Research Center for Quantum Information
Institute of Physics, Slovak Academy of Sciences

Multiple Observations of Quantum Systems

PhD Thesis

Peter Rapčan

Bratislava 2010

Fakulta matematiky, fyziky a informatiky
Univerzita Komenského

Multiple Observations of Quantum Systems

Dizertačná práca

Peter Rapčan

Študijný odbor:	4.1.2 Všeobecná fyzika a matematická fyzika
Externá vzdelávacia inštitúcia:	Centrum pre výskum kvantovej informácie Fyzikálny ústav Slovenskej akadémie vied
Školiteľ:	Prof. RNDr. Vladimír Bužek, DrSc. (CVKI, SAV)

Bratislava 2010

Abstract

Given a finite number of instances of a qudit, collectively carrying information on a unknown pure single-qudit state, that have already been measured optimally with the aim of gaining this information, can one still extract any information about the original unknown single-qudit state? Clearly, after a maximally informative measurement, the state of the system "collapses" into a post-measurement state from which the *same* observer cannot obtain further information about the original state of the system. However, the system still encodes a significant amount of information for a second observer who is (partially) unaware of the doings of the first one.

We study how a series of independent observers can obtain information about the unknown state of a system (quantified by the fidelity) when they sequentially measure it. We give closed expressions for qudit systems, when one or several qudits are available to carry information about the single-qudit state, and study the "classical" limit when a sizable number of observers can obtain (nearly) complete information on the system. In addition to the case when all observers perform most informative measurements we study the scenario when a finite number of observers estimate the state with equal fidelity and the scenario when they perform effectively the same measurement tailored so that the quality of the last observer's guess is as high as possible.

Abstrakt

Uvažujme systém pozostávajúci z konečného počtu elementárnych d-hladinových kvantových systémov (qu*d*itov), kolektívne nesúcich informáciu o neznámom čistom jedno–qu*d*itovom stave. Ak na nich bolo vykonané optimálne meranie s cieľom získať túto informáciu, je ešte možné z tohto systému extrahovať informáciu o pôvodnom neznámom jedno–qu*d*itovom stave? Očividne, po vykonaní maximálne informatívneho merania nastáva "kolaps" stavu meraného systému do stavu, z ktorého *ten istý* pozorovateľ nemôže byť schopný extrahovať žiadnu dodatočnú informáciu o pôvodnom stave systému. Avšak pre iného pozorovateľa, ktorý nepozná (niektoré) detaily merania prvého pozorovateľa, môže systém stále túto informáciu obsahovať.

Skúmame do akej miery môžu navzájom nezávislí pozorovatelia, jeden po druhom merajúci kvantový systém, získať informáciu o jeho neznámom stave (kvantifikovnú fidelitou). Uvádzame explicitné výrazy pre prípad, keď na uloženie informácie o jedno-qu*d*itovom stave je k dispozícii jeden alebo viac qu*d*itov a vyšetrujeme "klasickú limitu," t.j. situáciu, kedy je veľký počet pozorovateľov schopný získať (takmer) úplnú informáciu o tomto jedno-qu*d*itovom stave. Okrem prípadu, keď všetci pozorovatelia vykonávajú maximálne informatívne merania sa zaoberáme aj situáciou, kedy konečný počet pozorovateľov robí estimáciu stavu s rovnakou fidelitou a situáciou, kedy je estimácia každého pozorovateľa vykonávaná rovnakým meraním navrhnutým tak, aby bola kvalita odhadu stavu posledného pozorovateľa čo najvyššia.

Table of contents

Chapter 1

Introduction

In our everyday lives we observe what we refer to as classical, macroscopic, objects. Although these are composed of sub-systems which themselves may, typically, be still considered classical, we are, most of the times, not interested in the detailed configuration of these within the larger object. Instead, we assign certain properties – values of observables like velocity, position, magnetisation direction, etc. – to the whole object without caring at all about its constituent structure even though, often, we could be in position to assign the property of interest to the constituent sub-objects as well, if we wanted to.

We are assigning a single value of a property (perhaps an average value of the observable over all the sub-objects, a value of the observable at some special "part of the object", like its center of mass, etc.) to a whole class of configurations of the sub-objects. The dynamics of the macroscopic object could, in principle, consist of any change of configuration of any of the constituent parts. However, under the most common circumstances when we speak of a value of some observable of a *specific* object (like the magnetization direction of a particular magnet, or the position of a particular rock), what we have in mind is that the internal structure of the object (like that of the domains within a magnet, or the internal structure of the rock) is fixed and different values of the observable (magnetization direction, position) are ascribed to states related by "rigid" transformations of the macroscopic object as such.

Let us imagine measuring some observable of such a macroscopic physical object. That is, determining which of the possible "rigidly transformed" configurations is actually taking place. Further, let us imagine the observable under consideration is, for instance, the direction of magnetization of a magnet.

Being an observable of a classical system, one can, at least in principle, measure, or observe, the magnetization direction with arbitrary precision. Moreover, the measurement may be performed, at least in principle, without disturbing (the relevant properties of) the state of the system – in this case the magnetization direction. This implies that a successive observer will be able to perform the (same) measurement on the once-measured system and obtain the estimation of the observable with the same precision as the first observer. The same will hold for as many observers, who are ignorant of the previous observations' outcomes, and thus need to observe, i.e. measure, the system for themselves, as one wishes.

On the other hand, the situation in quantum mechanics of "small" objects is quite different. Consider a spin-1/2 particle in a pure state – the smallest quantum-mechanical "analogue" of a magnet as an object with a direction that is, in an analogous way, special with respect to the object's behavior in a magnetic field (the direction of the magnetization vector or the spin vector, respectively). Clearly, trying to estimate this direction – a particular pure state $\psi\colon = |\psi\rangle\langle\psi| \in \mathcal{S}(\mathcal{H}_2)$, in general leads to different measurement outcomes if the observer is not lucky to, by chance, guess the correct (or the opposite) orientation of a probing magnetic field, which he won't be able to do over many runs of the same experiment with a randomly oriented spin to measure. Thus, on average, the observer's precision of measuring the spin direction will not reach the one available when determining the direction of magnetization of the classical magnet. An obvious way to increase the observer's precision is to have each experiment run performed on many, say N, copies of the same spin.

Moreover, if the spin leaves a measuring apparatus, e.g. a filter, a second observer may estimate the post-measurement state's spin direction. Clearly, he may be at most as successful as the first observer has been in determining the original spin's direction, given the first observer has measured as precisely as he could. Also, the second observer may be able to conclude something about the original spin's direction. Assuming the first observer's apparatus does not unnecessarily disturb the post-measurement state (or disturbs it in a way known to the second observer) and the apparatus is (partially) known to him, the second observer may safely, according to the quantum theory, exclude some directions of the original spin. From the rest of the possible directions, some directions will be more probable to have led to the observed measurement outcome than others. The precision of the second observer's estimates of the original spin direction will be worse than the first observer's precision, but it will be better that random guessing. A third observer's situation will be similar but yet worse as far as precision of determining the original spin's direction is concerned.

Could two or, say, K observers, each measuring their predecessor's post-measurement state, reach the precision of the first observer? Again, an obvious way to achieve this is to have K copies of the original spin, and have each observer measure one of the copies and pass the rest (plus the ones already measured) to the next observer. In fact, as a post-measurement state even after a "most-informative" measurement is allowed to contain a relevant average overlap with the pre-measurement state (for a spin-1/2, cf. [7]), in such a case the second observer could even outperform the first one by measuring his unmeasured copy and the one already measured, without harming the next observers' yet undisturbed copies. The third one could then outperform both the first and the second, an so forth.

Certainly, producing $N \times K$ copies of the quantum system (N clean copies per observer to measure and $N(K-1)$ to pass) would then, with N sufficiently large, lead to the situation possessing a feature of a classical magnet's magnetization direction measurement – the possibility of an arbitrary (for $N \to \infty$) measurement precision and availability of the same precision to all K observers. Should the observers measure their clean copies plus the ones already measured, all would reach at least some fixed but arbitrary (for $N \to \infty$) precision and the precision would increase with the tally number, k, of the observer, the last observer achieving the largest precision. Finally, should the observers be "greedy" and each of them measure all the copies (one by one), arbitrary (for $N \to \infty$) measurement precision would be available to the first observer and the precision would decrease for the other observers as $k \to K$ (with $N \to \infty$ it would still be arbitrary for any finite K, though).

The above naive ways to mimic the *recyclability*, or re-usability with respect to (reasonable) measurements, of classical systems, using truly quantum systems, work but certainly are not the best that could be done. In the present Thesis we develop a picture of how exactly the above classical-like recyclability emerges in the quantum world. Namely, we look at the dependence of the ultimate achievable "precision" of determining the complete[1.1] description of a pure d-dimensional elementary quantum system (a pure qu*d*it), encoded in the state space of N qu*d*its (a signal state), on the number of successive observers who have already measured the *same* signal quantum system and on the size, N, of the signal quantum system. We are *not* interested in the trivial scenario, where a classical-like recyclability is achieved simply by using a distinguishable alphabet of orthogonal states from the Hilbert space of the signal state, known (perhaps probabilistically) to the observers. Therefore, we demand that knowledge of an observer about actions of the predecessing observers and of the preparer is always up to a rigid unitary $g^{\otimes N}, g \in SU(d)$.

The present Thesis is organized as follows. In Chapter 2 we collect basic mathematical knowledge relevant for the rest of the the preset work. The two most important mathematical objects introduced are two levels of description of a measurement – a generalized observable, POVM, and a quantum instrument. While a POVM provides the most general description of a strategy for an estimation of a physical parameter, a (compatible) instrument adds a description of the most general transformation rule from pre- to post-measurement states, compatible with a POVM which is also included in the instrument-description. In Chapter 3 we introduce the problem of optimal extraction of information from families of signal states that are invariant under operations from the representation space of a symmetry group. We present the covariant-measurement (covariant-POVM) approach, which is always among the optimal ones, as well as the approach utilizing measurements with finite, minimal, number of outcomes. We in particular elaborate on the case of a direction information (or, more generally, a pure qubit state) being carried by N parallel spin-1/2 systems (or qubits).

The original part of this presentation – the problem of recycling, or re-using, (qu*d*it) quantum information in a sequence of independent observations – is contained in Chapter 4. The Chapter begins with an introductory Section, 4.1, giving a short motivation and two simple examples of sequential measurements of a single spin-1/2 and a system consisting of two copies of a spin-1/2 via (minimal) projective measurements. We continue by Section 4.2 where the problem we consider is stated more precisely. In Section 4.3 we show that, due to the limited (up to a rigid unitary) knowledge of an observer about actions of the predecessing observers and of the preparer, we may restrict our attention to encodings and measurement apparata (quantum instruments) which are covariant with respect to the $g \mapsto g^{\otimes N}$ representation of SU(d), while still being guaranteed to reach the (unrestricted) optimal encoding/multiple measurements performance within this restricted set. Thus, while many optimal-estimation-problem papers simply start with the assumption that the parameter to be estimated is encoded covariantly, we provide a clear justification for working with this restricted set in our case, while not limiting the actual encodings (as described by their executor with unlimited knowledge of the encoding process) to be necessarily covariant.

Subsequently we consider two scenarios of the multiple-observations problem. In the first one which we refer to as "greedy" scenario, analyzed in Section 4.4, each of the observers, wishing to access the single-qu*d*it information, proceeds so that the fidelity of his estimate is maximized. We provide expressions for the kth observer's maximal average fidelity as a function of k and of the number of qu*d*its, N, encoding the single-qu*d*it to estimate. For general d we restrict ourselves to the finite ensemble case, i.e. symmetric, or parallel, encoding of a pure qu*d*it into the state space of N qu*d*its. For $d = 2$ we evaluate the best achievable fidelity of a kth observer's estimate without the symmetric encoding restriction. The Section concludes with a brief excursus to a more information-theory based approach to the problem in the simplest case.

1.1. up to an overall phase

In Section 4.5 we discuss the second, weak-measurements, scenario, where the observers optimize their measurements to pursue goals different from, in the first place, mere maximization of the quality of their own guesses. In particular, we study the case where K observers estimate the encoded qu*d*it state with equal, but maximal, fidelity (equalitarian strategy, Subsection 4.5.1) and the case where the observers use the same measurement apparatus such that the quality of the last observer's estimate is maximized (Subsection 4.5.4). In both cases the measurements performed are weak, i.e. in general not extracting all of the extractable information which enabled less disturbance to be undergone by the measured state. In the former case we quantify how the measurements approach more and more the ones from the "greedy" scenario with increasing tally number, k, until the last observer performs a "greedy" measurement as well as the maximal achievable, equal, fidelity of the observers' guesses. In the latter case we show there exists, and we calculate, the optimal "strength" of the measurement to be performed by all the observers as well as the observers' fidelities achieved.

Chapter 4 concludes with a discussion of the results of Section 4.4, which is the contents of Section 4.6. The scaling of the system size, N, necessary to provide mutually consistent observations to $k \to \infty$ successive independent observers, i.e. a classical-like recyclability of a quantum system, is discussed (Subsection 4.6.1). Finally, in Subsection 4.6.2 we give a simple operational interpretation of the figure of merit of the quality of estimation, the estimation fidelity, of a kth observer – one which is related to the problem of longevity of a directional reference frame, previously studied in the literature. A few suggestions for directions of related future research are given.

The presentation concludes with Appendices. Several more technical calculations of Chapter 4, are presented in Appendices A to E. Appendix F collects definitions of mathematical objects referred to throughout the present Thesis, which are not given within the main text, in order to make the presentation as self-contained as possible.

Chapter 2
Preliminaries

2.1 Elements of probability theory

In this section we will briefly summarize the basic concepts of the probability theory. The material is based on books [54] and [20].

The central object of the probability theory is the sample space (probability space) $(\Omega, \mathcal{F}, p)$ (Def. F.25). Ω is the set of elementary events. A set of all possible events forms a σ-algebra $\mathcal{F}(\Omega) \subset P(\Omega)$ (Def. F.20), where $P(\Omega)$ is the potential set (power set) – the set of all possible subsets of Ω. Finally, p is a probability (measure) (Def. F.22).

As an example, let us consider toss of a fair dice. In this case $\Omega = \{1, 2, 3, 4, 5, 6\}$. An admissible set of possible events $\mathcal{F}(\Omega)$ is, for instance, $P(\Omega)$. For a fair dice, we have $p(i) = 1/6$ ($i = 1, ..., 6$, the remaining probabilities being defined by the condition $p(\Omega) = 1$). On the other hand, if we were interested only in the parity of the toss, the set of all possible events $\mathcal{F}'(\Omega)$ would read $\{\varnothing, \{1, 3, 5\}, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\}$). In the latter case, we could have defined a new set Ω' of elementary events $\Omega' = \{\text{even number}, \text{odd number}\}$ with $\mathcal{F}(\Omega') = \mathcal{F}'(\Omega) \equiv P(\Omega') = \{\varnothing, \text{even}, \text{odd}, \Omega'\}$. Here, for a fair dice, we have $p'(\text{even}) = p'(\text{odd}) = 1/2$, which defines p' completely. Note that the probability measure (also called the image measure (Def. F.26)) on Ω' is induced by the probability measure on the original set Ω of elementary events.

Often we would like to compute some numerical quantities, like expectation values, variances, etc., that characterize the sample space. To do so, we need to represent events by (possibly tuples of) real numbers. This is achieved via the concept of random variable $X\colon \Omega \to \mathbb{R}$ (Def. F.30). In our dice example we could have, for instance, defined $X\colon \{\text{even}, \text{odd}\} \to \{4, 3\}$. The image measure p_X (a probability measure on $\mathbb{R}$) induced by the random variable X is called the *distribution* of the random variable X. If $\{X_\alpha\}_{\alpha \in \mathcal{A}}$ is a family of random variables such that $p_{X_\alpha} = p_{X_\beta}$ for all $\alpha, \beta \in \mathcal{A}$, then we say that the random variables X_α are *identically distributed*.

For any finite sequence $X_1, ..., X_n$ of random variables one can consider a map $(X_1, ..., X_n)\colon \Omega \to \mathbb{R}^n$. The (image) measure $p_{(X_1,...,X_n)}$ on $\mathbb{R}^n$ is called the *joint distribution* of the random variables $X_1, ..., X_n$. All properties (relevant to probability theory) of random variables can be expressed in terms of their joint distribution.

In the theory of integrals, there are three central notions: i) the measure (Def. F.21) ii) the measurable function (Def. F.29) iii) the integral. In probability theory, as a special case, we have so far introduced the corresponding notions of i) and ii), i.e. i) the probability (measure) and ii) the random variable. The object corresponding to iii), i.e the integral of a measurable real-valued function, is the expectation value (mathematical hope), Eq. (2.1).

The expectation value of a random variable is defined as

$$\bar{X} = \int_\Omega X(\omega) \mathrm{d}p(\omega) \tag{2.1}$$

The following lemma holds:

Lemma 2.1. *Let X be an integrable random variable with the distribution function (Def. F.32) F, g be a Borel measurable function, i.e. $g^{-1}(A) \in B$ for each $A \in B$, where B is a system of all Borel subsets of $\mathbb{R}$. Then $g \circ X$ is a random variable and*

$$\int_{\Omega} g \circ X(\omega) \, \mathrm{d}p(\omega) = \int_{-\infty}^{\infty} g(x) \mathrm{d}F(x), \tag{2.2}$$

if at least one of the functions $g \circ X$, g is integrable.

Taking $g(x) = x$, we have $g \circ X(\omega) = g(X(\omega)) = X(\omega)$, i.e. $g \circ X = X$. Now we can evaluate the integral (2.1) as

$$\bar{X} = \int_{\Omega} X(\omega) \, \mathrm{d}p(\omega) = \int_{-\infty}^{\infty} x \mathrm{d}F(x). \tag{2.3}$$

In the case of a discrete random variable X taking values $x_i \in \mathbb{R}$ with probabilities $p(x_i) = p(\{\omega \in \Omega; X(\omega) = x_i\})$, $i = 1, 2, \ldots$, the above integral is equivalent to

$$\bar{X} = \sum_i x_i p(x_i), \tag{2.4}$$

if and only if the series $\sum_i x_i p(x_i)$ converges absolutely.

If the distribution function F is absolutely continuous, i.e. $F(x) = \int_{-\infty}^{x} \tilde{p}(t) \mathrm{d}t$, where $\tilde{p}: \mathbb{R} \to \mathbb{R}$ is a non-negative function (probability density), then

$$\bar{X} = \int_{-\infty}^{\infty} x \, \tilde{p}(x) \mathrm{dx}, \tag{2.5}$$

provided that the function on the right-hand side is integrable. (Remark: The absolute continuity of F implies that $\tilde{p}(x) = F'(x)$ almost everywhere).

A very important concept in probability theory is that of (stochastic) independence. Consider a probability space $(\Omega, \mathcal{F}, p)$ and events $E, F \in \mathcal{F}$, such that $p(E) > 0$. The set function $p_E(F) = p(E \cap F)/p(E)$ is a probability measure on Ω called the *conditional probability* on E. It represents the probability of the event F, given the event E has occurred. If the probability of event F is the same whether or not E has occurred, then E and F are said to be independent. Thus, F is independent of E if and only if $p(E \cap F) = p(E)p(F)$; the latter condition is symmetric in E and F and makes sense even if $p(E) = 0$. The concept of independence can be extended to more than two events:

Definition 2.2. *Independence of events.*

A collection $\{E_\alpha\}_{\alpha \in \mathcal{A}}$ of events in $\mathcal{F}$ is independent *if*

$$p(E_{\alpha_1} \cap \ldots \cap E_{\alpha_n}) = \prod_{i=1}^{n} p(E_{\alpha_i}) \tag{2.6}$$

for all $n \in \mathbb{N}$ and all distinct $\alpha_1, \ldots, \alpha_n \in \mathcal{A}$.

Remark: For the events E_{α_i} to be independent it does not suffice for them to be pairwise independent.

The concept of independence can be introduced also for random variables in a natural way: A collection $\{X_\alpha\}_{\alpha \in \mathcal{A}}$ of random variables on Ω is called independent if the events $\{E_\alpha; X_\alpha := X_\alpha(E_a) \in B_\alpha\} = X_\alpha^{-1}(B_\alpha)$ are independent for all Borel sets $B_\alpha \subset \mathbb{R}$. In such a case it can be shown that

$$p_{(X_1, \ldots, X_n)} = \prod_{i=1}^{n} p_{X_i}, \tag{2.7}$$

i.e. $\{X_\alpha\}_{\alpha\in\mathcal{A}}$ is an independent set of random variables if and only if the joint distribution of any finite set of X_α's is the product of their individual distributions.

2.2 Quantum mechanics

In its minimal interpretation, quantum mechanics is a probabilistic theory. It enables us to describe experiments, i.e. propose outcomes of measurements (given some preparation and evolution) and assign probabilities (interpreted as limits of relative frequencies of measurement outcomes) to them. The term *minimal* refers to the fact that there exist other interpretations claiming that quantum mechanics is more than the above. A generally accepted interpretation beyond the minimal one has not yet been proposed, though. In this section we summarize (for our purposes the relevant part of) the mathematical formalism of quantum mechanics. The book [15] and the paper [12] served as the main sources for the summary.

2.2.1 Mathematical tools

The central mathematical object of quantum mechanics is a separable complex Hilbert space $\mathcal{H}$ which is used for the description of quantum objects. In later sections we will consider only finite-dimensional Hilbert spaces, i.e. finite-dimensional quantum systems, although some of the definitions given in this section will be valid also for the infinite-dimensional case.

A *linear operator* A on a (possibly infinite-dimensional) Hilbert space $\mathcal{H}$ is a linear map $A: D(A)\rightarrow\mathcal{H}$. For bounded operators A the domain $D(A)$ is either the whole $\mathcal{H}$ or its closed subset. Bounded linear operators $A:\mathcal{H}\rightarrow\mathcal{H}$ on a complex Hilbert space $\mathcal{H}$ form a *Banach algebra with an involution*, $\mathcal{L}(\mathcal{H})$, i.e. a linear space endowed with i) the (operator) norm $\|A\|:=\sup(\|A\psi\|:\psi\in\mathcal{H}\|\psi\|\leq 1)$ (where $\|\psi\|:=\sqrt{\langle\psi|\psi\rangle}$ is the norm induced by the scalar product $\langle\varphi|\psi\rangle=\langle\psi|\varphi\rangle^*$ in $\mathcal{H}$ (where (*) is the complex conjugation)) ii) the product $\circ$, $A\circ B=:AB$, iii) adjoint-linear (= anti-linear) involution, i.e. an operation (†): $A\rightarrow A^\dagger$, $\langle A^\dagger\psi|\varphi\rangle=\langle\psi|A\varphi\rangle$ iv) (in addition to associative-linear-algebra and Banach-space properties):

$$(A^\dagger)^\dagger\equiv A,\ (AB)^\dagger=B^\dagger A^\dagger$$
$$\|AB\|\leq\|A\|.\|B\|,\ \|A^\dagger\|\equiv\|A\|,\ \|AA^\dagger\|\equiv\|A\|^2.$$

Certain subsets of the Banach algebra $\mathcal{L}(\mathcal{H})$ are particularly important in quantum mechanics. The elements A of $\mathcal{L}(\mathcal{H})$ such that $AA^\dagger=A^\dagger A$ are called *normal* operators. The elements $A\in\mathcal{L}(\mathcal{H})$ such that $A=A^\dagger$ are called *self-adjoint*. The set of self-adjoint operators will be denoted by $\mathcal{L}_S(\mathcal{H})$. The *identity* operator $I_\mathcal{H}$ is denoted by $\mathbb{1}$, i.e. $\mathbb{1}A=A\mathbb{1}=A$ ($\forall A\in\mathcal{L}(\mathcal{H})$). If for a given A there exists $A'\in\mathcal{L}(\mathcal{H})$ such that $AA'=A'A=\mathbb{1}$, then $A'=:A^{-1}$ is called the *inverse* of A and such an A is an *invertible* operator. The subset of all invertible elements of $\mathcal{L}(\mathcal{H})$ is denoted by $\mathrm{GL}(\mathcal{H})$. Operators $U\in\mathrm{GL}(\mathcal{H})$ such that $U^\dagger=U^{-1}$ are called *unitary* and compose a subset of $\mathcal{L}(\mathcal{H})$ – the unitary (infinite-dimensional Lie) group $\mathcal{U}(\mathcal{H})$ on $\mathcal{H}$.

Given $A\in\mathcal{L}(\mathcal{H})$, one can define a *resolvent set* r of A by $r(A):=\{\lambda\in\mathbb{C}\colon (A-\lambda\mathbb{1})\in \mathrm{GL}(\mathcal{H})\}$. It is an open set of $\mathbb{C}$. Its complement $\mathrm{spec}(A):=\mathbb{C}\backslash r(A)$ is called the *spectrum* of A. The numbers $\lambda_j\in\mathbb{C}$ for which there exist nonzero vectors $\varphi_j\in\mathcal{H}$ such that

$$A\varphi_j=\lambda_j;\quad j\in J, J \text{ is an index set} \tag{2.8}$$

are called *eigenvalues* of A. Dimension of $\mathcal{H}$ spanned by all φ_j for the same λ_j of Eq. (2.8) is called the *degeneracy* of λ_j $(:=\deg(\lambda_j))$. For a self-adjoint operator $A\in\mathcal{L}_S(\mathcal{H})$ the spectrum is real, i.e. $\mathrm{spec}(A)\subset\mathbb{R}$. The self-adjoint operators with non-negative spectra are called *positive*. The set of positive operators will be denoted by $\mathcal{L}_+(\mathcal{H}):=\{A\colon \mathcal{L}_S(\mathcal{H})\ni A\geq 0 \text{ (sometimes denoted as } A>0)\}$. Denoting the set of all eigenvalues of an $A\in\mathcal{L}_S(A)$ by $\mathrm{spec}_{\mathrm{eig}}(A)$, its closure $\overline{\mathrm{spec}_{\mathrm{eig}}(A)}=:\mathrm{spec}_p(A)\subset\mathrm{spec}(A)$ is called the *pure-point spectrum*. Otherwise A has also some *continuous spectrum*. If the vectors φ_j of Eq. (2.8) form a basis in $\mathcal{H}$ then $\mathrm{spec}(A)=\mathrm{spec}_p(A)$.

The operators $P\in\mathcal{L}_+(\mathcal{H})$ such that $P=P^\dagger=P^2$ are called *projectors* (or orthogonal projections). Projectors P_1, P_2 are *mutually orthogonal* if and only if $P_1P_2=O$, where O is the zero operator. The projector onto the subspace spanned by all eigenvectors of $A\in\mathcal{L}_S(\mathcal{H})$ corresponding to the same eigenvalue λ_j is its *eigenprojector* $E_A(\lambda_j)$. The (closed) subspace $\mathcal{H}_{\lambda_j}:=E_A(\lambda_j)\mathcal{H}$ is called the *eigenspace* corresponding to the eigenvalue λ_j. Its dimension is $\dim(\mathcal{H}_{\lambda_j})=\deg(\lambda_j)$.

Important sub-algebras of the Banach algebra $\mathcal{L}(\mathcal{H})$ are its ideals (Def. F.14). For a separable $\mathcal{L}(\mathcal{H})$ there is a norm-closed ideal $\mathcal{C}(\mathcal{H})$ consisting of all *compact operators*, i.e. linear operators on $\mathcal{H}$ that map any norm-bounded subset of $\mathcal{H}$ into a norm-compact subset of $\mathcal{H}$ (for a definition of compactness of a topological space see Def. F.16; the norm induced by the scalar product in $\mathcal{H}$ induces a metric on $\mathcal{H}$ which, in turn, gives the standard topology on $\mathcal{H}$). Other important ideals in $\mathcal{L}(\mathcal{H})$, which are subsets of $\mathcal{C}(\mathcal{H})$, are the *Hilbert-Schmidt operators*, $\mathcal{T}_2(\mathcal{H})$, and its subset $\mathcal{T}(\mathcal{H})$ – the *trace-class operators* (for definitions see Def. F.36). All these sets are (as all two-sided ideals) *symmetric*, i.e. invariant with respect to the involution $(^\dagger)$ and thus each of their elements A can be decomposed into a complex-linear combination of two self-adjoint elements $B=B^\dagger$, $C=C^\dagger$, i.e. $A=B+\mathrm{i}C$ where $B=(A+A^\dagger)/2$, $C=B=(A-A^\dagger)/(2\mathrm{i})$.

Let us now characterize the ideals $\mathcal{T}(\mathcal{H})$ and $\mathcal{T}_2(\mathcal{H})$. If and only if an operator $A\in\mathcal{T}(\mathcal{H})$ is self-adjoint, then it has pure point spectrum and its eigenvalues are absolutely summable, i.e.

$$A=A^\dagger\in\mathcal{T}(\mathcal{H})\;\Leftrightarrow\; A \text{ has pure point spectrum and } \sum_{\lambda\in\mathrm{spec}_{\mathrm{eig}}(A)}\deg(\lambda)|\lambda|=:\|A\|_1<\infty. \tag{2.9}$$

This allows us to define a functional $\mathrm{Tr}(\,.\,)$ on self-adjoint operators A form $\mathcal{T}(\mathcal{H})$ defined as

$$\mathrm{Tr}(A):=\sum_{\lambda\in\mathrm{spec}_{\mathrm{eig}}(A)}\deg(\lambda)\lambda\quad (A=A^\dagger, A\in\mathcal{T}(\mathcal{H})) \tag{2.10}$$

called the *trace* of A. The trace can be extended uniquely to the whole complex space $\mathcal{T}(\mathcal{H})$ and hence (since $\mathcal{T}(\mathcal{H})$ is an ideal in $\mathcal{L}(\mathcal{H})$) to all products AB where $A\in\mathcal{T}(\mathcal{H})$ and $B\in\mathcal{L}(\mathcal{H})$. The norm $\|A\|_1$ in (2.9) is called the *trace norm*.

The Hilbert-Schmidt ideal $\mathcal{T}_2(\mathcal{H})$ is defined by

$$A\in\mathcal{T}_2(\mathcal{H})\;\Leftrightarrow\; A\in\mathcal{L}(\mathcal{H}) \text{ and } A^\dagger A\in\mathcal{T}(\mathcal{H}).$$

There is a duality between the Banach spaces $\mathcal{T}(\mathcal{H})$ and $\mathcal{L}(\mathcal{H})$ (and between $\mathcal{C}(\mathcal{H})$ and $\mathcal{T}(\mathcal{H})$) through the evaluation

$$A \rightarrow l_B(A) := \mathrm{Tr}(AB) \equiv \langle A; B\rangle, \;\; A \in \mathcal{T}(\mathcal{H}), \;\; B \in \mathcal{L}(\mathcal{H}) \quad (\text{resp. } A \in \mathcal{C}(\mathcal{H}), \; B \in \mathcal{T}(\mathcal{H})), \tag{2.11}$$

i.e. the operators B represent (through the isomorphism $B \rightarrow \mathrm{Tr}(\;.\,B)$) all continuous linear functionals l_B on the space of the operators A. In this sense $(\mathcal{C}(\mathcal{H}))^* = \mathcal{T}(\mathcal{H})$, $(\mathcal{T}(\mathcal{H}))^* = \mathcal{L}(\mathcal{H}) = (\mathcal{C}(\mathcal{H}))^{**}$ where (*) stands for the topological dual (Def. F.11).

Let us now define two operator-valued measures that we will need later.

Definition 2.3. *POVM.*

A normalized positive operator valued measure *(POVM) $E: \mathcal{F} \rightarrow \mathcal{L}(\mathcal{H})$ on a measurable space $(\Omega, \mathcal{F})$ is defined by the following properties:*

1. *$E(X) \geq O$ for all $X \in \mathcal{F}$ (positivity)*
2. *If (X_i) is a countable collection of disjoint sets in $\mathcal{F}$ then $E(\cup_i X_i) = \sum_i E(X_i)$, the series converging in the weak operator topology (σ-additivity)*
3. *$E(\Omega) = I$ (normalization)*

Sometimes (e.g. [12]), the σ-algebra $\mathcal{F}$ is required to be the set of Borel sets of Ω, i.e. set of all subsets obtained by countable unions and/or intersections of open and closed subsets of Ω), i.e. a topology on Ω has to be given.

Definition 2.4. *PVM.*

A projection operator valued measure *(PVM) $E: \mathcal{F} \rightarrow \mathcal{L}(\mathcal{H})$ is a POVM for which the following holds:*
$E(X)^2 = E(X)$ for all $X \in \mathcal{F}$.

A POVM $E: \mathcal{F} \rightarrow \mathcal{L}(\mathcal{H})$ is multiplicative, if $E(X \cap Y) = E(X)E(Y)$ for all $X, Y \in \mathcal{F}$. A POVM is multiplicative if and only if it is a PVM.

2.2.2 States and observables

Put simply, an experiment is a process of performing a measurement of some physical properties on a physical system. The physical system is prepared in, (or evolved into) some *state*, while the measurement yields an outcome that should somehow characterize (be in some correlation to) the state of the system being measured. The probabilities of possible outcomes of a measurement given any state of the physical system are described through the concept of an *observable*.

Definition 2.5. *States.*

States of a quantum system S are identified with the elements of $\mathcal{T}(\mathcal{H}_S)_1^+ =: \mathcal{S}(\mathcal{H}_S)$ (positive, trace one operators on $\mathcal{H}$).

Let us now define observables, which we can do more conveniently if we first introduce effects.

Definition 2.6. *Effects.*

An effect *E_X is an affine state functional*

$$E_X: \rho \rightarrow E_X(\rho) := p(X\,|\,\rho)$$

where $X \in \mathcal{F}$ is an element of a σ-algebra $\mathcal{F}$, and $p(X\,|\,\rho)$ is the probability of obtaining an outcome within the set X given the state ρ.

To speak of probabilities of outcomes, a measurement (its POVM) has to be specified. Then

$$p(X\,|\,\rho) = \mathrm{Tr}[E(X)\rho], \tag{2.12}$$

where E is a POVM. Due to the (topological) duality of the Banach spaces $\mathcal{T}(\mathcal{H})$ and $\mathcal{L}(\mathcal{H})$ (i.e. $(\mathcal{T}(\mathcal{H}))^* = \mathcal{L}(\mathcal{H})$) Eq. (2.11), the state functional $\mathrm{Tr}[E(X)\,.\,]$ uniquely specifies (through the isomorphism $\mathcal{T}(\mathcal{H}))^* \leftrightarrow \mathcal{L}(\mathcal{H})$: $\mathrm{Tr}[E(X)\,.\,] \leftrightarrow E(X)$) an operator $\hat{E}_X \equiv E(X)$. In the sense of this isomorphism one can identify an effect E_X with its corresponding operator $\hat{E}_X$ from the range of the POVM E (i.e. $E_X \overset{\text{isomorf.}}{=} \hat{E}_X$). Therefore one can (e.g. [13, 15]) say that the range of some POVM consists of effects.

Definition 2.7. *Observable.*
An observable *is a map assigning each "outcome" $X \in \mathcal{F}$ its associated effect*

$$E\colon X \to E_X, \qquad X \in \mathcal{F}, E_X \in (\mathcal{T}(\mathcal{H}))^*.$$

Thus, an observable is an effect-valued measure.

Through an argument analogous to the one just before the Def. 2.7, one often (e.g. [13, 15, 12]) identifies observables with POVMs. Moreover, in physics literature, a POVM (and thus also an observable) is often identified with a resolution of the identity, i.e. with a set $\{E_i\}$ of operators E_i, $O \leq E_i \leq \mathbb{1}$, such that $\sum_i E_i = \mathbb{1}$. The operators E_i are the images (under the POVM map) of some (unspecified) disjoint sets X_i from the σ-algebra $\mathcal{F}$. Usually what is meant (without saying so explicitly) is that $X_i \equiv \Omega_i \in \Omega$, $\cup_i \Omega_i = \Omega$. In addition, one assumes (formally) that $\Omega_i \equiv \{i\}$, in which case $E_i\colon = E(\{i\})$ and the above resolution of identity really defines a POVM (uniquely). However, strictly speaking, without specifying the X_i's *and* $\mathcal{F}$ the resolution of identity only specifies (a subset of) the range of a POVM (from a class of compatible POVMs). For example, consider the POVMs: E:($\{0\} \to |0\rangle\langle 0|$, $\{1\} \to |1\rangle\langle 1|$, $\varnothing \to O$, $\{0,1\} \to \mathbb{1}$), E': $(\varnothing \to O, \{0\} \to \mathbb{1})$, and E'': $(\varnothing \to O, \{1\} \to O, \{0\} \to \mathbb{1}, \{1,0\} \to \mathbb{1})$. The (most trivial) resolution of identity $\{\mathbb{1}\}$ can be viewed as a subset of ranges of, for instance, E, E', or E''. If we assume that $X_1 = 1$, then $\{\mathbb{1}\}$ can be viewed (if we restrict ourselves only to POVMs E, E' and E'') as a subset of the range of E or of E'. If we further assume that $\mathcal{F} = \{\varnothing, \{0\}\}$, only then the resolution of identity $\{\mathbb{1}\}$ uniquely defines a POVM – E'. Thus, in our trivial example, if we speak of the "POVM"=resolution of identity $\{\mathbb{1}\}$, it should be understood as the POVM E'.

The measurable space $(\Omega, \mathcal{F})$ is called the value space of the observable, since it describes the possible outcomes of E. Usually $(\Omega, \mathcal{F})$ is (a subspace of) the Borel space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, or some of its Cartesian products. If it is the case and if, moreover, E is a PVM then E uniquely determines a self-adjoint operator $A = \int_{\mathbb{R}} \imath \mathrm{dE}$ in $\mathcal{H}$, where $\imath$ is the identity function on the real line $\mathbb{R}$. The converse is also true (via the spectral theorem), i.e. each self-adjoint operator A in $\mathcal{H}$ defines a unique PVM E such that $A = \int_{\mathbb{R}} \imath \mathrm{dE}$. The induced PVM E is then denoted as E^A. Both the operator A and the (spectral) measure E^A are referred to as (sharp) observables.

Let us now state few important theorems of quantum theory. The fact that the representation of states as positive trace-one operators and "observables" as POVM's is the most general one compatible with the probabilistic structure of quantum mechanics is guaranteed by the Gleason theorem.

Theorem 2.8. *(Gleason theorem) Let $\mathcal{P}(\mathcal{H})$ be the set of projectors on $\mathcal{H}$. Let $m{:}\,\mathcal{P}(\mathcal{H})\rightarrow <0,1>$ be a generalized probability measure. If $\dim(\mathcal{H})\geq 3$, there exists exactly one state $\rho\in\mathcal{S}(\mathcal{H})$ such that $m(P)=\mathrm{tr}(\rho P)$ for all projectors $P\in\mathcal{P}(\mathcal{H})$.*

Theorem 2.9. *(a variant of the Gleason theorem for "effects") For any generalized probability measure m: $\mathcal{E}(\mathcal{H})\rightarrow\langle 0,\,1\rangle$ there exists exactly one state $\rho\in\mathcal{S}(\mathcal{H})$ such that $m(A)=\mathrm{tr}(\rho A)$ for all "effects" $A\in\mathcal{E}(\mathcal{H})$.*

Another important theorem, which guarantees that any POVM can be realized as a PVM on an extended quantum system system, is the Naimark dilation theorem [49].

Theorem 2.10. *(Naimark dilation theorem) Let $F{:}\,\mathcal{F}\rightarrow\mathcal{L}(\mathcal{H})$ be a POVM. There exists a Hilbert space $\tilde{\mathcal{H}}\supset\mathcal{H}$ and a PVM $E{:}\,\mathcal{F}\rightarrow\mathcal{L}(\tilde{\mathcal{H}})$ such that*

$$F(X)|\varphi\rangle=PE(X)|\varphi\rangle$$

holds for all $|\varphi\rangle\in\mathcal{H}$ and for every $X\in\mathcal{F}$. The operator P is the orthogonal projection of $\tilde{\mathcal{H}}$ onto $\mathcal{H}$. The PVM E is called the spectral dilation *of F.*

There exists a minimal dilation (unique up to a unitary isomorphism), where a minimal E is such that it acts on the smallest Hilbert space $\tilde{\mathcal{H}}$ containing the union of the closed subspaces $E(X)\mathcal{H}$, $X\in\mathcal{F}$. The minimal dilation $(\tilde{\mathcal{H}},E)$ of a POVM F in general does not have a direct physical interpretation. However, it is possible to construct dilations such that $\tilde{\mathcal{H}}=\mathcal{H}\otimes\mathcal{H}_0$, where $\mathcal{H}_0$ can be interpreted as some environment or a measuring apparatus. The Naimark theorem can be then reformulated in the following form

Theorem 2.11. *(Naimark dilation theorem restated) Let F: $\mathcal{F}\rightarrow\mathcal{L}(\mathcal{H})$ be a POVM. There exists a Hilbert space $\mathcal{H}_0$ and a state $|\phi_0\rangle\langle\phi_0|\in\mathcal{S}(\mathcal{H}_0)$ and a PVM E: $\mathcal{F}\rightarrow\mathcal{L}(\mathcal{H}\otimes\mathcal{H}_0)$ such that*

$$\mathrm{Tr}_{\mathcal{H}\otimes\mathcal{H}_0}[\rho\otimes|\phi_0\rangle\langle\phi_0|E(X)]=\mathrm{Tr}_{\mathcal{H}}[\rho F(X)]$$

holds for any $\rho\in\mathcal{S}(\mathcal{H})$ and for every $X\in\mathcal{F}$. Moreover, E can always be chosen to be of either of the forms $U^\dagger E(\,.\,)\otimes\mathbb{1}U$ or $U^\dagger\mathbb{1}\otimes E(\,.\,)U$ where U is a suitable unitary.

Finally, let us show that any (mixed) state $\rho_A\in\mathcal{S}(\mathcal{H}_A)$ of a system A can be seen as a part of a pure state in an extended Hilbert space [57].

