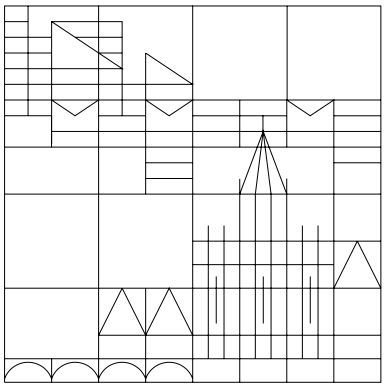
State and Parameter Estimation in Quantum Theory

Diplomarbeit

Marcus Kubitzki



AG Audretsch
Fachbereich Physik
Mathematisch-Naturwissenschaftliche Sektion
Universität Konstanz

Juni 2003

Contents

Zusammenfassung

Diese Arbeit beschäftigt sich mit der Schätzung von Parametern und Zuständen in der Quantentheorie. Die Theorie verallgemeinerter Quantenmessungen, vorgestellt in Kapitel 2, bildet hierzu den begrifflichen Rahmen.

Der erste Teil der Arbeit, Kapitel 3 und 4 umfassend, untersucht die Balance zwischen der durch Messung gewonnenen Information und der hierdurch am System verursachten Störung. Für Einzelmessungen an einem einzigen System wird dieser qualitative Kompromiß quantitativ in Form der Fidelities F und G untersucht, insbesondere für Qubits und Messung selbiger vermittels der Klasse unscharfer Messungen. Sie bestehen aus kommutierenden Effekten und können als 'verschmierte' Versionen gewöhnlicher hermitescher Observablen, wie z.B. Energie oder Spin, interpretiert werden. Es wird gezeigt, daß die optimale Balance zwischen Informationsgewinn und Störung eine einfache Nebenbedingung an die Parameter liefert, die diese unscharfen Qubit-Messungen charakterisieren.

Die FG-Ebene liefert eine anschauliche Darstellung der Fidelity-Balance. Für Qubits kann sie ausgedrückt werden durch jene Parameter, die die unscharfe Qubit-Messung repräsentieren. Diese Parametrisierung enthüllt eine einfache Struktur, die minimale von nicht-minimalen Messungen trennt (Kraus Operatoren nicht-minimaler Messungen enthalten einen nicht-trivialen unitären Anteil in ihrer polaren Zerlegung). Es zeigt sich, daß nicht-minimale Messungen die optimale Fidelity-Balance deutlich verschlechtern.

Ist vor einer Messung nichts über den zu schätzenden reinen Qubit-Zustand bekannt, kann eine unitäre 'back-action', d.h. eine nicht-minimale Messung, die Fidelity F nicht erhöhen. Existiert nun aber Vorinformation über den Zustand, dann verbessert – wie in Kapitel 4 gezeigt wird – diese Information zusammen mit einer passend gewählten 'back-action' die Fidelity, welche für minimale Messungen mit Vorinformation berechnet wurde. Letztere kann ihrerseits größer oder kleiner sein als die Fidelity, berechnet für den Fall daß keine Vorinformation über das Qubit existiert; Information bedeutet nicht unbedingt eine höhere Fidelity. Interessanterweise kann die Fidelity unter Einbeziehung von Vorinformation und diesen speziellen nicht-minimalen Messungen höher ausfallen als die Fidelity ohne Vorinformation.

Im zweiten Teil der Arbeit, bestehend aus den Kapiteln 5 und 6, verschiebt sich der Schwerpunkt auf die Schätzung von Parametern, die die Dynamik eines Qubits charakterisieren. Aufbauend auf füheren Arbeiten [Aud01, Aud02c] zur Echtzeit-Visualisierung von Rabi-Oszillationen mittels Sequenzen unscharfer Messungen (N-Serien), werden verschiedene Schätzverfahren für den Parameter $|c_1|^2$ entwickelt und verglichen. Die ursprünglich vorgeschlagene Schätzung [Aud01] wurde im Hinblick auf Erwartungstreue konstruiert, ein Kriterium aus der klassischen Schätztheorie. Ein zweites Verfahren wird mit Hilfe der 'maximum likelihood'-Methode abgeleitet. Diese Schätzung macht jedoch keinerlei Gebrauch von eventuell vorhandener Vorinformation über den Qubit-Zustand. Das dritte Schätzverfahren ('Bayesian estimator') bezieht

