}(a_i \theta a_i^*)}]^m) - \mathrm{tr}(\theta^m)]^2. \tag{25}$$

This represents the square difference between the expectation of the moment after measurement and the moment before the measurement. Using the submartingales, we can write

$$\delta_m(\Theta_n) = \mathbb{E}[(M_{n+1}^{(m)} - M_n^{(m)})^2 | \Sigma_n].$$

In the proof of lemma 6 we make use of the following property of this function.

Lemma 7. *Let $\delta_m(\Theta_n)$ represent the function as described above. Then we have that* $\lim_{n\to\infty} \mathbb{E}(\delta_m(\Theta_n)) = 0$.

Proof. To show this is the case it suffices to show that the function is square summable

$$\forall_{m\in\mathbb{N}} : \sum_{n=0}^{\infty} \mathbb{E}(\delta_m(\Theta_n)) \leq 1.$$

This is the case as

$$\mathbb{E}[(M_{n+1}^{(m)} - M_n^{(m)})^2|\Sigma_n] = \mathbb{E}[(M_{n+1}^{(m)})^2|\Sigma_n] - 2\mathbb{E}[(M_{n+1}^{(m)} M_n^{(m)})|\Sigma_n] + \mathbb{E}[(M_n^{(m)})^2|\Sigma_n].$$

Then $\mathbb{E}[(M_n^{(m)})|\Sigma_n] = M_n^{(m)}$, as if we are given all n outcomes, we know the n^{th} moment. Applying this to the negative term, we get

$$\mathbb{E}[(M_{n+1}^{(m)} - M_n^{(m)})^2|\Sigma_n] = \mathbb{E}[(M_{n+1}^{(m)})^2|\Sigma_n] - 2M_n^{(m)}\mathbb{E}[(M_{n+1}^{(m)})|\Sigma_n] + (M_n^{(m)})^2.$$

Now using the submartingale property $\forall_{m\in\mathbb{N}} : \mathbb{E}(M_{n+1}^{(m)}|\Sigma_n) \geq M_n^{(m)}$, we have that $-2M_n^{(m)}\mathbb{E}[(M_{n+1}^{(m)})|\Sigma_n] \leq -2(M_n^{(m)})^2$. Adding the last two terms, leading to

$$\mathbb{E}[(M_{n+1}^{(m)} - M_n^{(m)})^2|\Sigma_n] \leq \mathbb{E}[(M_{n+1}^{(m)})^2|\Sigma_n] - (M_n^{(m)})^2. \tag{26}$$

Now, to show square summability, we note that

$$\forall_{m\in\mathbb{N}} : \sum_{n=0}^{\infty} \mathbb{E}[(M_{n+1}^{(m)} - M_n^{(m)})^2] = \sum_{n=0}^{\infty} \mathbb{E}(\mathbb{E}[(M_{n+1}^{(m)} - M_n^{(m)})^2|\Sigma_n]).$$

Applying the found inequality 26, we have

$$\sum_{n=0}^{\infty} \mathbb{E}(\mathbb{E}[(M_{n+1}^{(m)} - M_n^{(m)})^2|\Sigma_n]) \leq \sum_{n=0}^{\infty} \mathbb{E}(\mathbb{E}[(M_{n+1}^{(m)})^2|\Sigma_n] - (M_n^{(m)})^2).$$

Splitting the expectation, we can solve the sum

$$\begin{aligned}\sum_{n=0}^{\infty} \mathbb{E}(\mathbb{E}[(M_{n+1}^{(m)})^2|\Sigma_n] - (M_n^{(m)})^2) &= \sum_{n=0}^{\infty} \mathbb{E}((M_{n+1}^{(m)})^2) - \mathbb{E}((M_n^{(m)})^2) \\ &= \mathbb{E}[(M_\infty^{(m)})^2] - \mathbb{E}[(M_0^{(m)})^2] \leq 1,\end{aligned}$$

where $M_\infty^{(m)}$ is the random variable to which the martingales converge, and as the martingales are between zero and one, their difference is bounded by one. We have thus shown square summability

$$\forall_{m\in\mathbb{N}} : \sum_{n=0}^{\infty} \mathbb{E}[(M_{n+1}^{(m)} - M_n^{(m)})^2] \leq 1,$$

completing the proof. □

With the proof of Lemma 7 complete, we can move on to the proof of Lemma 6.