Theorem 2.12. *(On purifications) Any (mixed) state $\rho_A\in\mathcal{S}(\mathcal{H}_A)$ can be viewed as a part of a pure state $|AR\rangle\langle AR|\in\mathcal{S}(\mathcal{H}_A\otimes\mathcal{H}_R)$, i.e.*

$$\rho_A=\mathrm{Tr}_R(|AR\rangle\langle AR|),$$

where R denotes a (possibly fictitious) reference *system of a minimal necessary dimension at most that of $\mathcal{H}_A$. The state $|AR\rangle\langle AR|$ is called a purification of ρ_A. The minimal necessary dimension of $\mathcal{H}_R$ is at most* $\dim(\mathcal{H}_A)$*.*

Proof. *(A constructive proof for the finite-dimensional case) Suppose ρ_A has orthonormal decomposition $\rho_A=\sum_i p_i|i_A\rangle\langle i_A|$. Introduce a system R with the same state space as that of A, with orthonormal basis $|i_R\rangle$ and define a pure state of the combined system*

$$|AR\rangle=\sum_i\sqrt{p_i}|i_A\rangle|i_R\rangle.$$

Calculating the reduced density operator for the system A

$$\mathrm{Tr}_R(|AR\rangle\langle AR|) \;=\; \sum_{i,j}\sqrt{p_i p_j}|i_A\rangle\langle j_A|\mathrm{Tr}_R(|i_R\rangle\langle j_R|)=\sum_i p_i|i_A\rangle\langle i_A|=\rho_A$$

we see that $|AR\rangle$ *is a purification of* ρ_A*. Moreover, since* $\mathrm{Tr}(B)=\mathrm{Tr}(UBU^\dagger)$*,* $B\in\mathcal{T}(\mathcal{H})$*, we see that there is unitary freedom for the reference system state, i.e.* $|AR'\rangle=(\mathbb{1}\otimes U)|AR\rangle$ *is also a purification of* ρ_A*.* □

2.2.3 Dynamics

The time evolution of an isolated quantum system initially (at the time t_0=0) prepared in a pure state $|\psi_0\rangle$ is governed by the Schrödinger equation

$$i\hbar\frac{\mathrm{d}}{\mathrm{dt}}|\psi(t)\rangle=H|\psi(t)\rangle, \qquad (|\psi(0)\rangle=|\psi_0\rangle), \tag{2.13}$$

where H is the Hamiltonian. The unit vector $|\psi\rangle$ then evolves according to the formal solution

$$|\psi(t)\rangle=U(t)|\psi_0\rangle=\mathrm{e}^{-\frac{i}{\hbar}Ht}|\psi_0\rangle. \tag{2.14}$$

The pure state $|\psi\rangle\langle\psi|$ then evolves as

$$(|\psi\rangle\langle\psi|)(t)=U(t)|\psi_0\rangle\langle\psi_0|U^\dagger(t). \tag{2.15}$$

General states can always be written (non-uniquely) as a convex combinations of pure states, i.e. $\rho=\sum_k\lambda_k|\psi_k\rangle\langle\psi_k|$. Due to linearity of the Schrödinger equation (and thus, of the operator U), a general state then evolves as

$$\rho(t)=U(t)\rho_0U^\dagger(t)=:\mathcal{U}_t(\rho_0). \tag{2.16}$$

By differentiating the Eq. (2.16) we get the Von Neumann equation

$$i\hbar\frac{\mathrm{d}}{\mathrm{dt}}\rho(t)=[H,\rho(t)] \tag{2.17}$$

2.2.4 Measurements

In quantum mechanics, there are three different levels of description of (non-destructive) measurements.

1. The complete description: A physical model of the measurements apparatus and its interaction with the object system is given. A part of the apparatus called ancilla (or a probe system), which is in a specified (pure[2.1]) initial state interacts (via a prescribed interaction) with the system to be measured and one of the (orthogonal) states (so-called pointer states) of the probe system (or, more generally, of the combined system) is then read out. This description yields the measurement outcomes statistics as well as the state of the system and of the probe after the measurement and is referred to as a *measurement model* (MeMo) in literature (cf. [37]).

[2.1] It is assumed that the ancilla is initially not entangled to the object system. Then it follows (through Theorem 2.12) that if the ancilla is in a mixed state, one can introduce a new pure-state ancilla that is an extension (purification) of the original ancilla such that this purification is again a part of an (enlarged) environment. For ancillas initially entangled with the object system see [58, 59].

2. The description of a part which is relevant to the object system: Often we are not interested in the exact measurement mechanism. At this level of description the state change and statistics of the measurement outcomes are given, leading to the notion of an *instrument* (a state transformer), introduced by Davies and Lewis [30].
3. Description of measurement-outcome distributions only: At this level of description we do not care even about the state change of the object system. The description is in this case given in terms of *observables*, i.e. positive-operator valued measures (POVMs).

In the following, it will suffice to consider the third and the second level of measurement description. Let us summarize the corresponding tools – observables and instruments.

2.2.4.1 Quantum operations

The mathematical formalism of quantum operations is an important tool for description of dynamics of open quantum systems (such as, e.g. a system that is being measured). It is especially well adapted to describe discrete state changes, i.e. transformations between the initial and the final state of a quantum system, without explicit reference to the time parameter.

Definition 2.13. *Operations (alternative name: state transformations).*

An operation *is a completely positive linear*[2.2] *mapping* $\Phi\colon \mathcal{T}(\mathcal{H}) \rightarrow \mathcal{T}(\mathcal{H})$ *satisfying*

$$0 \leq \mathrm{Tr}[\Phi(\rho)] \leq 1$$

for all $\rho \in \mathcal{S}(\mathcal{H})$*.*

Let us move on to the measurement description 2., i.e. introduce the notion of an instrument.

Definition 2.14. *Instruments. (alternative name: state transformers).*

An instrument *is an operation-valued (state-transformation valued) measure* $\mathcal{I}\colon \mathcal{F} \rightarrow \mathcal{L}(\mathcal{T}(\mathcal{H}))$, $X \mapsto \mathcal{I}_X$ *on a measurable space* $(\Omega, \mathcal{F})$ *defined by the properties*

1. $\mathcal{I}_X(\rho) \geq 0$ *for all* $X \in \mathcal{F}$, $\rho \in \mathcal{S}(\mathcal{H})$
2. $\mathrm{Tr}[\mathcal{I}_\Omega(\rho)] = \mathrm{Tr}(\rho)$ *for all* $\rho \in \mathcal{S}(\mathcal{H})$
3. $\mathcal{I}_{\cup X_i}(\rho) = \sum_i \mathcal{I}_{X_i}(\rho)$ *for all disjoint sequences* $(X_i) \subset \mathcal{F}$ *and all* $\rho \in \mathcal{S}(\mathcal{H})$*,*

where the series converges in the trace norm.

Comparing the above properties with the properties of a POVM, Definition 2.3 (resp. the corresponding observable, Def. 2.7), we see that the mapping $X \rightarrow E(X)$ defined by

$$\mathrm{Tr}[E(X)\rho] := \mathrm{Tr}[\mathcal{I}_X(\rho)] \quad X \in \mathcal{F}, \rho \in \mathcal{S}(\mathcal{H}) \tag{2.18}$$

[2.2] *Φ is convex-linear (or affine) as a map on $\mathcal{S}(\mathcal{H})$, which is not a linear space. Convex-linearity of Φ means that $\Phi(p\rho_1 + (1-p)\rho_2) = p\Phi(\rho_1) + (1-p)\Phi(\rho_2)$, where $\rho_1, \rho_2 \in \mathcal{S}(\mathcal{H})$ and $p \in \langle 0, 1\rangle$.*

is a normalized POVM on $(\Omega,\mathcal{F})$ (resp. $X\rightarrow\mathrm{Tr}[E(X)\ .\ ]$ defined by Eq. (2.18) is the corresponding observable), which is unique for an instrument $\mathcal{I}$. This induced POVM (observable) is called the *associate POVM (observable)*. However, as follows from Eq. (2.18), a different instrument $\mathcal{I}'$ defined by e.g. $\mathcal{I}'_X(\rho)=U(X)\mathcal{I}_X(\rho)U^{\dagger}(X)$, (where $U(X)$ is any unitary, possibly different for different sets $X\in\mathcal{F}$), and the original instrument $\mathcal{I}$ induce the same associate observable (POVM) E. Hence, while any instrument $\mathcal{I}$ induces a unique associate observable (POVM) E of $\mathcal{I}$, on the contrary, every observable (POVM) E has infinitely many *E-compatible* instruments (state transformers).

Remark 2.15. Why is it convenient to allow the operations to be trace-decreasing (i.e. the transformed "states" $\mathcal{I}_X(\rho)$ to be unnormalized with $0\leq\mathrm{Tr}[\mathcal{I}_X(\rho)]\leq 1$? This is to have in one mathematical object (the instrument) "encoded" both the measurement outcome with its probability (the observable (POVM)) and the physical-object-system state change due to the measurement (the operation).

Definition 2.16. *Lüders instrument.*

The Lüders instrument associated with a discrete POVM E defined on the power set $\mathcal{F}$ of Ω by $E(\{\omega_i\})=:E_i$ is defined by

$$\mathcal{I}(X)(\rho):=\sum_{\omega_i\in X}E_i^{1/2}\rho E_i^{1/2}=\sum_{\omega_i\in X}\Phi^{E_i}(\rho),$$

where Φ^{E_i} is the Lüders operation (state transformation, update rule) for E_i. More generally, for any POVM E and any value set $X\in\mathcal{F}$ one may define the associated Lüders operation as

$$\Phi_X(\rho):=E(X)^{1/2}\rho E(X)^{1/2}.$$

2.3 How close are two quantum states?

A useful measure of how close two quantum states are is the fidelity, selected properties of which we now summarize (for proofs, cf. [50]). The fidelity of two states $\rho,\rho'\in\mathcal{S}(\mathcal{H})$ is defined by[2.3]

$$f(\rho,\sigma)=\left(\mathrm{Tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}\right)^2 \tag{2.19}$$

Although the fidelity is symmetric under the exchange of the to systems, it is not a distance on density operators since it does not fulfill the other two required conditions (see the definition of a metric, Appendix F).

For two "classical", i.e commuting, states the fidelity, Eq. (2.19), reduces to the classical fidelity f_c

$$f_c(r,s)=\left(\sum_i\sqrt{r_i s_i}\right)^2 \tag{2.20}$$

of the "eigenvalue" distributions, r,s, where $i\in\{1,\ldots,\dim(\mathcal{H})\}$ label the diagonal elements of the density matrices ρ, ρ' in the basis in which they are simultaneously diagonal.

[2.3] Sometimes (e.g. [50]) the fidelity is defined as the square root of the quantity in Eq. (2.19).

If one of the states, say $\sigma = |\psi\rangle\langle\psi| =: \psi$, is pure then the fidelity, Eq. (2.19), reduces to

$$f(\rho, \psi) = \langle\psi|\rho|\psi\rangle = \mathrm{Tr}(\rho\psi), \tag{2.21}$$

i.e. the overlap of the states ρ and ψ.

One can arrive at the expression Eq. (2.19) in two noteworthy ways. The first is based on the classical fidelity, Eq. 2.20, in terms of which

$$f(\rho, \sigma) = \min_E f_c(p, q),$$

where the minimum is over all POVMs E and $p\colon m \mapsto p_m = \mathrm{Tr}(\rho E_m)$ and $q\colon m \mapsto q_m = \mathrm{Tr}(\sigma E_m)$ are the probability distributions describing probabilities of obtaining outcomes of a POVM E (for simplicity we use the notation of finite-outcome POVMs; the sum in Eq. 2.20 runs through all elementary outcomes of the particular POVM).

The second way to arrive at the expression Eq. (2.19) is based on purifications. The fidelity can be written as

$$f(\rho, \sigma) = \max_{|\varphi\rangle} |\langle\psi|\varphi\rangle|^2$$

where the maximization is over all purifications $|\varphi\rangle$ of σ for a fixed purification $|\psi\rangle$ of ρ (or vice versa). The fidelity being a square of the inner product product of two purification vectors, the angle between them reads

$$D_f = \arccos \sqrt{f(\rho, \sigma)} \tag{2.22}$$

and defines a distance on the set of density operators (Eq. (2.22) is symmetric in its inputs, vanishes if and only if $\rho = \sigma$ and fulfills the triangular inequality (cf. [50])).

2.4 Coherent states for any unimodular group

Consider a group $\mathcal{G}$ of N unitary matrices U_α. For any fixed $|\psi\rangle \in \mathcal{H}$, consider the states $|\psi_\alpha\rangle := U_\alpha|\psi\rangle$ (called coherent states of for the group $\mathcal{G}$) and the sum

$$A := \sum_\alpha |\psi_\alpha\rangle\langle\psi_\alpha| = \sum_\alpha U_\alpha|\psi\rangle\langle\psi|U_\alpha^\dagger \tag{2.23}$$

For any $U_\beta \in \mathcal{G}$

$$\begin{aligned} U_\beta A U_\beta^\dagger &= \sum_\alpha U_\beta U_\alpha |\psi_\alpha\rangle\langle\psi_\alpha| U_\alpha^\dagger U_\beta^\dagger \\ &= \sum_\gamma U_\gamma |\psi\rangle\langle\psi| U_\gamma^\dagger \\ &= A \end{aligned} \tag{2.24}$$

because the set of all $U_\gamma = U_\beta U_\alpha$ runs (for fixed U_β and all $U_\alpha \in \mathcal{G}$) over all the elements of $\mathcal{G}$. Multiplying Eq. (2.24) by U_β from right we get $U_\beta A = A U_\beta$ for all $U_\beta \in \mathcal{G}$. Then, from Schur's lemma, if the matrices U_β are elements from (the range of) an irreducible representation on a complex vector space, we have $A = c\mathbb{1}$. Taking trace of Eq. (2.23), we obtain $\mathrm{Tr}(A) = cd = \sum_\alpha \langle\psi_\alpha|\psi_\alpha\rangle = N$, where d is the dimension of $\mathcal{H}$. It follows that

$$\sum_\alpha |\psi_\alpha\rangle\langle\psi_\alpha| = \frac{N}{d}\mathbb{1}. \tag{2.25}$$

Thus, the matrices $(d/N)|\psi_\alpha\rangle\langle\psi_\alpha|$ form a POVM.

The above definitions can be extended to continuous groups, given a suitable measure is defined to replace the discrete sum over α. As an example that we will use later on, consider the (for 2j+1 even, projective) representation of the group of rotations, SO(3), on the $(2j+1)$-dimensional Hilbert space of a spin-j particle. The angular momentum coherent states are defined as

$$|\boldsymbol{n}\rangle := |\theta,\phi\rangle = \sum_{m=-j}^{j} |m\rangle \binom{2j}{j+m}^{\frac{1}{2}} \cos^{j+m}(\theta/2)\sin^{j-m}(\theta/2)\, e^{-\mathrm{i}m\phi}, \tag{2.26}$$

where θ, ϕ are the polar and azimuthal angles, respectively. The integral

$$\begin{aligned}\int_{\mathbb{S}^2} |\boldsymbol{n}\rangle\langle\boldsymbol{n}|\mathrm{d}\boldsymbol{n} &= \int_0^{2\pi}\int_0^{\pi} |\theta,\phi\rangle\langle\theta,\phi|\frac{1}{4\pi}\sin\theta\,\mathrm{d}\theta\mathrm{d}\phi \\ &= \frac{1}{2j+1}\sum_m |m\rangle\langle m| \\ &= \frac{1}{2j+1}\mathbb{1}.\end{aligned} \tag{2.27}$$

enables us to form a “resolution of identity with (infinitely many) elements”

$$(2j+1)|\boldsymbol{n}\rangle\langle\boldsymbol{n}|\mathrm{d}\boldsymbol{n}, \tag{2.28}$$

i.e. a POVM

$$M(B) = \int_{\boldsymbol{n}\in B} (2j+1)|\boldsymbol{n}\rangle\langle\boldsymbol{n}|\mathrm{d}\boldsymbol{n}, \tag{2.29}$$

where $\mathrm{d}\boldsymbol{n} = \frac{1}{4\pi}\sin\theta\,\mathrm{d}\theta\mathrm{d}\phi$ and B is some (Borel) subset – (an interval) from the Borel (σ-)algebra of intervals on $\mathbb{S}^2$ (i.e. of intervals $(\theta,\theta')\times(\varphi,\varphi')$; $\theta,\theta'\in\langle 0,\pi)$, $\varphi,\varphi'\in\langle 0,2\pi)$).

2.5 Elements of information theory

In this section we will briefly recall some basic concepts of the information theory.

Let us begin with the notion of information. In everyday life, (a piece of) information is some event (represented by a message) through which we have learned something expected to happen with less than unit probability (confirming something we have expected with probability one is not very informative). From this it follows that to talk about “information” content of an event, we have to be able to quantify how unexpected (surprising) the event was. More generally, in order to talk about the “information” content of an event E_2 about some other event E_1 we have to quantify how unexpected the event E_1 was before and after the realization of the event E_2.

We put “information” in quotation marks, since in the theory of information (or communication) the term information is used to denote a different (although related) concept – loosely speaking, information of a random variable Y (an object that can be in a collection of states (random events), where each state (event) is expected to be realized (to happen) with some probability) about a (possibly) different random variable X is the difference between the average “unexpectedness” (surprise) of an event from the set X (the uncertainty of X) and the average “unexpectedness” of an event from the set X given the knowledge of which event of Y was realized (average (over Y) uncertainty of X given Y). Simply, (mutual) information of Y about X is the average reduction of uncertainty in X due to getting to know Y.

Using constructions of the form “information content of an event E_2 about some other event E_1” we should expect that information is not an absolute (i.e. it is relative) notion, but always relates to “something” – the “object” unexpectedness of whose state we are quantifying.

As far as the relativity of “information” carried by a message (representing an event, or state of an object) is concerned, it is twofold. Firstly, a message often has some meaning, i.e. it refers to, or is correlated with, some real or conceptual object (e.g. the string “1” means nothing (carries no “information” in the “everyday-life sense”) per se (even if we specified that there were two possibilities “0” and “1” with probabilities 1/2), but it certainly has high “information” content as the binary answer to the question: “Will you marry me?”). The semantic aspects (like the one above) of messages, however, are *not* of interest in the theory of information (or communication).

The notion of relativity of information that is important in the theory of information (communication) is of a different nature. Consider, for instance, the following situation: Let us have two matches – a football match F and a tennis match T. Suppose either match can end up by victory “1” or loss “0” of one specific party (let’s say the guest team). I.e., we can introduce a random variable “FT” of the combined results:

$$FT\colon \{\text{victory}, \text{victory}\} \rightarrow 1, \{\text{victory}, \text{loss}\} \rightarrow 2, \{\text{loss}, \text{victory}\} \rightarrow 3, \{\text{loss}, \text{loss}\} \rightarrow 0,$$

and a random variable of let’s say the result “F” of the football match:

$$F\colon \{\text{victory}, \text{anything}\} \rightarrow 1, \{\text{loss}, \text{anything}\} \rightarrow 0.$$

Let us suppose $p_1^F = 9/10, p_i^T = 1/2$ and $p_1^{FT} = p_2^{FT} = 9/20, p_3^{FT} = p_0^{FT} = 1/20,\ (i = 0, 1)$ are the probabilities of the match(es) outcomes, representing our prior knowledge about the match results, as well as about the correlation (in this case independence) between the outcome of F and the outcome of T.

Now obtaining the message – value of the random variable F(match outcome) “1”, what is the information content of it? As we have stated in the previous paragraph, the semantic aspects (F stands for a football match, “1” means victory, which particular football match are we talking about) are irrelevant. The relevant aspect (with respect to the relativity of information content of the message “1”) is: What is the object (which random variable) the information should refer to? Is it

a) F

b) FT

c) other random variable ?

In the next paragraph we will see that given one of the above options, what is of course also relevant are the probabilities (the prior knowledge) containing the event (outcome) corresponding to $F(\text{match outcome}) = 1$, that is, for the above random variables,

a) p_1^F

b) p_1^{FT}, p_2^{FT}

c) other probabilities,

respectively. Why are these important? Consider the case a), i.e. the "information" content of the message "1" with respect to the random variable F. In our example the probability of the message "1" with respect to F is $9/10$ while the probability of the message "0" is $1/10$. This means that receiving the message "1" is something we expect to happen, while the opposite result is quite unexpected. One could intuitively say that in a sense the message "1" carries less "information" than the message "0" in this case. We feel that this "information" contained in each of the two messages depends on the probabilities p_1^F and p_0^F, respectively, i.e it should be some function $I_i = f(p_i)$. Also, I_i should be zero if $p_i = 1$ (if we know something, learning it for the second time should carry no information at all) and I_i should increase with decreasing p_i (making p_i smaller and smaller the "information" content of the corresponding message should be larger and larger). It is also reasonable that f be continuous. Moreover, if we learn two messages "1" about two independent random variables F and F' values with, e.g., the same probabilities $p_1^F = p_1^{F'}$, we would probably say the information content of the two messages is twice the content of a single message (this is related to the fact that humans intuitively measure entities by linear comparison with common standards [55]), while the joint probability of the event of receiving "11" is $p_{11}^{FF'} = (p_1^F)^2$. This leads us to the requirement that f should be a function of the probability that transforms products (of probabilities) into sums (of information contents). A function that fulfills the above intuitive requirements is[2.4]

$$I_i = \log \frac{1}{p_i},$$

where the base of the logarithm fixes the units in which we measure I_i and is usually taken to be 2 (corresponding to the units called *bits*) or $\mathbb{e}$ (corresponding to "natural units", or *nats*). The above concept of "information" is sometimes referred to as *self-information*.

We have seen that the "information" content of a (one-symbol) string was proportional to how unlikely the state of the "object" had been before we learned the string's content, i.e. the strings with low "information" content are more probable, while the strings with high "information" content are improbable. This will be important if, rather than being interested in a particular realization of our model situation (particular events or messages), we were interested in the average "information" content of a message over many realizations of the situation (many independent instances of the same match, i.e. messages with the result, with the same a-priori probabilistic knowledge on the winner).

In information theory the events are represented by symbols (usually bits – classical or quantum) originating from a (stochastic and ergodic) source fully characterized by probabilities of symbols (possibly depending on previously emitted symbols). In the case of a memory-less source (a source fully characterized by probabilities of symbols which do not change if we get to know any previously generated symbols), the entropy of the source is given by the average (over all symbols) "information" content (expected self-information per symbol).

Formally,

Definition 2.17. *Shannon information (Shannon entropy, information entropy).*

The quantity

$$H(X) = -\sum_{x_i \in X} p(x_i) \log_2 p(x_i) \tag{2.30}$$

[2.4] Of course also a $K \in \mathbb{R}^+$-multiple of I_i would do.

where $p(x_i)$ is the probability of the event x_i occurring, is called the Shannon entropy *of the random variable*[2.5,2.6] *X.*

Remark 2.18. Note that the interpretation of the Shannon information Def. 2.17 as the average "information" content (uncertainty decrease) of a message (symbol from an alphabet) only makes sense if the probabilities p_i of the symbols are the only knowledge we have about the source (the source is memory-less, i.e. knowledge of previously emitted symbols does not change the probabilities of the current symbol, or the receiver of the symbols from the source has no memory, i.e. when receiving symbols, the receiver cannot estimate the probabilities of symbols conditional on the previously received symbols; moreover, even if he had the conditional probabilities at hand, he cannot make any use of them, since he does not remember the previously received symbols). For instance, consider the sequence ABABABAB... . Given the alphabet $\{A, B\}$, we have $p(A) = p(B) = 1/2$ and, consequently, $H(X) = 1$. Certainly, we would not like to claim that each symbol carries (on average) 1 bit of information in this case (that is, given the additional knowledge about the sequence – the whole sequence itself in this particular case), as our lack of knowledge has not been decreased at all. In fact, choosing a different alphabet $\{\mathrm{AB}\}$, we have $p(\mathrm{AB}) = 1$ and $H(X) = 0$, which is just another way of saying what we had the extra information (which enabled us to introduce the new alphabet in the first place). If we have some additional knowledge about the sequence (on top of the probabilities of the symbols), the correct figure of merit to measure the average "information content" (average unexpectedness, or "surprise") of a message is the quantity

$$H(X|\text{knowledge}) = -\sum_{x_i \in X} p(x_i|\text{knowledge}) \log_2 p(x_i|\text{knowledge}), \tag{2.31}$$

which is just a special case of the conditional entropy:

Definition 2.19. *Conditional entropy.*

The quantity

$$\begin{aligned} H(X|Y) &= -\sum_{x \in X, y \in Y} p(y) p(x|y) \log_2 p(x|y) \\ &= \sum_{y \in Y} p(y) H(X) \\ &= H(X,Y) - H(Y) \end{aligned} \tag{2.32}$$

[2.5] *Even though according to the definition (F.30), a random variableX is a map from the set of events to the real numbers, one can often see it used to denote the range of the map (the random variable by definition), or the distribution (the probability measure p_X on the range of the random variable X induced (form the probability space $(\Omega,\ \mathcal{F},\ p)$) by the random variable X). A nice example of this "abuse of notation" is Eq. (2.30), which should be understood as follows: X in the sum is the range (more precisely support) of the random variable, i.e. $X = \{x_i\} := X(\Omega) = \{X(\omega_i); \omega_i \in \Omega\}$ while the X in the argument of H refers to the distribution p_X (defined by $p_X(x_i) \equiv p(x_i) = \sum_{j=1}^{|\{\omega_j : X(\omega_j) = x_i\}|} p(\omega_j)$, where $|\{\ .\ \}|$ is the number of elements of a set). We will often also "abuse" the notation in the above sense in what follows.*

[2.6] *In principle it is not necessary that X be (the range of, or a distribution on the range of) a random variable in the above definition of entropy (and in the entropic quantities to appear later), as entropy depends merely on probabilities of events and so the events need not be represented by numerical values, i.e. it would suffice for X (in the sum) to be a set of elementary events Ω (and for X in the argument of H to be probability from the probability space $(\Omega, \mathcal{F}, p)$). Of course, then the entropy $H(Y)$ of a (possible) random variable $Y : \omega_i \rightarrow y_i$ need not be the same as $H(\Omega)$: consider $\Omega = \{3, 4\}$, a random variable Y such that $Y(3) = Y(4)$, $\mathcal{F} = \{\varnothing, \{3\}, \{4\}, \{3, 4\}\}$ (also $\mathcal{F}' = \{\varnothing, \{3, 4\}\}$ would do for the above specific Y), $p_{\omega_i} = 1/2$, and . Now $Y(\Omega) = \{1\}$,$p_{y_1} = 1$ where $y_1 = Y(\omega_i)$, $(i = 1, 2)$, and so $H(Y) \neq H(\Omega)$ in this case.*

is called the conditional entropy *of the discrete random variable X with respect to (conditioned on) the random variable Y.*

Note that Eq. (2.32) is a generalization of the intuitively introduced quantity Eq. (2.31) where the generalization is that our knowledge may be of probabilistic nature, i.e. our additional (on top of $p(x_i)$'s) knowledge about the source of the random variable X depends on some random event y_i (from a set given here by (the range of) the random variable Y), probability of the event y_i being $p(y_i)$. The knowledge *about* (the source of) X is the key word here, as for the conditional entropy to differ from the Shannon information, the knowledge we have must have some relevance to the random variable X, i.e. the random variable Y has to have at least some correlation to the random variable X (i.e. $p(x_i,\ y_j) \neq p(x_i)p(y_j)$), otherwise $p(x_i|\text{knowledge}) = p(x_i|\text{irrelevant knowledge}) = p(x_i)$ (formally, $p(x_i|y_j) = p(x_i, y_j)/p(y_j) = p(x_i)p(y_j)/p(y_j) = p(x_i)$).

Remark 2.20. Let us have a source with memory and let the (range of the) random variable Y_n be composed of strings of symbols (numbers) of length n emitted by the source prior to the current symbol represented by the (the range of the) random variable X. Then, the limit of the conditional entropies

$$\lim_{n\rightarrow\infty} H(X|Y_n) \tag{2.33}$$

is the entropy of the stochastic ergodic source with memory [55]. If one knows, that the memory of the source goes back only up to N symbols, then the limit Eq. (2.33) is equal to $H(X|Y_N)$. In the case of a memory-less source we recover the Shannon entropy, Eq. (2.30).

Instead of the Shannon entropy, in the case of a continuous random variable X one can define the differential entropy:

Definition 2.21. *Differential entropy (continuous entropy).*

The quantity

$$h(X) = -\int_S \tilde{p}(x)\log_2 \tilde{p}(x)\mathrm{dx}, \tag{2.34}$$

where the integration is performed over the support S of $\tilde{p}(x)$, i.e. $\{x \in X; \tilde{p}(x) > 0\}$, is called the differential entropy *of the random variable X.*

The differential entropy Eq. (2.34), similarly to the Shannon entropy Eq. (2.30) "measures the spreading" of the probability density $\tilde{p}$. However, let us notice that:

i. The "spreading" is "measured" in the absolute sense, causing the differential entropy to be relevant (as a measure of uncertainty of X) only up to a constant. To see this, let us rescale the random variable X by a real constant a, i.e. introduce a new random variable $Y = aX$. Then, we have $\tilde{p}(y) = \frac{1}{|a|}\tilde{p}(x)$ and, consequently,

$$h(Y) = -\int_{S_Y} \tilde{p}(y)\log_2 \tilde{p}(y)\mathrm{dy} = -\int_{S_X} \frac{\tilde{p}(x)}{|a|}\log_2 \frac{\tilde{p}(x)}{|a|} a\,\mathrm{dx} = h(X) + \log_2 |a|.$$

Therefore, the differential entropy has no absolute meaning; a meaningful quantity is the difference of differential entropies (provided the supports of the two probability densities, i.e. $\{x \in X; \tilde{p}(x) > 0\}$, $\{x' \in X'; \tilde{p}(x') > 0\}$, are the same).

ii. As a consequence, for a very "narrow" probability density $\tilde{p}$, i.e. a small $|a|$, the differential entropy can be negative.

Therefore it is convenient to define a different quantity:

Definition 2.22. *Relative entropy (Kullback-Leibler divergence).*

The quantity

$$D_{\mathrm{KL}}(\tilde{p}\|\tilde{p}') = \int_S \tilde{p}(x)\log_2\frac{\tilde{p}(x)}{\tilde{p}'(x)}\mathrm{d}x, \tag{2.35}$$

where $\tilde{p}(x)$ and $\tilde{p}'(x) = \tilde{p}'(y)$ have same supports (S) is called the relative entropy *of the distribution (probability density) $\tilde{p}(x)$ with respect to (relative to) the distribution $\tilde{p}'(x)$.*

Note that $D(\tilde{p}\|\tilde{p}') \geq 0$ with equality if and only if $\tilde{p}(x) = \tilde{p}'(x)$ almost everywhere. Relative entropy measures the "difference" between the two distributions that appear as its arguments. However, it is not a distance between them, as it is not symmetric (with respect to the exchange $X \leftrightarrow Y$, i.e. $\tilde{p}(x) \leftrightarrow \tilde{p}'(x)$), nor does it satisfy the triangle inequality.

For discrete probability distributions we have

$$D_{\mathrm{KL}}(p\|p') = \sum_i p(x_i)\log_2\frac{p(x_i)}{p'(x_i)}, \tag{2.36}$$

where $x_i \in S$.

Remark 2.23. Sometimes (e.g. [55]) relative entropy (of a source) is defined as the ratio of the entropy of the source to the maximum value it could have while still restricted to the same alphabet[2.7] (this quantity will be referred to as H_{relat} in what follows). That is (for the discrete case), for a N-letter alphabet,

$$\begin{aligned} H_{\mathrm{relat}} &= \frac{-\sum_i p(x_i)\log_d[p(x_i)]}{\underbrace{-\sum_i \frac{1}{N}\log_d\left[\frac{1}{N}\right]}_{-\log_d[N]}} = \sum_i p(x_i)(-\log_d[N])^{-1}\log_d[p(x_i)] \\ &= \sum_i p(x_i)\log_d\left[p(x_i)^{-(\log_d[N])^{-1}}\right]. \end{aligned} \tag{2.37}$$

Should the $D_{\mathrm{KL}} \overset{!}{=} H_{\mathrm{relat}}$ hold, it must hold (taking the logarithm basis in H_{relat} to be the same as in the definition of D_{KL}, i.e. measuring in bits per symbol) that

$$\frac{p(x_i)}{p'(x_i)} \overset{!}{=} p(x_i)^{-(\log_2[N])^{-1}},$$

i.e.

$$p'(x_i) \overset{!}{=} p(x_i)^{1+(\log_2[N])^{-1}}. \tag{2.38}$$

Had we defined D_{KL} with the logarithm with the basis N (number of symbols in the alphabet) then Eq. (2.38) would read $p'(x_i) \overset{!}{=} p(x_i)^2$, which would imply $D_{\mathrm{KL}}^{\mathrm{base}\,N}(p\|p^2) = H_{\mathrm{relat}}^{\mathrm{base}\,N}(X) = H^{\mathrm{base}\,N}(X)$ as expected (from the definition of the relative entropy at the very beginning of the Remark 2.23). The quantity $1 - H_{\mathrm{relat}}$ is called *redundancy* of a source.

[2.7] Note that having used the Shannon entropy, Eq. (2.30), in the Eq. (2.37) we are restricting ourselves to memory-less stochastic ergodic sources. The verbal definition holds for sources with memory as well, one only has to work with the entropy suitable for the source at hand – at least if one wishes to interpret the calculated quantities in a standard way.

One can of course introduce a symmetrized version of the Kullback-Leibler divergence $D_{\mathrm{KL}}(\tilde{p}\|\tilde{p}') + D_{\mathrm{KL}}(\tilde{p}'\|\tilde{p})$ (which is the original K-L divergence as it was proposed by Kullback and Leibler). Another quantity that can be expressed through K-L divergence and which is symmetric under the exchange of the random variables in its argument is the mutual information between the random variables X and Y,

$$I(X;Y) = D_{\mathrm{KL}}(\tilde{p}(x,y)\|\tilde{p}(x)\tilde{p}(y)) \tag{2.39}$$

(the discrete case analogously).

Usually, the following formal definition is given:

Definition 2.24. *Mutual information.*

$$\begin{aligned} I(X;Y) &= h(X) - h(X|Y) \\ &= h(X) + h(Y) - h(X,Y) \\ &= \int_{S_X}\int_{S_Y} \tilde{p}(x,y)\log_2\frac{\tilde{p}(x,y)}{\tilde{p}(x)\tilde{p}(y)}\mathrm{d}x\mathrm{d}y, \end{aligned} \tag{2.40}$$

or, for discrete random variables X,Y,

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y) \\ &= \sum_{i,j} p(x_i,y_j)\log_2\frac{p(x_i,y_j)}{p(x_i)p(y_j)}, \end{aligned} \tag{2.41}$$

where $x_i \in S_X$, $y_j \in S_Y$, is called the mutual information between (distributions on) the random variables X and Y.

Mutual information is the correct measure of how much information two random variables possess about each other – how correlated they are.

2.5.1 Channels and capacities

Even though information is a theoretical concept, in practice, its carrier is always some kind of a physical system. As we have seen, in quantum mechanics physical systems are described by states $\rho \in \mathcal{S}(\mathcal{H})$. Any information processing device (a device transferring, storing, or otherwise manipulating information) then transforms the input states (states of some physical system) into output states of some (possibly different) physical system. Such transformations of states (and hence, the underlying devices) are described through the notion of a channel.

Definition 2.25. *Channels.*

A channel is a completely positive trace-preserving convex-linear map $\mathcal{E}$: $\mathcal{S}(\mathcal{H}_1) \rightarrow \mathcal{S}(\mathcal{H}_2)$ between the state spaces $\mathcal{S}(\mathcal{H}_i)$.

Convex-linearity (linearity as a map $\mathcal{T}(\mathcal{H}_1) \rightarrow \mathcal{T}(\mathcal{H}_2)$) is a consequence of requiring preservation of convex mixtures, the trace-preserving is due to the requirement that normalized states should be transformed to normalized states. Recalling the Section 2.2.4.1, Def. 2.13, channels are trace preserving quantum operations (additionally allowing for the input and output Hilbert spaces to be different). The above definition includes quantum channels as well as classical channels.

An important special case is a channel whose output is the tensor product of a classical and quantum output. Let e_i, $i = 1, ..., d$ be a classical basis in $\mathcal{H}_2'$. The general form of such a channel then is $\mathcal{E}: \mathcal{S}(\mathcal{H}_1) \rightarrow \mathcal{S}(\mathcal{H}_2) \otimes \mathcal{S}(\mathcal{H}_2')$ with

$$\mathcal{E}(\rho) = \sum_i \mathcal{E}_i(\rho) \otimes |e_i\rangle\langle e_i|, \tag{2.42}$$

where each of the $\mathcal{E}_i$: $\mathcal{T}(\mathcal{H}_1) \to \mathcal{T}(\mathcal{H}_2)$ is an operation. Channels of the type Eq. (2.42) are sometimes (e.g. [13]) called instruments[2.8]. Since there are two outputs, we get two "marginals" (i.e. channels obtained if one of the outputs is ignored): Ignoring the classical output we get a quantum channel $\mathcal{E}' = \sum_i \mathcal{E}_i$, while not looking at the quantum output, we have $\mathrm{Tr}_{\mathcal{H}_2}\mathcal{E}(\rho) = \sum_i \mathrm{Tr}[\mathcal{E}_i(\rho)]|e_i\rangle\langle e_i|$, i.e. for each i we have a realization of the effect $\mathrm{Tr}[\mathcal{E}_i(\,.\,)] =: E_i(\,.\,)$ (specifying for each i the corresponding outcome (element $\omega_i \in \Omega$) we would get an observable $E{:}\, \omega_i \to E_i$, or the corresponding POVM $E{:}\, \omega_i \to \mathcal{E}_i$).

Let us now introduce the von Neumann entropy that will be used later on.

Definition 2.26. *Von Neumann entropy.*

The von Neumann entropy *S of a state ρ is defined as*

$$S(\rho) = -\,\mathrm{Tr}(\rho \log_2 \rho) = -\sum_i \lambda_i \log_2 \lambda_i = H(\{\lambda_i\}),$$

where λ_i are the "eigenvalues"[2.9] of ρ, i.e. the Von Neumann entropy is the Shannon entropy of the distribution p: $p(i) = \lambda_i$.

What is the meaning of the von Neumann entropy? In general states ρ_x emitted by the source (the preparer) are non-orthogonal (not perfectly distinguishable). As an example, let us consider the states $\rho_{x_1} = |0\rangle\langle 0|$, $\rho_{x_2} = 1/2(|0\rangle\langle 0| + |1\rangle\langle 1|)$ emitted each with a probability $p_{x_1} = p_{x_2} = 1/2$. Then $\rho = 3/4|0\rangle\langle 0| + 1/4|1\rangle\langle 1|$. Due to indistinguishability of preparations of a mixture in quantum mechanics, the source given by ρ can be equally well interpreted as one emitting states $\rho_{y_1} = |0\rangle\langle 0|$ and $\rho_{y_2} = |1\rangle\langle 1|$ with probabilities $p_{y_1} = 3/4 = \lambda_1$ and $p_{y_2} = 1/4 = \lambda_2$.

We have derived two different descriptions of the source, each with some distribution of emitted "symbols". Having a distribution, we can calculate the Shannon entropy of the source, but which is the correct Shannon entropy of the source (of ρ)? Is it that given by $\{p_{x_i}\}$, or by $\{p_{y_i}\}$? Let us look at the classical case, where we can construct an analogous case. In the classical case, we could have a source emitting totally non-distinguishable (which is the extremal case of imperfect distinguishability) symbols a_1, a_2 with probabilities $p_{a_1} = p_{a_2} = 1/2$. Since a_1, a_2, are indistinguishable, we have $a_1 = a_2 = b_1$ with $p_{b_1} = 1$ (to have the same number of symbols, we could also introduce b_2 with $p_{b_2} = 0$). Certainly, the correct Shannon entropy of the above classical source is given by $H(\{p_{b_i}\}) = 0$, i.e. by the entropy of the source when we consider it as a source of *perfectly distinguishable* symbols. Analogously, in the quantum case we define the entropy as the entropy of the distribution of perfectly distinguishable states, which is just the von Neumann entropy, Def. 2.26. The fact that even for $p_{x_i} = 1/d$ (d being the alphabet size) with non-orthogonal symbols ρ_x the von Neumann entropy is less than the maximum entropy of a source randomly emitting d symbols is referred to as quantum redundancy of the non-orthogonal encoding (since a compression to $2^{S(\rho)} < 2^{\log_2 d}$ bits per symbol is possible in this case).

[2.8] According to our definition of instruments (Def. 2.14) this is not precise. Although the right-hand side of Eq. (2.42) defines an operation for each of the projectors $|e_i\rangle\langle e_i|$, to define an instrument we would need to specify a measurable space $(\Omega, \mathcal{F})$ of outcomes and also to which elements of the σ-algebra $\mathcal{F}$ the projectors $|e_i\rangle\langle e_i|$ correspond (i.e. a POVM needs to be specified). Only then, strictly speaking, Eq. (2.42) defines a mapping $\mathcal{F} \to \mathcal{L}(\mathcal{T}(\mathcal{H}))$, i.e. an instrument according to our definition Def. 2.14.

[2.9] *By eigenvalues we mean here the diagonal elements of ρ in the basis in which it is diagonal, i.e., for instance, $\rho = 1/2(|0\rangle\langle 0| + |1\rangle\langle 1|)$ has "eigenvalues" $\lambda_1 = \lambda_2 = 1/2$ here. This jargon is often used in the literature and we will also happen to use it. Strictly speaking, according to our definition in Sec. 2.2.1, ρ would have a single eigenvalue $\lambda = 1/2$ (of degeneracy two).*

2.5.1.1 Capacity of the noiseless (ideal) quantum channel.

While classical channels can be used to transfer only classical information (classical states), quantum channels can transfer both classical and quantum information (quantum states encoding a piece of classical information, or quantum states as such). Later on, we will be considering a quantum channel to transmit classical information. For this we need to specify an encoding procedure which is a map $C\colon \mathcal{A} \to \mathcal{S}(\mathcal{H})$ from the set of classical messages represented by an alphabet $\mathcal{A}$ into the set of quantum states $\mathcal{S}(\mathcal{H})$

$$\{p_x, x\} \to \{p_x, \rho_x\} = \rho = \sum_{x \in \mathcal{A}} p_x \rho_x. \tag{2.43}$$

Let us consider the following situation, Figure (2.1):

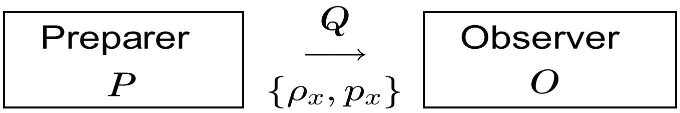

Figure 2.1. Communication of classical information over noiseless quantum channel.

A preparer P wants to transmit classical information (bits) to some other party O. He encodes the message (a ($\log_2 n$)-bit string, which can be viewed as a letter x in a n-letter alphabet) into a state of a quantum system (one of the n states ρ_x used to encode the different letters). The prepared state is then delivered (formally) via a noiseless quantum channel to the other party. The other party O then measures the quantum state and, based on one of the possible outcomes labeled by $y = 1, \ldots, m$, estimates which of the states ρ_x (which of the letters x) has been delivered (more precisely, rather than estimation, this task is called discrimination of quantum states in the estimation theory). The task is to quantify the maximum amount of classical information, i.e. the mutual information of the joint input-output distribution ("correlation" between the input and output random variables), that can be transferred via the noiseless quantum channel, i.e. accessed by the observer O. The upper bound on the accessible information is given by the Holevo bound:

Theorem 2.27. *(The Holevo bound) Consider states ρ_x, where $x = 0, \ldots, n$, prepared with probabilities p_x. Further consider a measurement with outcomes $y = 0, \ldots, m$ is performed on the prepared state. For any such measurement,*

$$I(X;Y) \leq S(\rho) - \sum_x p_x S(\rho_x) =: \chi_H, \tag{2.44}$$

where $\rho = \sum_x p_x \rho_x$. Where S is the Von Neumann entropy (see Def. 2.26). The quantity χ_H on the right-hand side of the inequality (2.44) is called the Holevo (χ_H) quantity.