solche Informationen über den Satz von Bayes mit ein. Um alle Schätzungen vergleichen zu können, wird ein mittleres Fehlerquadrat konstruiert um ein Gütemaß zu erhalten. Mit Hilfe dieses Maßes kann gezeigt werden, daß für eine N-Serie das Schätzverfahren von Bayes sowohl dem 'maximum likelihood'- als auch dem ursprünglichen Verfahren überlegen ist. Letzteres ist seinerseits dem 'maximum likelihood'-Verfahren unterlegen, d.h. die ursprüngliche Schätzung hat ein größeres mittleres Fehlerquadrat.

Chapter 1

Introduction

Roughly 25 years ago, physicists began to realize that quantum mechanical systems could be employed to accomplish information processing tasks beyond classical limitations. Since then, quantum information theory became an ever growing field, boosted by theoretical breakthroughs as well as the steadily increasing ability to control single quantum systems.

Two elementary processes are of great importance to quantum information and quantum computation: state estimation and the characterization of the dynamics of a quantum system. This thesis is dealing with both of these aspects, whose applications range from assessing the performance of quantum gates or quantum communication channels, to the determination of types and magnitudes of different noise processes in a system. In each case, information about a quantum state is necessary, which means that measurements have to be made.

Projective measurements are often useless in this respect, especially when superpositions of states occur. Hereby, generalized measurements provide a tool better suited. This concept of measurement, originating in the late 1980s, became the new paradigm of quantum measurement theory [Bus95, Bus91, Hol82, Kra83], superseding the old formulation of von Neumann which is included in this approach.

The mathematical notion of observables as positive operator valued measures (POVMs) – instead of hermitian operators with a projective valued measure (PVM) – had manifold motivations, ranging from practical issues to foundational interests in quantum theory. Due to their experimental realization as indirect projection measurements, the disturbing influence of generalized measurements on a state can be chosen according to special requirements. This feature predestines them for state estimation and monitoring of dynamics.

As aforementioned, both state estimation and the characterization of dynamics are discussed in this work which consists of two parts: In the first part, the relation between information gain and disturbance is discussed. Emphasis is put on qubits, the simplest non-trivial quantum system, and the special class of unsharp measurements. The second part is concerned with the characterization of dynamics. Rabi oscillations of a qubit provide an exactly solvable model system for studying the real-time visualization of dynamics. Here, different methods for the estimation of parameters characterizing this dynamical behavior can be tested.

This work is structured as follows: Chapter 2 introduces the language of generalized measurements, the fundamental tool for investigations into quantum state estimation and parameter estimation theory. After a short motivation the mathematical theory of generalized measurements in terms of positive operator valued measures is laid out as well as their experimental realization.

Chapters 3 and 4 are concerned with the influence unsharp measurements have on quantum mechanical systems. In chapter 3 the trade-off between information gain and disturbance is the main item of interest. A quantitative description of these qualitative terms is provided by the well-known concept of fidelity. For one measurement on a single system the optimal fidelity balance is investigated, especially for the class of minimal unsharp measurements of a qubit. Non-minimal qubit measurements are subject of chapter 4. There, the question of a partial reversal of non-unitary operations (i.e. measurements) by means of unitary back-action is discussed.

Chapters 5 and 6 deal with parameter estimation and the visualization of dynamics, the latter being subject of chapter 5. There, the real-time monitoring of a qubit's Rabi oscillations by means of a sequence of unsharp measurements motivates estimation of the parameter $|c_1|^2$. In analogy to the fidelity, a quantitative measure defining the quality of a guess is developed. Chapter 6 exclusively deals with guesses (estimators) for $|c_1|^2$ and their evaluation. In addition to the originally proposed estimator of chapter 5, two new estimators are developed, followed by a comparison of all estimators.

Chapter 7 summarizes all results and gives some ideas for further work.

Appendices A, B and C contain detailed calculations omitted in chapters 3 and 4, the Bloch sphere representation of qubits and basic definitions of probability theory and classical estimation theory, respectively.