Proof of Lemma 6. To prove Lemma 6, we assume that (i) is not the case and show that we must have (ii). In this case we have that for one $m \geq 2$, $\mathbb{E}(M_\infty^{(m)}) = \mu_m < 1$. By Observation 1 this means we are considering a mixed state, and thus by Remark 1, we have that for all $m \geq 2$, $\mathbb{E}(M_\infty^{(m)}) = \mu_m < 1$. To this end consider the following set for any $n \in \mathbb{N}$.

$$A_n := \{\omega \in \Omega | M_n^{(2)} \leq \frac{\mu_2 + 1}{2}\}.$$

By Nielsen's inequality 24 for $m = 2$, if we take the expectation on both sides, we must have that $\mathbb{E}(M_{n+1}^{(2)}) = \mathbb{E}(\mathrm{tr}(\Theta_{n+1}^2)) \geq \mathbb{E}(\mathrm{tr}(\Theta_n^2)) = \mathbb{E}(M_n^{(2)})$. Meaning that the expectation $\mathbb{E}(M_n^{(2)})$ is a non-decreasing sequence, such that:

$$\mu_2 = \mathbb{E}(M_\infty^{(2)}) \geq \mathbb{E}(M_n^{(2)}).$$

This leads to the following string of inequalities:

$$\mu_2 \geq \mathbb{E}(M_n^{(2)}) \geq \mathbb{E}(M_n^{(2)} \mathbf{1}_{M_n^{(2)} > \frac{\mu_2+1}{2}}) \geq \frac{\mu_2 + 1}{2} \mathbb{P}(M_n^{(2)} > \frac{\mu_2 + 1}{2}) = \frac{\mu_2 + 1}{2}(1 - \mathbb{P}(A_n)),$$

from which we conclude

$$\mathbb{P}(A_n) \geq \frac{1 - \mu_2}{1 + \mu_2}. \tag{27}$$

We use this shortly. First, A_n is a Σ_n measurable set, as it consists of moments, which we have shown are Σ_n measurable. This means A_n consists of unions of cylinder sets. On these cylinder sets $\Lambda_{i_1,\dots i_n}$, the random variable Θ_n is constant, which we define as $\Theta_n(i_1, \dots i_n)$. Then the expectation of the function δ_m defined before is certainly greater or equal to the expectation only on the set A_n.

$$\frac{1}{\mathbb{P}(A_n)} \sum_{\Lambda_{i_1,\dots i_n} \subset A_n} \mathbb{P}(\Lambda_{i_1,\dots i_n}) \sum_{m=1}^{d} \delta_m(\Theta_n(i_1, \dots i_n)) \leq \frac{1}{\mathbb{P}(A_n)} \sum_{m=1}^{d} \mathbb{E}[\delta_m(\Theta_n)].$$

As the left 'expectation' is a weighted average, there must be one $\rho_n = \Theta_n(i_1, \dots i_n)$ such that the inequality also holds,

$$\frac{1}{\mathbb{P}(A_n)} \sum_{m=1}^{d} \delta_m(\rho_n) \leq \frac{1}{\mathbb{P}(A_n)} \sum_{m=1}^{d} \mathbb{E}[\delta_m(\Theta_n)].$$

Furthermore, as $\frac{1}{\mathbb{P}(A_n)} \geq 1$, this factor can be removed from the left also. Finally, making use of equation 27 on the right:

$$\sum_{m=1}^{d} \delta_m(\rho_n) \leq \frac{1+\mu_2}{1-\mu_2} \sum_{m=1}^{d} \mathbb{E}[\delta_m(\Theta_n)] \tag{28}$$

As this holds for every n, we have a sequence $(\rho_n)_{n\in\mathbb{N}}$, which lies in the compact set of states

$$\{\theta \in \mathcal{S} | \mathrm{tr}(\theta^2) \leq \frac{1+\mu_2}{2}\}. \tag{29}$$

Meaning that we have a limit point ρ of the sequence. Using this limit point combined with Lemma 7 to equation 28, we get

$$\sum_{m=1}^{d} \delta_m(\rho) \leq \lim_{n\to\infty} \frac{1-\mu_2}{1+\mu_2} \sum_{m=1}^{d} \mathbb{E}[\delta_m(\Theta_n)] = 0. \tag{30}$$

Next, using that $\delta_m(\rho) \geq 0$, because it is a squared difference, we see that all elements of the sum $\sum_{m=1}^{d} \delta_m(\rho)$ are non-negative. Then we must have that the individual elements go to zero, so

$$\delta_m(\rho) = 0.$$

Furthermore, from the compact set 29, we get that

$$\mathrm{tr}(\rho^2) \leq \frac{1+\mu_2}{2} < 1,$$

meaning that ρ is a mixed state. If we now substitute the definition of the function 25 into equation 6, we see that

$$\mathrm{tr}(a_i \rho a_i^*)[\mathrm{tr}([\frac{a_i \rho a_i^*}{\mathrm{tr}(a_i \rho a_i^*)}]^m) - \mathrm{tr}(\rho^m)]^2 = 0.$$