Proof. Let us, analogously to the classical case, introduce the quantum mutual information (mutual Von Neumann entropy)

$$S(P;Q) = S(\rho_P) + S(\rho_Q) - S(\rho_{PQ}),$$

where

$$\begin{aligned} \rho_{PQ} &= \sum_x p_x |x\rangle\langle x| \otimes \rho_x && \text{(with } \{|x\rangle\} \text{ orthonormal)} \\ \rho_P &= \sum_x p_x |x\rangle\langle x| &\Rightarrow\ & S(\rho_P) = H(X) \\ \rho_Q &= \sum_x p_x \rho_x \equiv \rho &\Rightarrow\ & S(\rho_Q) = S(\rho). \end{aligned}$$

Since ρ_{PQ} is a block-diagonal matrix we have

$$S(\rho_{PQ}) = H(X) + \sum_x p_x S(\rho_x).$$

Thus,

$$S(P;Q) = S(\rho) - \sum_x p_x S(\rho_x) \ \equiv \ \chi_H,$$

i.e. the Holevo bound is the quantum mutual information between the preparer (as a physical system) and the quantum system Q.

Before the measurement is performed by the observer, the state of the total system is (if the prepared "symbol" is unknown)

$$\rho_{PQO} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x \otimes |y\rangle\langle y|,$$

where $|x\rangle\langle x|$ and $|y\rangle\langle y|$ are the possible internal (classical) states of the preparer and the observer, respectively. If a measurement described by POVM elements E_y (subset of the range of the POVM E such that $\sum_y E_y = \mathbb{1}$) is performed by the observer, the state after the measurement (with outcome y unknown) is

$$\rho'_{PQO} = \sum_{x,y} p_x |x\rangle\langle x| \otimes \mathcal{I}_y(\rho_x) \otimes |y\rangle\langle y|,$$

where $\mathcal{I}$ is an E-compatible (i.e. $\mathrm{Tr}[E(y)\rho] = \mathrm{Tr}[\mathcal{I}_y \rho]$) instrument (given by the internal workings of the measurement apparatus) associated with the measurement outcome corresponding to the eigenstate $|y\rangle\langle y|$. Tracing over the system Q (which is not of interest since we want to quantify correlations only between P and O irrespective of Q) we get

$$\rho'_{PO} = \sum_{x,y} p_x |x\rangle\langle x| \otimes \underbrace{\mathrm{Tr}[E(y)\rho_x]}_{p(y|x)} \otimes |y\rangle\langle y|,$$

i.e. ρ'_{PO} is a diagonal matrix with eigenvalues $p(x,y) = p(y|x)p(x)$. For the joint distribution and marginals we get $S(\rho'_{PO}) = H(X,Y)$, $S(\rho'_P) = H(X)$, and $S(\rho'_O) = H(Y)$. Hence

$$S(P';O') = I(X;Y) \ = \ \textit{accessed}\ \text{information}.$$

The Holevo bound then says that the accessed information is upper-bounded by the accessible information χ_H, i.e.

$$S(P';O') \leq S(P;Q), \tag{2.45}$$

(which we still need to show). We will show (2.45) only schematically:

$$S(P;Q) = S(P;QO) = S(P;QOA) = S(P';Q'O'A') \geq S(P';Q'O) \geq S(P';O'),$$

where we have used the following facts

- adding uncorrelated pure state (O and then A) conserves entropy (the ancillary system A is added so that the POVM (most general measurement) performed on Q is explicitly seen as PVM on QA, the coupling of the apparatus O with QA being expressed by a unitary evolution U on the combined system QOA)

- a unitary operation conserves entropy; thus

$$\begin{aligned}S(P;QOA) &= S(\rho_P)+S(\rho_{QOA})-S(\rho_{PQOA})\\ &= S(\rho_P)+S(U\rho_{QOA}U^\dagger)-S(\mathbb{1}\otimes U\rho_{PQOA}\mathbb{1}\otimes U^\dagger)\\ &= S(\rho_{P'})+S(\rho'_{QOA})-S(\rho'_{PQOA})\\ &= S(P';Q'O'A')\end{aligned}$$

- discarding a subsystem (A and then O) cannot increase the mutual information □

Remark 2.28. On the meaning of the Holevo bound. We have seen that the Holevo bound is given by the (quantum) mutual information between the classical state of the preparer (who can distinguish all symbols of the alphabet) and the quantum state used to encode the symbols. The (pure) quantum states used to represent (encode) the symbols may be indistinguishable – either just because such encoding has been chosen, or simply because the dimension of the quantum system is smaller than the number of symbols – causing $S(\rho_Q) < S(\rho_P)$. Moreover, encoding could be done into non-pure states, causing $\sum_x p_x S(\rho_x) > 0$. Both of the above may[2.10] decrease distinguishability of the originally distinguishable encoded symbols (causing $S(\rho_P) \geq S(P;Q)$)). Even if Q had sufficient dimensionality and we encode into pure states, still, if we e.g. do not know the basis used for the encoding into distinguishable pure states of Q, a measurement cannot distinguish these states perfectly. This leads to a lower value of $S(P';O')$ than the value of $S(P';Q')$, that is, the accessed information ("correlation" between P' and O') is smaller than the accessible information ("correlation" between P' and Q'). Note that due to the central role of a decreased distinguishability in the interpretation of the Holevo bound, a formal analogue of the Holevo quantity (and bound) can be constructed also in the classical case. Consider a classical channel (i.e. transmission of states diagonal in some fixed basis). Consider, for example, a source of symbols z_i, $(i=1,...,4)$, with probabilities $p_{z_i}=1/4$. However, suppose both z_1 and z_2 are interpreted as x_1 at the output and, likewise, z_3 and z_4 are interpreted as x_3 (for instance due to the fact that there is a noise in the channel such that errors exchanging $z_1 \leftrightarrow z_2$ and $z_3 \leftrightarrow z_4$ happen with probability 1/2 causing the meaning of z_j and z_{j+1} $(j=1,3)$ to be indistinguishable, i.e. $I(Z^{j,j+1}_{\text{input}};Z^{j,j+1}_{\text{output}})=0$). Hence, we have, $p_{x_j}=1/2$, $(j=1,3)$. This means that part of the entropy of the source is "wasted" within the encoding of the symbols x_j, and thus the entropy of the source as a source of symbols x_j (transmitted formally through a noiseless channel) is

$$\begin{aligned}H(X) &= I(X;X)\\ &= H(\{p_{z_i}\})-\sum_{j=1,3} p_{x_j}H(\{p_{x_j},p_{x_{j+1}}\})\\ &= H(Z)-\sum_x p_x H(X), \qquad (2.46)\end{aligned}$$

[2.10] Note that encoding into mixed states does not necessarily decrease distinguishability. One could, e.g., encode each of two (classical) symbols $\{i\}=\{1,3\}$ into two of four pure orthogonal states ρ_j of a four-dimensional quantum system with probability 1/2, i.e. $i \to \rho_i$ with probability 1/2 and $i \to \rho_{i+1}$ with probability 1/2. This is described formally by the encoding $i \to 1/2(\rho_i + \rho_{i+1}) =: \rho'_i$, i.e. encoding into two mixed states ρ'_i. Then $\sum_i p_i S(\rho'_i) > 0$, however this only compensates (in the expression for the Holevo bound, Eq. (2.44)) for the increase of $S(\rho) \equiv S(\rho_Q)$ because in our example $S(\rho_Q) > S(\rho_P)$ due to our way of encoding which introduces extra entropy into ρ_Q which is only "wasted" within the mixed states ρ'_i. The distinguishability of the states i and ρ'_i stays perfect (as the states ρ'_i are from orthogonal subspaces), thus $1 = S(\rho_P) = S(P;Q) = S(\rho) - \sum_x p_x S(\rho_x) = 2 - 1 = 1$ in this specific example.

which is the mutual information of a random variable X with itself. Should X (physically the original variable Z) be further transmitted via a noisy channel output of which we denote by a random variable Y (or should X be, for any other reason, only partially correlated with some other random variable Y), then the mutual information of X and Y is of course bounded by the mutual information of X with itself, i.e.

$$I(X;Y) = H(X) - \underbrace{H(X|Y)}_{\geq 0} \leq H(X). \tag{2.47}$$

Hence,

$$I(X;Y) \leq H(Z) - \sum_x p_x H(X). \tag{2.48}$$

Rewriting the Eq. (2.48) in density-matrix formalism, we have

$$I(X;Y) \leq S(\rho) - \sum_x p_x S(\rho_x),$$

where $\rho = \mathrm{diag}\{p_{z_i}\}$ and $\rho_{x_1} = \mathrm{diag}\{p_{z_1}, p_{z_2}, 0, 0\}/p_{x_1} = \mathrm{diag}\{p_{x_1}, p_{x_1}, 0, 0\}$, $\rho_{x_3} = \mathrm{diag}\{0, 0, p_{z_3}, p_{z_4}\}/p_{x_3} = \mathrm{diag}\{0, 0, p_{x_3}, p_{x_3}\}$ in our example.

Chapter 3

Extraction of information from finitely many instances of a quantum system

3.1 Introduction

In the minimal interpretation of quantum mechanics, there is a consensus that all information on a quantum system is contained in the quantum state (in the sense that it provides the right outcome probabilities for each conceivable measurement performed on the system). Since all this information is not accessible by a measurement on a single copy of the system (see, e.g., [26]), the meaning of quantum state has been traditionally associated to an infinite ensemble of identically prepared quantum systems (something which cannot be taken literally, but only as a conceptual notion). Advanced experiments with individual quantum systems (see, e.g., [15, 16]) and the advent of quantum information have brought the focus to individual systems, away from the infinite ensemble picture.

Due to the work of Helstrom [38] (see also Ref. [39]) we have means to quantify the amount of information we can obtain by performing measurements on individual quantum systems. In particular, we understand a limit of how much information on an elementary quantum system's state can be obtained if we perform a measurement on a *signal state* consisting of a single copy of the system or of a finite-size ensemble[3.1]. Naturally, going from a single copy to larger and larger ensembles, we should be able to get closer and closer to the complete information available in an infinite ensemble. How exactly the extractable information grows with the number of constituents of the ensemble as well as the question of what are the most informative measurements have been subject of research for different types of systems [47, 14, 31, 44].

Once we consider a finite ensemble (i.e. copies of a state), we suppose a collection of the finite number of *instances* (i.e. copies of a Hilbert space) of a quantum system is available. An interesting question to ask is then what are the signal states and measurements, taking advantage of the full Hilbert space provided by the instances, that yield extraction of maximum of information on a single instance's state, which is carried collectively by all the instances. These questions have been studied for different scenarios [17, 25, 3, 22, 23].

The above problems of (optimal) storing / retrieval of an elementary quantum system in / from the state space of finitely many of its instances will be of relevance for the subject of the present Thesis. Specifically, the relevant scenario of the estimation part is the following:

[3.1] 3.1. Sometimes, the term ensemble is used exclusively to denote an "infinite set of *conceptual* replicas of the same system, used for statistical argumentation" [51]. To denote a large number of identically prepared particles, then, the term *assembly* is used [ibid]. In this sense the term "finite ensemble" is an oxymoron. However, the term finite ensemble seems to be commonly used in other works, e.g. [47, 31], its denotation being that of the above-defined term assembly. We will use it in the same sense.

Suppose we have a quantum system in a state $\psi = |\psi\rangle\langle\psi|$, $|\psi\rangle \in \mathcal{H}$, $|\psi\rangle = g|\psi_0\rangle$, $g \in \mathcal{G}$. Suppose we have a system consisting of N instances of the quantum system. The overall state is prepared in a signal state $\rho_\psi \in \mathcal{S}(\mathcal{H}^{\otimes N})$ from a fiducial state $\rho_0 \in \mathcal{S}(\mathcal{H}^{\otimes N})$, via a unitary operation $U(g)$ which is an element from the range of a particular unitary finite-dimensional representation of a compact Lie group $\mathcal{G}$. An observer's task is to estimate the original state ψ by performing a measurement on the signal state ρ_ψ. The identity of the state ψ within the family $\{g|\psi_0\rangle,\ g \in \mathcal{G}\}$ is unknown to him, which implies that all the elements $g \in \mathcal{G}$ are equally probable. Given the class of signal states $\{U(g)\rho_0 U^\dagger(g),\ g \in \mathcal{G}\}$ and their (flat) a-priori probability distribution, the task is to find a measurement which, when applied to the system of *all N* instances, provides the best possible estimate of ψ. Such optimal measurements, given a particular (class of) figure(s) of merit, have been widely studied in the literature for various sets of signal states and can be of different nature. Two important types of optimal measurements are optimal covariant measurements and measurements with finite number of outcomes – an interesting subclass of the latter being minimal measurements, i.e. those with the minimal possible number of measurement outcomes.

The importance of the former is that given any optimal measurement, one can always[3.2] construct an optimal measurement that is covariant and attains the same value of the figure of merit with respect to which the optimality is considered. From that it follows that if one wants to find the value of the figure of merit at the optimum (an extremum of the figure of merit), one can restrict himself to the covariant measurements. The covariant quantum estimation problem is the subject of the Section 3.2.

From the point of view of experimental realizability of measurements, the measurements with finite (particularly with the minimal) number of POVM elements are important. In Section 3.3 the case of N-copies ensemble and finite measurements is overviewed. A universal algorithm for constructing optimal POVM's with finite number of elements for arbitrary finite-dimensional system of N identically prepared subsystems was introduced by Derka, Bužek, and Ekert [31]. For systems of two qubits (spin-1/2 systems), projective measurements shown to be optimal have been constructed [47] by Massar and Popescu. Acín, Latorre, and Pascual provide an optimal and minimal measurement for systems of two spin-j systems, and a lower bound on the number of POVM elements in the case of three spin-j systems [1].

In Section 3.4 known results [3] on the optimal N-instances state and measurements, which we will take advantage of later, are summarized.

3.2 Covariant quantum estimation

In this section we describe in detail the covariant quantum estimation problem, i.e. the problem of estimating the state of a quantum system from a set of states related by elements of the representation space of a unitary representation of a compact Lie group. We elaborate on the case of an irreducible representation and, in particular, we describe the optimal covariant measurement of parameters of orientation (a direction encoding). The whole section, apart from the Remark 3.9 and the proof of the Theorem 3.7, essentially follows the exposition of the Holevo's book [39]. An excellent and up-to-date overview of the covariant estimation problem, including more general settings than the ones discussed in this presentation, can be found in [21].

[3.2] More precisely, always when the figure of merit has certain (rather natural) properties (see Sec. 3.2.1).

Let us begin with a definition of a covariant POVM.

Definition 3.1. *Covariant POVM.*

Let $\mathcal{G}$ be a parametric group of transformations of a set Θ and $g \mapsto U_g$ be a continuous projective unitary representation of $\mathcal{G}$ in a Hilbert space $\mathcal{H}$. Let M be a measurement[3.3] *$\mathcal{B}(\Theta) \to \mathcal{L}(\mathcal{H})$, where $\mathcal{B}(\Theta)$ is a σ-algebra of Borel subsets of Θ. The measurement M is* covariant *with respect to the representation $g \mapsto U_g$ if*[3.4]

$$U_g M(B) U_g^{\dagger} = M(B_g), \quad g \in \mathcal{G} \tag{3.1}$$

for any $B \in \mathcal{B}(\Theta)$, where $B_g = \{\theta: \theta = g\theta', \theta' \in B\}$ is the image of the set B under the transformation g.

In other words this means that if we transform a state so that $\rho \mapsto U_g^{\dagger} \rho U_g$, for all $g \in G$ the probabilities of (all the) outcomes after the transformation should be the same as the probabilities of (all of the) outcomes "shifted" via the action of g^{-1} before the transformation (provided the measuring apparatus has not been touched). Equivalently, covariant-POVM effects corresponding to outcomes related by (any) $g \in \mathcal{G}$ should only be transformed by conjugation of U_g, if we transformed the apparatus and did not touch the state.

Let us now recall some general knowledge on parametric groups of transformations:

Definition 3.2. *Transitive action of a group.*

A group $\mathcal{G}$ acts transitively on Θ if any point θ_0 can be transformed into any other point θ by some $g \in \mathcal{G}$. In such a case the continuous mapping $g \longrightarrow \theta(g) = g\theta_0$ maps $\mathcal{G}$ onto the whole Θ. This mapping is one-to-one if and only if the stationary subgroup $\mathcal{G}_0$ ($\mathcal{G}_0 = \{g \in \mathcal{G}: g\theta_0 = \theta_0\}$) consists only of the identical transformation.

In what follows we consider the above transitivity.

Definition 3.3. *Left-(right-)invariant measure.*

A measure $\mu(\mathrm{d}g)$ on the σ-algebra $\mathcal{B}(\mathcal{G})$ of Borel subsets of $\mathcal{G}$ is left-(right-)invariant if

$$\mu(gA) = \mu(A) \quad (\,\mu(Ag) = \mu(A)\,), \;\; g \in \mathcal{G}, \tag{3.2}$$

where $gA = \{g': g' = gg'', g'' \in A\}$, ($Ag = \{g': g' = g''g, g'' \in A\}$), and $A \in \mathcal{B}(\mathcal{G})$.

Any compact group has a finite, invariant (i.e. left- and right-invariant), measure, which we will normalize so that $\mu(\mathcal{G}) = 1$. Assuming the stationary subgroup $\mathcal{G}_o$ is compact, then μ induces an invariant measure ν on the σ-algebra $\mathcal{B}(\Theta)$ by the relation

$$\nu(B) = \mu(\theta^{-1}(B)), \tag{3.3}$$

where $\theta^{-1}(B) = \{g: g\theta_0 \in B\}$ is the pre-image of the Borel set $B \in \mathcal{B}(\Theta)$. By invariance of ν we mean

$$\nu(B_g) = \nu(B); \;\; g \in \mathcal{G}, \;\; B \in \mathcal{B}(\Theta). \tag{3.4}$$

[3.3] *Here we adopt the terminology used by Holevo in Ref. [39]. $M(B) = \int_{B \in \mathcal{B}(\Theta)} M(d\theta)$ defines a POVM.*

[3.4] *Note that, equally well, the condition could have been $U_g^{\dagger} M(B) U_g = M(B_g)$, had we chosen the representation $g \rightarrow V_g = U_g^{\dagger} = U_{g^{-1}}$*

The invariance of ν follows from the left-invariance of μ, since $\nu(B_g) = \mu(\theta^{-1}(B_g)) = \mu(\theta^{-1}(gB)) \equiv \mu(g\theta^{-1}(B)) = \mu(\theta^{-1}(B)) = \nu(B)$. The right-invariance of μ implies that ν is the same for any choice of θ_0 since if $\theta_0' = g'\theta_0$, then

$$\nu(B) \;=\; \mu(\{g\colon g\theta_0 \in B\}) = \mu(\{gg'\colon gg'\theta_0 \in B\}) = \mu(\{g\colon g\theta' \in B\}). \tag{3.5}$$

Next, we look at the structure of the covariant measurements:

Lemma 3.4. *(Radon-Nikodym theorem for operator-valued measures) Let $\{M(B); B \in \mathcal{B}(\Theta)\}$ be an additive operator-valued set function dominated by the scalar measure $\{m(B); B \in \mathcal{B}(\Theta)\}$ in the sense that*

$$|\langle\varphi|M(B)|\psi\rangle| \leq m(B)\|\varphi\|\|\psi\|, \quad B \in \mathcal{B}(\Theta),$$

for all φ, $\psi \in \mathcal{H}$. Then there exists an operator-valued function $\tilde{M}(.)$ defined uniquely for m-almost all $\theta \in \Theta$ (i.e. for all θ with possible exception of a set of zero measure m), satisfying $\left\|\tilde{M}(\theta)\right\| \leq 1$ and such that

$$\langle\varphi|M(B)|\theta\rangle = \int_B \left\langle\varphi\middle|\tilde{M}(\theta)\middle|\psi\right\rangle m(d\theta), \quad B \in \mathcal{B}(\Theta), \tag{3.6}$$

for all $\varphi, \psi \in \mathcal{H}$. If $M(B) \geq 0$ for all $B \in \mathcal{B}(\Theta)$, then $\tilde{M}(\theta) \geq 0$ for m-almost all θ.

The function $\tilde{M}(\theta)$ is called the operator density of $M(d\theta)$ with respect to $m(d\theta)$. To form a POVM, we have $M(B) = \int_B \tilde{M}(\theta)m(d\theta)$, i.e. $M(d\theta) = \tilde{M}(\theta)m(d\theta)$ (with weak convergence of the integral).

With the aid of the above Lemma, it is possible to set up an affine one-to-one correspondence (Eq. (3.9)) between the convex set of all covariant measurements $M(d\theta)$ and the convex set of Hermitian operators satisfying the conditions of the following theorem:

Theorem 3.5. *If $\tilde{M}_{\mathcal{G}_0}$ is a Hermitian positive operator in the representation space, commuting with the operators U_g from the stationary subgroup representation space ($\{U_g; g \in \mathcal{G}_0\}$) satisfying*

$$\int_{\mathcal{G}} U_g \tilde{M}_{\mathcal{G}_0} U_g^\dagger \mu(dg) = \mathbb{1}, \tag{3.7}$$

then the operator-valued function

$$\tilde{M}(\theta) := \tilde{M}(g\theta_0) = U_g \tilde{M}_{\mathcal{G}_0} U_g^\dagger, \tag{3.8}$$

induces a POVM

$$M(B) = \int_B \tilde{M}(\theta)\nu(d\theta), \quad B \in \mathcal{B}(\Theta) \tag{3.9}$$

that is covariant with respect to the representation $g \mapsto U_g$.

The converse is also true, i.e. for any covariant measurement $M(d\theta)$ there is a unique operator $\tilde{M}_{\mathcal{G}_0}$, satisfying the conditions of the theorem, such that $M(B)$ is expressed through $\tilde{M}_{\mathcal{G}_0}$ by Eqs. (3.8) and (3.9).

The operator $\tilde{M}_{\mathcal{G}_0}$ satisfying the conditions of the Theorem 3.5 is sometimes [21] called the *seed* of the POVM M.

3.2.1 Optimal covariant measurements

To talk about optimality of measurements, a set of states to estimate and a figure of merit have to be specified. The following (covariant estimation) problem is considered: Let θ be a, possibly multidimensional, parameter describing (some) aspects of preparation of a quantum system. To each value of $\theta \in \Theta$ there corresponds a quantum state ρ_θ from $\mathcal{S}(\mathcal{H})$. Assume there is a symmetry group $\mathcal{G}$ acting transitively on Θ, which has a representation $g \mapsto U_g$ in $\mathcal{H}$. The family of states $\{\rho_\theta\}$ is invariant[3.5] with respect to the representation if

$$\rho_{g\theta} = U_g \rho_\theta U_g^\dagger; \;\; \theta \in \Theta, \;\; g \in \mathcal{G}. \tag{3.10}$$

Fixing a reference state $\rho_{\theta_0} = \rho_0$ we can write

$$\rho_\theta = U_g \rho_0 U_g^\dagger, \quad (\theta = g\theta_0). \tag{3.11}$$

Assuming the object is prepared in one of the states ρ_θ with the parameter θ unknown, the task is to estimate the parameter as accurately as possible.

To evaluate how good the estimate is, a deviation function $W_\theta(\theta_{\text{est}})$, where θ_{est} is the estimated value of the real value of the parameter θ, is used. It is assumed that it is a continuous function of its arguments, with

$$W_\theta(\theta_{\text{est}}) \geq W_\theta(\theta). \tag{3.12}$$

It is also assumed that the deviation function is invariant, i.e.

$$W_{g\theta}(g\theta_{\text{est}}) = W_\theta(\theta_{\text{est}}); \;\; g \in \mathcal{G}; \;\; \theta, \theta_{\text{est}} \in \Theta. \tag{3.13}$$

For a specific value of θ the mean deviation of the measurement $M = \{M(d\theta_{\text{est}})\}$ reads

$$\mathcal{R}_\theta(M) = \int W_\theta(\theta_{\text{est}}) \mu_\theta(d\theta_{\text{est}}), \tag{3.14}$$

where $\mu_\theta(d\theta_{\text{est}}) = \text{Tr}(\rho_\theta M(d\theta_{\text{est}}))$. In general, minimizing the expression Eq. (3.14) for different values of θ might yield different optimal measurements (although it is not the case if one restricts himself to covariant measurements, as will be shown below – Theorem (3.6)). Therefore, in general, an optimal measurement should minimize a single functional of the quantities $\mathcal{R}_\theta(M)$, $\theta \in \Theta$. In classical estimation theory one can form two different functionals – Bayes mean deviation (in the Bayes' approach)

$$\mathcal{R}_\pi(M) = \int \mathcal{R}_\theta(M) \pi(d\theta), \tag{3.15}$$

where $\pi(\theta)$ is a known distribution of the random parameter θ, and maximal mean deviation (in the minimax approach)

$$\mathcal{R}(M) = \max_\theta \mathcal{R}_\theta(M). \tag{3.16}$$

The minimizing measurements are then called Bayesian and minimax, respectively. Recalling that in the problem definition the actual value of θ is totally random and Θ is compact, the distribution $\pi(d\theta)$ is the uniform distribution (normalized invariant measure) $\nu(d\theta)$.

[3.5] In some texts, e.g. [39], such family is referred to as covariant. What is meant then is that the group action may be considered as a an encoding of the group parameters into the signal states. It is this encoding that is covariant (in the same way as a covariant POVM, Def. 3.1).

Theorem 3.6. *In the quantum covariant statistical estimation problem described above (i.e. the problem of estimating a state from the invariant family of states (3.11)) the minima of the Bayes' mean deviation $\mathcal{R}_\nu(M)$ and the maximal mean deviation $\mathcal{R}(M)$ for all Θ-measurements are achieved on a covariant measurement. Moreover, for a covariant measurement M*

$$\mathcal{R}_\nu(M) = \mathcal{R}(M) = \mathcal{R}_\theta(M), \quad \theta \in \Theta. \tag{3.17}$$

Hence, if one restricts himself to covariant measurements, minimizing the mean deviation for any fixed value of θ, e.g. θ_0, does the minimization for all for all values of θ simultaneously.

Assuming a finite-dimensional representation, then, via the Theorem (3.5), the mean deviation (functional) to minimize has the form

$$\begin{aligned}\mathcal{R}_{\theta_0}(M) \quad &= \quad \int_\Theta W_{\theta_0}(\theta_{\text{est}})\mu_{\theta_0}(d\theta_{\text{est}}) = \int_\Theta W_{\theta_0}(\theta_{\text{est}})\text{Tr}(\rho_0 M(d\theta_{\text{est}})) \\ &\overset{\text{Th. (3.5)}}{=} \int_\Theta W_{\theta_0}(\theta_{\text{est}})\text{Tr}\Big(\rho_0 U_g \tilde{M}_{\mathcal{G}_0} U_g^\dagger\Big)\nu(d\theta_{\text{est}}) = \text{Tr}(W_0 \tilde{M}_{\mathcal{G}_0}),\end{aligned} \tag{3.18}$$

where the operator of posterior deviation

$$W_0 = \int_\Theta W_{\theta_0}(\theta_{\text{est}}) U_g^\dagger \rho_0 U_g \nu(d\theta_{\text{est}}) = \int_{\mathcal{G}} W_{\theta_0}(g\theta_0) U_g^\dagger \rho_0 U_g \mu(dg) \tag{3.19}$$

is an operator commuting with all $\{U_g; g \in \mathcal{G}_0\}$. Thus, the optimal measurement is the one achieving (through its corresponding operator $\tilde{M}_{\mathcal{G}_0}$) the minimum of the mean deviation (functional)

$$\mathcal{R}_{\min}(M) = \min \text{Tr}(W_0 \tilde{M}_{\mathcal{G}_0}), \tag{3.20}$$

where the minimization is done over all Hermitian operators $\tilde{M}_{\mathcal{G}_0}$ fulfilling the conditions of the Theorem (3.5).

3.2.2 The case of an irreducible representation

In the case of irreducible representation, the following theorem holds:

Theorem 3.7. *Let $g \mapsto U_g$ be an irreducible representation of a compact group $\mathcal{G}$ of transformations of the set Θ on a complex Hilbert space $\mathcal{H}$. Then there is a one-to-one affine correspondence between the covariant measurements M and the density operators $\rho_{\mathcal{G}_0}$ commuting with $\{U_{g'}; g' \in \mathcal{G}_0\}$, namely*

$$M(d\theta) = d \cdot U_g \rho_{\mathcal{G}_0} U_g^\dagger \nu(d\theta) \quad (\theta = g\theta_0, \theta \in \Theta), \tag{3.21}$$

where $d < \infty$ is the dimension of $\mathcal{H}$.

Proof. Compactness of the group $\mathcal{G}$ and irreducibility of the representation $g \to U_g$ imply d finite. Inspired by the Section (2.4), we construct the operator

$$A = \int_{\mathcal{G}} U_g \rho_{\mathcal{G}_0} U_g^\dagger \mu(dg). \tag{3.22}$$

For any fixed $g' \in \mathcal{G}$,

$$U_{g'} A U_{g'}^\dagger = \int_{\mathcal{G}} U_{g'g} \rho_{\mathcal{G}_0} U_{g'g}^\dagger \mu(dg) = \int_{\mathcal{G}} U_{g'g} \rho_{\mathcal{G}_0} U_{g'g}^\dagger \mu(d(g'g)) = A, \tag{3.23}$$

where we used the invariance of the measure $\mu(dg)$ with respect to the group $\mathcal{G}$. Following the arguments as in the Section (2.4), utilizing the irreducibility of the representation, A has to be a multiple of the identity operator, i.e.

$$A = c\mathbb{1}. \tag{3.24}$$

Taking trace of Eq. (3.22) and (3.24) together with the normalization of μ ($\mu(\mathcal{G}) = 1$) we get $cd = 1$. Hence $A = d^{-1}\mathbb{1}$ and from the Eq. (3.22) we have

$$\mathbb{1} = d \cdot A = \int_{\mathcal{G}} U_g d \cdot \rho_{\mathcal{G}_0} U_g^{\dagger} \mu(dg). \tag{3.25}$$

Comparing the last equation with the condition Eq. (3.7) of the Theorem (3.5), $d \cdot \rho_{\mathcal{G}_0}$ can be identified with the seed $\tilde{M}_{\mathcal{G}_0}$. The statement of the Theorem (3.5) is then exactly the expression Eq. (3.21). □

The covariant estimation problem of the Subsection (3.2.1) for the states Eq. (3.11), i.e. finding the minimum (3.20), has the following solution: Let $w_{\min}$ be the smallest eigenvalue of the operator of posterior deviation (3.19), W_0, and $E_{\min}$ be the projection onto the corresponding subspace. Then $W_0 \geq w_{\min}\mathbb{1}$ and

$$\mathrm{Tr}(W_0 \tilde{M}_{\mathcal{G}_0}) \geq w_{\min} \mathrm{Tr}(\tilde{M}_{\mathcal{G}_0}) = w_{\min} d, \tag{3.26}$$

with equality when the seed $\tilde{M}_{\mathcal{G}_0}$ is proportional to the projection onto the smallest-eigenvalue eigenspace, i.e. when

$$\tilde{M}_{\mathcal{G}_0} = \frac{d}{d_{\min}} E_{\min}, \tag{3.27}$$

where $d_{\min}$ is the dimension of the smallest-eigenvalue eigenspace. $E_{\min}$ commutes with $\{U_g; g \in \mathcal{G}_0\}$ since W_0 does. The operator (3.27) fulfills the conditions of the Theorem (3.5), which then gives us the optimal covariant measurement:

Theorem 3.8. *Let $g \to U_g$ be an irreducible representation of a compact group $\mathcal{G}$ of transformations of the set Θ on a complex Hilbert space $\mathcal{H}$. The optimal covariant measurements $M_{\mathrm{opt}}(d\theta)$ of the parameter $\theta \in \Theta$ is given by*

$$M_{\mathrm{opt}}(d\theta) = \frac{d}{d_{\min}} U_g E_{\min} U_g^{\dagger} \nu(d\theta) \qquad (\theta = g\theta_0), \tag{3.28}$$

where $E_{\min}$ is the projection onto the smallest-eigenvalue ($w_{\min}$) eigenspace of the operator of posterior deviation (3.19), W_0. The minimal mean deviation is

$$\mathcal{R}_{\min}(M) = w_{\min} d. \tag{3.29}$$

3.2.3 Measuring orientation – spin representations

Consider the estimation of an orientation of a quantum object by measurements involving only spin degrees of freedom. Assume a spin-j object prepared in a fiducial (seed) state

$$\rho_0 = \sum_{m=-j}^{j} s_m |m\rangle\langle m|, \tag{3.30}$$

where $|m\rangle$ are the eigenstates of the spin angular momentum operator $J_{\boldsymbol{n}_0}$ corresponding to the axis $\boldsymbol{n}_0$. The state Eq. (3.30) is invariant under rotations about the axis $\boldsymbol{n}_0$, since $J_{\boldsymbol{n}_0}$ is the generator of rotations about this symmetry axis in $\mathbb{R}^3$, i.e. $\mathcal{G}_0 = \mathrm{SO}(2)$ around the axis $\boldsymbol{n}_0$.

If the preparation apparatus is rotated so that the symmetry axis is transformed to $\boldsymbol{n} = g\boldsymbol{n}_0$, where g is an element of the rotation group, the new prepared state will read $\rho_{\boldsymbol{n}} = U_g \rho_0 U_g^\dagger$, where $g \mapsto U_g$ is the irreducible projective unitary representation of the rotation group in the $(2j+1)$-dimensional Hilbert space spanned by $\{|m\rangle\}$. Orientation of the object is now described by the unit vector $\boldsymbol{n} \in \mathbb{S}^2$ pointing in the direction of the symmetry axis, where $\mathbb{S}^2$ is the unit sphere in $\mathbb{R}^3$.

Assuming the direction $\boldsymbol{n}$ is unknown, the task is to estimate it using the quantum measurement M.

According to the Theorem (3.7) any covariant measurement of the orientation $\boldsymbol{n}$ reads

$$M(\mathrm{d}\boldsymbol{n}) = (2j+1) U_g \rho_{\mathcal{G}_0} U_g^\dagger \nu(\mathrm{d}\boldsymbol{n}) \qquad (\boldsymbol{n} = g\boldsymbol{n}_0), \tag{3.31}$$

where $\rho_{\mathcal{G}_0}$ commutes with $\{U_g; g \in G_0\}$.

Choosing the (invariant) deviation function

$$W(\boldsymbol{n}, \boldsymbol{n}') = |\boldsymbol{n} - \boldsymbol{n}'|^2 = 2(1 - \boldsymbol{n} \cdot \boldsymbol{n}') \tag{3.32}$$

as the figure of merit of the quality of the estimation, we can express the operator of posterior deviation as

$$W_0 = 2 \int_G (1 - \boldsymbol{n}_0 . g\boldsymbol{n}_0) U_g \rho_0 U_g^\dagger \mu(\mathrm{d}g), \tag{3.33}$$

where $\mu(\mathrm{d}g)$ is the normalized invariant measure on the rotation group. It can be shown ([39], page 211) that the operator of the posterior deviation Eq. (3.33) is equal to

$$W_0 = \frac{2}{2j+1} \left(\mathbb{1} - \frac{\mathrm{Tr}(\rho_0 J_{\boldsymbol{n}_0})}{j(j+1)} J_{\boldsymbol{n}_0} \right). \tag{3.34}$$

According to the Theorem (3.8), the optimal covariant measurement of direction of the symmetry axis $\boldsymbol{n}$ is

$$\begin{aligned} M_{\mathrm{opt}}(\mathrm{d}\boldsymbol{n}) &= (2j+1) U_g | \pm j, \boldsymbol{n}_0 \rangle \langle \pm j, \boldsymbol{n}_0 | U_g^\dagger \nu(\mathrm{d}\boldsymbol{n}) \\ &= (2j+1) | \pm j, \boldsymbol{n} \rangle \langle \pm j, \boldsymbol{n} | \nu(\mathrm{d}\boldsymbol{n}) \qquad (\boldsymbol{n} = g\boldsymbol{n}_0), \end{aligned} \tag{3.35}$$

where $|m, \boldsymbol{n}\rangle$ denotes an eigenstate of the spin angular momentum operator in the direction $\vec{n}$ and the ± sign corresponds to the sign of $\overline{J_{\boldsymbol{n}_0}} := \mathrm{Tr}(\rho_0 J_{\boldsymbol{n}_0})$.

According to the previous result Eq. (3.29), one needs to find the minimum eigenvalue of the operator of posterior deviation W_0, Eq. (3.34), which is achieved via the largest eigenvalue of the operator $J_{\boldsymbol{n}_0} = \sum_{m=-j}^{j} m |m\rangle\langle m|$ of the spin component along the $\boldsymbol{n}_0$ direction, i.e. j. The minimal mean deviation is then equal to

$$\mathcal{R}_{\min}(M) = w_{\min} d = 2 \left(1 - \frac{|\overline{J_{\boldsymbol{n}_0}}|}{j+1} \right), \tag{3.36}$$

where $d = 2j+1$.

Remark 3.9. Choosing the "deviation" function $W' =: f$ (which we will call (one-qubit) fidelity)

$$W'(\boldsymbol{n}, \boldsymbol{n}') = 1 - \frac{1}{4} W(\boldsymbol{n}, \boldsymbol{n}') = \frac{1}{2}(1 + \boldsymbol{n} \cdot \boldsymbol{n}') := f(\boldsymbol{n}, \boldsymbol{n}'), \tag{3.37}$$

one gets

$$W_0' = \frac{1}{2(2j+1)}\left(\mathbb{1} + \frac{\mathrm{Tr}(\rho_0 J_{\boldsymbol{n}_0})}{j(j+1)} J_{\boldsymbol{n}_0}\right) := f_0. \tag{3.38}$$

In analogy with the minimal mean deviation Eq. (3.34), we can construct maximal mean fidelity

$$\bar{f}_{\max}(M) = \max \mathrm{Tr}(f_0 \tilde{M}_{\mathcal{G}_0}) = \lambda_{\max} d = \frac{1}{2}\left(1 + \frac{|\overline{J_{\boldsymbol{n}_0}}|}{j+1}\right), \tag{3.39}$$

where $\lambda_{\max}$ is the largest eigenvalue of the operator f_0, Eq. (3.38), which is again achieved via the largest eigenvalue of the operator $J_{\boldsymbol{n}_0}$. Therefore, the measurement realizing $\bar{f}_{\max}$, Eq. (3.39), is the one realizing the minimal mean deviation Eq. (3.36), i.e. the M_{opt} of the Eq. (3.35).

3.3 Estimation via finite measurements and minimal optimal measurements

We have seen in the previous Section that, for the covariant estimation problem, covariant measurements are especially useful for computing the extremal values of the figures of merit with respect to which they are optimal. The drawback of covariant measurements is the number of possible outcomes, which can be (uncountably) infinite (e.g. for a Lie group $\mathcal{G}$), which is the reason why they are sometimes considered unrealizable (e.g. [31]). Quite recently, Chiribella et al. have shown [24] that for any finite-level system any quantum measurement with a continuous set of outcomes is equivalent to a continuous random choice of measurements with a finite number of outcomes. (That is we have an apparatus with finitely many outcomes that we, for instance, rotate prior to the measurement, where the rotation parameters are chosen at random from a continuous set[3.6].) Moreover, the authors prove that any continuous measurement that optimizes some convex figure of merit (e.g., maximizes the mutual information or the Fisher information or, alternatively, minimizes a Bayes cost [38, 39]) can be always replaced by a single measurement with finite outcomes, without affecting optimality (see also [21]).

Hence, at least from the practical point of view, optimal measurements with finitely many outcomes are very important. A universal algorithm for constructing such finite optimal measurements on a finite number of identically prepared finite-dimensional systems has been derived by Derka et al. [31].

3.3.1 Measuring orientation – spin representations

The task of finding the minimal number of outcomes an optimal measurement may have, as well as the explicit form of the minimal optimal measurements, i.e. optimal measurements with minimal number of outcomes, are the subject of the papers by Massar and Popescu [47] and by Latorre et al. [43]. In the rest of the current section we briefly summarize their results.

[3.6] The realizability of (this particular implementation of) continuous-number-of-outcomes measurements then reduces to the possibility of randomly choosing an element from a *continuous* set in practice.

The paper [47] considers N copies of a spin-(1/2) particle in a pure state. As it's name suggests, a spin-(1/2) system in a pure state has a spin component 1/2 with regard to some axis. Hence, given a convention on its orientation, it encodes a direction (given by the axis) in a three-dimensional space. Naturally, N copies of such a system also encode the direction.

The following situation is considered. Given the above direction encoding, how well can one, based on a measurement outcome, estimate the encoded direction? Denoting by $\boldsymbol{n}$ the encoded direction and by $\boldsymbol{n}'$ the estimated direction, they consider the fidelity

$$f = \frac{1}{2}(1 + \boldsymbol{n} \cdot \boldsymbol{n}') \tag{3.40}$$

as the figure of merit of how good the estimate has been. The following questions are answered:

1. What is the maximum of the average score (maximum of the average of the fidelity over all encoded and estimated directions) that is achievable?
2. Which are the optimal measurements?
3. Does the optimal measurement have to treat the system as a whole?

The authors find equations that have to be fulfilled by an optimal measurement and evaluate the maximum of the achievable average fidelity as a function of the number of copies, N,

$$\bar{f}_N^{\max}(\boldsymbol{n}, \boldsymbol{n}') = \frac{N+1}{N+2}. \tag{3.41}$$

They argue there exist optimal measurements (ones attaining the bound Eq. (3.41)) with finitely many outcomes. In the case of $N = 2$ they explicitly construct the optimal measurement which is minimal, i.e. with least measurement outcomes (in this case four). The constructed optimal measurement is one with the eigenstates

$$\begin{aligned} |\psi_1\rangle &= \frac{1}{2}|S\rangle + e^{i\phi}\frac{\sqrt{3}}{2}|\boldsymbol{m}_1\boldsymbol{m}_1\rangle \\ |\psi_j\rangle &= \frac{1}{2}|S\rangle - e^{i\phi}\frac{\sqrt{3}}{2}|\boldsymbol{m}_j\boldsymbol{m}_j\rangle,\ j = 2, 3, 4, \end{aligned} \tag{3.42}$$

where $|S\rangle$ is the singlet, and $|\boldsymbol{m}_i\rangle$ are vectors pointing to the corners of a tetrahedron on a unit sphere. The corresponding eigenvalues have to be non-degenerate so that all four eigenstates are distinguished.