Chapter 2

Generalized Measurements

State and parameter estimation in quantum mechanics are connected by the theory of generalized measurements, laid out conceptually and mathematically in this chapter. Generalized measurements as a necessary conceptual leap beyond the long accepted formulation of quantum measurements given by von Neumann are motivated[1] in section 2.1. Following up on this, section 2.2 provides the mathematical implementation in terms of positive operator valued measures on Hilbert space. The physical realization of generalized measurements together with the Stern-Gerlach experiment as a prime example is depicted in the concluding section 2.3.

2.1 Motivation

Since its emergence in the late 1920s, quantum theory on Hilbert space has been the basis of fruitful and deep research into virtually all branches of physics. There seems to be no instance of conflict between theoretical predictions and experimental results. In view of this success it is remarkable that a few conceptual problems have resisted any attempted resolution even until now. The most prominent of these is known as the 'measurement problem': It paraphrases the fact that a superposition of quantum states $\sum_i c_i|i\rangle$ reduces upon measurement instantaneously with probability $|c_k|^2$ to the eigenstate $|k\rangle$ of the measured observable. This 'collapse of the wave function' (or 'reduction of the state vector') badgered physicists since the work of John von Neumann on mathematical foundations of quantum theory [vN32], where he introduced quantum observables as self-adjoint operators in Hilbert space and the aforementioned projection postulate to describe the reduced quantum states for discrete observables.

For over 40 years this paradigm stood at the heart of quantum measurement theory together with other conceptual shortcomings (like, for example, the possibility of interpreting quantum mechanics as a theory of individual systems with definite real properties, or certain limitations on measurability discovered by Wigner. For details, see chapter one of [Bus95]). Some of these became tractable in the 1980s once the probabilistic structure of quantum mechanics was appreciated in its full generality. Besides Gleason's theorem the introduction of observables as *positive operator valued measures* (POVMs) were a crucial step in this development. Interestingly, the latter discovery was made independently in a variety of rather disparate areas of physics, motivations ranging from foundational interests to fairly practical needs. This wide scope of the

[1] Here, as well as in subsections 2.2.1 and 2.3.2, we will closely follow [Bus95].

concept of POVMs demonstrates its status as an integral part of the basic structure of quantum theory.

Now what are those motivations which led to the incorporation of generalized[2] measurements (represented by POVMs) into the quantum vocabulary? A quite thorough account on these matters can be found in [Bus95]. Here, only one specific conceptual problem is sketched, followed by a more in-depth analysis of the Stern-Gerlach experiment in subsection 2.3.2 once the mathematical formalism is available.

Some puzzles in the foundations of quantum theory appeared in the form of a conflict between familiar classical physical ideas and some 'strange' implications of the quantum formalism. In each case the resolution consisted of rephrasing a strict no-go verdict excluding certain sharp (projective) measurements into a positive statement expressing the possibility of unsharp measurements subject to some limitations.

One of these no-go verdicts is the non-commutativity of certain pairs of self-adjoint operators, commonly interpreted as the root of the – classically unknown – incommensurability of the corresponding observables. From the fundamental commutator relation

$$[Q, P] = i\hbar\mathbb{1} \tag{2.1}$$

it was argued that measurements of position and momentum are mutually exclusive and cannot be performed together on single systems. Even worse, Heisenberg's interpretation of the uncertainty relation

$$\Delta Q \cdot \Delta P \geq \frac{\hbar}{2}\,, \tag{2.2}$$

being a consequence of (2.1), limits the accuracy of position and momentum measurements performed on ensembles of systems prepared in one and the same state; here, ΔQ and ΔP are the standard deviations of Q and P in some state ρ.

Instead of accepting the mere incommensurability of position and momentum, the pioneers of quantum theory considered various thought experiments (such as the gamma ray microscope) to demonstrate that joint measurements of these complementary observables should be possible in principle. The crucial idea hereby was that such measurements must not be too accurate, the limits of precision given by (2.2). While the measurement indeterminacy interpretation of the uncertainty relation is commonly accepted, its tenability was nevertheless long questioned due to the lacking rigorous incorporation of the idea of inaccurate measurements into the quantum formalism. Thus, a strict inclusion of unsharp measurements represented by positive operator valued measures paved the way for a solution and deeper understanding of this and related problems.