So we have either $\mathrm{tr}(a_i \rho a_i^*) = 0$, proving (ii) for $\lambda_i = 0$, or $\mathrm{tr}([\frac{a_i \rho a_i^*}{\mathrm{tr}(a_i \rho a_i^*)}]^m) = \mathrm{tr}(\rho^m)$ for all $m \in \mathbb{N}$. As all moments are equal, and we know that the moments are also equal to the sums of m^{th} powers of the eigenvalues $\mathrm{tr}(\rho^m) = \sum_i \lambda_i(\rho)^m$, for eigenvalues $\lambda_i(\rho)$ of ρ. We must have that the eigenvalues of ρ and $\frac{a_i \rho a_i^*}{\mathrm{tr}(a_i \rho a_i^*)}$ are equal. As both are density matrices they must be unitarily diagonalizable with the same diagonal matrix and thus we must have unitary equivalence, again proving (ii) for $\lambda_i = \mathrm{tr}(a_i \rho a_i^*)$. Showing that if option (i) is not the case, we must have option (ii), proving the Lemma. □

With Lemma 6 in place, proving Theorem 1 requires us to show that point (ii) in Lemma 6 is equivalent to saying that there exists a projection $p \in \mathcal{A}$ with a dimension of at least 2, such that

$$\forall_{i\in\{1,\ldots,k\}} \exists_{\lambda_i \geq 0} : p a_i^* a_i p = \lambda_i p.$$

This shows that a_i acts as a scaled isometry restricted to the space of p, when acting on the space of p. What we can gather from this is that the measurement

process scales the part of the state that is in this projection space. The part of the state that is outside this projection space, we know less about. This part might purify and get sent to this projection space. However, we do know that we cannot gather any more information from this projection space. This notation specifically highlights the space which we cannot purify, which we cover in more detail later.

Lemma 8. *Let ρ be a mixed state density matrix $\rho \in \mathcal{S}$ such that*

$$\forall_{i=1,\ldots,k} \exists_{\lambda_i \geq 0} : a_i \rho a_i^* \sim \lambda_i \rho.$$

Then there exists a projection $p \in \mathcal{A}$ with a dimension of at least 2, such that

$$\forall_{i \in \{1,\ldots,k\}} \exists_{\lambda_i \geq 0} : p a_i^* a_i p = \lambda_i p.$$

Proof. We make use of the fact that the support projection of the mixed state density matrix is a projection with a dimension of at least two.

Starting with the following statement, if p is the support projection of matrix x, then we have that

$$\det_{pos}(x) = \det_{pos}(\lambda p) \Rightarrow \mathrm{tr}(xp) \geq \mathrm{tr}(\lambda p), \tag{31}$$

with equality if and only if $x = \lambda p$. The term $\det_{pos}(x)$ represents the positive determinant of x, the product of the strictly positive eigenvalues of x. This statement is derived from the Arithmetic Mean-Geometric Mean inequality $\frac{x_1 + \cdots + x_n}{n} \geq (\Pi_{j=1}^n x_j)^{\frac{1}{n}}$, where there is equality, and thus a minimum, if and only if $x_i = (\Pi_{j=1}^n x_j)^{\frac{1}{n}}$ for all i. In other words, a sum of n numbers with the same product is minimal if and only if all numbers have the same value.

Let p be the support projection of the density matrix ρ. We wish to apply statement 31 to prove Lemma 8, so we start by showing that p is the support projection of $pa_i^* a_i p$. By the relation $a_i \rho a_i^* \sim \lambda_i \rho$, we must have that the image of $a_i \rho a_i^*$ must have the same dimension as the image of $\lambda_i \rho$. This means that a_i must map the entire image of ρ to some space with the same dimension. Then a_i^*, being the adjoint, must send this space to the entire image of ρ. Thus we must have that $pa_i^* a_i p$ maps vectors to the entire image of p, as any vector not in this space gets sent to zero by p. For this reason, p is the support projection of $pa_i^* a_i p$.