In the case of arbitrary N, it is shown that the covariant measurements Eq. (3.35) achieve the maximum of the average fidelity. This result, together with the bound Eq. (3.41), however, follows from the results by Holevo presented in the Section 3.2.3. It suffices to realize (which the authors do, as they utilize this fact) that it suffices to restrict oneself to the (($N+1$)-dimensional) totally symmetric subspace of the Hilbert space of N spins which the states of N copies of a spin-(1/2) system are restricted to. Then the problem can be mapped to the problem of estimating orientation of a spin-j system ($j = N/2$), which is discussed in the Section 3.2.3 (see also the Remark 3.9 for adapting the result by Holevo the case of the figure of merit Eq. (3.40) considered here).

The main result of the paper [47] is that there exist no optimal measurements that consist of separate measurements on each copy of the spin even if adaptive measurements are allowed (a measurement can be adapted based on the outcomes of the previous measurements). Thus, an optimal measurement of copies of a spin has to treat the system as a whole.

For the sake of completeness, let us now summarize the results of Latorre et al. presented in the paper [43]. The authors propose optimal and minimal quantum measurements for N copies of a two-level system. They explicitly construct the measurements up to $N=7$, the minimality is proved only up to $N=5$, however. They also suggest an expression for the minimal number of measurement outcomes, which reproduces all their results, and is supported by further arguments.

3.4 Measuring orientation – the optimal covariant encoding

It is well known that the finite ensemble of spin-1/2 particles is not the best choice of a signal state for transmitting the state of a single spin-1/2 system. It has been pointed out by Gisin and Popescu [36] that if two instances of a spin-1/2 system are available, the anti-parallel encoding

$$|\boldsymbol{n}\rangle\langle\boldsymbol{n}| \;\mapsto\; |\boldsymbol{n}\rangle\langle\boldsymbol{n}|\otimes|-\boldsymbol{n}\rangle\langle-\boldsymbol{n}| \tag{3.43}$$

is better than the symmetric one considered in the previous Section. Using a POVM with effects whose eigenstates are

$$|\psi_j\rangle = \alpha|\boldsymbol{n},-\boldsymbol{n}\rangle - \beta\sum_{k\neq j}|\boldsymbol{n}_k,-\boldsymbol{n}_k\rangle, \tag{3.44}$$

where $\alpha = 13/(6\sqrt{6}-2\sqrt{2})$ and $\beta = \alpha(5-2\sqrt{3})/13$, the average fidelity reads

$$\bar{f}_N^{\uparrow\downarrow} = (3+\sqrt{3})/6, \tag{3.45}$$

which is greater than the 3/4 given by Eq (3.41) for two parallel spins. The reason why this is so is the larger dimensionality, d, of the Hilbert space spanned by the set of two arbitrarily rotated anti-parallel spin states ($d=4$) as opposed to the set of two arbitrarily rotated parallel spin states ($d=3$) [3]. The fidelity Eq. (3.45) is the maximal possible for a U-covariant encoding of a single qubit into two qubits [3, 46], U being the symmetric representation $g\mapsto g\otimes g, g\in\mathrm{SU}(2)$.

A generalization to N instances of a spin-1/2 system was given by Bagan et al. [3]. The signal states that lead to the maximal fidelities are among those that have the smallest non-negative values of the total-spin component along $\boldsymbol{n}$, namely, $m=0$ for N even and $m=1/2$ for N odd, but still span the largest Hilbert space under rotations. For odd N

$$\mathcal{F}_N = \frac{1}{2}\Big(1 + x^{0,1}_{N/2+1/2}\Big), \tag{3.46}$$

where $x^{0,1}_{N/2+1/2}$ is the largest zero of the Jacobi polynomial $P^{0,1}_{N/2+1/2}(x)$, the optimal covariant encoding reads

$$\boldsymbol{n} \;\mapsto\; U(\boldsymbol{n})|A\rangle\langle A|U^\dagger(\boldsymbol{n})$$

where

$$|A\rangle = \sum_{j=1/2}^{N/2} A_j|j,1/2\rangle, \tag{3.47}$$

the coefficients A_j being such that $|A\rangle$ is the eigenvector corresponding to the maximal eigenvalue of the tridiagonal matrix

$$\begin{pmatrix} d_l & c_{l-1} & 0 & \dots & 0 \\ c_{l-1} & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & c_2 & 0 \\ \vdots & \ddots & c_2 & d_2 & c_1 \\ 0 & \cdots & 0 & c_1 & d_1 \end{pmatrix}, \tag{3.48}$$

where

$$l = \frac{N+1}{2}$$

and

$$\begin{aligned} d_i &= \frac{1}{4\left(i+\frac{1}{2}\right)\left(i-\frac{1}{2}\right)} \\ c_i &= \frac{\sqrt{i(i+1)}}{2\left(i+\frac{1}{2}\right)}. \end{aligned}$$

The optimal covariant POVM has the density

$$\widetilde{\mathcal{M}}(\boldsymbol{n}) = U(\boldsymbol{n})|B\rangle\langle B|U^{\dagger}(\boldsymbol{n}) \tag{3.49}$$

where

$$|B\rangle = \sum_{j=1/2}^{N/2} \sqrt{2j+1}|j, 1/2\rangle.$$

Note that $U(\boldsymbol{n}) = U(g(\boldsymbol{n})) = g^{\otimes N}$, where g is an element of SU(2) such that $|\boldsymbol{n}\rangle\langle\boldsymbol{n}| = g|\boldsymbol{z}\rangle\langle\boldsymbol{z}|g^{\dagger}$. The case of even N is treated in Section 4.4.1. The seed of the signal state, Eq. (3.47), and the POVM, Eq. (3.49), are specified only on the relevant subspace, i.e. for a spin-j representation with multiplicity greater than one, on the representation space of only one of the equivalent representations (see [3] for details). On the rest of the Hilbert space $\mathcal{H}_2^{\otimes N}$ the signal state and the measurement can be defined so that the encoding and the POVM remains U-covariant.

Chapter 4

Recycling of quantum information: Multiple observations of quantum systems

4.1 Introduction

4.1.1 Motivation

In classical physics, states of physical systems are, at least in principle, perfectly distinguishable. Consequently, there is no reason why any number of observers should not be able, at least in principle, to observe the same values of the parameters describing the state of a system. One could reason as follows: Even if there was a change of the state of the system due to a measurement, the observer could, at least in principle, prepare the measured system in the very same state that he has measured, i.e. in the state identical to the state before the measurement, and pass the prepared state to a subsequent observer.

On the other hand, in quantum physics, non-orthogonal states are allowed, i.e. if a physical system is in a state $|a\rangle$, an observer can still, via a measurement[4.1], infer that the system is in a non-orthogonal state $|b\rangle \neq |a\rangle$. Consequently, even if there was no change of the state of the system due to the measurement, any subsequent observer could end up inferring a different state of the physical system with respect to a preceding observer although, without a state change, each observer's estimate would be, on average, equally good. However, in quantum mechanics, any measurement that yields some gain of information on a state that could have been in non-orthogonal configurations is accompanied, in general, by a disturbance of the measured system (see e.g. [6, 33]).

Clearly, if a given measurement extracts the maximum information on the state of a system, then the same observer cannot obtain additional information by performing further measurements on the system. This almost tautological statement has the staring consequence: whenever the set of possible states of the measured system is non-orthogonal and thus the obtainable information is, in general, incomplete, such most informative quantum measurements, no matter how cautious they are, in general inevitably disturb the state of the system and thereby erase any information on the original state of the system as far as the same observer is concerned.

[4.1] We mean a single-shot measurement, i.e. a measurement on a single copy of the system, here. The statement holds for a finite number of copies as well.

However, these observations need to be revised if a second measurement is performed by a second observer who independently aims at gaining information on the original state of the system. This follows, e.g., from results on information-disturbance trade-off due to measurements – see Refs. [34], [6] (single pure qubit case), [5] (single pure qudit), [7, 42] (ensemble of pure qubits, the latter also conjecture for qudits), [48] (coherent states), [45] (entropic, instead of the common fidelity-based, approach), (see also [8] , [41]) – essentially stating that a post-measurement state even after a "most-informative" measurement is allowed to contain, on average, a relevant overlap with the pre-measurement state (and a less-informative one even more so). Indeed, we will see that a second (and any further) independent observer, who does not know the precise actions nor measurement outcomes of the previous one, can still obtain *some* information on the original state of the system.

Let us illustrate this on the simplest example – estimation of a unknown (pure) single qubit state[4.2] by two successive observers, followed by an example with one extra copy of the qubit.

4.1.2 A spin-1/2

It is convenient to use the Bloch vector parametrization, which to each pure-state density matrix assigns a Bloch vector $\boldsymbol{n}(\vartheta,\,\varphi)$ of length one. To be specific, let us suppose the qubit is realized by a spin-1/2 system, in which case the Bloch vector can be associated with a direction in physical three-dimensional space, i.e. the qubit state can be viewed as encoding of a direction. An observer's task is to access this information (knowledge about the encoded direction), and state his estimation, $\boldsymbol{n}_k$, as close to the encoded direction as possible. The observers have no knowledge whatsoever about the direction being encoded, i.e. they assume a uniform a-priori probability density $\tilde{p}$: $\boldsymbol{n}\mapsto\tilde{p}(\boldsymbol{n})=1$ over a (unit) sphere $\mathbb{S}^2$ (the density is defined with respect to the measure $\mathrm{d}\boldsymbol{n}=\frac{1}{4\pi}\sin\,(\vartheta)\mathrm{d}\varphi\mathrm{d}\vartheta$). For simplicity, suppose there are two non-communicating observers who use the same measuring apparatus: a non-demolition "analogue of a Stern-Gerlach apparatus" (SG), i.e. a device performing a projective measurement of spin component along a direction, described by the quantum instrument

$$J^{g_k}\ :\ \{\{1/2,-1/2\},\{1/2\},\{-1/2\},\{\emptyset\}\}\rightarrow\mathcal{L}(\mathcal{T}(\mathcal{H}))\tag{4.1}$$

defined by

$$\pm 1/2\ \mapsto\ J^{g_k}_{\pm 1/2}(\,\cdot\,)\!:=|\pm\boldsymbol{n}_k\rangle\langle\pm\boldsymbol{n}_k|(\,\cdot\,)|\pm\boldsymbol{n}_k\rangle\langle\pm\boldsymbol{n}_k|,\tag{4.2}$$

where $g_k\in\mathrm{SU}(2)$ parametrizes the orientation of the SG and $\boldsymbol{n}_k\equiv\boldsymbol{n}_k(g_k)$ is the orientation of the projection axis of the SG (i.e. $|\boldsymbol{n}_k\rangle\langle\boldsymbol{n}_k|=g_k|\boldsymbol{z}\rangle\langle\boldsymbol{z}|g_k^{-1}$) and $\pm 1/2$ are the "finest-grained" measurement outcomes of the SG.

We will, as in, e.g., [31, 36, 47], measure the observer's success in gaining knowledge about the encoded direction in terms of the (single-qubit) fidelity Eq. (2.21) between the true and estimated states, which for two pure qubits reads

$$f(\boldsymbol{n}_k,\boldsymbol{n})=|\langle\boldsymbol{n}_k|\boldsymbol{n}\rangle|^2=\cos^2\left(\frac{\vartheta}{2}\right)=\frac{1}{2}(1+\cos\vartheta)=\frac{1}{2}(1+\boldsymbol{n}_k\cdot\boldsymbol{n}),\tag{4.3}$$

[4.2] 4.2. Let us note that our model problem is a generalization of the problem of optimal encoding and estimation of phase into (and from) a quantum system, which is referred to as the problem of optimal quantum clocks in the literature (with the "complication" that, naturally, a time evolution of the quantum clock (the encoded phase) is also present). Optimal quantum clocks and multiple (sequential) observations of the encoded phase have been studied in [17] and [16], respectively.

ϑ being the angle between the estimated and the encoded direction.

In particular, we are interested in the mean of the fidelity over all encoded and estimated directions:

$$F_k = \int_{\text{events}} \mathrm{d}p_k(\boldsymbol{n}_k, \boldsymbol{n}) f(\boldsymbol{n}_k, \boldsymbol{n}) \tag{4.4}$$

$$= \int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n} \sum_{n' = \pm \boldsymbol{n}_k(g_k)} p_k(\boldsymbol{n}'|\boldsymbol{n}) f(\boldsymbol{n}', \boldsymbol{n}) \tag{4.5}$$

where $\mathrm{d}\boldsymbol{n} = \frac{1}{4\pi}\sin(\vartheta)\mathrm{d}\varphi\mathrm{d}\vartheta$, the index k signifies quantities concerning the k'th observer (no index or, sometimes, $k=0$, denoting quantities related to preparation), $\mathrm{d}p_k(\boldsymbol{n}_k, \boldsymbol{n})$ is the joint probability of of the event of the estimated direction being $\boldsymbol{n}_k$ and the encoded direction being $\boldsymbol{n}$; $p_k(\boldsymbol{n}'|\boldsymbol{n})$ is the corresponding conditional probability for a discrete set of estimates $\boldsymbol{n}' \in \{\boldsymbol{n}_k(g_k), -\boldsymbol{n}_k(g_k)\}$.

An observer's estimate, or guess, i.e. an assignment of a particular estimate $\boldsymbol{n}_k$ to the obtained measurement outcome, $+1/2$ or $-1/2$, which maximizes the Bayesian mean fidelity Eq. (4.4) is plus or minus the orientation of the SG, depending on the sign of the outcome of his measurement. For convenience we can re-label the outcomes of the SG to include the process of assigning estimates to the original measurement outcomes, $\pm 1/2$, leading to an instrument

$$I^{g_k} : \{\{\boldsymbol{n}_k, -\boldsymbol{n}_k\}, \{\boldsymbol{n}_k\}, \{-\boldsymbol{n}_k\}, \{\emptyset\}\} \to \mathcal{L}(\mathcal{T}(\mathcal{H})) \tag{4.6}$$

defined by

$$\pm\boldsymbol{n}_k \mapsto I^{g_k}_{\pm\boldsymbol{n}_k}(\,.\,) := |\pm\boldsymbol{n}_k\rangle\langle\pm\boldsymbol{n}_k|(\,.\,)|\pm\boldsymbol{n}_k\rangle\langle\pm\boldsymbol{n}_k|. \tag{4.7}$$

The induced POVM

$$M^{g_k} : \{\{\boldsymbol{n}_k, -\boldsymbol{n}_k\}, \{\boldsymbol{n}_k\}, \{-\boldsymbol{n}_k\}, \{\emptyset\}\} \to \mathcal{L}(\mathcal{H})$$

is defined by

$$\pm\boldsymbol{n}_k \mapsto M^{g_k}_{\pm\boldsymbol{n}_k} := |\pm\boldsymbol{n}_k\rangle\langle\pm\boldsymbol{n}_k|.$$

Note that $I^{g_k}_{\pm\boldsymbol{n}_k}(\,.\,)$ and $M^{g_k}_{\pm\boldsymbol{n}_k}$ are non zero only if $|\boldsymbol{n}_k\rangle\langle\boldsymbol{n}_k| = g_k|\boldsymbol{z}\rangle\langle\boldsymbol{z}|g_k^{-1}$.

Since the measurement is of Lüders type and rank-one, after the measurement the qubit is left in a state that is again a valid direction encoding. A subsequent observer then, using the same procedure, estimates the measurement outcome (guess) of his or her predecessor. The point is that the probability of an estimated direction depends on the (unknown) guess of the predecessing observer in such a manner, that it is larger for better estimates. Consequently, on average, an observer knows something about the direction guessed and "encoded" by the preceding observer, namely a better-than-a-random-guess estimation of the direction. Following the above logic from the first to the k-th observer, we conclude that, on average, also the last observer has partial information about the direction originally encoded by the preparer. Since, say, the k-th observer has, on average, a partial knowledge of the prior estimate which, in turn, is only a partial knowledge of its predecessor's estimate, and so on, the k-th observer knowledge will be diminishing with increasing k.

To express the above quantitatively, we need to evaluate the integral Eq. (4.4) as a function of k. Let us do this explicitly for the two observers. The fidelity of the first observer, is given by Eq. (4.5) where

$$\begin{aligned} p_1(\boldsymbol{n}'|\boldsymbol{n}) &= \mathrm{Tr}[M^{g_1}_{\boldsymbol{n}'}|\boldsymbol{n}\rangle\langle\boldsymbol{n}|] \\ &= |\langle\boldsymbol{n}'|\boldsymbol{n}\rangle|^2 \\ &= \cos^2(\vartheta/2), \end{aligned} \tag{4.8}$$

ϑ being the angle between the vectors $\boldsymbol{n}$ and $\boldsymbol{n}'$. Performing the integration in Eq. (4.5) we have

$$\begin{aligned} F_1 &= \frac{1}{4\pi}\int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \mathrm{d}\vartheta \sin(\vartheta) \sum_{\vartheta'=\vartheta,\vartheta+\pi} \cos^4(\vartheta'/2) \\ &= \frac{2}{3}. \end{aligned} \tag{4.9}$$

To evaluate the average fidelity of the second observer

$$F_2 = \int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n} \sum_{\boldsymbol{n}'=\pm\boldsymbol{n}_2(g_2)} p_2(\boldsymbol{n}'|\boldsymbol{n}) f(\boldsymbol{n}',\boldsymbol{n}), \tag{4.10}$$

we need to calculate the conditional probability $p_2(\boldsymbol{n}'|\boldsymbol{n})$ of the event of the second observer obtaining an estimate $\boldsymbol{n}'$ given the original signal state $|\boldsymbol{n}\rangle\langle\boldsymbol{n}|$ and given the second observer's apparatus orientation g_2 which is known and fixed from his point of view, from which we perform the calculation.

Considering a sequence of events – the first observer's choice of apparatus orientation, g_1, and obtaining a measurement outcome o_1, both unknown to the second observer – and summing probabilities of all possible first observer's actions we have

$$p_2(\boldsymbol{n}'|\boldsymbol{n}) = \sum_{o_1=\pm 1/2} \int_{g_1\in SU(2)} \mathrm{d}p(\boldsymbol{n}', o_1, J^{g_1}|\,\boldsymbol{n}). \tag{4.11}$$

where

$$\mathrm{d}p(\boldsymbol{n}', o_1, J^{g_1}|\,\boldsymbol{n}) = p(\boldsymbol{n}', o_1|J^{g_1}, \boldsymbol{n})\mathrm{d}p(J^{g_1}|\,\boldsymbol{n}).$$

The first observers's SG orientation must be independent of the unknown $\boldsymbol{n}$ and is uniformly distributed from the second observer's viewpoint, i.e. $\mathrm{d}p(J^{g_1}|\,\boldsymbol{n}) = \mathrm{d}p(J^{g_1}) = \mathrm{d}\mu(g_1)$, which is the invariant measure over $SU(2)$. Quantum mechanics gives

$$p(\boldsymbol{n}', o_1|J^{g_1}, \boldsymbol{n}) = \mathrm{Tr}\big[M^{g_2}_{\boldsymbol{n}'} J^{g_1}_{o_1}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|)\big], \tag{4.12}$$

where g_2 is the actual orientation of the second observer's SG should his estimate be $\boldsymbol{n}'$, i.e. g_2 is such that the projection axis is along the $\pm\boldsymbol{n}'$, that is $g_2|\boldsymbol{z}\rangle\langle\boldsymbol{z}|g_2^{-1} = |\boldsymbol{n}'\rangle\langle\boldsymbol{n}'|$ or $g_2|\boldsymbol{z}\rangle\langle\boldsymbol{z}|g_2^{-1} = |-\boldsymbol{n}'\rangle\langle-\boldsymbol{n}'|$.

Putting everything together, Eq. (4.11) reads

$$p_2(\boldsymbol{n}'|\boldsymbol{n}) = \sum_{o_1=\pm 1/2} \int_{g_1\in SU(2)} \mathrm{d}\mu(g_1)\mathrm{Tr}\big[M^{g_2}_{\boldsymbol{n}'} J^{g_1}_{o_1}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|)\big] \tag{4.13}$$

Noticing that in an isotropic space any apparatus has the property

$$\forall\hat{\rho}, g, o;\quad J^g_o(\hat{\rho}) = g J^e_o(g^{-1}\hat{\rho}\, g)g^{-1}, \tag{4.14}$$

(e being the unit element in the group), i.e. a post-measurement state (and the probability of any outcome) of a rotated apparatus performing a measurement on a rotated input state is the same as if a non-rotated apparatus acted on a non-rotated input state and the post-measurement state has been then rotated, Eq. (4.13) reads

$$p_2(\boldsymbol{n}'|\boldsymbol{n}) \;=\; \sum_{o_1=\pm 1/2}\int_{g_1\in SU(2)} \mathrm{d}\mu(g_1)\mathrm{Tr}\Big[M^{g_2}_{\boldsymbol{n}'} g J^e_{o_1}(g_1^{-1}|\boldsymbol{n}\rangle\langle\boldsymbol{n}|g_1)g_1^{-1}\Big]. \tag{4.15}$$

Observing that the Kraus operators in the definition of the instrument Eq. (4.2) for the $\pm 1/2$ outcomes are unitarily related, i.e. $J^e_{1/2}(\,\cdot\,) = g' J^e_{-1/2}(g'^{-1}\cdot g')g'^{-1}$, $g'\in SU(2)$, we see that the group integral, with the invariant measure, for the two summands in Eq. (4.15) is the same, i.e

$$p_2(\boldsymbol{n}'|\boldsymbol{n}) \;=\; 2\int_{g_1\in SU(2)} \mathrm{d}\mu(g_1)\mathrm{Tr}\Big[M^{g_2}_{\boldsymbol{n}'} g_1 J^e_{1/2}(g_1^{-1}|\boldsymbol{n}\rangle\langle\boldsymbol{n}|g_1)g_1^{-1}\Big]. \tag{4.16}$$

Choosing the z axis so that $J^e_{1/2}$ corresponds to projecting onto the $\boldsymbol{z}$ direction and looking at Eq. (4.7) we see that $J^e_{1/2}\equiv I^e_{\boldsymbol{z}}$ and $g_1 J^e_{1/2}(g_1^{-1}\cdot g_1)g_1^{-1}\equiv I^{g_1}_{\boldsymbol{n}_1(g_1)}(\,\cdot\,)$. Summing over unitary "rotations" around the z axis in Eq. (4.16), for which the quantum operation associated with the 1/2 outcome is always the same, we can write

$$p_2(\boldsymbol{n}'|\boldsymbol{n}) \;=\; 2\int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n}_1\mathrm{Tr}\Big[M^{g_2}_{\boldsymbol{n}'} I^{g(\boldsymbol{n}_1)}_{\boldsymbol{n}_1}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|)\Big], \tag{4.17}$$

where $g(\boldsymbol{n}_1)$ is the shortest rotation bringing $\boldsymbol{z}$ to $\boldsymbol{n}_1$. Note that

$$\boldsymbol{N} \;\mapsto\; \int_{\boldsymbol{n}\in\boldsymbol{N}} 2 I^{g(\boldsymbol{n})}_{\boldsymbol{n}}(\,\cdot\,)\mathrm{d}\boldsymbol{n} = \mathcal{I}_{\boldsymbol{N}}(\,\cdot\,),$$

where $\boldsymbol{N}$ is a set of vectors, defines a (covariant) instrument $\mathcal{I}$ with density $\tilde{\mathcal{I}}_{\boldsymbol{n}} = 2 I^{g(\boldsymbol{n})}_{\boldsymbol{n}}$.

Thus the SG plus the first observer (ignorance of the parameter g_1) can be viewed as new effective measurement apparatus. If the outputs of this apparatus are labeled by the first observer's estimates, $\boldsymbol{n}_1$, we can interpret Eq. (4.17) as a decomposition into possible histories – estimates of $\boldsymbol{n}_1$ given the use of the effective "apparatus" $\mathcal{I}$, i.e.

$$p_2(\boldsymbol{n}'|\boldsymbol{n}) \;=\; \int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n}_1 \tilde{p}(\boldsymbol{n}', \boldsymbol{n}_1|\boldsymbol{n}) \equiv \int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n}_1 p(\boldsymbol{n}'|\boldsymbol{n}_1, \boldsymbol{n})\tilde{p}(\boldsymbol{n}_1|\boldsymbol{n}), \tag{4.18}$$

where

$$\tilde{p}(\boldsymbol{n}_1|\boldsymbol{n}) = \mathrm{Tr}[\mathcal{I}_{\boldsymbol{n}_1}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|)] \tag{4.19}$$

and

$$p(\boldsymbol{n}'|\boldsymbol{n}_1, \boldsymbol{n}) = \frac{\mathrm{Tr}[M^{g_2}_{\boldsymbol{n}'}\mathcal{I}_{\boldsymbol{n}_1}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|)]}{\mathrm{Tr}[\mathcal{I}_{\boldsymbol{n}_1}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|)]}. \tag{4.20}$$

Since the measurement $J_{o_1}^{g_1}$, Eq. (4.2), is rank-one and projective (Lüders), the post-measurement state

$$\mathcal{I}_{\boldsymbol{n}_1}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|/\mathrm{Tr}[\mathcal{I}_{\boldsymbol{n}_1}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|)] = |\boldsymbol{n}_1\rangle\langle\boldsymbol{n}_1| \tag{4.21}$$

is $\boldsymbol{n}$ independent (while its probability is not). Thus $p(\boldsymbol{n}'|\boldsymbol{n}_1,\boldsymbol{n}) \equiv p(\boldsymbol{n}'|\boldsymbol{n}_1)$. Summarizing, Eq. (4.18) now reads

$$p_2(\boldsymbol{n}'|\boldsymbol{n}) \;=\; \int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n}_1 p(\boldsymbol{n}'|\boldsymbol{n}_1)\tilde{p}(\boldsymbol{n}_1|\boldsymbol{n}), \tag{4.22}$$

Performing the integration in the Eq. (4.22) we get

$$p_2(\boldsymbol{n}_2|\boldsymbol{n}) = \frac{1}{2}\left(1 + \frac{1}{3}\boldsymbol{n}_2\cdot\boldsymbol{n}\right) \tag{4.23}$$

and, finally, evaluating the integral Eq. (4.10),

$$F_2 = \frac{5}{9}. \tag{4.24}$$

One can proceed analogously for any number k of observers. The integral Eq. (4.22) then becomes

$$\begin{aligned} p_k(\boldsymbol{n}_k|\boldsymbol{n}) \;=\; & \int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n}_{k-1} p(\boldsymbol{n}_k|\boldsymbol{n}_{k-1}) \int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n}_{k-2}\tilde{p}(\boldsymbol{n}_{k-1}|\boldsymbol{n}_{k-2})... \\ & ... \int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n}_1 \tilde{p}(\boldsymbol{n}_2|\boldsymbol{n}_1)\tilde{p}(\boldsymbol{n}_1|\boldsymbol{n}) \end{aligned} \tag{4.25}$$

Finding a recurrence rule for the integrals in the Eq. (4.25) we get

$$p_k(\boldsymbol{n}_k|\boldsymbol{n}) = \frac{1}{2}\left(1 + \frac{1}{3^{k-1}}\boldsymbol{n}_k\cdot\boldsymbol{n}\right) \tag{4.26}$$

and, from Eq. (4.5),

$$F_k = \frac{1}{2}\left(1 + \frac{1}{3^k}\right). \tag{4.27}$$

Note that if the post-measurement state after the first measurement contained no information (from the second observer's viewpoint) on the original direction $\boldsymbol{n}$, i.e. the *average* first observer's post-measurement state, given any particular fixed pre-measurement signal state, was the total mixture, the second observer's mean fidelity would be determined by

$$\begin{aligned} p_2(\boldsymbol{n}_2|\boldsymbol{n}) \;&=\; \mathrm{Tr}\left[M^{g_2}_{\boldsymbol{n}_2(g_2)}\frac{1}{2}\mathbb{1}\right] \\ &=\; \frac{1}{2} \end{aligned} \tag{4.28}$$

and would read

$$F_2^{\mathrm{mix}} = \frac{1}{2}.$$

Note that this is also the value of the average fidelity of a particular observer's guess achieved should the observer use a strategy of mere guessing without actually measuring the state to estimate, i.e. 1/2 is clearly the fidelity achieved[4.3] when no information (reduction of uncertainty about the state to be estimated) is collected by the measurement process.

4.1.3 Two copies of a spin-1/2

Let us now proceed with a slightly more complicated situation, where the direction information of the single spin-1/2 system is carried by two copies of it. For two copies of a qubit, a POVM that is optimal and minimal for a single observer is known [47] (see Section 3.3 for an overview) to have eigenstates given by Eq. (3.42). These are not the states for which we know an optimal measurement (to be performed by the second and subsequent observers). Therefore, we will consider two situations:

i. All observers will measure by the apparatus defined by the eigenstates Eq. (3.42), i.e by the resolution of identity $\mathbb{1} = \sum_i \psi_i$, $\psi_i = |\psi_i\rangle\langle\psi_i|$ (with $\boldsymbol{m}_i$'s, i.e., the apparatus rotated by a rotation, $g_k \in \mathrm{SO}(3)$, chosen at will by the observers), together with the Lüders update rule for the states after the measurement (the apparatus will be described by a Lüders instrument (Def. 2.16)). Considering non-degenerate measurements, the post-measurement states will then be (up to an overall phase) given by one of the eigenstates Eq. (3.42). As these are not the two-copies states, only the first observer will be measuring optimally.

ii. We define a new apparatus which, in addition to performing the POVM given by the same orthonormal resolution of identity as before, upon obtaining the outcome $\boldsymbol{n}_k$, prepares the system in the desired two-copies state $|\boldsymbol{n}_k\boldsymbol{n}_k\rangle$. This boils down to defining a new update rule (normalized operation, Def. 2.13) for the states coming out of the apparatus, given a particular eigenvalue has been observed. Since the POVM description of the apparatus is the same as before, the new apparatus remains optimal for the single-observer scenario. Moreover, it will enhance the estimation performance for the observers to follow because the apparatus performs measurements optimal for its output states.

We now proceed with calculations of the average fidelities for the two cases. As in the single spin-1/2 case, we label the measurement outcomes by the best guesses, i.e. we identify $|\psi_{\boldsymbol{z}}\rangle$ with the $|\psi_1\rangle$ in Eq. (3.42) for $\boldsymbol{m}_1 = \boldsymbol{z}$. Beginning with the latter case (ii.), evaluating an analogue of Eq. (4.19) for general k, we get:

$$\begin{aligned}\tilde{p}(\boldsymbol{n}_k|\boldsymbol{n}_{k-1}) &= 4\mathrm{Tr}[\psi_{\boldsymbol{n}_k}|\boldsymbol{n}_{k-1}\rangle\langle\boldsymbol{n}_{k-1}|] \\ &= 4\frac{3}{4}\cos^4(\frac{\vartheta}{2}) \\ &= \frac{3}{4}+\frac{3}{2}\boldsymbol{n}_k\cdot\boldsymbol{n}_{k-1}+\frac{3}{4}(\boldsymbol{n}_k\cdot\boldsymbol{n}_{k-1})^2\end{aligned} \qquad (4.29)$$

[4.3] more generally $1/d$ if the state to estimate is from a d-dimensional Hilbert space

The factor four in Eq. (4.29) comes in by the same mechanism as the factor two in Eq. (4.16) of the single-spin case, however now the summation in an analogue of Eq. (4.15) is over four measurement outcomes.

Performing the integrations in the Eq. (4.25) (finding a recurrence rule) we have

$$\tilde{p}_k(\boldsymbol{n}_k|\boldsymbol{n}) = \frac{1}{2}\left(1 + \frac{5^{k-1} 6\, \boldsymbol{n}_k \cdot \boldsymbol{n} + 3(\boldsymbol{n}_k \cdot \boldsymbol{n})^2 - 1}{2^{k+1} 5^{k-1}}\right) \tag{4.30}$$

and, finally, from Eq. (4.4) in the form[4.4]

$$F_k = \int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n} \int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n}_k\, \tilde{p}_k(\boldsymbol{n}_k|\boldsymbol{n}) f(\boldsymbol{n}_k, \boldsymbol{n}), \tag{4.31}$$

we have

$$F_k = \frac{1}{2}\left(1 + \frac{1}{2^k}\right). \tag{4.32}$$

Next, we proceed with the former case (i.). For the first observer the situation is identical to that of the previous case, i.e. $\tilde{p}(\boldsymbol{n}_1|\boldsymbol{n})$ is given by Eq. (4.29), (we define $\boldsymbol{n}_0 := \boldsymbol{n}$). For $k > 1$ we have

$$\begin{aligned}
\tilde{p}(\boldsymbol{n}_k|\boldsymbol{n}_{k-1}) &= 4\mathrm{Tr}\big[\psi_{\boldsymbol{n}_k}\psi_{\boldsymbol{n}_{k-1}}\big] \\
&= 4\left(\frac{1}{16} + \frac{9}{16}\cos^2\left(\frac{\vartheta}{2}\right)\right)^2 \\
&= \frac{13}{16} + \frac{9}{8}\boldsymbol{n}_k \cdot \boldsymbol{n}_{k-1} + \frac{9}{16}(\boldsymbol{n}_k \cdot \boldsymbol{n}_{k-1})^2
\end{aligned}$$

Performing the integration in the Eq. (4.25) we have

$$\tilde{p}(\boldsymbol{n}_k|\boldsymbol{n}) = \frac{1}{2}\left(1 + 3^{k-1}\frac{5^{k-1} 6\, \boldsymbol{n}_k \cdot \boldsymbol{n} + 3(\boldsymbol{n}_k \cdot \boldsymbol{n})^2 - 1}{2^{3k-1} 5^{k-1}}\right) \quad ; \quad k > 1 \tag{4.33}$$

and, finally, from Eq. (4.31),

$$F_k = \frac{1}{2}\left(1 + \frac{4}{3}\left(\frac{3}{8}\right)^k\right). \tag{4.34}$$

4.1.4 Implications

In Section 4.1.2 we have considered a non-destructive version of a Stern-Gerlach-like projective measurement of a single spin-1/2 system. The measurement has been optimal in two ways.

[4.4] The average fidelity is invariant under an additional averaging over last observer's apparatus orientation. Therefore we may perform the calculation as if the orientation was random, i.e. from the viewpoint of someone not aware of apparata-orientation choices made by any observer, including the last one.

First of all, it is one of the measurements maximizing the first observer's average fidelity of estimation, since there is no better way to encode a pure qubit state[4.5] into a qubit state other than the pure qubit state itself and the considered POVM is optimal for a unknown pure qubit (cf. [5]).

Secondly, the considered update rule due to the measurement guarantees the best possible second observer's fidelity of estimation. Intuitively, a measurement yielding optimal estimation must produce a post-measurement state which carries no pre-measurement-state information other than that already revealed by the measurement outcome, otherwise an additional measurement could be done by the same observer to reveal this extra information, which would be in contradiction with the measurement-optimality assumption. The post-measurement state may of course carry information on the last measurement outcome and it should carry this (pure qubit) information as efficiently as possible, with respect to its estimability by a next observer. This is the case, e.g., if the same (optimal) encoding that has been used for the original signal state is used to encode the estimates, which is exactly what the instrument I^{g_k}, Eqs. (4.6) and (4.7), does.

We have seen quantitatively that, even with only a single instance of a quantum system, even though an observer extracted all the information he could from the system, there is still information left in the post-measurement state, given the measurement has been made carefully, i.e. the POVM has been realised via a wisely chosen compatible instrument. The information left in the post-measurement state vanishes exponentially with the tally number, k, of the observation (note that what we considered was a situation in absence of a (directional) reference frame shared among the observers, presence of which would enable to re-observe a system without a decrease of the estimation fidelity since with a common frame an alignment of the SG apparata is possible).

As discussed in Chapter 3, it is known that using a larger system to encode a state from a state space of an elementary system may help in terms of estimation fidelity of the parameters of the elementary system. The previous Section indicates that while, as expected, the estimation fidelity can be enhanced for *all* observers by adding copies of the elementary system, its decrease remains exponential with the number of observations, k. In what follows we continue with a general treatment.

4.2 Multiple observations of qu*d*its

Let us consider the situation depicted in Figure 4.1. A single qu*d*it (d-dimensional quantum system) in a pure state $\psi \in \mathcal{S}(\mathcal{H}_d)$ is encoded into a state from the state space of a qu*D*it (system Q) via an encoding ρ

$$\rho\colon\ \mathcal{S}(\mathcal{H}_d) \rightarrow \mathcal{S}(\mathcal{H}_D),\ \ \psi \mapsto \rho(\psi). \tag{4.35}$$

We may imagine that the specific encoding is "chosen and performed" by nature or, if we wish, by a preparer who knows (the classical description of) the pure state ψ. The state ψ is, however, unknown (uniformly distributed over pure states from $\mathcal{S}(\mathcal{H}_d)$) from the point of view of an experimentalist, or observer, who wishes to estimate it. Consider K observers, who make measurements on the system Q, one after another, i.e. the kth observer performs a measurement on the post-measurement state of the $(k-1)$th observer, $k = 1, \ldots, K$.

[4.5] up to a phase, i.e. by a *state* we mean its density matrix

Since this, apparently, requires that there remains a system to be re-measured, we consider non-destructive measurements, which also includes destructive measurements followed by preparation of a new system. A description of such measurements implies at least an instrument description of the measurement process (see Definition 2.14) which is also sufficient.

If one wanted, it is always possible to imagine an indirect measurement scheme described by a specific measurement model which is compatible with the instrument description (cf. [37]). In such a model, each measurement is done indirectly by coupling a fresh ancilla system A (the kth measurement apparatus) to the system Q and subsequently measuring upon the ancilla state by observing some "actual" apparatus reading o_k (see Figure 4.1).

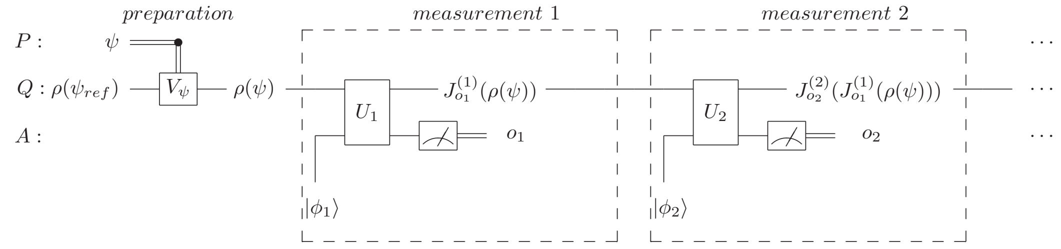

Figure 4.1. Circuit model of the sequential measurement scheme. The scheme depicts snapshots of one particular joint event of successive events of one particular preparation and one particular measurement outcome per observer. Time goes from left to right. The state at the output of each apparatus is normalized to the conditional joint-event probability of the events of obtaining measurement outcomes which happened up to the given measurement, conditioned on the particular state, $\rho(\psi)$, prepared and on the particular choices of apparata $J^{(k)}$ and their initial ancilla states, $|\phi_k\rangle$.

Remark 4.1. Note that the initial states of the ancilla, $|\phi_k\rangle$, for different observers are not allowed to be identical, since this would require a shared "directional" reference in the Hilbert spaces of the ancillas which, if the observers used the same apparatus (couplings, i.e., U_k were the same for $k = 1, 2, ...$), would provide a special "direction" in the signal state's Hilbert space shared by the observers, which we disallow. In such a case, if a most informative measurement was required, the observers could simply do an identical von Neumann (projective, rank-one) measurement and the fidelity of each one's estimate would be identical, given by the first observer's estimation fidelity. This trivial situation is not of our interest. The multi-user problem along these lines, but with generalized (weak) measurements saturating the single-user information-disturbance trade-off (c.f. Ref. [5]) has been studied in Ref. [35] for a single-qu*d*it system.

The goal of each observer is to measure the state given to him, then, based on the measurement outcome, estimate the original state ψ and finally leave the system Q in a post-measurement state for the next observer to come. The POVM performed by a particular observer, will depend on the observer's goal (e.g., in one of the studied scenarios, maximize his fidelity of estimation) and on the observer's choice, if more POVMs are optimal for the goal. The post-measurement state will have to be one allowed by quantum mechanics for the particular POVM used – i.e., if we denote the kth observer's POVM by $M^{(k)}$, the POVM and the state change due to measurement will be described by a $M^{(k)}$-compatible instrument $J^{(k)}$. Our goal is to find out to what extent the outcome of a measurement performed on a signal state already measured $(k-1)$ times can tell us something about the original encoding pre-image, ψ.

The figure of merit of the correctness of a measurement outcome ψ' with respect to the actual qu$d$it state ψ is denoted by $f(\psi', \psi)$. Since the objects that are encoded to, and estimated from, the state of the system Q are qu$d$its, we use the fidelity, Eq. (2.21), of the single-event fidelity of the estimated state with respect to the true input state, as the figure of merit. For pure-state qu$d$its the fidelity reads

$$f(\psi, \psi') = \mathrm{Tr}(\psi\psi') = \frac{1}{d}(1 + (d-1)\boldsymbol{n}(\psi) \cdot \boldsymbol{n}(\psi')), \tag{4.36}$$

where the vectors $\boldsymbol{n}$ are the generalized Bloch vectors for states of a single d-dimensional system (see Appendix (B) for details). The fidelity corresponds to a particular choice of a cost function in the context of detection and Bayesian estimation theory [38] which is common in related works, limited to a single observation [47, 31, 3, 4, 36]. Analogously to the single-observer state estimation, the performance of an encoding/estimation strategy is measured in terms of the average fidelity F_k of the kth observer's estimate with respect to the actual qu$d$it state ψ (from now on referred to just as fidelity). It reads

$$F_k = \int_{\text{events}} \mathrm{d}p(\psi_k, \psi) f(\psi_k, \psi), \tag{4.37}$$

where $\mathrm{d}p(\psi_k, \psi)$ is the probability of the joint event of the state ψ being encoded and obtaining the estimate ψ_k. The integration is carried over all such events, i.e. over all pure states $\psi \in \mathcal{S}(\mathcal{H}_d)$ and all pure states $\psi_k \in \mathcal{S}(\mathcal{H}_d)$.

Mathematically, our goal is to evaluate

$$\mathcal{F}^{(k)}(d, D) \;=\; \max F^{(k)}(d, D), \tag{4.38}$$

where the maximization is over the encoding and measurements of all the observers, compatible with our assumptions and the particular goal that has to be achieved by each observer. The goal will be defined in terms of requirements on the achieved performance of F_k for each $k = 1, \ldots, K$.