2.2 Mathematical Theory

Generalized measurements were motivated in the preceeding section among other things by the fact that they arise naturally in the theoretical description of many experiments. Now we are providing the main ideas behind POVMs in compact form in subsection 2.2.1 using an experiment as our guide [Bus95]. Subsequently, the mathematical formulation in familiar terms of operators in Hilbert space is given.

[2]Although generalized measurements are only introduced in the next section we use the term here to point out the necessity for a theory of measurement going beyond von Neumann.

2.2.1 Statistical Analysis of an Experiment

In analyzing the general features of any physical experiment one is able to specify those mathematical structures that are relevant to the theoretical description of an experiment. Any type of physical system is characterized by means of a collection of *preparation* procedures, the application of which prepare the system in a *state* ρ. The set of states is not a simplex, thus accounting for the fact that the same mixed state can be prepared by different mixtures of pure states. Given a system prepared in a state ρ, a measurement can be applied, leading to the *registration* of some outcome ω_i. For illustrative purposes, we assume a finite[3] set of pointer readings $\Omega = \{\omega_1, \ldots, \omega_n\}$. The very existence of physical experience is due to the fact that one is able to observe regularities in the event sequences occurring in nature. In particular, physical experimentation as sketched above would lose its meaning, were there not a *probabilistic* connection between the occurrence of a registration and the preceeding preparation. Hence, any pair (ρ, ω_i) of a state ρ and an outcome ω_i should determine a conditional probability $p(\omega_i|\rho)$,

$$(\rho, \omega_i) \;\mapsto\; p(\omega_i|\rho) \tag{2.3}$$

which in a long run of repeated experiments (N trials) is approximated by the relative frequency $N(\omega_i)/N$ of the occurrence of the outcome ω_i. It should be noted that different preparation procedures may be statistically equivalent in that they yield the same statistics for all possible measurements. Therefore the states ρ correspond, strictly speaking, to equivalence classes of preparation procedures. Similarly, different registration procedures may be statistically equivalent in the sense of yielding the same probabilities in every state. This gives rise to the definition of an observable as an equivalence class of measurements. In fact, the map (2.3) can be viewed in two ways. First, any outcome ω_i induces a state functional E_i,

$$E_i : \quad \rho \;\mapsto\; E_i(\rho) := p(\omega_i|\rho) \tag{2.4}$$

called an *effect*. Now the measured *observable* may be defined as the map assigning to each outcome ω_i its associated effect:

$$E : \quad \omega_i \;\mapsto\; E_i \,. \tag{2.5}$$

According to the second reading of (2.3), any state ρ fixes a probability distribution

$$p_\rho : \quad \omega_i \;\mapsto\; p_\rho(\omega_i) := p(\omega_i|\rho) \,.$$

In the simple case of a discrete experiment the properties of such a probability measure are summarized in the positivity ($p_\rho(\omega_i) \geq 0$) and normalization ($\sum_i p_\rho(\omega_i) = 1$) conditions. In view of (2.4) the mapping $\rho \;\mapsto\; p_\rho$ is defined by the observable E. Since the properties of p_ρ are naturally transferred to E, an observable will appropriately be called an *effect valued measure*.

It is natural to assume that any state functional E_i preserves the convex structure of the set of states, that is, it associates with any mixture of states the corresponding convex combination of probability. This is taken as a reflection of the statistical independence of a long run of identical measurements performed on an ensemble of mutually independent systems. Effects are thus represented as linear functionals on the space of states.

[3]This assumption is reasonable and will be used throughout the thesis. See also chapter 5.2 for a justification.

2.2.2 Hilbert Space Formulation

The general statistical analysis sketched above gives a glimpse on the probabilistic structure underlying generalized measurements. In the following we shall not be concerned[4] with a rigorous mathematical treatment of the introduced effect valued measures. Instead we consider how these operationally defined objects enter quantum mechanics in terms of operators in Hilbert space.

One preliminary remark: Throughout this thesis we are solely dealing with *closed*[5] quantum systems of finite dimension and observables with discrete nondegenerate spectra. Hence, there is no need to introduce the general operator-sum formalism whereby trace and non-trace preserving quantum operations are represented. See, for example, chapter 8 of [Nie01].