Now to show that the determinants are equal, denote the polar decomposition of $a_i p$ as $v_i \sqrt{p a_i^* a_i p}$, with v_i a partial isometry. Then by the relation $a_i \rho a_i^* \sim \lambda_i \rho$, we have that

$$\begin{aligned} \det_{\mathrm{pos}}(\lambda_i p)\det_{\mathrm{pos}}(\rho) = \det_{\mathrm{pos}}(\lambda_i \rho) = \det_{\mathrm{pos}}(a_i \rho a_i^*) = \\ \det_{\mathrm{pos}}(a_i p \rho p a_i^*) = \det_{\mathrm{pos}}(v_i \sqrt{p a_i^* a_i p} \rho \sqrt{p a_i^* a_i p} v_i^*) = \\ \det_{\mathrm{pos}}(p a_i^* a_i p \rho) = \det_{\mathrm{pos}}(p a_i^* a_i p)\det_{\mathrm{pos}}(\rho), \end{aligned}$$

where in the second-to-last equality we used that the partial isometries can be removed as they do not change the positive determinant; their purpose is rotation. In this same step, as p is the support projection of $pa_i^*a_ip$, it is also the support projection of $\sqrt{pa_i^*a_ip}$. Then as $\sqrt{pa_i^*a_ip}$ and ρ have the same support projection p, the positive determinant is the normal determinant over the image of the support projection p, and thus we can make use of the multiplicativity of the determinant. As $\det_{\text{pos}}(\rho) > 0$, we conclude that $\det_{\text{pos}}(\lambda_i p) = \det_{\text{pos}}(pa_i^*a_ip)$. Now we can apply statement 31 to obtain that

$$\text{tr}(pa_i^*a_ip) \geq \text{tr}(\lambda_i p), \tag{32}$$

where we used the projection property $p^2 = p$. This result holds for all $i \in \{1, \ldots, k\}$. Now, to prove the lemma, we want to show equality of the trace. Using the relation $a_i\rho a_i^* \sim \lambda_i\rho$, we have that

$$\sum_{i=1}^{k} \lambda_i = \sum_{i=1}^{k} \text{tr}(\lambda_i\rho) = \sum_{i=1}^{k} \text{tr}(a_i\rho a_i^*) = \text{tr}(\rho \sum_{i=1}^{k} (a_i^*a_i)) = \text{tr}(\rho) = 1,$$

where we used that $\sum_{i=1}^{k}(a_i^*a_i) = \mathbf{1}$, as described in section 3.2. This means that

$$\text{tr}(p) = \sum_{i=1}^{k} \text{tr}(pa_i^*a_i) = \sum_{i=1}^{k} \text{tr}(p^2a_i^*a_i) = \sum_{i=1}^{k} \text{tr}(pa_i^*a_ip),$$

where after the second equality we again used the projection property $p^2 = p$. Combining this result with the inequality 32, we get

$$\text{tr}(p) = \sum_{i=1}^{k} \text{tr}(pa_i^*a_ip) \geq \sum_{i=1}^{k} \text{tr}(\lambda_i p) = \text{tr}(p).$$

Showing equality of the sums, combined with the inequality 32, we must have equality of the individual terms,

$$\text{tr}(pa_i^*a_ip) = \text{tr}(\lambda_i p).$$

Then by the if and only if relation of statement 31, we can then conclude that

$$pa_i^*a_ip = \lambda_i p.$$

□

Thus we have also proven Theorem 1 as it follows directly from the combination of Lemma 6 and Lemma 8.

6.1 Extension

Following the work of Maassen and Kümmerer [3], Lemma 8 can be further extended over multiple measurements; that is, there exists a projection $p \in \mathcal{A}$ with a dimension of at least 2, such that for any $l \in \mathbb{N}$

$$\forall_{i_1,\ldots,i_l \in \{1,\ldots,k\}} \exists_{\lambda_{i_1,\ldots,i_l} \geq 0} : p a_{i_1} \ldots a_{i_l} a_{i_l}^* \ldots a_{i_1}^* p = \lambda_{i_1,\ldots,i_l} p. \tag{33}$$

As we can consider $a_{i_1} \ldots a_{i_l}$ to be one measurement operator. Maassen and Kümmerer then call the projections satisfying this 'dark projections' and the range a 'dark subspace'.