We will consider the problem explicitly in three scenarios, in which the experimentalists pursue different goals, i.e. state three sets of conditions on F_k for all $k = 1, \ldots, K$. These will be discussed in Sections 4.4, 4.5.1 and 4.5.4. Before that, let us arrive at a few general results which will simplify our analysis.

4.3 General considerations

Let us begin with a small interlude. In previous sections we discussed the (single-observer) covariant estimation problem, i.e. the encoding ρ, Eq. (4.35), was assumed to be covariant with respect to a representation U of SU(d), i.e. $\rho\colon \mathcal{S}(\mathcal{H}_d) \to \mathcal{S}(\mathcal{H}_D)$, $\psi \mapsto \rho(\psi)$ such that $\rho(g\,\psi g^{-1}) \mapsto U_g\, \rho(\psi) U_g^\dagger$, $U\colon \mathrm{SU}(d) \to \mathrm{SU}(D)$, $g \to U_g$. In the general formulation of our problem, we drop this explicit requirement.

Instead, we allow the encoding ρ to be arbitrary, as seen from the perspective of the one who performs it. We assume that the encoding is "known" (or allowed to be communicated) by the observers, however, this knowledge is limited to a description that makes no reference to information or, put differently, makes no use of a resource, of the very same type as the property the observers will be trying to estimate. In the case of spins, this is equivalent to absence of a real-space reference direction (and thus a reference frame) common to the preparer and the observers.

The "motivation" we have in mind is something along the lines of a situation, where an astronaut arrives at a directional sign, placed somewhere in space, realized by a collection of spins. Given a particular convention for assigning the north pole to a magnetic field (2-form), the astronaut is able to estimate the "north" direction and follow it. With only the convention, another astronaut will be able to do the same in, say, hundred years with the original directional sign left by the previous astronaut after having estimated its state. If the preparer of the directional sign and the astronauts were able to agree on a common (reference) direction beforehand, there would obviously be no need to communicate one via a (quantum) directional sign. Moving from spins to general qu*d*its (and from directions to pure states from $\mathcal{S}(\mathcal{H}_d)$) the lack of a pre-agreed reference will lead to being sufficient to consider encodings and measurements which are covariant with respect to the $g \mapsto g^{\otimes N}$ representation of $\mathrm{SU}(d)$ (see also Remark 4.2). We will show this explicitly in the reminder of the present Section.

To say something about the fidelity, Eq. (4.37), we need to evaluate the conditional probability, $\mathrm{d}p(\psi_k|\psi_0)$, that the kth observer makes an estimation ψ_k given the original single-qu*d*it state ψ_0. Naturally, this quantity depends on the preparation (encoding of ψ_0 into a signal state), on the kth observer's measurement and guessing strategies, and on whatever happened in between.

We may decompose each particular conditional joint event into a sum over histories – intermediate events such as measurements apparata choices, measurement outcomes obtained (the two prescribing the post-measurement states realized), or guesses made based on the outcomes – which may have led to the event $(\psi_k|\psi_0)$. More precisely the decomposition is into events that may have led to the event $(\psi_k|\psi_0)$ given the (lack of) knowledge of the observers about the other observers' actions.

We may choose to perform calculations from any observer's perspective; we choose the point of view of the kth observer, i.e. the (currently) last one in the sequence. This implies that, naturally, the actions of the kth observer are considered to be fully known and the actions of the previous observers and of the preparer may be known at most up to a rigid unitary $g^{\otimes N}$, $g \in \mathrm{SU}(d)$.

Remark 4.2. For the spins-1/2 scenario, a $g^{\otimes N}$ unitary is the transformation undergone by the N-spin-1/2 system if single copy undergoes a real-space rotation associated with $g \in \mathrm{SU}(2)$. For other systems one could, in principle, imagine more general representations, but the $g \mapsto g^{\otimes N}$ representation is probably the most natural choice when talking about N *copies* of a Hilbert space. In any case, it is the one considered in the present work.

Let us perform such a decomposition (lower index denotes the tally number of an observer, index 0 stands for the preparer, the integration is over all joint events, i.e. over all variables of the probability $\mathrm{d}p(\,.\,)$ to the left of "|", with the exception of ψ_k, which is apparent from the left-hand side of the below equation):

$$\mathrm{d}p(\psi_k|\psi_0) \;=\; \int \mathrm{d}p(\psi_k, g_k, \hat{\rho}_{k-1}, \psi_{k-1}, g_{k-1}, \hat{\rho}_{k-2}, \ldots, \hat{\rho}_1, \psi_1, g_1, \hat{\rho}_0, g_0|\psi_0), \qquad (4.39)$$

where the integral is over the chosen intermediate events, i.e. over all possible post measurement (or post-preparation) states $\hat{\rho}_0, \hat{\rho}_1, ..., \hat{\rho}_{k-1} \in \mathcal{S}(\mathcal{H}_d)$, over all guesses, i.e. pure states $\psi_1, ..., \psi_{k-1} \in \mathcal{S}(\mathcal{H}_d)$ and over all apparata "orientations"[4.6] $g_0, g_1, ..., g_k$ (g_0 is an orientation associated with the preparer). The orientation parameters, $g_i \in \mathrm{SU}(d)$, are the $\mathrm{SU}(d)$ transformation such that $g^{\otimes N}$ transforms the reference frame attached to the ith observer's apparatus (its Hilbert space) into the reference frame of the kth observer's Hilbert space.

Note that a decomposition into actual measurement outcomes, which we denote o_i, is not made in Eq. (4.39).

For the sake of simplicity let us work out the integral Eq. (4.39) explicitly for $k=2$, i.e. imagine there are only two observers of interest. Expressing joint probabilities in terms of conditional ones several times we have

$$\begin{aligned} \mathrm{d}p(\psi_2|\psi_0) &= \int \mathrm{d}p(\psi_2|g_2, \hat{\rho}_1, \psi_1, g_1, \hat{\rho}_0, g_0, \psi_0)\mathrm{d}p(g_2|\hat{\rho}_1, \psi_1, g_1, \hat{\rho}_0, g_0, \psi_0) \\ &\quad \times \mathrm{d}p(\hat{\rho}_1|\psi_1, g_1, \hat{\rho}_0, g_0, \psi_0)\mathrm{d}p(\psi_1|g_1, \hat{\rho}_0, g_0, \psi_0)\mathrm{d}p(g_1|\hat{\rho}_0, g_0, \psi_0) \\ &\quad \times \mathrm{d}p(\hat{\rho}_0|g_0, \psi_0)\mathrm{d}p(g_0|\psi_0). \end{aligned} \tag{4.40}$$

The probability of a particular, ith, observer's guess, ψ_i, may depend on the pre-measurement state $\hat{\rho}_{i-1}$ since the probability of assigning a guess ψ_i depends on the ith measurement outcome o_i, probability of which, in turn, depends on the state at the output of the $(i-1)$th observer's apparatus, $\hat{\rho}_{i-1}$. The probability of a guess ψ_i may also depend on the orientation of the ith observer's apparatus which is parametrized by $g_i \in \mathrm{SU}(d)$. (Note that this does not mean the ith observer is aware of g_i with respect to the kth observer's reference frame, but he is aware of the objective orientation of his apparatus – e.g. he can test that, for instance, upon performing a measurement on some particular objective state ψ_i, his apparatus always shows, say, the outcome $o_i=3$. Transforming the apparatus, e.g. physically rotating it if working with spin-1/2s, $o_i=3$ always clicks for a different state ψ_i'. This of course influences the guess assigned to the outcome "3" but does not require to be aware of the corresponding g_i). Dependence on anything else – previous observers' guesses, the original state ψ_0, orientations of previous observers' apparata and of the preparer – is only through the state $\hat{\rho}_{i-1}$ that is being measured, i.e.

$$\mathrm{d}p(\psi_1|g_1, \hat{\rho}_0, g_0, \psi_0) \equiv \mathrm{d}p(\psi_1|g_1,\hat{\rho}_0)$$

and

$$\mathrm{d}p(\psi_2|g_2, \hat{\rho}_1, \psi_1, g_1, \hat{\rho}_0, g_0, \psi_0) \equiv \mathrm{d}p(\psi_2|g_2, \hat{\rho}_1).$$

Note that any observer's apparatus orientation (including the preparer) from the point of view of any other observer is totally random, i.e. from the kth observer's point of view, $\forall i < k, \mathrm{d}p(g_i|...) \equiv \mathrm{d}\mu(g_i)$. As for the kth observer's $\mathrm{d}p(g_k|...) \equiv \mathrm{d}p(g_k)$, it may be any distribution as the observer may choose any probabilistic strategy of measurement apparatus orientations. However, picking (any of) the best performing orientation(s), $\tilde{g}_k$, cannot decrease fidelity of the guess, hence we may put $\mathrm{d}p(g_k|...) \equiv \delta(g_k - \tilde{g}_k)$.

[4.6] We use this terminology as a reminiscence of the examples of Section 4.1, where we considered a spin-1/2 system's state to be carried by a collective state of N spin-1/2 systems, in which case the group parameters $g_i \in \mathrm{SU}(2)$ have been in a (two-to-one, $\mathrm{SO}(3) \ni h \leftrightarrow \pm g \in \mathrm{SU}(2)$) correspondence with spatial rotations, SO(3), in real space. In the case of $d > 2$, or for qubit systems not physically realized by spins, there is no such correspondence but we use the term nevertheless, for brevity. We will (most of the times) drop the quotes in what follows.

Thus we have

$$\begin{aligned}\mathrm{d}p(\psi_2|\psi_0) &= \int \mathrm{d}p(\psi_2|\tilde{g}_2, \hat{\rho}_1)\mathrm{d}p(\hat{\rho}_1|\psi_1, g_1, \hat{\rho}_0, g_0, \psi_0)\mathrm{d}p(\psi_1|g_1, \hat{\rho}_0)\mathrm{d}\mu(g_1)\\ &\quad \times \mathrm{d}p(\hat{\rho}_0|g_0, \psi_0)\mathrm{d}\mu(g_0). \end{aligned} \tag{4.41}$$

where we have already integrated over g_2; $\int\ \mathrm{d}p(\psi_2|g_2, \hat{\rho}_1)\delta(g_2 - \tilde{g}_2) = \int\ \mathrm{d}p(\psi_2|\tilde{g}_2, \hat{\rho}_1)$ [the integral on the LHS is over ψ_2 and g_2, on the RHS over ψ_2 only].

Note that the state on the output of any (in this case the first) observer's apparatus can depend on the quantum operation (i.e. transformation of the input) performed by the apparatus given an outcome observed (in this case o_1) and on the state at the input of the apparatus. Moreover, the post-measurement state may depend on what is known to the observer – this information may be used for any post-processing of the output state (which we may formally include in the apparatus). In particular, for $i = 1$,

$$\mathrm{d}p(\hat{\rho}_1|\psi_1, g_1, \hat{\rho}_0, g_0, \psi_0) = \mathrm{d}p(\hat{\rho}_1|\psi_1, g_1, \hat{\rho}_0),$$

i.e. the probability of any post-measurement state is not a function of the unknown parameters nor states prior to the input state (this would be important for $i > 1$). [$\mathrm{d}p(\hat{\rho}_1|\psi_1, g_1, \hat{\rho}_0)$ is a distribution where the probability of $\hat{\rho}_1$ given the guess ψ_1 is the sum of probabilities of $\hat{\rho}_1$ given any outcome o_1 leading to the guess ψ_1.]

Thus Eq. (4.41) becomes

$$\begin{aligned}\mathrm{d}p(\psi_2|\psi_0) &= \int \mathrm{d}p(\psi_2|\tilde{g}_2, \hat{\rho}_1)\mathrm{d}p(\hat{\rho}_1|\psi_1, g_1, \hat{\rho}_0)\mathrm{d}p(\psi_1|g_1, \hat{\rho}_0)\mathrm{d}\mu(g_1)\mathrm{d}p(\hat{\rho}_0|g_0, \psi_0)\mathrm{d}\mu(g_0)\\ &=: \int \mathrm{Tr}\Big[I^{(2),\tilde{g}_2}_{\mathrm{d}\psi_2}\Big(I^{(1),g_1}_{\mathrm{d}\psi_1}(\hat{\rho}_0)\Big)\Big]\mathrm{d}p(\hat{\rho}_0|g_0, \psi_0)\mathrm{d}\mu(g_1)\mathrm{d}\mu(g_0), \end{aligned} \tag{4.42}$$

where we have introduced a quantum operation $I^{(i),g_i}_{\mathrm{d}\psi_i}$ defined by

$$I^{(i),g_i}_{\mathrm{d}\psi_i}(\hat{\rho}_{i-1}) = \hat{\rho}_i \mathrm{d}p(\psi_i|g_i, \hat{\rho}_{i-1}).$$

The quantum instrument $I^{(i),g_i}$ fully characterizes the (ith observer's) measurement apparatus in terms of the guesses – providing both the probability $\mathrm{d}p(\psi_i|g_i,\ \hat{\rho}_{i-1}) = \mathrm{Tr}[I^{(i),g_i}_{\mathrm{d}\psi_i}(\hat{\rho}_{i-1})]$ of obtaining the guess ψ_i given the measured state has been $\hat{\rho}_{i-1}$, and the post-measurement state

$$\hat{\rho}_i = \frac{I^{(i),g_i}_{\mathrm{d}\psi_i}(\hat{\rho}_{i-1})}{\mathrm{Tr}[I^{(i),g_i}_{\mathrm{d}\psi_i}(\hat{\rho}_{i-1})]}$$

(note that if a guess never occurs, i.e. if $\mathrm{Tr}[I^{(i),g_i}_{\mathrm{d}\psi_i}(\hat{\rho}_{i-1})] = 0$, there is no need to specify the post-measurement state, i.e. $\hat{\rho}_i$ is well defined in all situations that can actually take place).

Integrating over $\hat{\rho}_0$ we may introduce a deterministic encoding $\rho_0^{g_0}$ defined by

$$\rho_0^{g_0}(\psi_0) = \int \hat{\rho}_0 \mathrm{d}p(\hat{\rho}_0|g_0, \psi_0).$$

Thus without loss of generality we can assume deterministic encodings only. We now have

$$\mathrm{d}p(\psi_2|\psi_0) \;=\; \int \mathrm{Tr}\Big[I^{(2),\tilde{g}_2}_{\mathrm{d}\psi_2}\Big(I^{(1),g_1}_{\mathrm{d}\psi_1}\big(\rho_0^{g_0}(\psi_0)\big)\Big)\Big] \mathrm{d}\mu(g_1)\mathrm{d}\mu(g_0), \tag{4.43}$$

Performing integration over $g_1 \in \mathrm{SU}(d)$ we obtain an effective instrument $\mathcal{I}^{(1)}$ which is covariant (see Appendix A for details). The deterministic encoding $\rho_0^{g_0}$ can also be viewed as an instrument (independent on its input):

$$I^{(0),g_0}\colon \quad I^{(0),g_0}_{\mathrm{d}\psi_0}(\,.\,) = \rho_0^{g_0}(\psi_0)\mathrm{d}\psi_0.$$

Like in the case of $I^{(1)}$, integrating over its "orientations", $g_0 \in \mathrm{SU}(d)$, we obtain an effective covariant encoding

$$\varrho_0\colon \quad \varrho_0(\psi_0) = \int \rho_0^{g_0}(\psi_0) d\mu(g_0),$$

i.e. without loss of generality it is enough to consider covariant initial encodings only. We arrive at

$$\mathrm{d}p(\psi_2|\psi_0) \;=\; \int \mathrm{Tr}\Big[I^{(2),\tilde{g}_2}_{\mathrm{d}\psi_2}\Big(\mathcal{I}^{(1)}_{\mathrm{d}\psi_1}(\varrho_0(\psi_0))\Big)\Big]. \tag{4.44}$$

Performing the integral over first observer's guesses, ψ_1,

$$\begin{aligned} \mathrm{d}p(\psi_2|\psi_0) \;&=\; \mathrm{Tr}\Big[I^{(2),\tilde{g}_2}_{\mathrm{d}\psi_2}(\chi_1(\varrho_0(\psi_0)))\Big] \\ &=\; \mathrm{Tr}\Big[M^{(2),\tilde{g}_2}_{\mathrm{d}\psi_2} \chi_1(\varrho_0(\psi_0))\Big] \end{aligned} \tag{4.45}$$

where χ_1 is a channel induced by the first observer's measurements, i.e. the quantum operation that takes place if the outcome of the measurement (or its associated guess) is unknown. $M^{(2),\tilde{g}_2}$ is the (unique) POVM induced by the instrument $I^{(2),\tilde{g}_2}$ (cf. [37] for details).

Due to covariance of the effective measurement $\mathcal{I}^{(1)}$ the channel χ_1 is invariant. It follows that the "encoding" $\chi_1(\varrho_0(\,.\,))$, given by the covariant encoding ϱ_0 and the invariant channel χ_1, is covariant. For such situation there always exists a covariant POVM, which we denote $\mathcal{M}^{(2)}$, which reaches any given value of average fidelity achieved by any POVM $M^{(2),\tilde{g}_2}$ [39]. The arguments presented in this Section can be extended to any k. Thus, without loss of generality, it suffices to consider

$$\begin{aligned} \mathrm{d}p(\psi_k|\psi_0) \;&=\; \mathrm{Tr}\Big[\mathcal{M}^{(k)}_{\mathrm{d}\psi_k} \chi_{k-1} \circ \ldots \circ \chi_1(\varrho_0(\psi_0))\Big], &(4.46)\\ &=\; \underbrace{\int \ldots \int}_{(k-1)\text{ times}} \mathrm{Tr}\Big[\mathcal{M}^{(k)}_{\mathrm{d}\psi_k}\Big(\mathcal{I}^{(k-1)}_{\mathrm{d}\psi_{k-1}} \circ \ldots \circ \mathcal{I}^{(1)}_{\mathrm{d}\psi_1}\Big)(\varrho_0(\psi_0))\Big] &(4.47) \end{aligned}$$

where all measurements are covariant and all channels are invariant and in the second line the integrations are over all the possible guesses $\psi_1, \ldots, \psi_{k-1}$.

The Eqs. (4.46) and (4.47) tell us how to proceed further. In general, we search for pairs of covariant encodings ϱ_0 and POVM(s) $\mathcal{M}^{(1)}$ fulfilling a desired property of the average fidelity F_1 (e.g. maximizing F_1, possibly given additional constraints) for the invariant family of states

$$\{\varrho_0(\psi_0) = U_{g_0}\varrho_0(\psi_{\text{ref}})U_{g_0}^\dagger, g_0 \in \text{SU}(d), U_{g_0} = g_0^{\otimes N}\},$$

distributed according to $\text{d}p(g_0) = d\mu(g_0)$. Having at least one such pair $\{\varrho_0, \mathcal{M}_1\}$, one can evaluate F_1 to get its value when it fulfills the desired property (e.g. get its maximal achievable value).

Next, consider all covariant quantum instruments $\mathcal{I}^{(1)}$ compatible with POVM(s) $\mathcal{M}^{(1)}$, obtained in the previous step, and calculate the set of invariant channels which are induced by any of those instruments. Next, search for covariant POVM(s) $\mathcal{M}^{(2)}$ fulfilling a desired property of the average fidelity F_2 for the average states from the invariant families

$$\{\chi_1(\varrho_0(\psi_0)) = U_{g_0}\chi_1(\varrho_0(\psi_{\text{ref}}))U_{g_0}^\dagger, g_0 \in \text{SU}(d), U_{g_0} = g_0^{\otimes N}\},$$

distributed as governed by $\text{d}p(g_0) = d\mu(g_0)$, given by the actions of all channel(s) χ_1 from the previous step, and so on.

The task is greatly simplified by the possibility to restrict oneself to covariant apparata (invariant channels). If the optimal covariant apparata turn out to be unique at each step, the task becomes yet much simpler. However, still, using this general approach one has to calculate the induced channel at each step to obtain the set of *average* states for the next optimization.

Alternatively, one can equivalently work with (optimize for) the ensemble of states

$$\{\hat{\rho}' = \mathcal{I}^{(1)}_{\text{d}\psi_1}(\hat{\rho})/\text{Tr}[\mathcal{I}^{(1)}_{\text{d}\psi_1}(\hat{\rho})], \;\; \psi_1 = |\psi_1\rangle\langle\psi_1|, \;\; |\psi_1\rangle \in \mathcal{H}_d\},$$

i.e. without the need to calculate the channels induced by the measurements. One would wish that the states $\{\hat{\rho}'\}$ form an invariant, equiprobable family which, in general (weak measurements, see Section 4.5), is not the case. It is the case if the observers are "greedy" (see the next Section), in which case it turns out that the family of states $\{\hat{\rho}'\}$ is (or can be chosen to be) identical to, and equally distributed as, the family of the original signal states $\{\varrho_0(\psi_0)\}$ (and so on for all observers). In such a case we have to optimize over the encodings / measurements and calculate the optimal fidelity only once, i.e. for the first observer, since we will provide an expression for the maximal fidelities $\mathcal{F}_i$, $i > 1$, in terms of (a quantity derived from) the maximal fidelity $\mathcal{F}_1$.

4.4 "Greedy" observers

Let us now specialize to the case of "greedy" observers who primarily want to maximize the fidelity of their own guesses. The results of this Section for qutit systemts have been published in the paper [53]. Our aim here is to arrive at a result without the need of calculating action of the channels induced by the observers' measurements, i.e. to be able to reduce the task at hand to the problem of single-observer encoding/estimation of quantum states into/from larger systems. Solutions to the latter problem are often known ([39, 31, 47, 3, 43, 1]).

In the special case of "greedy" observers, each observer performs the best estimation he can – in other words, there exists no additional measurement he could perform that would increase the fidelity of his guess that was obtained based on his original measurement. It follows that the post-measurement state after the ith measurement can depend on the original state ψ_0 only indirectly, through the obtained guess ψ_i (more precisely, this has to hold up to a set of measure zero which has no effect on any observer's average fidelities). Thus the dependence of the probability of $\hat{\rho}_1$ on the pre-measurement state $\hat{\rho}_0$ may be dropped – either a particular $\hat{\rho}_0$ reveals something about ψ_0 not revealed by ψ_1 and then the post-measurement state after optimal greedy measurement cannot depend on such $\hat{\rho}_0$, or $\hat{\rho}_0$ reveals nothing about ψ_0 in addition to ψ_1, in which case omitting such dependence will not decrease the fidelity of the second (and any other) observer's guess (more precisely, this argumentation has to be made at the level of actual measurement outcomes $\{o_1\}$; adding-up events o_1 which lead to a guess ψ_1 will make the statement work for ψ_1s). Thus without affecting any observers' fidelities we may write $\mathrm{d}p(\hat{\rho}_1|\psi_1, g_1, \hat{\rho}_0) \equiv \mathrm{d}p(\hat{\rho}_1|\psi_1, g_1)$ (the identity holding up to a set of measure zero) and we have

$$\begin{aligned}\mathrm{d}p(\psi_2|\psi_0) &= \int \mathrm{d}p(\psi_2|\tilde{g}_2, \hat{\rho}_1)\mathrm{d}p(\hat{\rho}_1|\psi_1, g_1)\mathrm{d}p(\psi_1|g_1, \hat{\rho}_0) \\ &\quad \times \mathrm{d}\mu(g_1)\mathrm{d}p(\hat{\rho}_0|g_0, \psi_0)\mathrm{d}\mu(g_0).\end{aligned} \tag{4.48}$$

Since $\mathrm{d}p(\psi_i|g_i, \hat{\rho}_{i-1})$ is linear in the state variable (probabilities of particular guesses are sums of probabilities of measurement outcomes leading to the given guess, the probabilities of outcomes depend on the measured states through the trace formula), integrating over states we have

$$\mathrm{d}p(\psi_2|\psi_0) = \int \mathrm{d}p(\psi_2|\tilde{g}_2, \rho_1^{g_1}(\psi_1))\mathrm{d}p(\psi_1|g_1, \rho_0^{g_0}(\psi_0))\mathrm{d}\mu(g_1)\mathrm{d}\mu(g_0), \tag{4.49}$$

where

$$\rho_i^{g_i}(\psi_i) := \int \hat{\rho}_i \mathrm{d}p(\hat{\rho}_i|g_i, \psi_i), \quad i = 0, 1.$$

The map $\rho_0^{g_0}$ describes an effective deterministic encoding of ψ_0 performed by the preparer (who actually may have used any encoding given by the most general dependence described by the distribution $\mathrm{d}p(\hat{\rho}_0|g_0, \psi_0)$). Similarly, the map $\rho_1^{g_1}$ can be viewed as an effective encoding, applied on the first observer's guess ψ_1, performed by his measurement apparatus plus any post processing (again the actual most general encoding is described by the distribution $\mathrm{d}p(\hat{\rho}_1|\psi_1, g_1)$).

Integrating over $g_0 \in \mathrm{SU}(d)$ ($\mathrm{d}p(\psi_1|g_1, \rho_0^{g_0}(\psi_0))$ is linear in $\rho_0^{g_0}(\psi_0)$) we get an average, deterministic, encoding ϱ_0 defined by

$$\varrho_0(\psi_0) := \int \rho_0^{g_0}(\psi_0)\mathrm{d}\mu(g_0).$$

Note that

$$\mathrm{d}p(\psi_i|g_i, \rho_{i-1}^{g_{i-1}}(\psi_{i-1})) = \mathrm{Tr}[M_i^{g_i}(\mathrm{d}\psi_i)\rho_{i-1}^{g_{i-1}}(\psi_{i-1})],$$

where $M_i^{g_i}$ is a POVM performed by the ith observer which includes the post-processing to produce a guess (the actual POVM is, say, $E_i^{g_i}$ with arbitrary outcome space $\{o_i\}$; the POVM including the guessing is defined by

$$M_i^{g_i}(\mathrm{d}\psi_i) := \int E_i^{g_i}(\mathrm{d}o_i)$$

where the integral is over all o_i which lead to the guess ψ_i).

Performing the integral over $g_1 \in \mathrm{SU}(d)$ we get

$$\int \rho_1^{g_1}(\psi_1) \mathrm{d}p(\psi_1|g_1, \varrho_0(\psi_0)) \mathrm{d}\mu(g_1) =: \varrho_1(\psi_1) \mathrm{d}p(\psi_1|\varrho_0(\psi_0)),$$

where

$$\mathrm{d}p(\psi_1|\varrho_0(\psi_0)) =: \mathrm{Tr}[\mathcal{M}_1(\mathrm{d}\psi_1)\,\varrho_0(\psi_0)]$$

defines an effective POVM $\mathcal{M}_i$ performed by the ith observer from the point of view of the kth observer – i.e. an averaged POVM given by any POVM $M_i^{g_i}$, chosen by the ith observer and optimal for the ensemble of states on the input of his apparatus, and by averaging over the unknown parameters of the POVM $M^{(i)}$ from the point of view of the kth observer, as discussed in the Introduction. Likewise, the encodings ϱ_i are averaged.

Following the same line of reasoning for all $i = 1, ..., k-1$, and using the short-handed notation $p(\psi_i|\varrho_{i-1}(\psi_{i-1})) =: p(\psi_i|\psi_{i-1})$, we obtain

$$p(\psi_k|\psi_0) \;=\; \int \mathrm{d}\psi_{k-1}\, p(\psi_k|\psi_{k-1}) ... \int \mathrm{d}\psi_0\, p(\psi_1|\psi_0), \tag{4.50}$$

with

$$p(\psi_i|\psi_{i-1}) \;=\; \mathrm{Tr}[\mathcal{M}^{(i)}(\psi_i)\,\varrho_{i-1}(\psi_{i-1})].$$

Using the Bloch-vector formalism we may rewrite the average fidelity Eq. (4.37) as

$$F_k \;=\; \frac{1}{d}\left(1 + (d-1) \int \mathrm{d}\psi_0 \mathrm{d}\psi_k \boldsymbol{n}(\psi_k) \cdot \boldsymbol{n}(\psi_0) \tilde{p}(\psi_k|\psi_0)\right) \tag{4.51}$$

where $\boldsymbol{n}(\varphi)$ stands for the generalized Bloch vector of a pure state $\varphi \in \mathcal{S}(\mathcal{H}_d)$ (see Appendix B for details).

We argue in Appendix A that both the (averaged) POVM $\mathcal{M}^{(i)}$ and the (averaged) encoding ϱ_i are covariant. The performance, in terms of average fidelity of the guess, of the covariant POVM is by definition the same as the performance of the actual POVM performed. Hence, without loss of generality, we may restrict our attention to covariant POVMs optimal for the set of states equiprobable from the invariant family

$$\{\varrho_{i-1}(\psi_{i-1}) = U_{g_{i-1}} \varrho_{i-1}(\psi_{\mathrm{ref}}) U_{g_{i-1}}^{\dagger}, \psi_{i-1} = g_{i-1} \psi_{\mathrm{ref}} g_{i-1}^{\dagger}, g_{i-1} \in \mathrm{SU}(d), U_{g_{i-1}} = g_{i-1}^{\otimes N}\}.$$

For such a situation, we show in Appendix (B) that

$$\int \mathrm{d}\psi_{i-1}\, \boldsymbol{n}(\psi_{i-1}) p(\psi_i|\psi_{i-1}) = \Delta_i \boldsymbol{n}(\psi_i), \tag{4.52}$$

where Δ_i is a number.

Plugging Eqs. (4.50) and (4.52) into Eq. (4.51) we have

$$\begin{aligned} F_k \;&=\; \frac{1}{d}\left(1 + (d-1) \prod_{i=1}^{k} \Delta_i \int \mathrm{d}\psi_k \boldsymbol{n}(\psi_k) \cdot \boldsymbol{n}(\psi_k)\right) \\ &=\; \frac{1}{d}\left(1 + (d-1) \prod_{i=1}^{k} \Delta_i\right). \end{aligned} \tag{4.53}$$

Thus, successive maximizations of $F_1, F_2, ..., F_k$ are achieved via successive maximizations of $\Delta_1, ..., \Delta_k$. The maximization of Δ_i is over the pair – covariant encoding ϱ_{i-1}, covariant POVM $\mathcal{M}_i$ optimal for the set of states $\varrho_{i-1}(\psi_{i-1})$ with unknown, hence equiprobable, previous observer's guess $\psi_{i-1} = |\psi_i\rangle\langle\psi_i| \in \mathcal{S}(\mathcal{H}_d)$.

If the initial encoding ϱ_0 has been optimal, then one cannot achieve a better performance than if we take $\forall i$, $\varrho_i \equiv \varrho_0$, i.e. $\Delta_i = \max \Delta_1 =: \Delta$. Hence, the maximum $\mathcal{F}_k$ of the average fidelity F_k if all F_i, $i < k$ are, one-after-another, maximal reads

$$\mathcal{F}_k = \frac{1}{d}\left[1 + (d-1)\Delta^k\right], \tag{4.54}$$

where by taking $k = 1$ we get

$$\Delta = \frac{\mathcal{F}_1 d - 1}{d-1}. \tag{4.55}$$

The situation is different if the initial encoding has been restricted by some additional requirements, e.g. encoding into copies of the state ψ_0, leading to some suboptimal encoding ϱ_0'. Then, for $i \geq 1$, the best strategy is, naturally, to take ϱ_i equal to the unrestricted optimal ϱ_0. However, it makes more sense to assume that the restrictions on the initial encoding are valid for all observers – see discussion at the beginning of Section 4.4.2. With this additional requirement, the result, Eq. (4.54) holds for suboptimal encodings, too.

What is left to do in order to evaluate $\mathcal{F}_k$, in the greedy observers case, is to calculate $\mathcal{F}_1$, which we will do shortly.

4.4.1 The Fidelity for the optimal N-qubit encoding

For the qubit case, $d = 2$, the optimal procedure of encoding/estimation of a single qubit in a pure state into/from the state space of N qubits is known [3] (see also Section 3.4).

The optimally prepared state ($k = 0$) as well as the state after the kth measurement, $k > 1$, reads

$$\varrho_k(\boldsymbol{n}_k) = U(\boldsymbol{n}_k)|A\rangle\langle A|U^\dagger(\boldsymbol{n}_k); \quad k \geq 0, \tag{4.56}$$

where (for simplicity we assume that N is even)

$$|A\rangle = \sum_{j=0}^{N/2} A_j |j, 0\rangle \tag{4.57}$$

with the coefficients A_j such that $|A\rangle$ is the eigenvector corresponding to the maximal eigenvalue of the matrix Eq. (3.48) with

$$l = \frac{N}{2} + 1$$

and

$$\begin{aligned} d_i &= 0 \\ c_i &= \frac{i}{\sqrt{(2i+1)(2i-1)}}. \end{aligned}$$

Note that the state (and the measurement below) is specified only on the relevant subspace, i.e. if a spin-j representation has multiplicity, only on one occurrence.

The operator density of the optimal measurement is given by Eq. (3.49) with

$$|B\rangle = \sum_{j=0}^{N/2} \sqrt{2j+1}|j,0\rangle.$$

In this case

$$\Delta = x_{N/2+1}, \tag{4.58}$$

where $x_{N/2+1}$ is the largest zero of the Legendre polynomial $P_{N/2+1}(x)$. Thus

$$\mathcal{F}_k^{\text{opt}} = \frac{1}{2}\Big[1 + x_{N/2+1}^k\Big]. \tag{4.59}$$

Asymptotically, it is known that

$$x_n = 1 - \frac{\xi_0^2}{2n^2} + \cdots,$$

where $\xi_0 \doteq 2.4$ is the first zero of the Bessel function $J_0(x)$. Hence, asymptotically,

$$\Delta \cong 1 - \frac{2\xi_0^2}{N^2} \tag{4.60}$$

and, asymptotically,

$$\mathcal{F}_k^{\text{opt}} \cong \frac{1}{2}\left[1 + \left(1 - \frac{2\xi_0^2}{N^2}\right)^k\right]. \tag{4.61}$$

4.4.2 The Fidelity for N parallel qu*d*its

Let us now consider the situation when the preparer and all observers' post-measurement states are restricted to use only the totally symmetric subspace of the state space $\mathcal{S}(\mathcal{H}_D)$, i.e. encode the qu*d*it into N copies of itself. This may be due to experimental hurdles or simply because nature might "use" such an encoding, which we then desire to leave in place for all observers, at least in the greedy observer's case. In the same manner, measuring magnetisation of a magnet may disturb the magnet as a whole but does not reorganize its domain structure.

To further motivate the requirement to keep, in the greedy observers case, the same encoding of ψ_k for all k, note that it also follows, in the greedy observers case, from requiring that the measured state remains undisturbed by a measurement in the case of a sequence of correct estimates. With this motivation in mind, we will now study what happens in this sub-optimal-encoding scenario with the restricted, symmetrical, state space for all k.

One of the optimal POVMs, $\mathcal{M} = \mathcal{M}^{\text{sym}} + \mathcal{M}^{\text{sym}^\perp}$, for the encoding into copies is known to be the extremal covariant POVM [39] (see Section 3.2.2), with the operator density on the relevant, symmetric, subspace given by

$$\widetilde{\mathcal{M}}^{\text{sym}}(\psi) = d_N^{\text{sym}} |\psi\rangle\langle\psi|^{\otimes N} \tag{4.62}$$

where

$$|\psi\rangle^{\otimes N} = (g|\psi_{\text{ref}}\rangle)^{\otimes N}, \;\; g \in SU(d), \;\; |\psi_{\text{ref}}\rangle \in \mathcal{H}_d. \tag{4.63}$$

The resolution of identity on the orthocomplement of the symmetric subspace $\mathcal{S}(\mathcal{H}_D^{\text{sym}})$ may be arbitrary, since all states $|\psi\rangle^{\otimes N}$ belong to $\mathcal{H}_D^{\text{sym}}$ and so the contribution of $\mathcal{M}^{\text{sym}^\perp}$ to any outcome's observation probability vanishes.

The maximal single-observation fidelity is

$$\begin{aligned}\mathcal{F}_1^{\mathrm{par}} &= \int \mathrm{d}\psi\, \mathrm{d}\hat{\psi}\ |\langle\psi|\hat{\psi}\rangle^N|^2 p(\hat{\psi}|\psi)\\ &= d_N^{\mathrm{sym}} \int \mathrm{d}\psi\ |\langle\psi|\psi_0\rangle|^{2(N+1)}\\ &= d_N^{\mathrm{sym}} \langle\psi_{\mathrm{ref}}^{\mathrm{sym}}|\Big[\int \mathrm{d}\mu(g)\ U(g)|\psi_{\mathrm{ref}}^{\mathrm{sym}}\Big\rangle\Big\langle\psi_{\mathrm{ref}}^{\mathrm{sym}}|U(g)^{\dagger}\Big]|\psi_{\mathrm{ref}}^{\mathrm{sym}}\rangle\\ &= \frac{d_N^{\mathrm{sym}}}{d_{N+1}^{\mathrm{sym}}},\end{aligned} \tag{4.64}$$

where $|\psi_{\mathrm{ref}}^{\mathrm{sym}}\rangle = |\psi_{\mathrm{ref}}\rangle^{\otimes(N+1)}$ belongs to the representation space of the symmetric representation

$$\underbrace{\square\square\cdots\square\square}_{N+1},$$

U, whose dimension, d_{N+1}^{sym}, can be computed using the formula

$$d_{N+1}^{\mathrm{sym}} = \frac{\prod_{i<j}^{d}\ (l_i - l_j)}{1!2!3!\cdots(d-1)!},$$

where $l_i = \lambda_i + d - i + 1$. Here $\lambda_1 = N + 1$ and $\lambda_k = 0$ for $k > 1$, hence $l_1 = N + d + 1$ and $l_k = d - k + 1$ for $k > 1$. We have

$$l_1 - l_j = N + j$$

and

$$l_i - l_j = j - i; \qquad i > 1,$$

which implies

$$d_N^{\mathrm{sym}} = \binom{N + d - 1}{N}. \tag{4.65}$$

Substituting Eq. (4.65) into Eq. (4.64) we get

$$\begin{aligned}\mathcal{F}_1^{\mathrm{par}} &= \frac{(N+1)!(d-1)!}{(N+d)!}\frac{(N-1+d)!}{N!(d-1)!}\\ &= \frac{N+1}{N+d}.\end{aligned}$$

Using Eq. (4.55) we have

$$\mathcal{F}_k^{\mathrm{par}} = \frac{1}{d}\left[1 + (d-1)\left(\frac{N}{N+d}\right)^k\right]. \tag{4.66}$$

4.4.3 A (brief) look at the mutual information

Let us make a small detour at this point. In the preceding sections we have often used expressions like extraction of *information*, whereas the quantity studied has been the average estimation fidelity. In the information theory, information (content of a message) has a well defined meaning and is measured in terms of mutual-information, Def. 2.24. Although both of the two quantities capture certain aspects of how well input and output of a channel are correlated, in general they may lead to different solutions of concrete problems, e.g. that of some cases of optimal detection [27].

The two quantities are usually relevant in different contexts. A Bayes cost approach with fidelity as the cost function is suitable in situations when a decision is to be made after a single measurement. The mutual-information-based approach is more relevant in information-theoretical problems, when one has a sequence of runs of the same "experiment" with coding and decoding as in communication systems [27, 56, 2]. In the information-theoretic (detection) problem the number of measurement outcomes is free to vary whereas in the Bayes cost problem it is often, unlike in our case however, fixed [27].

Although the fidelity-based approach is more relevant in our context, we can consider the problem at hand in terms of information theory, too. Our (naive, as we sill see shortly) motivation is to look for a simple relationship between the behavior, with respect to the number of observations, k, of the maximal achievable mutual information and of other quantities like the already evaluated maximal achievable average fidelity, or the Holevo quantity, Eq. (2.44).

In particular with regard to the latter it seems not to be totally unreasonable to hope for a dependence of both on k that would be mutually related in some simple way. The hope stems from the following: We consider, for each $\boldsymbol{n}$, the average (over measurement outcomes) state at the input of the kth observer's apparatus, (i.e. the state $\phi^{\circ(k-1)}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|$, ϕ being the channel induced by an observer's measurement) as a (quantum) letter encoding the letter $\boldsymbol{n}$ from the classical (distinguishable) alphabet $\{\boldsymbol{n}\}$. For each k the quantum letters are formally passed through an ideal channel and then estimated by the kth observer. Such situation is discussed in Section 2.5.1.1. The Holevo quantity is the quantum (von Neumann) mutual information between a classical system P (its states given by the alphabet $\{\boldsymbol{n}\}$) and the signal quantum system (its states given by $\{\phi^{\circ(k-1)}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|\})$. The maximal achievable mutual information is the quantum mutual information (in this case equal to its classical counterpart) between the system P and set estimates $\{\boldsymbol{n}_k\}$ (more precisely between the corresponding distributions).

Roughly speaking, the Holevo quantity captures the decreased distinguishability of the symbols $\{\boldsymbol{n}\}$ due to encoding into mixed states (by subtracting weighted entropies of the letters). On the other hand, the mutual information captures all the sources of decreased distinguishability – non-orthogonality of $\{|\boldsymbol{n}\rangle\langle\boldsymbol{n}|\}$ as well as mixedness of each $\phi^{\circ(k-1)}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|)$. Our hope would require that there be some simple k-independent transformation (which we would be in position to "guess") which, applied to the Holevo quantity would lead to some expression in terms of the mutual information.

Let us consider the simplest possible case of a single spin-1/2 particle, which is sufficient to see that that our hopes have been indeed naive. In terms of the information theory, the problem at hand can be seen as communication of classical information over a quantum channel. Mathematically, what is being communicated is (given a reference frame and a parametrization of the unit vectors) the tuple (θ, φ), where $\theta \in \langle 0, \pi)$ and $\varphi \in \langle 0, 2\pi)$, i.e. a "random variable" $\mathbb{n}$. Denoting by $\mathbb{n}_k$ the random variable of the k-th observer's measurement outcomes, we wish to evaluate the mutual information

$$H(\mathbb{n};\mathbb{n}_k) = \int \mathrm{d}\boldsymbol{n} \int \mathrm{d}\boldsymbol{n}_k \tilde{p}(\boldsymbol{n}_k, \boldsymbol{n}) \log_2 \frac{\tilde{p}(\boldsymbol{n}_k, \boldsymbol{n})}{\tilde{p}(\boldsymbol{n}_k)\tilde{p}(\boldsymbol{n})}, \tag{4.67}$$

which quantifies the correlation between the prepared and estimated directions distributions.