Generalized Measurement Postulate

From what we have seen in subsection 2.2.1, we can phrase one aim of quantum measurement theory: Given the initial state of a system, we want to be able to specify the probability of a particular measurement result and the state of the system immediately after that measurement. Therefore, we can formulate the postulate for generalized measurements:

Postulate 2.1 (measurement). *Quantum measurements are described by a collection* $\{M_s\}$ *of* measurement operators *(often called* Kraus operators*), acting on the Hilbert space* $\mathcal{H}$ *of the system being measured. The index* s *refers to the measurement outcomes that may occur*[6] *in the experiment. If the state of the system is* $|\psi\rangle$ *(respectively* ρ *for mixed states) immediately before the measurement then the probability that result* s *occurs is given by*

$$p_s = \langle\psi|M_s^\dagger M_s|\psi\rangle \quad \textit{resp.} \quad p_s = \mathrm{tr}[M_s^\dagger M_s \rho]\,; \tag{2.6}$$

after a measurement with result s*, the system's state is*

$$|\psi'\rangle = \frac{M_s|\psi\rangle}{\sqrt{\langle\psi|M_s^\dagger M_s|\psi\rangle}} \quad \textit{resp.} \quad \rho' = \frac{M_s\rho M_s^\dagger}{\mathrm{tr}[M_s^\dagger M_s\rho]}\,.$$

The measurement operators satisfy the completeness relation

$$\sum_s M_s^\dagger M_s = \mathbb{1}\,, \tag{2.7}$$

expressing the fact that probabilities add up to one.

The novelty of this postulate (compared to the one of von Neumann) lies in the fact that *any* set of operators $\{M_s\}$ satisfying (2.7) can represent a measurement. In particular, the M_s do not have to be projectors, $M_sM_s \neq M_s$, nor do they need to be hermitian, i.e. no reference is necessarily made to any observable.

[4] The omitted theoretical parts are not significant for the understanding of this thesis, where only state changes inflicted by unsharp measurements are important. Answering fundamental questions requires more detailed mathematical background knowledge, an exhaustive account of which can be found e.g. in [Bus95] and [Bus91].

[5] closed, except for the duration of a measurement which we assume to be arbitrary short.

[6] In general, *different* Kraus operators M_{si} can be associated to one specific measurement outcome s, e.g. $p_s = \sum_i \langle\psi|M_{si}^\dagger M_{si}|\psi\rangle$ and $|\psi'\rangle = \sum_i M_{si}|\psi\rangle/p_s$. In this way noise processes are modeled which we neglect downright.

This raises the question what kind of properties these measurements represent. Often, the property a certain POVM[7] represents can be determined by making reference to some known *sharp* observable (i.e. a hermitian observable represented by a projection valued measure (PVM); e.g. energy, spin, angular momentum, etc.). Indeed many POVMs derive from some PVM by a *coarse-graining* procedure. For example, a POVM associated with the position variable Q arises if one performs a convolution of the spectral measure with some confidence function [Bus95]. We shall refer to such an unsharp observable as a *smeared* (position) observable (it then becomes possible to associate with a pair of incommensurable sharp observables a new pair of coexistent unsharp observables which are smeared versions of the original ones. Whether two such unsharp observables are coexistent or not depends on the degree of smearing involved. In the case of position and momentum it is precisely the uncertainty relation (2.2) which serves to characterize the amount of smearing required for their joint measurability).

Terminology

In order to avoid misunderstandings, let me define our terminology: *generalized* measurements denote all possible POVMs. The term 'observable' stands synonymous for a generalized observable, described by some arbitrary POVM. Unlike generalized, hermitian observables represented by some spectral measure are called *sharp* or *ordinary* observables. Smeared (coarse-grained) versions of sharp observables will be named *unsharp*. There, the spectral measure of the sharp observable has been convoluted with some confidence function; thus, from a given measurement result, no reliable conclusion regarding the initial state can be drawn. Unsharp measurements are often called *weak*, referring to the influence such a measurement has on a state compared to projective measurements. However, the unsharpness in question should in general not only be taken as an imperfect perception (like a loose pointer of some measurement apparatus) of an underlying more sharply determined property. On the contrary, this term is also intended to describe possible elements of reality whose preparation and determination are subject to inherent limitations.

Speaking of sharp observables and spectral measures, we see that projective measurements are a special case of the above postulate: by setting

$$M_s = P_s \quad \text{with} \quad P_r P_s = \delta_{rs} P_s \,,$$

the hermitian observable $M = \sum_s s P_s$ is measured in accordance with von Neumann's postulate.