Maassen and Kümmerer [3] then explain the following regarding these dark projections. Suppose p is a dark projection, and let $a_i p$ have the polar decomposition $v_i \sqrt{p a_i^* a_i p} = \sqrt{\lambda_i} v_i p$, with v_i a partial isometry. If we then take the projection $p_i' = v_i p v_i^*$, we get that

$$\begin{aligned} \lambda_i p_i' a_{i_1}^* \ldots a_{i_l}^* a_{i_l} \ldots a_{i_1} p_i' &= \lambda_i v_i p v_i^* a_{i_1}^* \ldots a_{i_l}^* a_{i_l} \ldots a_{i_1} v_i p v_i^* \\ &= v_i p a_i^* a_{i_1}^* \ldots a_{i_l}^* a_{i_l} \ldots a_{i_1} a_i p v_i^* \\ &= \lambda_{i,i_1,\ldots,i_l} p_i'. \end{aligned}$$

This means that if $\lambda_i \neq 0$, then p_i' is also a dark projection with constants $\lambda'_{i_1,\ldots,i_l} = \frac{\lambda_{i,i_1,\ldots,i_l}}{\lambda_i}$.

What we see is that using the polar decomposition of $a_i p$, a measurement of i results in the part of the state that is in the dark subspace of p is sent to some space with the same dimension determined by the partial isometry v_i. A projection of this new dark space is p_i', which is itself a dark projection, meaning that a measurement from here leads to this new space being sent to some same dimensional dark subspace again, as described before. The probability of being sent to these specific dark subspaces overlaps with the respective measurement output, so have probabilities given by λ_i. If we have that the projection is of dimension one, then we have purification.

7 Conclusion

In conclusion, we have shown that for a perfect indirect measurement, a quantum trajectory follows one of two options. The first is that the trajectory purifies almost surely, meaning that after performing repeated indirect measurements, we eventually expect the outcome state to be pure. The alternative is that the measurement operation is not capable of extracting all the information from the state. The consequence of this is that after purifying as much as possible, the state switches through different subspaces of the same dimension with every measurement. With these measurements, no more information is revealed, so the state remains mixed. Further work could check the effects of repeated measurements with imperfect measurements; this would give the results wider use. Another follow-up could look at the rate at which trajectories purify and on which factors this depends. Similarly, in the case that the trajectory does not purify, at what rate the trajectory is confined to these dark subspaces.

References

[1] T. A. Brun, "A simple model of quantum trajectories," *American Journal of Physics*, vol. 70, no. 7, pp. 719–737, Jun. 2002, ISSN: 1943-2909. DOI: 10.1119/1.1475328. [Online]. Available: http://dx.doi.org/10.1119/1.1475328.

[2] M. A. Nielsen, "Characterizing mixing and measurement in quantum mechanics," *Physical Review A*, vol. 63, p. 022 114, 2 Jan. 2001, ISSN: 1094-1622. DOI: 10.1103/physreva.63.022114. arXiv: quant-ph/0008073 [quant-ph]. [Online]. Available: http://dx.doi.org/10.1103/PhysRevA.63.022114.

[3] H. Maassen and B. Kümmerer, *Purification of quantum trajectories*, 2005. arXiv: quant-ph/0505084 [quant-ph]. [Online]. Available: https://arxiv.org/abs/quant-ph/0505084.

[4] D. J. Griffiths and D. F. Schroeter, *Introduction to Quantum Mechanics*, 3rd ed. Cambridge University Press, Aug. 2018. DOI: 10.1017/9781316995433. [Online]. Available: https://www.cambridge.org/highereducation/books/introduction-to-quantum-mechanics/990799CA07A83FC5312402AF6860311E#overview.

[5] M. Nielsen and I. Chuang, *Quantum computation and quantum information*. Cambridge University Press, 2010, ch. Quantum noise and quantum operations. [Online]. Available: https://profmcruz.files.wordpress.com/2017/08/quantum-computation-and-quantum-information-nielsen-chuang.pdf.

[6] S. Karlin and H. M. Taylor, *A first course in stochastic processes*. Gulf Professional Publishing, Apr. 1975.

[7] R. Durrett, *Probability: Theory and Examples*, 5th ed. Cambridge University Press, Jan. 2019, Online version from Duke university. [Online]. Available: https://sites.math.duke.edu/~rtd/PTE/PTE5_011119.pdf.

[8] M. A. Nielsen, "Probability distributions consistent with a mixed state," *Phys. Rev. A*, vol. 62, p. 052 308, 5 Oct. 2000, ISSN: 1094-1622. DOI: 10.1103/PhysRevA.62.052308. arXiv: quant-ph/9909020 [quant-ph]. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.62.052308.

[9] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge University Press, Apr. 2013. DOI: https://doi.org/10.1017/CBO9781139020411. [Online]. Available: https://www.cambridge.org/highereducation/books/matrix-analysis/FDA3627DC2B9F5C3DF2FD8C3CC136B48#overview.

[10] D. Williams, *Probability with Martingales*. Cambridge University Press, 1991.