Hence the only thing we need is the joint probability density $\tilde{p}(\boldsymbol{n}_k, \boldsymbol{n})$ for a sequence of k measurements optimal in the spirit of our greedy scenario, Section 4.4, however with respect to averaging of a new "fidelity" function

$$g(\boldsymbol{n}, \boldsymbol{n}_k) = \log_2 \frac{\tilde{p}(\boldsymbol{n}_k, \boldsymbol{n})}{\tilde{p}(\boldsymbol{n}_k)\tilde{p}(\boldsymbol{n})}. \tag{4.68}$$

Due to its non-linearity in measurements (and states) the mutual-information-based optimal estimation is, in general, a more difficult problem. However, using results by Davies [27] (which also hold for the case of compact groups [21]) in our case we can again restrict ourselves to covariant POVMs (following the same line of reasoning as in Sections 4.3 and 4.4 which ensure us that at each step we can restrict ourselves to a covariant estimation problem). It is easy to convince oneself that the optimal covariant POVMs, update rules and induced channels are the same as in the fidelity-based problem thus we can use, for each k, the corresponding joint probability (see also [60]).

Having the joint probability density, Eq. (4.26), at hand (for the single qubit case), one can calculate the mutual information I_1 of the encoded and estimated direction distributions if measurements optimal with respect to the one-qubit fidelity are performed by the observers. A direct calculation gives:

$$I_1(\mathbb{n}; \mathbb{n}_k) \;=\; \begin{cases} 1 - \frac{1}{2}\log_2 \mathrm{e} & k = 1 \\ 1 - \frac{1}{2}\log_2 \mathrm{e} + \log_2\left(p^{(k-1)}\right) + \frac{1}{2}p_\perp^{(k-1)}(3^{k-1} - 1)\log_2\left(\frac{p^{(k-1)}}{p_\perp^{(k-1)}}\right) & k > 1, \end{cases} \tag{4.69}$$

where $p^{(i)} = \frac{1}{2}\left(1 + \frac{1}{3^i}\right)$ and $p_\perp^{(i)} = \frac{1}{2}\left(1 - \frac{1}{3^i}\right)$.

Having the explicit expression Eq. (4.69) we are tempted to seek for a (possibly simple) relation of I_1 to the Holevo quantity (accessible information) Eq. (2.44). We perform a direct calculation of the Holevo bound for the k-th observer

$$\chi_H^{(k)} = S(\hat{\rho}_{k-1}) - \int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n}\tilde{p}(\boldsymbol{n})S(\hat{\rho}_{k-1,\boldsymbol{n}}),$$

where (from Eq. (4.34)) $\hat{\rho}_{k-1,\boldsymbol{n}} = c^k\hat{\rho}_{0,\boldsymbol{n}} + \frac{1-c^k}{2}\mathbb{1} = p^{(k-1)}|\boldsymbol{n}\rangle\langle\boldsymbol{n}| + p_\perp^{(k-1)}|-\boldsymbol{n}\rangle\langle-\boldsymbol{n}|$.

Since the average (over possible signal states) state on the input of the k'th observer, $\hat{\rho}_{k-1} = \int \mathrm{d}\boldsymbol{n}\tilde{p}(\boldsymbol{n})\hat{\rho}_{k-1,\boldsymbol{n}}$, is a total mixture (i.e. $S(\hat{\rho}_{k-1}) = 1$ for all k), and the eigenvalues $p^{(k-1)}$, $p_\perp^{(k-1)}$ are independent of $\boldsymbol{n}$, and $\int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n}\tilde{p}(\boldsymbol{n}) = 1$ (probability sums to one), we have

$$\chi_H^{(k)} = 1 + p^{(k-1)}\log_2 p^{(k-1)} + p_\perp^{(k-1)}\log_2 p_\perp^{(k-1)}. \tag{4.70}$$

Although the expressions (4.69) and (4.70) are both expressed in terms of the eigenvalues of $\hat{\rho}_{k-1,\boldsymbol{n}}$, it seems that one is not expressible in terms of the other in a simple manner.

To summarize the present subsection, our hope to find a simple mutual relationship connecting the behavior of maximal achievable mutual-information with that of the maximal average estimation fidelity or with the Holevo quantity under a increasing number of observations has turned up to be naive, even for the simplest, single-qubit, system. As a by-product, we have argued that, at least for the "greedy" scenario considered in this presentation, also in the mutual-information-based problem it suffices to restrict oneself to the covariant estimation/instruments approach giving the same possible optimal encoding/estimation strategies as in the fidelity based approach. The maximal achievable mutual-information for the single-qubit case, as a function of k, has been evaluated. In the remainder of the present Thesis we will consider exclusively the fidelity-based approach.

4.5 Weak measurements

In this Section we generalize the problem to include situations where the observers optimize their measurements to pursue goals different from, in the first place, mere maximization of the quality of their own guesses. In particular, we study the case where K observers estimate the original state with equal, but maximal, fidelity (equalitarian strategy) and the case where the observers use the same measurement apparatus such that the quality of the last observer's estimate is maximized. In both cases the measurements performed are weak, i.e. in general not extracting all of the extractable information which enables less disturbance to be applied to the measured state. In the former case we quantify how the measurements are more and more greedy with increasing tally number, k, until the last observer performs a greedy measurement as well as the maximal achievable, equal, fidelity of the observers' guesses. In the latter case we show there exists, and we calculate, an optimal "strength" of the measurement to be performed by all the observers as well as the observers' fidelities achieved.

4.5.1 Equalitarian observers

We devise a protocol such that the fidelity of the guess of the state estimate obtained by each observer is the same and maximal, i.e. we want to find and reach the maximum $\mathcal{F}_k \equiv \mathcal{F}_{\rm eq}$ of F_k under the constraint $\forall k \in \{2, ..., K\}$, $F_k = F_1$. The overall number of observers, K, is fixed beforehand and each observer knows his tally number k in the sequence. Alternatively, one can imagine that the observers use the same, up to a "measurement-strength," apparatus whose measurement-strength is adjusted automatically before a measurement. We again do not allow communication between observers.

Within the conditions of our problem, it is clear that the last observer will perform an optimal measurement for the ensemble of states on the input of his apparatus, while going backwards each of his predecessor's measurement will be weaker and weaker, i.e. less and less demolishing.

As in the greedy observers scenario, it suffices to consider covariant measurements (see Appendix A). All U-covariant POVMs (with the estimates as outcomes) are of the form

$$\mathcal{M}(\mathrm{d}\psi) \sim U_g S_{\rm ref} U_g^\dagger \mathrm{d}\psi \qquad (\psi = g\psi_{\rm ref} g^\dagger,\, g \in \mathrm{SU}(d)) \tag{4.71}$$

where $S_{\rm ref}$ can be any density operator commuting with $\{U_g; g \in G_{\rm ref}\}$ where $G_{\rm ref} \subset \mathrm{SU}(d)$ is the set of unitaries which leave the reference state $\varrho_0(\psi_{\rm ref})$ invariant [39].

It is clear that, for optimal weak measurements, the post-measurement states will not in general be pure states anymore. They will depend not only on the measurement outcome (more precisely guess) of the current observer but on particular guesses of all predecessing observers and the preparation parameter ψ_0. Thus, we have to start from scratch with the histories decomposition Eq. (4.44) which, for general k, reads

$$\mathrm{d}p(\psi_k|\psi_0) \;=\; \int \mathrm{Tr}\Big[\mathcal{M}^{(k)}(\mathrm{d}\psi_k)\mathcal{I}^{(k-1)}_{\mathrm{d}\psi_{k-1}} \circ ... \circ \mathcal{I}^{(1)}_{\mathrm{d}\psi_1}(\varrho_0(\psi_0))\Big]. \tag{4.72}$$

In particular, the above integral does not simplify to Eq. (4.50) as now probability density of obtaining measurement outcome leading to a guess ψ_k, given the previous observer has obtained the guess ψ_{k-1}, is not independent of previous observers' guesses, i.e. $p_k(\psi_k|\psi_{k-1}, ..., \psi_0) = p(\psi_k|\psi_{k-1})$ does not hold in general. However, the average fidelity Eq. (4.37) is independent on the intermediate guesses, ψ_1, ..., ψ_{k-1}, trace is linear and quantum channels are (convex)-linear in their input states. Thus, we can integrate through ψ_{k-1}, then by ψ_{k-2} and so on in the decomposition Eq. (4.72). We get

$$\mathrm{d}p(\psi_k|\psi_0) \;=\; \int \mathrm{Tr}\Big[\mathcal{M}^{(k)}(\mathrm{d}\psi_k)\chi_{k-1}\circ ...\circ\chi_1(\varrho_0(\psi_0))\Big], \tag{4.73}$$

where χ_i is the channel induced by the ith observer's covariant measurement (given the actual measurement, the inclusion of guessing and the averaging over its unknown parameters).

That is, we can view the whole situation in the following equivalent way: from the point of view of the preparer the state after the first measurement, with unknown outcome, is

$$\varrho_1(\psi_0) = \chi_1(\varrho_0(\psi_0)), \tag{4.74}$$

i.e. the preparer plus the first observer who keeps the outcome of his measurement for himself act together effectively as a new preparation apparatus producing covariant encodings of the state ψ_0 for the second observer. As far as the average fidelity of his estimation is concerned, the second observer's measurement optimization for uniformly (over $\psi_0 = |\psi_0\rangle\langle\psi_0| \in \mathcal{S}(\mathcal{H}_d)$) distributed states Eq. (4.74), is equivalent to measurement optimization for states

$$\tilde{\varrho}_1(\psi_1, \varrho_0(\psi_0)) = \frac{\mathcal{I}^{(1)}_{\mathrm{d}\psi_1}(\varrho_0(\psi_0))}{\mathrm{Tr}[\mathcal{I}^{(1)}_{\mathrm{d}\psi_1}(\varrho_0(\psi_0))]}$$

distributed according to

$$\mathrm{d}p(\psi_1, \psi_0) = \mathrm{Tr}[\mathcal{I}^{(1)}_{\mathrm{d}\psi_1}(\varrho_0(\psi_0))]\mathrm{d}\psi_0.$$

The situation for the rest of the observers goes in the same spirit.

To proceed further, we need to calculate actions of the channels $\chi_k, k = 1, ..., K-1$. We will do that in what follows for the qubit case restricted to encoding into copies and for the single qu$d$it case.

4.5.2 Single copy, any dimension

Let us start with the case of any dimension, d, of a single copy of a unknown pure state $\psi_0 = |\psi_0\rangle\langle\psi_0|$ that is, the $d = D$ case. A qu$d$it being measured using a SU(d)-covariant instrument undergoes, if the measurement outcome is unknown, dynamics given by a channel χ which is SU(d)-invariant, i.e. a convex combination of the identity channel and the contraction to total mixture, acting as

$$\chi(\hat{\rho}) = r\hat{\rho} + (1-r)\mathbb{1}/d. \tag{4.75}$$

The kth observer's fidelity of the guess of an original reference state ψ_{ref}, for an effectively encoded state $\hat{\rho}^{(k-1)}_{\mathrm{ref}} = \chi_{k-1} \circ ... \circ \chi_1(|\psi_{\mathrm{ref}}\rangle\langle\psi_{\mathrm{ref}}|)$ – the result of sending $|\psi_{\mathrm{ref}}\rangle\langle\psi_{\mathrm{ref}}|$ through the SU(d)-invariant channels $\chi_1, ..., \chi_{k-1}$ – is given by

$$F_k \;=\; \sum_{o_k} \int \mathrm{d}U\; \mathrm{Tr}(U|\psi_{\mathrm{ref}}\rangle\langle\psi_{\mathrm{ref}}|U^\dagger U_{o_k}|\psi_{\mathrm{ref}}\rangle\Big\langle\psi_{\mathrm{ref}}|U^\dagger_{o_k}\Big)\mathrm{Tr}\Big(U\hat{\rho}^{(k-1)}_{\mathrm{ref}}\,U^\dagger M^{(k)}_{o_k}\Big), \tag{4.76}$$

$U \in \mathrm{SU}(d)$.

The state we wish to estimate is $U|\psi_{\mathrm{ref}}\rangle\langle\psi_{\mathrm{ref}}|U^\dagger$ and our guess we write as $\psi_{o_k} = U_{o_k}|\psi_{\mathrm{ref}}\rangle\langle\psi_{\mathrm{ref}}|U_{o_k}^\dagger$. Using Eq. (E.1) of Appendix E we obtain

$$F_k = \frac{(dO_S^{(k)} - 1)O_M^{(k)}}{d(d+1)(d-1)} + \frac{d - O_S^{(k)}}{(d+1)(d-1)}, \tag{4.77}$$

where $O_S^{(k-1)}$ is the overlap

$$O_S^{(k-1)} = \mathrm{Tr}\Big(\psi_{\mathrm{ref}}\hat{\rho}_{\mathrm{ref}}^{(k-1)}\Big) \tag{4.78}$$

of the states and $O_M^{(k)}$ is the overlap

$$O_M^{(k)} = \sum_{o_k} \mathrm{Tr}\Big(\psi_{o_k} M_{o_k}^{(k)}\Big). \tag{4.79}$$

For a general SU(d)-invariant qudit channel, Eq. (4.75), induced by the kth measurement and the averaging due to lack of knowledge about it, one has

$$\begin{aligned} F_{k+1} &= r_k \sum_{o_{k+1}} \int \mathrm{d}U\ \mathrm{Tr}(U|\psi_0\rangle\langle\psi_0|U^\dagger U_{o_{k+1}}|\psi_{\mathrm{ref}}\rangle\Big\langle\psi_{\mathrm{ref}}|U_{o_{k+1}}^\dagger\Big)\mathrm{Tr}\Big(U\hat{\rho}_{\mathrm{ref}}^{(k-1)}\,U^\dagger M_{o_{k+1}}^{(k+1)}\Big) \\ &\quad + \frac{1-r_k}{d} \\ &= r_k\Big(F - \frac{1}{d}\Big) + \frac{1}{d}, \end{aligned}$$

F being the average fidelity of the $(k+1)$th observer's guess based on a measurement as if performed on the state $\hat{\rho}^{(k-1)}$ – i.e. independent of the post-measurement state after the kth measurement. The fidelity has the property $1/d \le F \le 1$, where $F = 1/d$ corresponds to pure guessing without actually measuring anything. It follows that in order to maximize the possible F_{k+1} for any fixed measurement $M^{(k+1)}$ (other than mere guessing) one has to have r_k as large as possible. Naturally r_k will be ultimately limited by the achieved F_k but also by choice of the kth observer's measurements given their performance, F_k.

We can always introduce a new variable c defined by

$$r_k = \frac{c-1}{(d+1)(d-1)}, \tag{4.80}$$

where the above equation specifies c if a SU(d)-invariant qudit channel χ_k, i.e. r_k, is given.

Now there are two options for the actually performed measurement whose POVM description $M^{(k)}$ appears in Eqs. (4.76) and (4.79). The first option is that the quantum operation performed, upon obtaining any outcome o_k, is given by a single-term Kraus decomposition ($\forall o_k$, $\hat{\rho}_{\mathrm{out}} = A_{o_k}^\dagger\hat{\rho}_{\mathrm{in}}A_{o_k}$). The second options is that there exist some outcomes for which the operation has multiple Kraus operators in its decomposition ($\exists\alpha$; $\hat{\rho}_{\mathrm{out}} = \sum_i B_{\alpha,i}^\dagger\hat{\rho}_{\mathrm{in}}B_{\alpha,i}$). In this case we formally redefine the POVMs used in Eqs. (4.76) and (4.79) – we simply use the language of a fine-grained measurement with POVM elements $M_{o_k} \equiv M_{\alpha,i} := B_{\alpha,i}^\dagger B_{\alpha,i} \equiv A_{o_k}^\dagger A_{o_k}$ and operations defined by $\hat{\rho}_{\mathrm{out}} = A_{o_k}^\dagger\hat{\rho}_{\mathrm{in}}A_{o_k}$ for all αs where a multi-term Kraus decomposition would otherwise take place. If the additional labels i are not used for anything (they are not really accessible to the observer and thus can't influence his guess), these new formal apparata provide an equivalent description. Thus we always end up with a description of the measurement process in terms of an apparatus with single-term Kraus decomposition for each outcome.

For such apparatus, averaged over its unknown “orientation” as always, the parameter c of Eq. (4.80) is given, in terms of its Kraus operators, by Eq. (E.3) (see Appendix E). Recall that we wish to have r_k as large as possible given the kth observer’s achieved fidelity F_k. It follows that we wish, in the language of the single-Kraus-term apparata, the c of Eq. (E.3) as large as possible.

For a given value of F_k, one of measurements both reaching F_k (i.e. the required $O_M^{(k)}$) and maximizing c of Eq. (E.3) is known to be given by [5]

$$A_a^{(k)} = \sqrt{\frac{O_M^{(k)}}{d}}|a\rangle\langle a| + \sqrt{\frac{d-O_M^{(k)}}{d(d-1)}}(\mathbb{1}-|a\rangle\langle a|), \tag{4.81}$$

where $a=1,\ldots,d$ and the projectors $|a\rangle\langle a|$ constructed using any orthonormal basis $\{|a\rangle\}$. Thus the largest c, given F_k (i.e. given $O_M^{(k)}$), is

$$c = \left[\sqrt{O_M^{(k)}} + \sqrt{(d-1)(d-O_M^{(k)})}\right]^2.$$

The corresponding POVM reads

$$\begin{aligned} M_a^{(k)} &= A_a^{(k)\dagger}A_a^{(k)} \\ &= \frac{O_M^{(k)}-1}{d-1}|a\rangle\langle a| + \frac{d-O_M^{(k)}}{d(d-1)}\mathbb{1}, \end{aligned}$$

i.e. for this particular POVM the optimal instrument in terms of Kraus operators is given by the Hermitian square-root $A_a^{(k)} = \sqrt{M_a^{(k)}}$. We could continue our analysis using the (unaveraged) d-outcome measurement, Eq. (4.81), optimal for any achievable value of F_k. However, we will proceed in terms of the effective covariant apparata with measurement “outcomes” given by the possible guesses. The covariant aparata may always be, mathematically, easily constructed, unlike the minimal optimal measurements, a representant of which is that of Eq. (4.81), which have to be laboriously searched for in each new situation even in the case of a pure-state estimation – e.g. adding one more copy of a system (see Ref. [43] for the case of N identical qubits).

It follows from Eq. (4.71) that any SU(d)-covariant POVM on a qu$d$it (with outcomes corresponding to guesses) has the operator density of the form

$$\widetilde{\mathcal{M}}^{(\varepsilon_k)}(\psi_k) = (1-\varepsilon_k)\mathbb{1} + \varepsilon_k\widetilde{\mathcal{M}}(\psi_k), \tag{4.82}$$

where $\widetilde{\mathcal{M}}(\psi_k) = \mathcal{M}(\mathrm{d}\psi_k)/\mathrm{d}\psi_k = d|\psi_k\rangle\langle\psi_k|$ defines the operator density $\widetilde{\mathcal{M}}$ of the optimal covariant POVM $\mathcal{M}$ of the greedy-observers scenario and ε_k parametrizes the strength of the measurement.

Eq. (4.82) leads to

$$O^{(k)}_{\mathcal{M}^{(\varepsilon_k)}} = 1 + \varepsilon_k(d-1), \tag{4.83}$$

where we have used Eq. (4.79) in the form

$$O_{\mathcal{N}}^{(k)} = \int \mathrm{d}\psi_k \mathrm{Tr}\Big(\psi_k \tilde{\mathcal{N}}_{\psi_k}^{(k)}\Big),$$

where $\mathcal{N}$ is a covariant POVM with operator density $\tilde{\mathcal{N}}$. Note that $1 \leq O^{(k)}_{\mathcal{M}^{(\varepsilon_k)}} \leq d$ depending on the "greediness", or strength, ε_k, $0 \leq \varepsilon_k \leq 1$, of the kth observer's measurement. Constructing the corresponding Hermitian-square-root Kraus operators $\mathcal{A}^{(\varepsilon_k)}$ defined by

$$\tilde{\mathcal{A}}^{(\varepsilon_k)}_{\psi} = \sqrt{O^{(k)}_{\mathcal{M}^{(\varepsilon_k)}}}|\psi\rangle\langle\psi| + \sqrt{\frac{d - O^{(k)}_{\mathcal{M}^{(\varepsilon_k)}}}{(d-1)}}(\mathbb{1} - |\psi\rangle\langle\psi|) \tag{4.84}$$

($\tilde{\mathcal{A}}$ is the operator density of the Kraus operator[4.7] $\mathcal{A}$, i.e. $\mathcal{A}_{(\mathrm{d}\psi)^{1/2}} = (\mathrm{d}\psi)^{1/2}\tilde{\mathcal{A}}_{\psi}$), we may verify that for given F_k it implies the same channel as the minimal optimal measurement, Eq. (4.81). Thus, he Hermitian-square-root realization of the general weak covariant POVM, Eq. (4.82), gives the optimal covariant instrument.

Using Eqs. (4.83) and (4.77) we have

$$F_k = \frac{1}{d} + \frac{(d\,O^{(k-1)}_S - 1)\varepsilon_k}{d(d+1)}.$$

Using Eq. (4.75) with a pure initial state, Eq. (4.78) reads

$$O^{(k)}_S = \frac{1}{d} + \frac{d-1}{d}\prod_{\beta=1}^{k} r_{\beta}.$$

Substituting the above equation into the forelast we obtain

$$F_k = \frac{1}{d} + \frac{\epsilon_k(d-1)}{d(d+1)}\prod_{\beta=1}^{k-1} r_{\beta}. \tag{4.85}$$

The condition $F_k = F_{k-1}$ translates into

$$\varepsilon_{k-1} = \varepsilon_k r_{k-1},$$

or, explicitly

$$\varepsilon_{k+1} = \frac{\varepsilon_k(d+1)}{d - 1 + (2-d)\varepsilon_k + 2\sqrt{1 + \varepsilon_k(d-1)}\sqrt{1 - \varepsilon_k}}, \tag{4.86}$$

where the initial condition $\varepsilon_K = 1$ follows from the fact that the last, Kth, observer can measure greedily as there is no subsequent observer to care about. A closed expression seems to be hard to obtain and this is as far as we get for finite K.

For $K \gg 1$ we expect the first measurements to be very weak, i.e. with $\varepsilon \ll 1$. Doing a Taylor expansion in the recursion relation Eq. (4.86) we obtain a simplified difference relation which is then converted into a differential equation. Defining $a_j = \varepsilon_{K+1-j}$, it's solution reads

$$\varepsilon_1 = a_K \simeq \frac{1}{d}\sqrt{\frac{2(d+1)}{K}} \quad (K \gg 1). \tag{4.87}$$

Inserting the above into F_1 of Eq. (4.85) we have, for large K, the maximal average fidelity of each equalitarian observer

$$\mathcal{F}_{\mathrm{eq}}(K,d) \simeq \frac{1}{d}\left[1 + \frac{d-1}{d}\sqrt{\frac{2}{(d+1)K}}\right]. \tag{4.88}$$

[4.7] The square-root of a measure here is only a formal notation. In expressions where something is actually calculated, the measure always appears to the first power. A rigorous treatment of Radon-Nikodym derivatives of quantum instruments can be found in [29] and [40].

4.5.3 N copies of a qubit

Let us continue with the situation where the state to encode into is a state of N copies of a two-dimensional system, which is again known to be in a pure state. This situation can be mapped to the problem of estimating the state of a single D-dimensional system ($D = d_N^{\rm sym} = N + 1$) which is, however, known to be in the restricted set of states from the orbit of a reference pure N-copy state generated by elements of the range of the symmetric $SU(2)$ representation.

From Eq. (4.71) it follows that for the first observer, who is estimating a state $|j, j\rangle \oplus |0\rangle_{\overline{\rm sym}} = |\psi\rangle^{\otimes N}$, any covariant POVM will be a convex combination of the optimal greedy POVM and pure guessing strategy, with operator density given, on the relevant, symmetric, subspace by ($k = 1$)

$$\widetilde{\mathcal{M}}^{(k),\varepsilon_k}(\psi) = (1 - \varepsilon_k)\mathbb{1} + \varepsilon_k \widetilde{\mathcal{M}}^{(k)}(\psi), \tag{4.89}$$

where $\widetilde{\mathcal{M}}^{(1)} \equiv \widetilde{\mathcal{M}}^{\rm sym}$ is the operator density of the optimal covariant POVM of the greedy-observers problem, Eq. (4.62) and $0 \leq \varepsilon_1 \leq 1$ parametrizes the strength of the first observer's measurement. From now on we will drop the irrelevant part of the Hilbert space – the orthocomplement to the symmetrical subspace, $\overline{\rm sym}$.

As far as the post-measurement states are concerned, we again consider the Hermitian-square-root[4.8] dynamics which is given, on the symmetric subspace, by Kraus operators densities ($k = 1$)

$$\tilde{\mathcal{A}}_k^{\varepsilon_k}(\psi) \;=\; \underbrace{\left(\sqrt{1 + (d_N^{\rm sym} - 1)\varepsilon_k} - \sqrt{1 - \varepsilon_k}\right)}_{b_k} \frac{\widetilde{\mathcal{M}}^{(k)}(\psi)}{d_N^{\rm sym}} + \underbrace{\sqrt{1 - \varepsilon_k}}_{a_k}\mathbb{1}. \tag{4.90}$$

It is shown in Appendix C that such evolution leads to a channel which leaves the post-measurement state, after averaging over guesses, diagonal in the $\{|j,\, m\rangle\}$ basis with ψ defining the z axis is uniformly distributed over pure states. For such ensemble all covariant POVMs are formed using $S_{\rm ref}$ of Eq. (4.71), diagonal in the $\{|j, m\rangle\}$ basis where now $|j,\, j\rangle \oplus |0\rangle_{\overline{\rm sym}} = |\psi_{\rm ref}\rangle^{\otimes N}$ thus a general weak covariant measurement is parametrized by $d_N^{\rm sym} - 1$ independent "measurement-strength" parameters. Motivated by the single-qu*d*it case we will, however, consider only convex combination of the optimal greedy POVM and pure guessing strategy, with operator density given by Eq. (4.89), where the optimal strategy for reference states $|j,\, j\rangle$ or those diagonal in the $\{|j, m\rangle\}$ is the same, i.e. $\widetilde{\mathcal{M}}^{(2)} \equiv \widetilde{\mathcal{M}}^{\rm sym}$ (see [39]). We again consider the update rule given by Eq. (4.90), i.e. a coherent superposition optimal greedy measurement and guessing, with single measurement-strength parameter ε_2. The situation for $k > 2$ remains the same, hence the observers' measurements differ exclusively by the corresponding measurement strength, ε_k, i.e $\widetilde{\mathcal{M}}^{(k),\varepsilon_k} =: \widetilde{\mathcal{M}}^{\varepsilon_k}$, $\tilde{\mathcal{A}}_k^{\varepsilon_k} =: \tilde{\mathcal{A}}^{\varepsilon_k}$.

[4.8] If we consider only the symmetric subspace, i.e. a POVM giving a resolution of identity only on the symmetric subspace only.

We emphasize that although the above two restrictions seem to be a reasonable guess for a generalization of the optimal apparatus from single-copy case, we do not have a proof that, for $N>1$, such apparata really are among the optimal ones. Therefore, it is only guaranteed that we obtain a lower bound $\mathbb{F}_{\mathrm{eq}}$ on the maximum $\mathcal{F}_{\mathrm{eq}}$, i.e. $\mathbb{F}_k\leq\mathcal{F}_{\mathrm{eq}}$.

In other words we seek the maximum of the constrained maximization of $F_k(\varepsilon_k)$ containing the conditional probability

$$\begin{aligned} p(\psi_k|\psi_0) &= p(\psi_k|\psi_0,\varrho_{k-1}(\psi_0)\\ &= \chi^{\varepsilon_{k-1}}\circ\ldots\circ\chi^{\varepsilon_1}\varrho_0(\psi_0)), \end{aligned}$$

where χ^{ε_i} are the channels induced by the POVMs $\mathcal{M}^{\varepsilon_i}$, Eq. (4.89), with the Hermitian-square-root dynamics Eq. (4.90). We start by rewriting the average fidelity $\mathbb{F}_1^{\varepsilon_k}$ of single estimation using the (ε_k-strong) apparatus, Eq. (4.89), of the kth observer measuring on arbitrary set of states $\{\rho(\psi)\}$ in terms of estimation fidelity using the apparatus Eq. (4.62) of the greedy observers problem, $\mathcal{F}_1$,

$$F_1^{\varepsilon_k}=\frac{(1-\varepsilon_k)}{2}+\varepsilon_k\mathcal{F}_1. \tag{4.91}$$

If the states $\rho(\psi)$ are from the invariant family of states

$$\{\rho(\psi)=U(g)^{\otimes N}\rho(\psi_0),U(g)\in SU(2)\},\quad \text{such that} \tag{4.92}$$

$$\mathrm{Tr}\big[J_{n(\psi_0)}\rho(\psi_0)\big]=:\big\langle J_{\boldsymbol{n}(\psi)}\big\rangle_{\rho(\psi)}>0$$

then $\mathcal{F}_1(\rho)$ can be quickly evaluated as

$$\mathcal{F}_1=\frac{1}{2}\left(1+\frac{2\big\langle J_{\boldsymbol{n}(\psi)}\big\rangle_{\rho(\psi)}}{N+2}\right)$$

(see e.g. [39], page 209, taking $F=1-\frac{1}{4}W$, $N+1=2j+1$). Then, by Eq. (4.91),

$$F_1^{\varepsilon_k} \;=\; \frac{1}{2}\left(1+\varepsilon_k\frac{2\big\langle J_{\boldsymbol{n}(\psi)}\big\rangle_{\rho(\psi)}}{N+2}\right). \tag{4.93}$$

Appendix C gives us the post measurement states for each step in a sequence of weak measurements Eq. (4.90) and thus, for each k, we can evaluate the average fidelity of the kth observer $F_k^{\boldsymbol{\varepsilon}}$ where $\boldsymbol{\varepsilon}=(\varepsilon_k,\ldots,\varepsilon_1)$. Formally this is done according to Eq. (4.93) with $\rho(\psi)\mapsto\hat{\rho}_{k-1}^{\boldsymbol{\varepsilon}}=\chi^{\varepsilon_{k-1}}\circ\ldots\circ\chi^{\varepsilon_1}(\rho(\psi))$, i.e.

$$F_k^{\boldsymbol{\varepsilon}}=F_1^{\varepsilon_k}(\hat{\rho}_{k-1}^{\boldsymbol{\varepsilon}})=\frac{1}{2}\left(1+\varepsilon_k\frac{2\langle J_{\boldsymbol{n}}\rangle_{\hat{\rho}_{k-1}^{\boldsymbol{\varepsilon}}}}{N+2}\right). \tag{4.94}$$

For every observer to have the same fidelity ($\mathrm{F}_k^{\boldsymbol{\varepsilon}}=F_l^{\boldsymbol{\varepsilon}}\;\forall l,k$ s.t. $0<l<k\leq K$) it must hold that

$$\varepsilon_k\langle J_n\rangle_{\hat{\rho}_{k-1}^{\boldsymbol{\varepsilon}}}=\varepsilon_l\langle J_{\boldsymbol{n}}\rangle_{\hat{\rho}_{l-1}^{\boldsymbol{\varepsilon}}}. \tag{4.95}$$

To proceed further we need to evaluate how the channels χ^{ε_k} transform $\langle J_{\boldsymbol{n}}\rangle$ for the relevant states, which we do in Appendix C. Comparing Eq. (4.95), with $l = k + 1$, to Eq. (D.3) of Appendix C we get a recurrent relation for the measurement-strength parameter

$$\varepsilon_{k+1} = \frac{(N+1)(N+2)\varepsilon_k}{(N+1)^2 + (N-1)(1-2\varepsilon_k) + 4\sqrt{(1-\varepsilon_k)(1+N\varepsilon_k)} - 2}, \tag{4.96}$$

where $\varepsilon_K = 1$. We need to solve the recurrence relation Eq. (4.96) for $k = 1$. Then, from Eq. (4.94), one gets

$$\mathbb{F}_{\text{eq}}(N,K) = \frac{1}{2}\left(1 + \varepsilon_1(K,N)\frac{N}{N+2}\right). \tag{4.97}$$

A closed expression for ε_1 seems to be very difficult to obtain, and this is as far as we can go for finite K. Therefore we will discuss leading order behaviors of ε_1 and $\mathbb{F}_{\text{fair}}$ in different regimes.

Let us first consider the situation $K \gg N$. For K very large we expect the first measurements to be very weak, i.e with $\epsilon_k \ll 1$, $k \leq k_0 \ll K$. Thus we can do a Taylor expansion to the third order in ε around the point $\varepsilon = 0$ in the recursion relation Eq. (4.96) and get the approximate recurrence relation

$$\varepsilon_{k+1} = \varepsilon_k + \frac{N+1}{2(N+2)}\varepsilon_k^3, \tag{4.98}$$

The difference equation Eq. (4.98) can be transformed into the differential equation

$$\frac{N+1}{2(N+2)}\varepsilon(k)^3 = \frac{\mathrm{d}\varepsilon(k)}{\mathrm{d}k} + \sum_{i=2}^{\infty}\frac{1}{i!}\frac{\mathrm{d}^i\varepsilon(k)}{\mathrm{d}k^i} \cong \frac{\mathrm{d}\varepsilon(k)}{\mathrm{d}k}, \tag{4.99}$$

where we require that $\forall i > 1, \frac{\mathrm{d}^i\varepsilon(k)}{\mathrm{d}k^i} \ll i!$ for $K \geq k \geq k_0$ where k_0 may be different for each K.

By defining $a_j = \epsilon_{K+1-j}$, we obtain the differential equation

$$\frac{\mathrm{d}a(j)}{\mathrm{d}j} = -\frac{N+1}{2(N+2)}a(j)^3$$

which yields

$$a(K) \simeq \frac{1}{\sqrt{\frac{N+1}{N+2}(K-j_0) + \frac{1}{a(j_0)^2}}}, \tag{4.100}$$

where $j_0 = K + 1 - k_0$ is a fixed (possibly different for each K) lower boundary of integration chosen such that the approximations Eq. (4.98) and Eq. (4.99) are valid. For $K \gg j_0$ and $K \gg a(j_o)^{-2}$ we have

$$a(K) \simeq \sqrt{\frac{N+2}{(N+1)K}}. \tag{4.101}$$

Inserting Eq. (4.101) into Eq. (4.97), we have

$$\mathbb{F}^{\text{eq}}(K,N) \simeq \frac{1}{2}[1+\Delta]. \tag{4.102}$$

with

$$\Delta \simeq \frac{N}{\sqrt{(N+1)(N+2)K}} \quad ; \ (K \gg N). \tag{4.103}$$

This is an interesting result since one would naively expect that $\Delta \sim 1/K$. The realization of the POVM Eq. (4.89) as an instrument given by Hermitian-square-root Kraus operators, Eq. (4.90), is crucial to obtain this square root degradation of Δ. Actually, had we used a more destructive realization we would indeed have obtained $\Delta \sim 1/K$.

In particular, if we realize the POVM Eq. (4.89) as a measurement such that with probability $(1-\epsilon_k)$ the outcome is just guessed, i.e. nothing is done to the state, and with probability ϵ_k the optimal greedy covariant measurement is performed, the (relevant part of the) channel induced by such measurement is $\chi'_{\varepsilon_k} = (1-\varepsilon_k)\mathrm{Id} + \varepsilon_k\chi$, where Id is the identity channel. In this case

$$\langle J_{\boldsymbol{n}}\rangle_{k+1} = \left(1 - \frac{\varepsilon_{k+1}}{N/2+1}\right)\langle J_{\boldsymbol{n}}\rangle_k. \tag{4.104}$$

The condition Eq. (4.95) then leads to the recurrence relation

$$\epsilon_k = \frac{N/2+1}{N/2+K-k+1} \tag{4.105}$$

that can easily be solved to give

$$\epsilon_1 = \frac{N+2}{N+2K},$$

which yields

$$\Delta = \frac{N}{N+2K}. \tag{4.106}$$

In this case the fidelity degrades linearly in $1/K$.

Now we proceed to the case $N \gg 1$. If one neglects all additive terms but those proportional to N and N^2 in Eq. (4.96) and solves the so obtained approximate recurrence relation directly one obtains

$$\begin{aligned} \varepsilon_1(K,N) &\simeq \frac{N+3}{N+2(K-1)} \\ &\simeq \frac{N}{N+2(K-1)}. \end{aligned} \tag{4.107}$$

The above solution can also be obtained by Taylor expanding Eq. (4.96) around $1/N = 0$ to the first order and solving the so obtained approximate recurrence relation,

$$\varepsilon_{k+1} = \varepsilon_k + 2\varepsilon_k^2/N, \tag{4.108}$$

using the method of converting it into an approximate differential equation which is then solved. One obtains

$$a(K) \simeq \frac{1}{\frac{2(K-k_0)}{N} + \frac{1}{a(k_0)}}.$$

Taking $k_0 = 1$ we recover the result Eq. (4.107).

Hence, we have

$$\mathbb{F}_{\text{eq}}(K,N)\simeq\frac{1}{2}\left[1+\frac{N+2}{N+2(K-1)}\right] \;\;;\;\; (N\gg K).$$

Note that, in addition to the condition $N\gg 1$, the condition $N\gg K$ has to be imposed, since otherwise we would have used the, for finite N only approximate, recurrence rule too many times.

The numerical evaluation of the (inverted) exact and approximate recurrence relations and approximate solutions for $N=10^3$ allowing to depict the $K\gg N$, $N\gg K$ and intermediate regimes are plotted in Fig. 4.2.

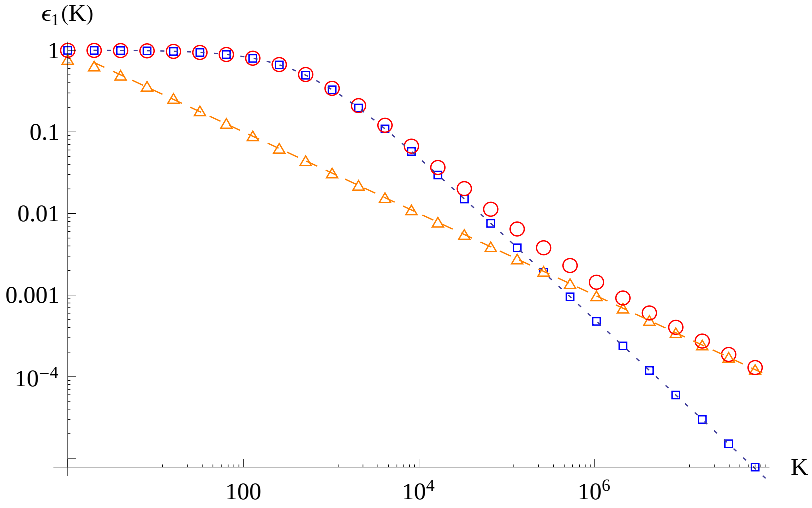

Figure 4.2. The first observer's measurement strength, ε_1, as a function of the number of observers, K. (Also: the kth observer's measurement strength, $\varepsilon_k(K+1-k)$, as function of K, $k=1$, ..., whole part$[(K+1)/2]$.) $N=10^3$ has been chosen to depict both the $K\gg N$ and $N\gg K$ regimes, as well es the intermediate regime. Logarithmic scale is used for both axes. Plotted quantities – selected points of numerical evaluation of the (inverted) recurrence relations: exact, Eq. (4.96), (circles), approximate as of Eq. (4.108) (squares), approximate as of Eq. (4.98) (triangles); solutions given by: Eq. (4.101) (dashed), Eq. (4.107) (dotted).

4.5.4 Measure identically, favor the last

In the present Section we solve the following problem: find a protocol such that after N *identical* measurements, the last observer has the maximal information possible. Find the optimal measurement strength ε and the fidelity.

4.5.4.1 Single qudit

For one copy the covariant POVMs with Hermitian-square-root update rule are (among) optimal. The covariant POVMs are from a one-parameter family given by Eq. (4.82). Based on the Eq. (4.85) the fidelity ($F_K=(1+(d-1)\Delta_K)/d$) of the last observer is determined by

$$\Delta_K=\frac{\varepsilon}{d+1}r^{K-1}, \tag{4.109}$$

where r is defined in Eqs. (4.80), (4.81) and (4.83). This is as far as we get for finite K.

For $K \gg 1$ we expect $\varepsilon \ll 1$. Taylor expanding Δ_K of Eq. (4.109) around $\varepsilon = 0$ and taking term up to third power in ε we get an approximate Δ_K, maximization of which gives

$$\Delta_{K,\max} \simeq \frac{4}{3d\sqrt{3(d+1)(K-1)}} \quad ; \quad (K \gg 1). \tag{4.110}$$

4.5.4.2 N copies of a qubit

As in the fair-observers case, we restrict our attention to weak measurements of the type Eq. (4.89) and the Hermitian-square-root[4.9] update rule. The task is then trivial using results of the previous sections. Based on the Eq. (D.3) the fidelity ($F_K = (1 + \Delta_K)/2$) of the last observer is determined by

$$\begin{aligned} \Delta_K &= \varepsilon \frac{\langle J_{\boldsymbol{n}} \rangle_{\rho_{k-1}}}{N/2+1} \\ &= \frac{2\varepsilon}{N+2} \left(\frac{A(\varepsilon)}{N+1} \right)^{K-1}, \end{aligned} \tag{4.111}$$

where

$$A(\varepsilon) = 2ab + (N+1)a^2 + \frac{N}{N+2} b^2$$

with a and b defined as in Eq. (D.1).

For $K \gg N$ we expect $\varepsilon \ll 1$. Taylor expanding Δ_K around $\varepsilon = 0$ and taking two lowest orders[4.10] in ε we get

$$\Delta_K \simeq \frac{N\varepsilon\left(2\,(N+2) - (K-1)\,(N+1)\varepsilon^2\right)}{2(N+2)^2}. \tag{4.112}$$

The maximum of Δ_K reads

$$\Delta_{K,\max} \simeq \frac{2\sqrt{2}N}{3\sqrt{3(K-1)(N+1)(N+2)}}. \tag{4.113}$$

Again the fidelity degrades as $1/\sqrt{K}$ instead of the naive $1/K$.

Let us now proceed to the case $N \gg K$. Since we expect $\varepsilon \to 1$ it this case, we Taylor expand Eq. (4.111) in the variable $(1-\varepsilon)$ around 0 and take terms up to the first power of $(1-\varepsilon)$. Maximization gives ε which, after a Taylor expansion in $1/N$ around 0, reads

$$\varepsilon = 1 - \frac{4\,(K-1)^2}{N^3}. \tag{4.114}$$

Plugging Eq. (4.114) into the approximate expansion of Eq. (4.111), the lowest order of expansion around $1/N = 0$ gives

$$\Delta_{K,\max} = 1 - \frac{2K}{N}.$$

[4.9] Again, if working onlywith the symmetric subspace.

[4.10] In the Taylor expansion we threw away terms $O(\varepsilon^4)$ and higher-order ones. Those terms have prefactors $\sim k^{\alpha}$ where $\alpha \leq \beta/2 - 1$ (β is the exponent of ε of the term ε^{β} we are throwing away). Hence, if the *exact* $\varepsilon_{\text{maximizing}}$ scales as $K^{-(1/2-\gamma)}$ with $\gamma < 1$ then our approximaiton is OK. This has to be checked, though, e.g. by looking at maximization of the exact Δ_K for different fixed N, K, which we did.