POVM

The quantum measurement postulate involves two elements. First, it gives a rule describing the measurement statistics, that is, the respective probabilities of the different possible measurement outcomes. Second, it gives a rule describing the post-measurement state of the system. However, for some applications the post-measurement state is of little interest (e.g. in an experiment where the system is measured only once, upon conclusion of the experiment), with the main item of interest being measurement statistics. If we define the operator

$$E_s := M_s^\dagger M_s \,, \tag{2.8}$$

[7] we will define a POVM more closely in the next paragraph.

the probability of measuring outcome s is (for a pure state) given from postulate 2.1 by $p_s = \langle\psi|E_s|\psi\rangle$, with $|\psi\rangle$ being the initial state before the measurement. Thus, the set of operators E_s is sufficient to determine the measurement statistics and we can restate (2.4) in terms of operators:

Definition 2.1 (POVM). *A set of operators $\{E_s\}$ is named a* positive operator valued measure (POVM) *if and only if the following two conditions are met:*

(i) each operator E_s is positive $\Leftrightarrow \langle\varphi|E_s|\varphi\rangle \geq 0 \ \ \forall\varphi$

(ii) the completeness relation $\sum_s E_s = \mathbb{1}$ *is obeyed.*

The elements of $\{E_s\}$ are called *effects* or *POVM elements*. On its own, a given POVM $\{E_s\}$ is enough to have complete knowledge of the probabilities of all possible outcomes; measurement statistics is the only item of interest. If knowledge about the post-measurement state is favored, we have to use definition (2.8) to connect POVM elements E_s with measurement operators M_s.

Minimal and Non-Minimal Measurements

Interestingly, this interconnection of POVM elements and Kraus operators is not unequivocal. In analogy to the decomposition of complex numbers $z = |z|e^{i\phi}$ into modulus and phase, a similar decomposition exists for linear operators on vector spaces (for a proof, see e.g. [Nie01]):

Proposition 2.1 (polar decomposition). *Let A be a linear operator on a vector space $\mathcal{V}$. Then there exists unitary U and positive operators J and K such that*

$$A = UJ = KU\,, \tag{2.9}$$

where the unique *positive operators J and K satisfying (2.9) are defined by $J := \sqrt{A^\dagger A}$ and $K := \sqrt{AA^\dagger}$. Moreover, if A is invertible then U is unique. The expression $A = UJ$ is called the* left polar decomposition *of A, and $A = KU$ the* right polar decomposition *of A.*

Hence, every Kraus operator on $\mathcal{H}$ can be written as $M_s = U_s|M_s|$, with unitary U_s and positive operator $|M_s|$. Plugging this into (2.8) we get[8]

$$\begin{aligned} E_s &= M_s^\dagger M_s && (2.10)\\ &= |M_s|U_s^\dagger U_s|M_s| && (2.11)\\ &= |M_s|^2 \end{aligned}$$

from what follows

$$\begin{aligned} |M_s| &= \sqrt{E_s}\\ M_s &= U_s\sqrt{E_s}\,. \end{aligned} \tag{2.12}$$

From (2.10)-(2.12) we see that there is exactly one effect corresponding to each Kraus operator, whereas the reverse is not true. To every POVM element E_s corresponds an infinite number of Kraus operators; different possible unitary parts, say U_s and $\tilde{U}_s$, are canceled in (2.11).

[8] Every positive operator is hermitian, i.e. $|M_s|^\dagger = |M_s|$.

This has important consequences on the interpretation of the interplay between disturbance caused by a measurement and information gained from it. Every effect E_s just determines the positive part of each Kraus operator, $\sqrt{E_s} = |M_s|$. On the other hand we need nothing more than the effects to determine all probabilities for every possible measurement result, see (2.6). Hence, the unavoidable disturbance inherent to every measurement is completely given by the positive part of each measurement operator. At the same time, $\sqrt{E_s}$ yields all information one can get with this measurement. Any non-trivial unitary part $U_s \neq \mathbb{1}$ in M_s does not contribute extra information (In general, this unitary rotation of the measured state in $\mathcal{H}$ causes an unwanted additional disturbance. But, as we will see in chapter 4, this unitary operation can lead as well to an attenuation of disturbance caused by the positive part $\sqrt{E_s}$). We therefore define:

Definition 2.2 (minimal, non-minimal). *A measurement is called* minimal *if the respective Kraus operators have a trivial unitary part, i.e.* $U_s = \mathbb{1}$ *for all possible measurement outcomes* s*:*

$$\textit{minimal measurement} \quad \Leftrightarrow \quad M_s = |M_s| = \sqrt{E_s}\,.$$

Consequently, non-minimal *measurements have a non-trivial unitary part,* $U_s \neq \mathbb{1}$*:*

$$\textit{non-minimal measurement} \quad \Leftrightarrow \quad M_s = U_s|M_s| = U_s\sqrt{E_s}\,.$$

For measurements on ensembles (i.e. N identically prepared copies of one state ρ – not to be confused with the meaning 'ensemble' has in statistical mechanics, where it denotes an infinite number of *conceptual* replicas of one system) this distinction is of no interest because each ensemble member is measured only once so any additional state transformation has no consequence on measurement statistics whatsoever. For successive measurements on a single system, the above distinction becomes quite important. In this case, the same system is measured (and thus disturbed) again and again. Thereby, unitary parts cause additional transformations of the measured state. We will encounter this problem again in section 5.2.

2.3 Physical Realization

2.3.1 Neumark's Theorem

Talking about generalized measurements is all very well, as long as there exists an operational scheme by which an experimenter in some laboratory can *realize* such a measurement. The existence of this scheme follows from a theorem due to Neumark, which not only ensures the physical realizability of *every* possible POVM but at the same time provides the operational rules necessary to carry out the measurement.

The theorem basically states that every POVM $\{E_s\}$ on a Hilbert space $\mathcal{H}$ can be extended to a projection valued measure (PVM) $\{P_s\}$ in a larger Hilbert space $\tilde{\mathcal{H}} \supset \mathcal{H}$.

Thus, every generalized measurement on a system $\mathcal{S}$ can be realized by coupling it unitarily to a second system $\mathcal{A}$ (ancilla), thereby entangling both. By means of a projection measurement on the ancilla, the desired POVM on the system $\mathcal{S}$ is realized. The number of measurable pointer readings s is given by the dimension of the ancilla $\mathcal{A}$, which can, in principle, be arbitrary.

While it is true that any POVM can be formally reduced to a PVM acting on a larger Hilbert space, this does not diminish the need for POVMs in the description of physical systems. If one does not want to stick to an account of experiments solely in terms of pointer observables, thus

dealing with phenomena on the level of measuring devices, one has got to perform the Neumark projection: It is this step that enables one to speak of the object under investigation and its measured observable.

2.3.2 Stern-Gerlach Experiment

Up to now the representation of quantum observables as positive operator valued measures has been elucidated in theoretical respects. We shall now discuss a simplified[9] version of the famous Stern-Gerlach experiment showing that the quantity measured is, even under quite idealized conditions, a smeared version of a sharp (i.e. projective) spin measurement.