4.6 Discussion and outlook

4.6.1 How large a quantum system must be to be considered "classical"?

The minimum size N for a system of qubits (e.g. spin-1/2 particles) to be considered "classical" as carriers of a single-instance pure state (e.g. direction), is related to the number of independent estimations of the direction we may perform on it and still get consistent outcomes. Since the average F_k can reach the maximum value of the quantity (the fidelity f) it is an average of only if the measure of the set of summands corresponding to incorrect estimates (those with $f \neq 1$) is vanishing, the requirement

$$\forall k \quad F_k \to 1. \tag{4.115}$$

guarantees that all observers' estimates are (very close to) correct with a probability approaching one.

Naturally, it is increasingly hard to fulfill the condition Eq. (4.115) for observers further and further in the sequence of observations, i.e. for larger and larger values of k. Therefore, for a classical-like observability by any number of observers we require that

$$\lim_{k\to\infty} F_k(N) = 1. \tag{4.116}$$

Inspecting the limit Eq. (4.116) with $N = ck^\alpha$, $c > 0$ being a constant, and the fidelity Eq. (4.66) we see that, for parallel qu*d*its, to observe the classical-like behavior for any large number of observers we need a minimum size of the order

$$N \sim k^\alpha; \quad \alpha > 1 \tag{4.117}$$

We also see that

$$F_k \to \frac{1}{d} \quad \text{if } N \sim k^\alpha,\ \alpha < 1. \tag{4.118}$$

For the optimal recycling of (qubit) information we need to inspect the limit Eq. (4.116), with the asymptotics of the maximum average fidelity for even N, Eq. (4.61). We conclude that, for qubits,

$$\mathcal{F}_k \to 1 \quad \text{if } N \sim k^\alpha,\ \alpha > 1/2, \tag{4.119}$$

hence, in this case we just need a size square root of the number of observations for a system of spins to be considered "classical." (Note that $1/d$ is the average fidelity of the random-guessing strategy. For $\alpha = 1$, i.e. $N = ck$, one gets $\lim_{k\to\infty} \mathcal{F}_k = \frac{1}{2}(1 + (d - 1)\mathrm{e}^{-d/c})$.)

Note that the scaling Eq. (4.117) is not in contradiction with the result [11] where the authors obtain $k = O(N^2)$ for what we call symmetric encoding into parallel spins. The quantity considered in [11] (see also Section 4.6.2), related to longevity, k, of a directional reference frame carried by a quantum system, is the first moment of the spin-projection operator for the state after k uses,

$$\left\langle J_{\boldsymbol{n}(\psi)} \right\rangle_{\rho_k(\psi)} = \frac{1}{2}(2\mathcal{F}_k - 1)(N+2),$$

which they require to stay above arbitrary but *fixed* threshold c. We require that the threshold approaches $(N+2)/2$ for all N and we take the limit $N \to \infty$.

The above applies to arbitrary realizations of qubits, of course. A generalization to qu*d*its requires knowledge of the performance of an optimal encoding/estimation procedure for qudits.

4.6.2 Implications for the longevity of a directional reference frame

A closely related problem to the problem studied in the present Thesis is that of degradation of a quantum directional reference frame, studied in Refs. [10] and [52] and generalized in Ref. [11]. The basic "setup" considered in Ref. [10] is depicted in Fig. 4.3.

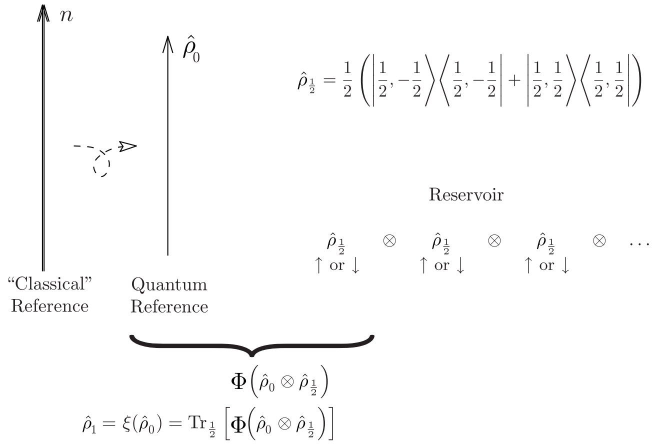
n
$\hat{\rho}_0$
$\hat{\rho}_{\frac{1}{2}} = \frac{1}{2}\left(\left|\frac{1}{2},-\frac{1}{2}\right\rangle\left\langle\frac{1}{2},-\frac{1}{2}\right| + \left|\frac{1}{2},\frac{1}{2}\right\rangle\left\langle\frac{1}{2},\frac{1}{2}\right|\right)$
Reservoir
$\hat{\rho}_{\frac{1}{2}} \otimes \hat{\rho}_{\frac{1}{2}} \otimes \hat{\rho}_{\frac{1}{2}} \otimes \ldots$
↑ or ↓
↑ or ↓
↑ or ↓
"Classical" Reference
Quantum Reference
$\Phi\left(\hat{\rho}_0 \otimes \hat{\rho}_{\frac{1}{2}}\right)$
$\hat{\rho}_1 = \xi(\hat{\rho}_0) = \mathrm{Tr}_{\frac{1}{2}}\left[\Phi\left(\hat{\rho}_0 \otimes \hat{\rho}_{\frac{1}{2}}\right)\right]$

Figure 4.3. The scenario studied by Bartlett et al. [10]. The quantum reference, realized via a spin-j system is used to measure the direction of a series of spin-1/2 particles in the completely mixed state by means of a projection onto the $j - 1/2$ or $j + 1/2$ subspaces. The figure is taken from Ref. [11] (the figure notation has been partially altered in order to be consistent / not interfere with ours). Note that the state of the spin-1/2 system is *not* assumed to be a pure state that is aligned or anti-aligned with the reference system (that being only one possible realization of the average reservoir state $\hat{\rho}_{1/2}$).

A reference direction $\boldsymbol{n}$, pointed in by a classical gyroscope, is encoded in a state $\hat{\rho} \equiv \hat{\rho}^{(0)}$ of a spin-j system (quantum reference). The quantum reference is repeatedly used to measure the spin component of a fresh particle drawn from a reservoir of spin-1/2 systems. The spin component of each reservoir particle along the reference direction is, prior to the measurement, totally random, i.e. the average pre-measurement state of the reservoir of K particles is $\hat{\rho}_{1/2}^{\otimes K}$, with $\hat{\rho}_{1/2}$ being the total mixture (see Fig. 4.3).

The measurement that provides the maximum information gain about the relative orientation between the quantum reference and a reservoir particle is a measurement of the magnitude of the total angular momentum $\hat{J}^2$ of the combined system [9], i.e. a two-outcome projective measurement with effects $\{\Pi_J\}$, $J = j \pm 1/2$, where the operator Π_J projects $\mathcal{H}_j \otimes \mathcal{H}_{1/2}$ on to the total $\hat{J}^2$ eigenspace with eigenvalue $J(J+1)$.

The state of the quantum reference is considered from the perspective of someone who has not kept a record of the outcome of previous measurements. Thus, at every measurement, an average over the possible outcomes with their respective weights is made to obtain the post-measurement density operator. The measurement $\{\Pi_J\}$ accompanied by the Lüders update induces, on overage, a rotationally invariant channel Φ on the combined system $\hat{\rho}_k \otimes \hat{\rho}_{1/2}$ which induces a channel ξ on the quantum reference (see Fig. 4.3; $\xi(\hat{\rho}_{k-1}) =: \hat{\rho}_k$ is the average state of the reference after k measurements).

(Re-)usability of the quantum reference for its purpose, is measured in what the authors refer to as longevity of the reference, defined as number of times the quantum reference can be used until the average (success) probability of correctly performing a given task drops below a prescribed level. One of the tasks considered is that of determining, in a hypothetical $(k+1)$th measurement, the spin component of a spin-1/2 reservoir particle along the direction $\boldsymbol{n}$ *as if* the $(k+1)$th reservoir particle was in a known pure state $|\boldsymbol{n}\rangle\langle\boldsymbol{n}|$ or $|-\boldsymbol{n}\rangle\langle-\boldsymbol{n}|$, each with probability 1/2. In this case the success probability $\bar{P}_s^{(k+1)}$, using the quantum reference already used k times, reads

$$\begin{aligned}\bar{P}_s^{(k+1)} &= \frac{1}{2}\sum_{\mu\in\left\{\frac{1}{2},-\frac{1}{2}\right\}}\mathrm{Tr}[\Pi_{j+\mu}(\hat{\rho}_k\otimes|\mu\rangle\langle\mu|)]\\ &= \frac{1}{2}\left(1+\frac{2}{2j+1}\mathrm{Tr}\Big[\hat{\rho}_k\hat{J}_{\boldsymbol{n}}\Big]\right).\end{aligned}\tag{4.120}$$

Comparing Eq. (4.120) to the maximal average fidelity $\mathcal{F}_k$, Eq. (4.94) ($\boldsymbol{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_k)=(1,\ldots,1)$, $j=N/2$), of our results we have

$$\bar{P}_s^{(k+1)} = \frac{1}{2}\left(1+\frac{2(j+1)}{2j+1}\Delta^{k+1}\right),\tag{4.121}$$

where Δ is defined as usual (see Eq. (4.54)) through

$$\mathcal{F}_k=\frac{1}{2}\big[1+\Delta^k\big].\tag{4.122}$$

In the context of the problem of Ref. [10] the success probability can be defined, in our original problem, as follows: suppose the true reference direction is $\boldsymbol{n}$ and the $(k+1)$th observer concluded it had been $\boldsymbol{n}_{k+1}$. The conclusion occurred with probability density $\mathrm{Tr}[\widetilde{\mathcal{M}}(\boldsymbol{n}_{k+1})\chi^{\circ k}(\varrho(\boldsymbol{n}))]$. He then measures a hypothetical reservoir particle which is in a known pure state $|\boldsymbol{n}\rangle\langle\boldsymbol{n}|$ or $|-\boldsymbol{n}\rangle\langle-\boldsymbol{n}|$, each with probability 1/2. Although the reservoir-particle in the state $|\boldsymbol{n}\rangle\langle\boldsymbol{n}|$ is pointing along the true reference direction, he makes this correct conclusion, i.e. the reservoir particle in (in unknown state) is pointing along the reference direction (whatever it is), only with probability $|\langle\boldsymbol{n}_{k+1}|\boldsymbol{n}\rangle|^2$ since, based on his measurement of the quantum reference, he believes the reference direction is $\boldsymbol{n}_{k+1}$ along which he then measures. Analogously, although the reservoir-particle in the state $|-\boldsymbol{n}\rangle\langle-\boldsymbol{n}|$ is pointing opposite to the reference direction, the observer makes this correct conclusion only with probability $|\langle-\boldsymbol{n}_{k+1}|-\boldsymbol{n}\rangle|^2$. Averaging over all estimates, $\boldsymbol{n}_{k+1}$, we arrive at the success probability

$$\begin{aligned}\bar{P}_s^{(k+1)} &= \frac{1}{2}\bigg(\int_{\mathbb{S}^2}\mathrm{d}\boldsymbol{n}_{k+1}\mathrm{Tr}\big[\widetilde{\mathcal{M}}(\boldsymbol{n}_{k+1})\chi^{\circ k}(\varrho(\boldsymbol{n}))\big]|\langle\boldsymbol{n}_{k+1}|\boldsymbol{n}\rangle|^2\\ &\quad+\int_{\mathbb{S}^2}\mathrm{d}\boldsymbol{n}_{k+1}\mathrm{Tr}\big[\widetilde{\mathcal{M}}(\boldsymbol{n}_{k+1})\chi^{\circ k}(\varrho(\boldsymbol{n}))\big]|\langle-\boldsymbol{n}_{k+1}|-\boldsymbol{n}\rangle|^2\bigg)\\ &= \int_{\mathbb{S}^2}\mathrm{d}\boldsymbol{n}_{k+1}\mathrm{Tr}\big[\widetilde{\mathcal{M}}(\boldsymbol{n}_{k+1})\chi^{\circ k}(\varrho(\boldsymbol{n}))\big]|\langle\boldsymbol{n}_{k+1}|\boldsymbol{n}\rangle|^2\\ &\equiv \frac{1}{2}\big[1+\Delta^{k+1}\big]=\mathcal{F}_{k+1}.\end{aligned}\tag{4.123}$$

Thus, in the case of spin-1/2 systems, we have an operational interpretation for the average fidelity $\mathcal{F}_{k+1}$ as the kth observer's (average) success probability of correctly determining the spin component, along the true reference direction $\boldsymbol{n}$, of a spin-1/2 particle aligned or anti-aligned along the $\boldsymbol{n}$.

As expected, an incoherent strategy (estimate $\boldsymbol{n}$ by a measurement and then execute a desired $\boldsymbol{n}$-dependent task using the estimate) performs worse than a coherent one (perform the same task without the intermediate estimation) – in this case by a factor $2(j+1)/(2j+1)$ in front of the relevant quantity, Δ^i, the whole term describing, on a scale $\langle -1, 1\rangle$, how much better (or worse) the current strategy is, compared to random guessing.

However, the coherent strategy, at least on a spin-j system (or, equivalently, on a N-spin-1/2 ensemble), is better only by a k-independent pre-factor, bounded by 3/2 and going to one as $N \rightarrow \infty$, while the optimal direction encoding on N instances *effectively* exploits, with growing N, a quadratically growing portion of the available Hilbert space [3], in contrast to the linearly growing dimension of the symmetric subspace spanned by the ensemble. Since the maximum achievable average single-observer fidelity $\mathcal{F}_1$, i.e. also Δ, depends on the dimensionality of the Hilbert space effectively utilized by the signal states [3] and the coherent pre-factor is k-independent, with the growing exponent, k, the quantity $\Delta^{(k+1)}_{\text{optimal, incoherent}} := \Delta^{k+1}_{\text{opt}}$ (Δ_{opt} being the Δ given by Eqs. (4.58) and (3.46) for N even and odd, respectively) will eventually, for some finite k, become greater than $\frac{2(j+1)}{2j+1}\Delta^{k+1} =: \Delta^{(k+1)}_{\text{ensemble, coherent}}$ for any N – see Figure 4.4.

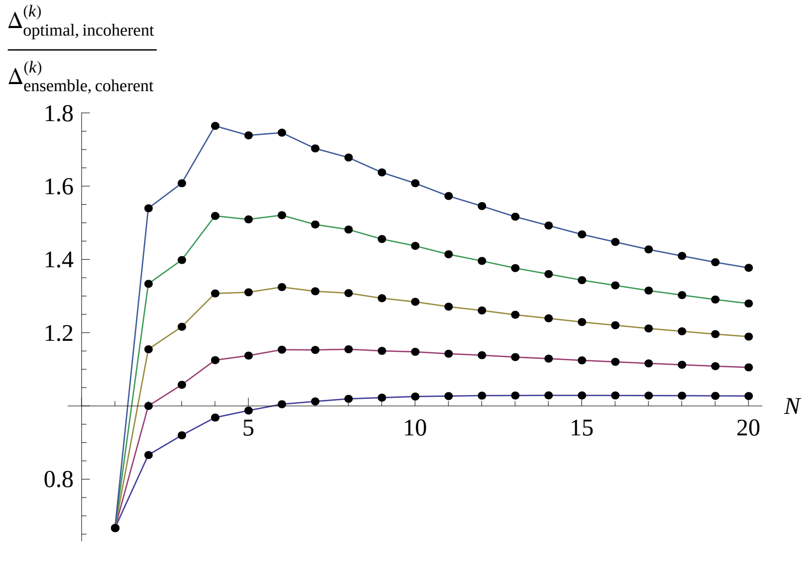

Figure 4.4. What is the "performance" of the optimal incoherent strategy for N instances of a spin-1/2 system in units of the "performance" of the best coherent strategy on a N-qubit ensemble? The "performance," $\Delta^{(k)}$, quantifies, on a scale $\langle -1, 1\rangle$, how much better (or worse) the current strategy is, compared to random guessing. Different colors stand for $k = 1, 2, \ldots, 5$, from bottom to top. Plot is joined for better readability. Since, for N even and large, $\frac{\Delta^{(k)}_{\text{optimal,incoherent}}}{\Delta^{(k)}_{\text{ensemble,coherent}}} \simeq \left(1 - \frac{2\xi_0^2}{N^2}\right) / \left(\left(\frac{N+2}{N+1}\right)^{-k} \frac{N}{N+2}\right)$ (see Eqs. (4.60) and (4.66)), one can check that $\underset{N \rightarrow \infty}{\text{Lim}} \frac{\Delta^{(k)}_{\text{optimal,incoherent}}}{\Delta^{(k)}_{\text{ensemble,coherent}}} = 1$.

In the light of our results, it would be interesting to consider an optimal coherent strategy using an optimal signal state to obtain, for a directional quantum reference made of N instances of a spin-1/2 system, the ultimate success probability of correctly identifying an aligned or anti-aligned spin-1/2 reservoir system and the implied longevity of the directional reference. Our results indicate that the scaling of its $\Delta^{(k)}$ with the number of observations will be better by at least a square-root as compared to the (spin-j) directional reference considered in Refs. [10, 11, 52].

Another direction of research along the lines of Refs. [10, 11, 52] would be to study, under the optimal signal state, the behaviour of the relevant parameters of generalized fidelity functions – analogues of the higher moments of the $\hat{J}_{\boldsymbol{n}}$ operator from the parallel encoding case – i.e. quantities which arise when the success probability of determining alignment of higher-spin reservoirs, or implementing a direction-dependent unitary gates (see Ref. [11]), is considered.

Chapter 5
Conclusion

The aim of the present Thesis was to provide an insight into the problem of emergence of classical features within a quantum description of physical systems. Specifically, we focused on a property of classical systems that they are observable, in principle, by as many careful independent observers as one wishes, in a consistent manner; that is all who observe the physical system are able to obtain essentially the same value of an observable.

Motivated by a previously studied problem of repeated use of quantum clocks, i.e. repeated estimation of a state of a (dynamically evolving) phase reference, we in particular considered a natural generalization thereof, namely repeated estimation of a quantum directional reference, or a "quantum gyroscope" – a single spin-$1/2$ pure state carried collectively by, or encoded into, a collection of N spin-$1/2$ systems. Dropping the requirement that physically the two-level systems be realized by spins we arrive at a problem directly generalizable to that of repeated estimation of a pure single-qu$d$it state encoded in the state space of N qu$d$its.

After having introduced, in Chapter 2, the basic mathematical tools relevant for the rest of the Thesis, we collected, in Chapter 3, the common knowledge on optimal extraction of information from families of signal states that are invariant under operations from the representation space of a symmetry group. The research program of the Thesis was presented in Chapter 4.

We began our analysis by explicitly showing, in Section 4.3, that due to the limitations on resources allowed to be in possession of the observers we may restrict our attention to encodings and measurement apparata which are covariant with respect to the $g \mapsto g^{\otimes N}$ representation of $\mathrm{SU}(d)$. Subsequently we considered two scenarios.

In the first one which we referred to as "greedy" scenario, analyzed in Section 4.4 and further discussed in Section 4.6, each of the observers, wishing to access the single-qu$d$it information, proceeded so that the fidelity of his estimate was maximized. We showed that in such a case each observer's measurement can be viewed, for each outcome, as a measure and prepare channel. From this point of view, each observer effectively encodes his estimate for the next one, irrespective of his input state. The above enabled us to solve the problem in terms of the first observer's estimation fidelity, without the need to calculate the average channels induced by the measurements and without a measurement optimization for each observer's average input state. Applying this result, we managed to express a kth observer's maximal average fidelity as a function of k and of the number of qu$d$its, N, encoding the single-qu$d$it to estimate, namely $\mathcal{F}_k = \left[1 + (d-1)\Delta^k\right]/d,$ where $\Delta = (\mathcal{F}_1 d - 1)/(d-1)$.

For general d we restricted ourselves to the finite ensemble case, i.e. symmetric, or parallel, encoding of a pure qudit into the state space of N qudits. The maximal average fidelity in this case is determined by $\Delta = N/(N+d)$ which, asymptotically, for large N, approaches 1 as $1 - d/N$. Since for $d = 2$ an optimal strategy without the parallel-encoding restriction is known, we were in position to evaluate the best achievable fidelity of a kth observer's estimate in the qubit case. It is determined by $\Delta = x_{N/2+1}$, where $x_{N/2+1}$ is the largest zero of the Legendre polynomial $P_{N/2+1}(x)$. Asymptotically, for large N, $\Delta \cong 1 - 2\xi_0^2/N^2$, where $\xi_0 \doteq 2.4$ is the first zero of the Bessel function $J_0(x)$.

These results enabled us to make quantitative conclusions about the question that had motivated us from the beginning: how fast does the classical-like recyclability with respect to observations arise when the size, i.e. the number of constituents, of the system is increased while the number of parameters of the system is given by the number of parameters of a single constituent? The answer we give is that, for a pure-qu$d$it information carried by N parallel qu$d$its, should $k \gg 1$ observers all observe the same encoded qu$d$it state it is required that $N \sim k^{\alpha}$; $\alpha > 1$. For a pure qubit encoded in N qubit instances optimally, $\alpha > 1/2$, i.e. only a size square root of the number of observations, is required for a system of qubits to be recyclable as if classical. Whether the enhancement that occurs for an optimal encoding, as compared to the parallel encoding is a square-root, or some other, one in the case of qu$d$its is an open question. Due to our results, this question has been reduced to the problem of optimal single-observer covariant encoding/estimation.

For qubit systems realized by spin-1/2s we found a nice operational interpretation of the estimation fidelity of a kth observer as the success probability of correctly identifying, via performing a spin-projection measurement onto the axis given by the $(k-1)$th observer's estimate, that an additional qubit, aligned or anti-aligned with respect to the encoded qubit state, has been aligned or anti-aligned, respectively.[5.1] This enabled us to see our (spins-1/2) problem as an incoherent-strategy version of the (part of the) problem of degradation of a quantum directional reference, previously studied in literature. Unlike the latter, our approach is not limited to the parallel encoding, which enabled us to compare the previously studied, coherent, strategy with parallel encoding to an incoherent one with both parallel and optimal direction encoding. Though inferior for the parallel case, the incoherent strategy for the optimal encoding is superior to the coherent parallel one, starting from six spin-1/2s for the first observation and for more than one spin for any[5.2] successive observation. The success probability of our incoherent strategy with the optimal encoding provides a lower bound on that of a coherent one. Finding the highest possible success probability and the optimal coherent strategy itself, i.e. optimal measurements, is an interesting problem for future research.

In the second scenario, discussed in Section 4.5, we considered a generalization of the first one, to include situations where the observers optimize their measurements to pursue goals different from, in the first place, mere maximization of the quality of their own guesses. In particular, we studied the case where K observers estimate the encoded qu$d$it state with equal, but maximal, fidelity (equalitarian strategy, Subsection 4.5.1) and the case where the observers use the same measurement apparatus such that the quality of the last observer's estimate is maximized (Subsection 4.5.4). In both cases the measurements performed were weak, i.e. in general not extracting all of the extractable information

[5.1] The interpretation for qubits which are not realized as spins is analogous only the "spin" vector, and thus the (anti-)alignment, does not correspond to a direction in real three-dimensional space.

[5.2] With the exception of the second observation and two spins, in which case the performance of the coherent strategy with parallel encoding is identical to that of the incoherent strategy with the optimal encoding.

which enabled less disturbance to be undergone by the measured state. In the former case we quantified how the measurements approach more and more the ones from the "greedy" scenario with increasing tally number, k, until the last observer performs a "greedy" measurement as well as the maximal achievable, equal, fidelity of the observers' guesses. In the latter case we showed there exists, and we calculated, the optimal "strength" of the measurement to be performed by all the observers as well as the observers' fidelities achieved.

Acknowledgment

First of all, I would like to express my thanks to my supervisor, Prof. Vladimír Bužek, for the opportunity to study under his supervision and for creating excellent working conditions and a pleasant and friendly atmosphere in his research group at the Research Center for Quantum Information in Bratislava. I also appreciate Mário Ziman and Teiko Heinosaari for many fruitful discussions and for their, always welcome, effort to make people involved in working on interesting problems. Last but not least, my thanks also belong to my parents, who have always supported me in every aspect of my life and to Alexandra, whose love, sense of humor and cute impatience have helped me survive the tedious process of finalizing this work.

Appendix A
Covariance of the average apparatus

As discussed in the Introduction, whenever a quantum operation $\$^g$ could have taken place (we use a group parameter g to parametrize the orientation of a device realizing the operation), due to lack of next observers' knowledge, the operation could have happened with the same probability for all $g\in\mathrm{SU}(d)$. I.e., in the expression for the average fidelity one should consider the average operation

$$\$_s(\hat{\rho})=\int \mathrm{d}g U_g \$^e(U_g^{\dagger}\hat{\rho}U_g)U_g^{\dagger},$$

where $\$^e$ is some reference operation corresponding to the identity of the group and $U_g=g^{\otimes N}$. Let us look at the ith observer's actions ($i<k$): let's introduce the quantum operation

$$I^g_{\mathrm{d}\psi}\colon I^g_{\mathrm{d}\psi}(\hat{\rho}):=\rho^g(\psi,\hat{\rho})\mathrm{Tr}(M^g(\mathrm{d}\psi)\hat{\rho}),$$

describing the actions of his apparatus I^g on a state $\hat{\rho}$ given his guess ψ (the operation $I^g_{\mathrm{d}\psi}$ is defined by $I^g_{\mathrm{d}\psi}(\hat{\rho})=\mathrm{d}\psi\sum_o p(\psi|o)J^g(o)$ where $p(\psi|o)$ is the probability density that an outcome o leads to the guess ψ and $J^g_o\colon J^g_o(\hat{\rho}):=\rho^g(o,\hat{\rho})\mathrm{Tr}(E^g(o)\hat{\rho})$ (by J^g and E^g we denote the instrument and POVM actually performed from the viewpoint of its executor – for simplicity we assume its outcomes are from a discrete set[A.1] {o}).

If we take the operations describing the actions of the observer's apparatus whenever a particular guess takes place, we have to realize that the action of the group on the guesses is $g(\psi)\equiv g\psi g^{\dagger}$ (so that a 'rotated' observer with a 'rotated' apparatus measuring a 'rotated' input outputs a rotated guess $g\psi_i g^{\dagger}$ every time a non-rotated observer with non-rotated apparatus measuring a non-rotated input would output ψ), i.e.

$$I^g_{\psi}(\hat{\rho})=\mathcal{U}_g I^{\mathrm{e}}_{g^{\dagger}\psi g}(\mathcal{U}_g^{\dagger}\hat{\rho}\mathcal{U}_g)\mathcal{U}_g^{\dagger}.$$

Since quantum operations are linear maps one may integrate over the apparata orientations and guesses in which case we would arrive at the channel approach to our problem. For our purposes, we want to keep the guesses, though. Collecting all operations that lead, for whatever value of the 'rotation' of the apparatus, g, to the same guess ψ we have

$$\mathcal{I}_{\mathrm{d}\psi}(\hat{\rho}):=\int \mathrm{d}\mu(g)I^g_{\mathrm{d}\psi}(\hat{\rho})$$

[A.1] One could formally consider the general case here but, at the end, any physically realized measurement will have a finite precision.

which defines an instrument $\mathcal{I}$ covariant with respect to the representation $\mathcal{U}$ of $\mathrm{SU}(d)$ and action of $\mathrm{SU}(d)$ on measurement outcomes ($\mathcal{U}$-covariant for short) [28, 18]. This can be checked directly by calculating

$$\begin{aligned}
\mathcal{I}_{h\psi h^{-1}}(\hat{\rho}) &= \int \mu(g) I^{g}_{h\psi h^{\dagger}}(\hat{\rho}) \\
&= \int \mu(g)\mathcal{U}_g I^{\mathrm{e}}_{g^{\dagger}h\psi h^{\dagger}g}(\mathcal{U}_g^{\dagger}\hat{\rho}\mathcal{U}_g)\mathcal{U}_g^{\dagger} \\
&= \int \mu(hg)\mathcal{U}_{hg} I^{\mathrm{e}}_{g^{\dagger}\psi g}(\mathcal{U}_{hg}^{\dagger}\hat{\rho}\mathcal{U}_{hg})\mathcal{U}_{hg}^{\dagger} \\
&= \mathcal{U}_h \int \mu(g)\mathcal{U}_g I^{\mathrm{e}}_{g^{\dagger}\psi g}(\mathcal{U}_g^{\dagger}\mathcal{U}_h^{\dagger}\hat{\rho}\mathcal{U}_h\mathcal{U}_g)\mathcal{U}_g^{\dagger}\mathcal{U}_h^{\dagger} \\
&= \mathcal{U}_h \int \mu(g) I^{\mathrm{e}}_{\psi}(\mathcal{U}_h^{\dagger}\hat{\rho}\mathcal{U}_h)\mathcal{U}_h^{\dagger} \\
&= \mathcal{U}_h \mathcal{I}_{\psi}(\mathcal{U}_h^{\dagger}\hat{\rho}\mathcal{U}_h)\mathcal{U}_h^{\dagger}.
\end{aligned}$$

Covariance of the instrument implies covariance of its induced POVM $\mathcal{M}$ as well as covariance of the output encoding given by

$$\tilde{\varrho}(\psi,\hat{\rho}) := \frac{\mathcal{I}_{\psi}(\hat{\rho})}{\mathrm{Tr}[\mathcal{I}_{\psi}(\hat{\rho})]}$$

since

$$\begin{aligned}
\tilde{\varrho}(h\psi h^{-1},\hat{\rho}) &= \frac{\mathcal{I}_{h\psi h^{-1}}(\hat{\rho})}{\mathrm{Tr}\big[\mathcal{I}_{h\psi h^{-1}}(\hat{\rho})\big]} \\
&= \frac{\mathcal{U}_h\mathcal{I}_{\psi}(\mathcal{U}_h^{\dagger}\hat{\rho}\mathcal{U}_h)\mathcal{U}_h^{\dagger}}{\mathrm{Tr}\Big[\mathcal{I}_{\psi}(\mathcal{U}_h^{\dagger}\hat{\rho}\mathcal{U}_h)\Big]} \\
&= \mathcal{U}_h\tilde{\varrho}(\psi_i,\mathcal{U}_h^{\dagger}\hat{\rho}\mathcal{U}_h)\mathcal{U}_h^{\dagger}.
\end{aligned}$$

In the special case of optimal greedy measurements the output state can be considered independent of the input state (see Section 4.4), giving us a covariant encoding of the guesses only, i.e.

$$\begin{aligned}
\forall\hat{\rho},\ \ \tilde{\varrho}(\psi,\hat{\rho}) &= \frac{\mathcal{I}_{\psi}(\,.\,)}{\mathrm{Tr}[\mathcal{I}_{\psi}(\,.\,)]} \\
&=:\ \varrho(\psi),
\end{aligned}$$

where the covariance explicitly reads $\varrho(h\psi h^{-1}) = \mathcal{U}_h\varrho(\psi)\mathcal{U}_h^{\dagger}$.

Appendix B

Evaluation of the integral Eq. (4.52)

Choosing, for the sake of calculations, an arbitrary reference state $\psi_0 \in \mathcal{S}(\mathcal{H}_d)$ we can parametrize the states by elements $g \in \mathrm{SU}(d)$ and replace the integration over the pure states by integration over the group $\mathrm{SU}(d)$. The integral Eq. (4.52) becomes

$$\int_{g \in SU(d)} \mathrm{d}\mu(g)\, \boldsymbol{n}(g) p(\hat{g}\,|g), \tag{B.1}$$

where $\boldsymbol{n}(g)$ is a d-dimensional Bloch vector parametrizing the state $|\psi(g)\rangle\langle\psi(g)|$ and

$$p(\hat{g}\,|g) = \mathrm{Tr}[\mathcal{M}(\hat{g})\rho(g)], \tag{B.2}$$

where $\mathcal{S}(\mathcal{H}_D) \ni \rho(g) = \mathcal{U}(g)\rho_0\mathcal{U}(g)^\dagger$. Note that due to covariance of both the measurement and the states $\rho(g)$ it holds that

$$\mathrm{Tr}[\mathcal{M}(\bar{g}\hat{g})\rho(\bar{g}g)] = \mathrm{Tr}[\mathcal{U}'(\bar{g})\mathcal{M}(\hat{g})\mathcal{U}'(\bar{g})^\dagger\mathcal{U}(\bar{g})\rho(g)\mathcal{U}(\bar{g})^\dagger].$$

For optimal covariant encoding-decoding schemes it holds that the representations are the same, i.e. $\mathcal{U}'(g) = \mathcal{U}(g)$, hence

$$p(\bar{g}\hat{g}\,|\bar{g}g) = p(\hat{g}\,|g). \tag{B.3}$$

A d-dimensional system in a pure state $\psi = |\psi\rangle\langle\psi|$ can be parametrized as

$$\psi = \frac{1}{d}\{\mathbb{1} + \kappa_d n^a T_a\},$$

where

$$\{T_a, T_b\} = \frac{\delta_{ab}}{d} + d^c_{ab}T_c,$$

with the generators defined as half the standard Gell-Mann matrices,

$$\kappa_d = \sqrt{2d(d-1)},$$

and n^a are the components of a (d^2-1)-dimensional unit vector: $\boldsymbol{n} = (n^1, n^2, ..., n^{d^2-1})$, to which we refer as Bloch vector. This follows from imposing on ψ the conditions $\mathrm{Tr}\,\psi = 1$ and $\mathrm{Tr}\,\psi^2 = 1$.

Not any unit vector $\boldsymbol{n}$ is allowed. By imposing the condition $\psi = \psi^2$ we get further constrains

$$n^a = \frac{\kappa_d}{2(d-2)} d^a_{bc} n^b n^c. \tag{B.4}$$

Any state can be obtained by applying a $SU(d)$ transformation to the reference state

$$|\psi_0\rangle = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Note that

$$\psi_0 = |\psi_0\rangle\langle\psi_0| = \frac{1}{d}\{\mathbb{1} - \kappa_d T_{d^2-1}\},$$

since

$$T_{d^2-1} = \frac{1}{\sqrt{2d(d-1)}}\begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & \ldots & 0 & 1-d \end{pmatrix}$$

(the normalization ensures that $\mathrm{Tr}\big[T_{d^2-1}^2\big] = 1/2$). Hence, the 'reference' Bloch vector is

$$\boldsymbol{n}_0 = (\underbrace{0, 0, \ldots, 0}_{d^2-2}, -1),$$

i.e., its components are

$$n_0^{d^2-1} = -1; \quad n_0^a = 0 \quad \text{if } a \neq d^2 - 1.$$

Note that $|\psi_0\rangle$ is invariant under $SU(d-1) \subset SU(d)$ transformation of the form

$$\tilde{U} \equiv U(\tilde{g}) = \begin{pmatrix} U_{1\,1} & U_{1\,2} & \ldots & U_{1\,d-1} & 0 \\ U_{2\,1} & U_{2\,2} & \ldots & U_{2\,d-1} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ U_{d-1\,1} & U_{d-1\,2} & \ldots & U_{d-1\,d-1} & 0 \\ 0 & 0 & \ldots & 0 & 1 \end{pmatrix}.$$

Hence,

$$\begin{aligned} \langle\psi_0|\psi(g)\rangle &\equiv \langle\psi_0|U(g)|\psi_0\rangle \\ &= \langle\psi_0|\tilde{U}U(g)|\psi_0\rangle = \langle\psi_0|\psi(\tilde{g}g)\rangle \\ &= \langle\psi_0|U(g)\tilde{U}\,|\psi_0\rangle = \langle\psi_0|\psi(g\tilde{g})\rangle. \end{aligned} \tag{B.5}$$

Moreover, due to covariance of the encoding, it also has to hold that

$$\mathrm{Tr}[\rho_0\rho(g)] = \mathrm{Tr}[\rho_0\rho(g\tilde{g})] \tag{B.6}$$

We use the group parameters g to label the different states according to:

$$|\psi(g)\rangle\langle\psi(g)| = U(g)|\psi_0\rangle\langle\psi_0|U^\dagger(g).$$

It follows that

$$n^a(g)T_a = U(g)\; n_0^a T_a\; U^\dagger(g) = A_a^b(g) n_0^a T_b,$$

where $A_b^a(\bar{g})$ belongs to the adjoint representation and we have used that

$$U(g)T_a U^\dagger(g) = A_a^b(g)T_b.$$

We see that

$$n^a(g) = A_b^a(g) n_0^b.$$

In general

$$n^a(g) = A^a_b(g) A^b_c(\bar{g}^{-1}) n^c(\bar{g}) = A^a_c(g\bar{g}^{-1}) n^c(\bar{g}),$$

from which

$$n^a(g\bar{g}) = A^a_b(g\bar{g}\bar{g}^{-1}) n^b(\bar{g}) = A^a_b(g) n^b(\bar{g}).$$

Let us now consider the integral

$$V^a(\hat{g}) \equiv \int \mu(g)\; n^a(g) p(\hat{g}\,|g).$$

Here $p(\hat{g}\,|g)$ is the conditional-probability density, Eq. (B.2). Let $\bar{U} = U(\bar{g})$ be any $SU(d)$ transformation. We have

$$\begin{aligned} V^a(\bar{g}\hat{g}) &= \int \mathrm{d}\mu(g)\; n^a(g) p(\bar{g}\hat{g}\,|g) \\ &= \int \mathrm{d}\mu(\bar{g}^{-1}g)\; n^a(\bar{g}\bar{g}^{-1}g) p(\bar{g}\hat{g}\,|\bar{g}\bar{g}^{-1}g) \\ &= A^a_b(\bar{g}) \int \mathrm{d}\mu(\bar{g}^{-1}g)\; n^a(\bar{g}^{-1}g) p(\hat{g}\,|\bar{g}^{-1}g) \\ &= A^a_b(\bar{g}) V^b(\hat{g}), \end{aligned}$$

where we have used the invariance of the Haar measure $\mathrm{d}\mu(g)$ and the invariance of the probability, Eq. (B.3).

We see that, in particular

$$V^a(\hat{g}) = A^a_b(\hat{g}) V^b(\mathbf{0}),$$

where $\mathbf{0}$ denotes the identity parameters. I.e.,

$$V^b(\mathbf{0}) = \int \mathrm{d}\mu(g)\; n^b(g) p(\mathbf{0}|g).$$

We now wish to show that, as expected, $V^b(\mathbf{0}) \propto n^b_0$. We proceed as follows. From

$$T_b V^b(\mathbf{0}) = \int \mathrm{d}\mu(g)\; T_b n^b(g) p(\mathbf{0}|g)$$

we observe that

$$\begin{aligned} \tilde{U} T_b V^b(\mathbf{0}) \tilde{U}^\dagger &= \int \mathrm{d}\mu(\tilde{g}g)\; T_b n^b(\tilde{g}g) p(\mathbf{0}|g) \\ &= \int \mathrm{d}\mu(\tilde{g}g)\; T_b n^b(\tilde{g}g) p(\mathbf{0}|\tilde{g}g) \\ &= T_b V^b(\mathbf{0}), \end{aligned}$$

where we have used Eq. (B.6) in the form $p(\mathbf{0}|g) = p(\mathbf{0}|\tilde{g}g)$. Hence, according to Schur's lemma, $T_b V^b(\mathbf{0})$ must be the identity in the subspace corresponding to SU$(d-1)$, i.e., proportional to T_{d^2-1}, from where the desired result follows immediately. Note that from this it also follows that

$$V^a(\hat{g}) \propto A^a_b(\hat{g}) n^b_0 = n^b(\hat{g})$$

or, more explicitly,

$$\int \mathrm{d}\mu(g)\; \boldsymbol{n}(g) p(\hat{g}\,|g) = \Delta\; \boldsymbol{n}(\hat{g}), \tag{B.7}$$

where Δ is a constant.

Appendix C

The channel induced by optimal greedy measurements of orientation on N copies of a qubit

Let us look what the channel induced by single SO(3)-covariant measurement of orientation optimal for the greedy strategy (with the outcome unknown) does to a state

$$\sum_{m=-j}^{j} s_m |m\rangle\langle m| \tag{C.1}$$

(general state invariant under elements from the range of spin-j representation of SO(3) around the "z" axes defined by $|j\rangle$). We recall that the optimal covariant greedy measurement of orientation for the state Eq. (C.1) with non-zero spin component along the "z" axis is of the form

$$(2j+1)|\pm\boldsymbol{n}\rangle\langle\pm\boldsymbol{n}|\mathrm{d}\boldsymbol{n},$$

where the sign corresponds to the sign of the spin component along the "z" axis. We will restrict ourselves to the case of a positive spin component, which only means that we assume that a direction $\boldsymbol{n}$ is encoded into the, more natural, state with positive, rather than negative, spin component along that direction.

Due to linearity of channels, it suffices to look at what happens to the basis states $|m\rangle\langle m'|$. In our case it will suffice to consider the action of the channel in the case $m=m'$ since, before the action of the channel, we begin with a state of the type Eq. (C.1) and, as we will see shortly, due to the nature of the channel the states on its output will remain to be of the same type. Hence studying the $m=m'$ case will suffice also for repeated uses of the channel. Let us first look at the transformation of the states $|m\rangle\langle m|$ under the action of the channel then:

$$\begin{aligned}
|m\rangle\langle m| &\rightarrow \chi(|m\rangle\langle m|) \\
&= \int_{\mathbb{S}^2} \mathcal{I}_{\mathrm{d}\boldsymbol{n_1}}(|m\rangle\langle m|) \\
&= \int_{\mathbb{S}^2} \mathrm{Tr}[\mathcal{M}(\mathrm{d}\boldsymbol{n_1})|m\rangle\langle m|]|\boldsymbol{n_1}\rangle\langle\boldsymbol{n_1}| \\
&= (2j+1)\int_{\mathbb{S}^2} |\langle m|\boldsymbol{n_1}\rangle|^2|\boldsymbol{n_1}\rangle\langle\boldsymbol{n_1}|\mathrm{d}\boldsymbol{n_1}.
\end{aligned}$$

Using

$$\begin{aligned}
|\boldsymbol{n}\rangle &= \sum_{m=-j}^{j} \binom{2j}{j+m}^{\frac{1}{2}} \mathrm{Cos}^{j+m}\Big(\frac{\theta}{2}\Big)\mathrm{Sin}^{j-m}\Big(\frac{\theta}{2}\Big)e^{i\varphi m}|m\rangle \\
&=: \sum_{m=-j}^{j} c_m|m\rangle
\end{aligned}$$

(cf. [51]), where $(1,\theta,\varphi)$ are the spherical coordinates of the endpoint of the unit vector $\boldsymbol{n}$ pointing from the origin of the coordinate system with the z axis defined by $|j\rangle$, and some fixed x axis, we have

$$\begin{aligned}
\chi(|m\rangle\langle m|) &= (2j+1)\sum_{m',m''=-j}^{j}\int_{\$^2}\underbrace{\left|\sum_{m'''=-j}^{j} c_{m'''}\langle m|m'''\rangle\right|^2}_{|c_m|^2} c_{m'}c^*_{m''}|m'\rangle\langle m''|\mathrm{d}\boldsymbol{n}_1\\
&= \frac{2j+1}{4\pi}\sum_{m',m''=-j}^{j}|m'\rangle\langle m''|\int_0^{\pi}|c_m|^2|c_{m'}||c^*_{m''}|\sin(\theta)\mathrm{d}\theta\underbrace{\int_0^{2\pi}e^{i\varphi(m'-m'')}\mathrm{d}\varphi}_{2\pi\delta_{m',m''}}\\
&= \frac{2j+1}{2}\sum_{m'=-j}^{j}|m'\rangle\langle m'|\int_0^{\pi}|c_m|^2|c_{m'}|^2\sin(\theta)\mathrm{d}\theta.
\end{aligned}$$

Evaluating the last integral

$$\begin{aligned}
\int_0^{\pi}|c_m|^2|c_{m'}|^2\sin(\theta)\mathrm{d}\theta &= \underbrace{\int_0^{\pi}\cos^{4j+2(m+m')}\left(\frac{\theta}{2}\right)\sin^{4j-2(m+m')}\left(\frac{\theta}{2}\right)\sin(\theta)\mathrm{d}\theta}_{2\frac{(2j+m+m')!(2j-m-m')!}{(4j+1)!}}\\
&\quad\times\binom{2j}{j+m}\binom{2j}{j+m'}\\
&= \frac{2}{(4j+1)!}(2j+m+m')!(2j-m-m')! \qquad \text{(C.2)}
\end{aligned}$$

we finally get

$$\begin{aligned}
\chi(|m\rangle\langle m|) &= \frac{2j+1}{(4j+1)!}\binom{2j}{j+m}\sum_{m'=-j}^{j}\binom{2j}{j+m'}(2j+m+m')!(2j-m-m')!\\
&\quad\times|m'\rangle\langle m'|. \qquad \text{(C.3)}
\end{aligned}$$

Now we want to show that if we have a state of the type Eq. (C.1) with a positive (or negative) spin component along the "z" axis before on the input of the channel, it stays positive (or negative) for the state at the output of the channel (by concatenation also after any finite number of uses of the channel).

The spin component $\langle J_z\rangle_i$ of a state ρ_i (state after i uses of the channel) along the "z" axis reads

$$\langle J_z\rangle_i = \mathrm{Tr}(\rho_i J_z) = \sum_{m=-j}^{j} m\langle m|\rho_i|m\rangle.$$

Having a state of the type Eq. (C.1), with additional index i labeling the observer's tally number, the spin component reads

$$\langle J_z\rangle_i = \sum_{m=-j}^{j} m s_{i,m}. \qquad \text{(C.4)}$$

After using the channel one more time we have at the output of the channel, according to Eq. (C.3), the state

$$\sum_{m,m'=-j}^{j} s_{i,m} c_{j,m,m'} |m'\rangle\langle m'|, \tag{C.5}$$

where $c_{j,m,m'} = \langle m'|\chi(|m\rangle\langle m|)|m'\rangle$. The "$z$"-spin component for the state Eq. (C.5) is

$$\begin{aligned} \langle J_z\rangle_{i+1} &= \sum_{m=-j}^{j} s_{i,m} \underbrace{\sum_{m'=-j}^{j} m' c_{j,m,m'}}_{m\frac{j}{j+1}} \\ &= \frac{j}{j+1}\langle J_z\rangle_i. \end{aligned} \tag{C.6}$$

Hence, for $j > 0$, $\langle J_z\rangle_i \gtrless 0 \Leftrightarrow \langle J_z\rangle_{i+1} \gtrless 0$, which is what we wanted to show.

Appendix D

The channel induced by the weak measurements, Eq. (4.90), on N copies of a qubit

Let us now proceed to the case of weak measurements. Having

$$M_{\boldsymbol{n}_k} = (1-\varepsilon_k)\mathbb{1} + (N+1)\varepsilon_k |\boldsymbol{n}_k\rangle\langle\boldsymbol{n}_k|$$

we want to evaluate

$$\begin{aligned} |m\rangle\langle m| \to \rho' &= \int_{\mathbb{S}^2} I_{\boldsymbol{n}_k}(|m\rangle\langle m|)\mathrm{d}\boldsymbol{n}_k \\ &= \int_{\mathbb{S}^2} A_{\boldsymbol{n}_k}|m\rangle\langle m|A^\dagger_{\boldsymbol{n}_k}\mathrm{d}\boldsymbol{n}_k, \end{aligned}$$

where

$$\begin{aligned} A_{\boldsymbol{n}_k} &= \sqrt{M_{\boldsymbol{n}_k}} \\ &= a_k\mathbb{1} + b_k|\boldsymbol{n}_k\rangle\langle\boldsymbol{n}_k| \end{aligned} \tag{D.1}$$

with $a_k = \sqrt{1-\varepsilon_k}$ and $b_k = \left(\sqrt{1+2j\varepsilon_k} - \sqrt{1-\varepsilon_k}\right)$. Thus, we have

$$\begin{aligned} \chi_{\varepsilon_k}(|m\rangle\langle m|) &= \int \mathrm{d}\boldsymbol{n}_k\big[a_k^2|m\rangle\langle m| + (|m\rangle\langle m|\boldsymbol{n}_k\rangle\langle\boldsymbol{n}_k| + H.C.)a_kb_k \\ &\quad + b_k^2|\langle\boldsymbol{n}_k|m\rangle|^2|\boldsymbol{n}_k\rangle\langle\boldsymbol{n}_k|\big] \\ &= a_k^2|m\rangle\langle m| + a_kb_k\alpha_1 + b_k^2\alpha_2, \end{aligned}$$

where $H.C.$ stands for the Hermitian-conjugated term,

$$\begin{aligned} \alpha_2 &= \int_{\mathbb{S}^2}\mathrm{d}\boldsymbol{n}_k \underbrace{\sum_{m'm''=-j}^{j} c_{m'}{c^*}_{m''}\langle m''|m\rangle\langle m|m'\rangle}_{|c_m|^2}|\boldsymbol{n}_k\rangle\langle\boldsymbol{n}_k| \\ &= \frac{\chi(|m\rangle\langle m|)}{(2j+1)} \end{aligned}$$

and

$$
\begin{aligned}
\alpha_1 &= \int_{\mathbb{S}^2} \mathrm{d}\boldsymbol{n}_k \left(\underbrace{\sum_{m'=-j}^{j} c_{m'} \langle m|m'\rangle}_{c_m} |m\rangle\langle \boldsymbol{n}_k| + H.C \right) \\
&= \frac{1}{4\pi} \sum_{m'} |m\rangle\langle m'| \int_0^{\pi} \mathrm{d}\theta \sin\theta |c_m||c^*_{m'}| \underbrace{\int_0^{2\pi} \mathrm{e}^{\mathrm{i}\varphi(m-m')} \mathrm{d}\varphi}_{2\pi\delta_{m,m'}} + H.C. \\
&= |m\rangle\langle m| \frac{1}{2} \binom{2j}{j+m} \underbrace{\int_0^{\pi} \mathrm{d}\theta \sin\theta \cos^{2(j+m)}\left(\frac{\theta}{2}\right) \sin^{2(j-m)}\left(\frac{\theta}{2}\right)}_{\frac{2(j+m)!(j-m)!}{(2j+1)!}} + H.C. \\
&= |m\rangle\langle m| \frac{2}{2j+1},
\end{aligned}
$$

where to evaluate the last integral one can utilize the result Eq. C.2 with substitutions $(4j \mapsto 2j)$ and $(m + m' \mapsto m)$.

Thus we have

$$
\chi_{\varepsilon_k}(|m\rangle\langle m|) = \frac{1}{2j+1} \big[\big(2a_k b_k + (2j+1) a_k^2\big) |m\rangle\langle m| + b_k^2 \chi(|m\rangle\langle m|) \big] \qquad \text{(D.2)}
$$

with $\chi(|m\rangle\langle m|)$ given by Eq. C.3.

Having a state of the type Eq. (C.1), with an additional index, k, labeling the observer's tally number, the expectation value $\langle J_z \rangle_k$ of the z spin component for the average state at the input of the kth observer's apparatus is given by Eq. (C.4). The expectation value of the $(k+1)$th state spin component reads

$$
\begin{aligned}
\langle J_z \rangle_{k+1} &= \sum_{m=-j}^{j} m \, \langle m| \sum_{m'} s_{k,m'} \chi_{\varepsilon_k + 1}(|m'\rangle\langle m'|) |m\rangle \\
&= \frac{1}{2j+1} \sum_{m=-j}^{j} m \, \langle m| \Bigg(\sum_{m'} s_{k,m'} \Big(\big[2a_{k+1} b_{k+1} + (2j+1) a_{k+1}^2 \big] |m'\rangle\langle m'| \\
&\qquad + b_{k+1}^2 \underbrace{\chi(|m'\rangle\langle m'|)}_{\sum_{m''} c_{j,m',m''} |m''\rangle\langle m''|} \Big) \Bigg) |m\rangle \\
&= \frac{1}{2j+1} \Bigg(\big[2a_{k+1} b_{k+1} + (2j+1) a_{k+1}^2 \big] \underbrace{\sum_{m=-j}^{j} m \, s_{k,m}}_{\langle J_z \rangle_k} \\
&\qquad + b_{k+1}^2 \underbrace{\sum_{m=-j}^{j} m \sum_{m'=-j}^{j} s_{k,m'} c_{j,m',m}}_{\frac{j}{j+1} \langle J_z \rangle_k \, (\text{as in } C.6)} \\
&\qquad + (2j+1) a_{k+1}^2 \underbrace{\sum_{m=-j}^{j} m |m\rangle\langle m| \underbrace{\sum_{m'=-j}^{j} s_{k,m'}}_{\mathrm{Tr}(\rho_k)=1}}_{0} \Bigg)
\end{aligned}
$$

$$= \frac{2a_{k+1}b_{k+1} + (2j+1)a_{k+1}^2 + b_{k+1}^2\frac{j}{j+1}}{2j+1}\langle J_z\rangle_k, \tag{D.3}$$

i.e., for $j > 0$ and $b_k \neq 0$, $\langle J_z\rangle_0 \lesseqgtr 0 \Leftrightarrow \langle J_z\rangle_k \lesseqgtr 0$.

Appendix E

The average channel induced by single-Kraus-operator measurements on a single qu*d*it

We will show that the optimal weak instrument, i.e. one maximizing next observer's fidelity given current observer's fidelity, for a qu*d*it induces a channel which is adding total mixture to the encoding state. We first collect some mathematical results concerning unitary group integrals that will be extensively used below. For matrices U belonging to the fundamental representation of $SU(d)$ and denoting by $\mathrm{d}U$ the corresponding Haar measure, we have

$$\int \mathrm{d}U\; U_i^j U_r^{\dagger\,s} = \frac{\delta_i^s \delta_r^j}{d}$$

and, similarly,

$$\int \mathrm{d}U\; U_i^j U_k^l U_r^{\dagger\,s} U_t^{\dagger\,v} = \frac{(\delta_i^s\delta_k^v + \delta_i^v\delta_k^s)(\delta_r^j\delta_t^l + \delta_r^l\delta_t^j)}{2d(d+1)} + \frac{(\delta_i^s\delta_k^v - \delta_i^v\delta_k^s)(\delta_r^j\delta_t^l - \delta_r^l\delta_t^j)}{2d(d-1)}. \tag{E.1}$$

The last result can be most easily seen by writing the integral above as

$$\int \mathrm{d}U \left(\yng(2) \oplus \yng(1,1)\right) \otimes \left(\yng(2) \oplus \yng(1,1)\right)^{\dagger}$$

and recalling the orthogonality relations of the irreducible representations of unitary groups, which state that

$$\int \mathrm{d}U\, \yng(2) \otimes \yng(1,1)^{\dagger} = \int \mathrm{d}U\, \yng(1,1) \otimes \yng(2)^{\dagger} = 0$$

$$\int \mathrm{d}U\, \yng(2) \otimes \yng(2)^{\dagger} \sim \mathbb{1}_{\yng(2)}; \; \int \mathrm{d}U\, \yng(1,1) \otimes \yng(1,1)^{\dagger} \sim \mathbb{1}_{\yng(1,1)}.$$

As we argued in Section 4.3, the effective apparatus, given by the actual one and the lack of knowledge about it, is covariant (with respect to SU(d) in this case). In terms of the Kraus operators, associated to measurement outcomes which do *not* transform upon a unitary "rotation" of the apparatus (e.g. LEDs or numbers of outcomes on a display), this means there is a unitary freedom in the next observers' possible knowledge of those Kraus operators for any given outcome and an average is performed over SU(d). We restrict our attention to measurements with a single term in the Kraus decomposition for any outcome – see discussion in Section 4.5.2.

Moreover, we assume that a given observer does not know the measurement outcomes of the previous observers, thus no other object, except the ith observer's output state, its probability, and guess, depends on his measurement outcome. Therefore we can perform the sum over all outcomes to get the channel induced by such measurement. Hence, one way to look at the measurement process is via the map

$$\begin{aligned}\hat{\rho} \mapsto \hat{\rho}' &= \chi(\hat{\rho}) \\ &= \sum_o \int \mathrm{d}U U A_o U^\dagger \hat{\rho} \quad U A_o^\dagger U^\dagger,\end{aligned}$$

where $\{o\}$ is the set of possible outcomes of the predecessing observer's apparatus (or the set enriched by additional outcomes so that a quantum operation performed given any outcome o has single Kraus operator in its Kraus decomposition).

Using Eq. (E.1) we get

$$\chi(\hat{\rho}) = \frac{c-1}{(d+1)(d-1)}\,\hat{\rho} + \frac{d^2-c}{(d+1)(d-1)}\frac{\mathbb{1}}{d}, \tag{E.2}$$

where

$$c = \sum_o |\mathrm{Tr}\, A_o|^2. \tag{E.3}$$

Appendix F
Definitions

Definition F.1. *Topology (open and closed sets). Topological space.*

A topology *on a set X is a system $\{\tau\}$ of subsets τ of the set X such that*

1. *$\emptyset \in \{\tau\}$, $X \in \tau$*
2. *union of (any number) of elements from $\{\tau\}$ is from $\{\tau\}$*
3. *intersection of finitely many elements of $\{\tau\}$ is from $\{\tau\}$.*

Elements of $\{\tau\}$ are called open sets *and the complements $X \backslash \tau$ are called* closed sets. *The tuple $(X, \{\tau\})$ is a* topological space.

Definition F.2. *Closure, interior, being dense, neighborhood.*

Consider a topological space $(X, \{\tau\})$. For any subset $\mathcal{M} \subset X$ there is a unique minimal (with respect to the set inclusion) closed subset $\overline{\mathcal{M}}$ of X containing $\mathcal{M}$, called the closure *of $\mathcal{M}$: $\overline{\mathcal{M}}$.*

Similarly, there is a unique maximal open subset of X contained in $\mathcal{M}$, called the interior *of $\mathcal{M}$, denoted $\mathcal{M}^\circ$. If the closure of $\mathcal{M} \subset X$ is the whole space X, then $\mathcal{M}$ is* dense in *X.*

Any subset $\mathcal{M}$ of X such that its interior $\mathcal{M}^\circ \ni x$ is a neighborhood *of $x \in X$.*

Definition F.3. *Continuity of maps.*

Let $(X, \{\tau_X\})$, $(Y, \{\tau_Y\})$ be two topological spaces. A map

$$f: X \to Y$$

is called continuous *if and only if for any $A \in \{\tau_Y\}$ the pre-image $f^{-1}(A) \in \{\tau_X\}$, i.e. if the pre-image of any open set is an open set. The map f is* continuous in the point *$x \in X$ if and only if for any open neighborhood $\mathcal{V}$ of $f(x) \in Y$, $f(x) \in \mathcal{V}$ there is an open neighborhood $\mathcal{U}$ of x, $x \in \mathcal{U}$ such that its image under f is contained in $\mathcal{V}$: $f(\mathcal{U}) \subset \mathcal{V}$.*

Definition F.4. *Homeomorphism (topological isomorphism) of topological spaces.*

Let $(X, \{\tau_X\})$, $(Y, \{\tau_Y\})$ be two topological spaces and let $f: X \to Y$ be a continuous map. If f is a bijection and if its inverse f^{-1} is also continuous, then f is a homeomorphism *of the spaces X and Y. Mutually homeomorphic spaces are indistinguishable from the topological point of view – they are* topologically isomorphic.

Definition F.5. *Topological linear space.*

Let $\mathcal{L}$ be a linear space over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, where $\mathbb{K}$ is considered with its canonical (usual) topology. Let a topology $\{\tau\}$ on $\mathcal{L}$ be given. Consider the multiplication of elements $x \in \mathcal{L}$ by scalars $\lambda \in \mathbb{K}$ as a mapping from the topological product-space $\mathbb{K} \times \mathcal{L}$ into $\mathcal{L}$: $(\lambda; x) \to \lambda x$ and the addition $(x; y)(\in \mathcal{L} \times \mathcal{L}) \to x + y(\in \mathcal{L})$ also with the product-topology of $\mathcal{L} \times \mathcal{L}$.

Then the topological space $(\mathcal{L},\ \{\tau\})$ *is a* topological linear space *if and only if the addition and multiplication by scalars are (everywhere) continuous functions. This allows to define any topology* $\{\tau\}$ *of a topological linear space* $(\mathcal{L},\{\tau\})$ *on the linear space* $\mathcal{L}$ *by giving just all open sets containing an arbitrary chosen point (e.g. x=0).*

Definition F.6. *Norm.*

Norm $\|.\|$ *on a linear space* V *over the field* $\mathbb{C}$ *is a map* $V\rightarrow\mathbb{R}$, $v\rightarrow\|v\|$, *where for all* $u,v\in V$ *and* $\lambda\in\mathbb{C}$ *the following conditions hold:*

1. *triangular inequality (sub-additivity):* $\|u+v\|\leq\|u\|+\|v\|$
2. *homogeneity:* $\|\lambda v\|=|\lambda\,|\,\|v\|$
3. *non-degeneracy:* $\|v\|=0\Leftrightarrow v=0$
4. *non-negativity:* $\|v\|\geq0$.

If the norm is induced by an inner product $\langle.|.\rangle$, *i.e.* $\|v\|=\sqrt{\langle v|v\rangle}$, *then the following relation holds:*

5. *the parallelogram rule:* $\|u+v\|^2+\|u-v\|^2=2\big(\|u\|^2+\|v\|^2\big)$.

Definition F.7. *Metric.*

Metric *on a set* Ω *is a function* $d{:}\,\Omega\times\Omega\rightarrow\mathbb{R}$ *(called the distance) such that for all* x, y,z *in* Ω, *this function is required to satisfy the following conditions:*

1. $d(x,y)=0$ *if and only if* $x=y$
2. $d(x,y)=d(y,x)$ *(symmetry)*
3. $d(x,z)\leq d(x,y)+d(y,z)$ *(subadditivity / triangle inequality).*

It follows that $d(x,y)\geq0$ *(non-negativity) since* $2d(x,y)=d(x,y)+d\,(y,x)\geq d(x,x)=0$.

Definition F.8. *Metric space.*

The tuple (Ω,d) *where* Ω *is a set and* d *is a metric on* Ω *is called* metric space.

Definition F.9. *Complete metric space.*

Metric space (Ω,d) *is* complete *if each Cauchy sequence of its elements (i.e. a sequence* $\{a_i\}$, $a_i\in\Omega$, *such that* $\forall\varepsilon\ \exists N$: $d(a_i,a_j)<\varepsilon$ *for* $\forall i,j>N$*) converges to an element of the space (i.e.* $\exists\omega\in\Omega$ *such that* $\forall\varepsilon\ \exists N$: $d(a_i,\omega)<\varepsilon$ *for* $\forall i>N$*).*

Definition F.10. *Banach space.*

Banach space B *is a linear space with a norm such that* B *is complete "in the norm" (the norm* $\|\,.\,\|$ *on a linear space with a norm, V, induces a metric* $d(u,v)\colon=\|u-v\|$, $u,v\in V$. *The convergence of sequences of elements of V is assumed with respect to this induced metric d).*

Definition F.11. *Topological dual (space).*

Let $\mathcal{L}$ *be a Banach space over* $\mathbb{K}\in\{\mathbb{R},\mathbb{C}\}$ *with a norm* $\|\,.\,\|$. *A linear mapping* m: $x(\in\mathcal{L})\rightarrow m(x)\equiv\langle m,x\rangle\in\mathbb{K}$ *is a* linear functional *on* $\mathcal{L}$. *On general (infinite-dimensional) Banach spaces, there are also discontinuous linear functionals. The set of all* continuous linear functionals *on* $\mathcal{L}$ *is denoted by* $\mathcal{L}^*$, *and is called the* topological dual (space) *of* $\mathcal{L}$.

In $\mathcal{L}^\star$, *there is a canonical norm-topology determined by that of* $\mathcal{L}$:

$$\|m\|\equiv\sup\left\{\frac{|m(x)|}{\|x\|}{:}\,0\neq x\in\mathcal{L}\right\},\qquad m\in\mathcal{L}^\star.$$

With the above norm, $\mathcal{L}^\star$ is a Banach space. Its dual space $\mathcal{L}^{\star\star}$ contains, as a canonically isometrically embedded subspace, the original Banach space $\mathcal{L}$: $x\in\mathcal{L}$ is interpreted as the mapping $m\rightarrow m(x)\equiv\langle m,x\rangle$, i.e. an element of $\mathcal{L}^{\star\star}$.

Definition F.12. *Hilbert space.*

Hilbert space *is a complex linear space with a scalar product $\langle .|.\rangle$ complete in the norm $\|\psi\|=\sqrt{\langle\psi|\psi\rangle}$. The scalar product needs to satisfy the following conditions*

1. *linearity:*
2. *symmetry:*
3. *non-degeneracy: $\langle\psi|\phi\rangle=0\ \forall\psi\ \Rightarrow\ \phi=0$*
4. *non-negativity: $\langle\psi|\psi\rangle\geq 0$.*

Definition F.13. *Separable space.*

A vector space is separable *if and only if its orthogonal basis is countable.*

Definition F.14. *Ideal.*

Let $\mathcal{A}$ be an algebra. A subset $\mathcal{I}$ ($\mathcal{I}\subset\mathcal{A},\{0\}\neq\mathcal{I},\mathcal{A}\neq\mathcal{I}$) such that multiplication of its elements by any $B\in\mathcal{A}$ from left/right/any side gives an elements of $\mathcal{I}$ ($\forall B\in\mathcal{A}$:$\mathcal{I}\cdot B\subset\mathcal{I}$, $\mathrm{resp}\,B\cdot\mathcal{I}\subset\mathcal{I},\mathrm{resp}\,B\cdot\mathcal{I}\cup\mathcal{I}\cdot B\subset\mathcal{I}$) is called a left/right/two-sided ideal. *Two-sided ideals are called simply ideals. An ideal $\mathcal{I}\subset\mathcal{A}$ is automatically a subalgebra of $\mathcal{A}$.*

Definition F.15. *Cover of a set. Subcover.*

Let Y be a subset of X, and C be a collection of subsets U_α of X, whose union contains Y, then C is said to be a cover *of Y. I.e. C is a cover of Y if $\cup_\alpha U_\alpha\supseteq Y$. If C is a cover of Y then a subset of C that still covers Y is a* subcover *of the cover C.*

Definition F.16. *Compactness of a topological space.*

A topological space is compact *if each of its open covers has a finite subcover.*

Definition F.17. *Topological group.*

A topological group *$\mathcal{G}$ is a topological space and a group such that the group operations*

$$\mathcal{G}\times\mathcal{G}\rightarrow\mathcal{G}\colon\ \ (g,h)\rightarrow gh$$

and

$$\mathcal{G}\rightarrow\mathcal{G}\colon\ \ g\rightarrow g^{-1}$$

are continuous functions ($\mathcal{G}\times\mathcal{G}$ is viewed as a topological space with the product topology).

Definition F.18. *Compact group.*

A compact group *is a topological group that is also a compact (topological) space.*

Definition F.19. *σ-additivity (of a set function).*

*Let Ω be a set. Let $P(\Omega)$ denote the set of all subsets of Ω. Let $E\subset P(\Omega)$ (where $P(\Omega)$ is the set of all subsets of Ω) be some nonempty system (set) of subsets of Ω, such that $\emptyset\in E$. Let $\mu\colon E\rightarrow\bar{R}$ be a function defined on the set Ω, where $\bar{R}:=R\cup\infty$ and $\emptyset$ is the empty set. The function μ is called σ-*additive *(or* countably additive*, or a* generalized measure*) if*

1. *$\mu(\emptyset)=0$*

2. $E_n \in E, n=1,2,\ldots, E_i \cap E_j = \varnothing$ for $i \neq j$, $\overset{\infty}{\underset{n=1}{\cup}} E_n \in E \Rightarrow \mu(\overset{\infty}{\underset{n=1}{\cup}} E_n) = \sum_{n=1}^{\infty} \mu(E_n)$.

Definition F.20. *σ-algebra (σ-field).*

Let Ω be a set. The set of subsets $\mathcal{F} \subset P(\Omega)$ of Ω (where $P(\Omega)$ is the set of all subsets of Ω) is a σ-algebra if it has the following properties:

1. $\emptyset, \Omega \in \mathcal{F}$
2. *If $X_j \in \mathcal{F}$, then $\cup_j X_j \in \mathcal{F}$*
3. *If $X \in \mathcal{F}$, then $\Omega \backslash X \in \mathcal{F}$*

Definition F.21. *Measure.*

If a generalized measure (σ-additive function) μ is non-negative, it is a measure.

Definition F.22. *Probability measure.*

Probability (measure) *p is a measure defined on a σ-algebra $\mathcal{F}$ of subsets of a set Ω such that $p(\Omega) = 1$.*

Definition F.23. *Measurable space.*

The tuple $(\Omega, \mathcal{F})$, where Ω is a set and $\mathcal{F} \subset P(\Omega)$ is a σ-algebra is called a measurable space.

Definition F.24. *Measure space.*

A measurable space $(\Omega, \mathcal{F})$ with a fixed measure μ form a measure space $(\Omega, \mathcal{F}, \mu)$.

Definition F.25. *Probability space (sample space).*

The triple $(\Omega, \mathcal{F}, p)$, i.e. a measure space where the measure is a probability measure is called a probability space.

Definition F.26. *Image measure (induced measure).*

Let $(\Omega,\ \mathcal{F},\ p)$ be a measure space (e.g. a probability space) and let $(\Omega',\ \mathcal{F}')$ be a measurable space, and let $\Phi: \Omega \rightarrow \Omega'$ be a $(\mathcal{F}, \mathcal{F}')$-measurable map. Then the measure p induces an image measure *p_Φ on Ω' by*

$$p_\Phi(F) = p(\Phi^{-1}(F)),$$

where $F \in \mathcal{F}$.

Definition F.27. *Characteristic function of a set.*

The characteristic function $\chi_A: X \rightarrow \mathbb{R}$ for a set $A \subset X$ is defined as

$$\chi_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

Definition F.28. *Simple measurable function.*

Let $(X, \mathcal{F})$ be a measurable space, $E_1, \ldots E_n$ a finite number of subsets of the set X, and $c_1, \ldots c_n$ a finite number of real numbers. A function $f: X \rightarrow \mathbb{R}$ such that for each $x \in X$

$$f(x) = c_1 \chi_{E_1}(x) + \ldots + c_n \chi_{E_n}(x),$$

where χ_{E_i} is the characteristic function of a set E_i, is called a simple function. *If $E_i \in \mathcal{F}$ for $i = 1, \ldots, n$, then f is called a* simple measurable function.

Definition F.29. *Measurable function.*

Let $(X, \mathcal{F})$ *be a measurable space. A function* $f: X \to \mathbb{R}$ *is called* measurable *(more precisely* $\mathcal{F}$*-measurable) if there exists a sequence* $(f_n)_{n=1}^{\infty}$ *of simple measurable functions such that for each* $x \in X$ $f(x) = \lim_{n\to\infty} f_n(x)$.

Definition F.30. *Random variable.*

Let $(\Omega, \mathcal{F}, p)$ *be a probability space. The map* $X: \Omega \to \mathbb{R}$ *is called a random variable if for each* $x \in \mathbb{R}$ $\{w \in \Omega; X(\omega) < x\} \in \mathcal{F}$.

Alternative definition: The map X: $\Omega \to \mathbb{R}$ *is a random variable* $\Leftrightarrow$ X *is a (*$\mathcal{F}$*-)measurable function.*

Definition F.31. *Distribution.*

The image measure p_X *(a probability measure on* $\mathbb{R}$*) induced by the random variable* X *is called the* distribution *of the random variable* X.

Definition F.32. *Distribution function.*

A distribution function $F: \mathbb{R} \to \mathbb{R}$ *of a random variable* X *is defined by the inequality* $F(x) = p(\{\omega; X(\omega) \leq x\})$.

Definition F.33. *Borel σ-algebra.*

A Borel σ-algebra *on a topological space* $(\Omega, \{\mathcal{T}\})$ *is the minimal* σ*-algebra containing all open sets (sets* $\mathcal{T} \in \{\mathcal{T}\}$*).*

Definition F.34. *Borel space.*

Borel space $(\Omega, \mathcal{F})$ *is a topological space* Ω *equipped with a Borel* σ*-algebra* $\mathcal{F}$.

Definition F.35. *Projective representation.*

A mapping $U: \mathcal{G} \to \mathrm{Aut}(V)$ *is a* projective *representation of a group* $\mathcal{G}$ *if and only if for* $\forall g, h \in \mathcal{G}$,

$$U(g \circ h) = m(g, h) U(g) U(h),$$

where $m: \mathcal{G} \otimes \mathcal{G} \to S^1 \subset \mathbb{C}$ *is a multiplier for the group* $\mathcal{G}$ *satisfying the following identities implied by associativity of group multiplication:*

$$m(g_1, g_2 \circ g_3) m(g_2, g_3) = m(g_1 \circ g_2, g_3) m(g_1, g_2)$$

$$m(g, e) = m(e, g) = 1, \ \forall g \in \mathcal{G}, \ e \circ g \equiv g$$

Definition F.36. *Commonly used sets.*

$\boldsymbol{\mathcal{L}(\mathcal{H})}$. *the set of bounded linear operators, i.e.* $\|\boldsymbol{A}\| < \infty$

$\boldsymbol{\mathcal{L}_S(\mathcal{H})}$. *the set of self-adjoint linear operators, i.e.* $\boldsymbol{A} = \boldsymbol{A}^\dagger$

$\boldsymbol{\mathcal{L}_+(\mathcal{H})}$. *the set of positive linear operators, i.e.* $\boldsymbol{A} \geq 0$

$\boldsymbol{\mathcal{T}(\mathcal{H})}$. *the set of trace class linear operators, i.e.* $\mathrm{Tr}|\boldsymbol{A}| < \infty$

$\boldsymbol{\mathcal{T}(\mathcal{H})_1^+ \equiv \mathcal{S}(\mathcal{H})}$. *the set of trace one positive linear operators (density matrices), i.e.* $\mathrm{Tr}(\boldsymbol{A}) = 1$, $\boldsymbol{A} \geq 0$

$\boldsymbol{\mathcal{T}_2(\mathcal{H})}$. *the set of Hilbert-Schmidt operators, i.e.* $\mathrm{Tr}(\boldsymbol{A}^\dagger \boldsymbol{A}) < \infty$

$\boldsymbol{\mathcal{L}(\mathcal{T}_2)}$. *the set of superoperators, i.e.* $\boldsymbol{\Phi}: \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H})$

$\boldsymbol{\mathcal{U}(\mathcal{H})}$. *the set of unitary operators, i.e.* $\boldsymbol{U^{-1}=U^\dagger}$

Mutual relationships of the above sets:

$$\mathcal{S}(\mathcal{H})\subset\mathcal{L}_+(\mathcal{H})\subset\mathcal{L}_S(\mathcal{H})\subset\mathcal{L}(\mathcal{H}),\qquad \mathcal{T}(\mathcal{H})\subset\mathcal{T}_2(\mathcal{H})\subset\mathcal{L}(\mathcal{H})$$

Topologies on $\mathcal{L}(\mathcal{H})$:

If $\mathcal{H}$ is a Hilbert space, the set $\mathcal{L}(\mathcal{H})$ carries three useful topologies

1. The *norm topology* is the topology induced by the operator norm $\|T\| = \sup_{\|u\|=1}\|Tu\|$, where $u\in\mathcal{H}$.
2. The *strong operator topology* is the topology induced by the seminorms $T\rightarrow\|Tu\|$, $u\in\mathcal{H}$. A net $\{T_\alpha\}$ in $\mathcal{L}(\mathcal{H})$ converges to T strongly if and only if $\|T_\alpha u-Tu\|\rightarrow 0$ for every $u\in\mathcal{H}$.
3. The *weak operator topology* is the topology induced by the seminorms $T\rightarrow|\langle v, Tu\rangle|$, u, $v\in\mathcal{H}$. A net $\{T_\alpha\}$ in $\mathcal{L}(\mathcal{H})$ converges to T weakly if and only if $\langle v, T_\alpha u\rangle\mapsto\langle v,Tu\rangle$ for every $u,v\in\mathcal{H}$.

Bibliography

[1] A. Acín, J. I. Latorre, and P. Pascual. Optimal generalized quantum measurements for arbitrary spin systems. *Phys. Rev. A*, 61(2):022113, January 2000.

[2] R. B. Ash. *Information Theory*. Interscience, 1965.

[3] E. Bagan, M. Baig, A. Brey, R. Muñoz Tapia, and R. Tarrach. Optimal encoding and decoding of a spin direction. *Phys. Rev. A*, 63(5):052309, Apr 2001.

[4] E. Bagan, M. Baig, and R. Muñoz Tapia. Communication of spin directions with product states and finite measurements. *Phys. Rev. A*, 64(2):022305, Jul 2001.

[5] K. Banaszek. Fidelity balance in quantum operations. *Phys. Rev. Lett.*, 86(7):1366–1369, Feb 2001.

[6] K. Banaszek. Information gain versus state disturbance for a single qubit. *Open Systems & Information Dynamics*, 13(1):1–16, 2006.

[7] K. Banaszek and I. Devetak. Fidelity trade-off for finite ensembles of identically prepared qubits. *Phys. Rev. A*, 64(5):052307, Oct 2001.

[8] H. Barnum. Information-disturbance tradeoff in quantum measurement on the uniform ensemble and on the mutually unbiased bases. *arXiv:quant-ph/0205155v1*, 2002.

[9] S. D. Bartlett, T. Rudolph, and R. W. Spekkens. Optimal measurements for relative quantum information. *Phys. Rev. A*, 70(3):032321, Sep 2004.

[10] S. D. Bartlett, T. Rudolph, R. W. Spekkens, and P. S. Turner. Degradation of a quantum reference frame. *New Journal of Physics*, 8(4):58, 2006.

[11] J.-C. Boileau, L. Sheridan, M. Laforest, and S. D. Bartlett. Quantum reference frames and the classification of rotationally invariant maps. *Journal of Mathematical Physics*, 49(3):032105, 2008.

[12] P. Bóna. Extended quantum mechanics. *Acta Physica Slovaca*, 50(1), February 2000.

[13] D. Bruss and G. Leuchs, editors. *Lectures on Quantum Information*. Wiley-vch, 2007.

[14] D. Bruss and C. Macchiavello. Optimal state estimation for d-dimensional quantum systems. *Physics Letters A*, 253(5-6):249 – 251, 1999.

[15] P. Busch, M. Grabowski, and P. J. Lahti. *Operational Quantum Physics*. New Series m: Monographs. Springer-Verlag, second edition, 1997.

[16] V. Bužek, P. L. Knight, and N. Imoto. Multiple observations of quantum clocks. *Phys. Rev. A*, 62(6):062309, Nov 2000.

[17] V. Bužek, R. Derka, and S. Massar. Optimal quantum clocks. *Phys. Rev. Lett.*, 82(10):2207–2210, March 1999.

[18] C. Carmeli, T. Heinosaari, and A. Toigo. Covariant quantum instruments. *Journal of Functional Analysis*, 257(11):3353–3374, 12 2009.

[19] N. J. Cerf. Classical and quantum information theory, from discrete to continuous variables. Lecture notes at the Quantum Information, Computation, and Complexity trimester, Paris, France, 2006.

[20] G. B. Folland. *Real Analysis, Modern Techniques and Their Applications*. John Wiley & sons, 1999.

[21] G. Chiribella. *Optimal estimation of quantum signals in the presence of symmetry*. PhD thesis, UNIVERSITA DEGLI STUDI DI PAVIA, 2006.

[22] G. Chiribella, G. M. D'Ariano, P. Perinotti, and M. F. Sacchi. Efficient use of quantum resources for the transmission of a reference frame. *Phys. Rev. Lett.*, 93(18):180503, Oct 2004.

[23] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi. Optimal estimation of group transformations using entanglement. *Phys. Rev. A*, 72(4):042338, Oct 2005.

[24] G. Chiribella, G. M. D'Ariano, and D. Schlingemann. How continuous quantum measurements in finite dimensions are actually discrete. *Physical Review Letters*, 98(19):190403, 2007.

[25] G. M. D'Ariano, C. Macchiavello, and M. F. Sacchi. On the general problem of quantum phase estimation. *Physics Letters A*, 248(2-4):103 – 108, 1998.

[26] G. M. D'Ariano and H. P. Yuen. Impossibility of measuring the wave function of a single quantum system. *Phys. Rev. Lett.*, 76(16):2832–2835, April 1996.

[27] E. Davies. Information and quantum measurement. *IEEE Transactions on Information Theory*, 24(5):596–599, September 1978.

[28] E. B. Davies. On the repeated measurement of continuous observables in quantum mechanics. *Journal of Functional Analysis*, 6(2):318–346, 10 1970.

[29] E. B. Davies. *Quantum Theory of Open Systems*. Academic Press, London, 1976.

[30] E. B. Davies and J. T. Lewis. An operational approach to quantum probability. *Communications in Mathematical Physics*, 17:239–260, 1970. 10.1007/BF01647093.

[31] R. Derka, V. Bužek, and A. K. Ekert. Universal algorithm for optimal estimation of quantum states from finite ensembles via realizable generalized measurement. *Phys. Rev. Lett.*, 80(8):1571–1575, February 1998.

[32] M. Fecko. *Diferenciálna geometria a Liove grupy pre fyzikov*. Iris, 2004.

[33] C. A. Fuchs and K. Jacobs. Information-tradeoff relations for finite-strength quantum measurements. *Phys. Rev. A*, 63(6):062305, May 2001.

[34] C. A. Fuchs and A. Peres. Quantum-state disturbance versus information gain: Uncertainty relations for quantum information. *Phys. Rev. A*, 53(4):2038–2045, Apr 1996.

[35] M. G. Genoni and M. G. A. Paris. Information/disturbance trade-off in single and sequential measurements on a qudit signal. *Journal of Physics: Conference Series*, 67(1):012029, 2007.

[36] N. Gisin and S. Popescu. Spin flips and quantum information for antiparallel spins. *Phys. Rev. Lett.*, 83(2):432–435, July 1999.

[37] T. Heinosaari and M. Ziman. Guide to mathematical concepts of quantum theory. *Acta Physica Slovaca*, 58(4):487–674, 2008.

[38] C. W. Helstrom. *Quantum Detection and Estimation Theory*. Academic Press, New York, 1976.

[39] A. S. Holevo. *Probabilistic And Statistical Aspects Of Quantum Theory*, volume 1 of *North-Holland Series In Statistics And Probability*. North-Holland Publishing Company, 1982.

[40] A. S. Holevo. Radon–nikodym derivatives of quantum instruments. *Journal of Mathematical Physics*, 39:1373, 1998.

[41] D. Kretschmann, D. Schlingemann, and R. F. Werner. The information-disturbance tradeoff and the continuity of stinespring's representation. *IEEE Transactions on Information Theory*, 54(4):1708–1717, March 2008.

[42] J. Ladislav Mišta and J. Fiurášek. Optimal partial estimation of quantum states from several copies. *Physical Review A (Atomic, Molecular, and Optical Physics)*, 74(2):022316, 2006.

[43] J. I. Latorre, P. Pascual, and R. Tarrach. Minimal optimal generalized quantum measurements. *Phys. Rev. Lett.*, 81(7):1351–1354, August 1998.

[44] C. Macchiavello. Optimal estimation of multiple phases. *Phys. Rev. A*, 67(6):062302, Jun 2003.

[45] L. Maccone. Entropic information-disturbance tradeoff. *Europhysics Letters*, 4(77):4002, Feb 2007.

[46] S. Massar. Collective versus local measurements on two parallel or antiparallel spins. *Phys. Rev. A*, 62(4):040101, Sep 2000.

[47] S. Massar and S. Popescu. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.*, 74(8):1259–1263, February 1995.

[48] L. Mišta. Minimal disturbance measurement for coherent states is non-gaussian. *Phys. Rev. A*, 73(3):032335, Mar 2006.

[49] M. A. Naimark. Spectral functions of symmetric operators. *Izv. Akad. Nauk SSSR, Ser. Mat.*, 3:277–318, 1940.

[50] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[51] A. Peres. *Quantum Theory: Concepts and Methods*, volume 57 of *Fundamental Theories of Physics*. Kluwer Academic Publishers, 1993.

[52] D. Poulin and J. Yard. Dynamics of a quantum reference frame. *New Journal of Physics*, 9(5):156, 2007.

[53] P. Rapčan, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, and V. Bužek. Recycling of qubits. *Physica Scripta*, 2010(T140):014059, 2010.

[54] B. Riečan and T. Neubrunn. *Teória miery*. Veda, 1992.

[55] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, July, October 1948.

[56] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.

[57] W. F. Steinspring. Positive functions on c*-algebras. *Proc. Amer. Math. Soc.*, 6(2):211–216, April 1955.

[58] P. Štelmachovič and V. Bužek. Dynamics of open quantum systems initially entangled with environment: Beyond the kraus representation. *Phys. Rev. A*, 64(6):062106, Nov 2001.

[59] P. Štelmachovič and V. Bužek. Erratum: Dynamics of open quantum systems initially entangled with environment: Beyond the kraus representation [phys. rev. a 64, 062106 (2001)]. *Phys. Rev. A*, 67(2):029902, Feb 2003.

[60] R. Tarrach and G. Vidal. Universality of optimal measurements. *Phys. Rev. A*, 60(5):R3339–R3342, Nov 1999.