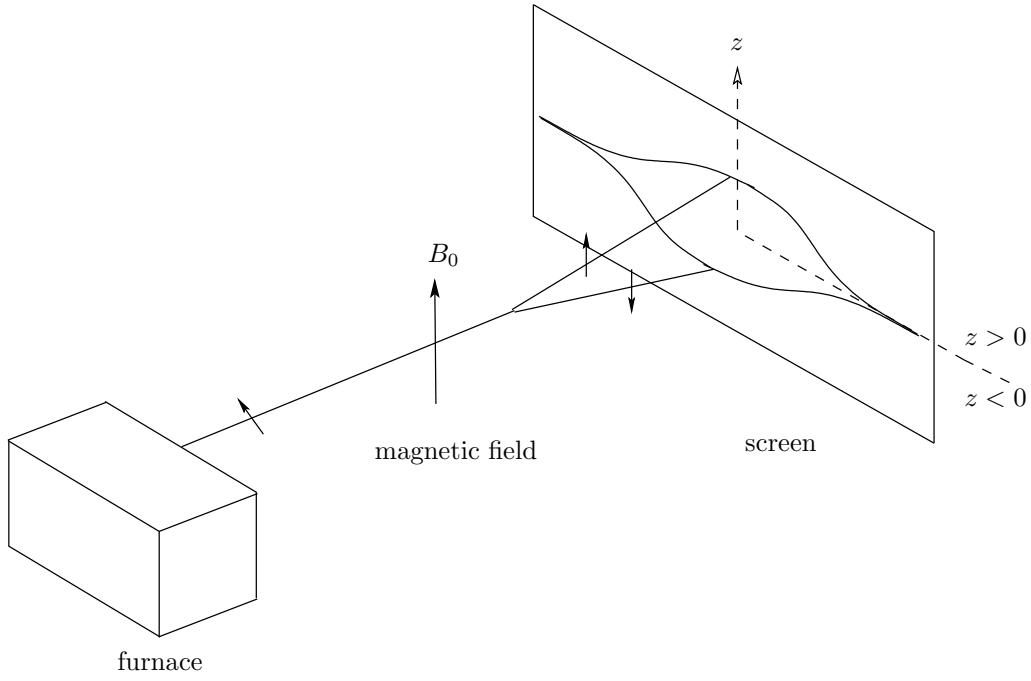


Figure 2.1: Scheme of the Stern-Gerlach experiment. Small arrows along the path of the silver atoms indicate their spin of $\frac{1}{2}$. Drawing taken from [Bus95].

A beam of neutral silver atoms, produced by a furnace, is directed through an inhomogeneous magnetic field unitarily coupling spin degrees of freedom with spatial degrees of freedom. Due to this coupling the atoms are deflected towards the upper ($z > 0$) and lower ($z < 0$) half of the screen depending on their respective spin state (cf. figure 2.1).

Let $\boldsymbol{B} = (0, 0, B_0 - bz)$ be the classical magnetic field produced by the Stern-Gerlach device pointing along the z-direction. It acts on the atom via the interaction Hamiltonian ($\hbar = 1$ in the following)

$$H = \boldsymbol{\mu} \cdot \boldsymbol{B}\,, \quad \boldsymbol{\mu} = \mu_0 \boldsymbol{\sigma}\,, \quad \mu_0 = \frac{g_s e}{2mc}$$

correlating spin and space degrees of freedom (g_s gyromagnetic ratio, e elementary charge, m mass of a silver atom, c speed of light, $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ with σ_i Pauli matrices, $\boldsymbol{\mu}$ magnetic

[9]The description given here actually renders the experiment inconsistent with Maxwell's equations. However, a realistic treatment (and one in accordance with electrodynamics) confirms the results stated here.

moment). Assume that the field strength and gradient are so strong that changes due to the 'free' evolution $H_0 = p^2/2m$ of the atom are negligible in comparison to the effect the interaction has on the particle. Suppose further that the interaction region is confined to the location of the magnetic field. The initial state of the atom entering the device is

$$|\psi_0\rangle = |\phi(z)\rangle \otimes |\varphi\rangle\,,$$

with the spatial[10] part $|\phi(z)\rangle$ fairly well localized relative to the extension of the field region. We take the spin state to be a superposition of σ_z eigenstates,

$$|\varphi\rangle = c_+|+\rangle + c_-|-\rangle\,.$$

Upon passage the initial state is transformed according to

$$\begin{aligned}|\psi_\tau\rangle &= U(|\phi(z)\rangle \otimes |\varphi\rangle)\\ &= e^{-i\tau\mu_0(B_0-bz)\sigma_z}(|\phi(z)\rangle \otimes |\varphi\rangle)\\ &= c_+|\phi_+(z)\rangle \otimes |+\rangle + c_-|\phi_-(z)\rangle \otimes |-\rangle\end{aligned}$$

where $|\phi_\pm(z)\rangle = e^{\mp i\tau\mu_0(B_0-bz)}|\phi(z)\rangle$ are the deflected wave functions. Writing these functions in the momentum representation

$$\tilde{\phi}_\pm(p) = \langle p_z|\phi_\pm(z)\rangle = e^{\mp i\tau\mu_0 B_0}\tilde{\phi}(p_z \mp \tau\mu_0 b)$$

shows that the inhomogeneous part of the magnetic field produces shifts of magnitudes $\mp\tau\mu_0 b$ in the z-component of the center-of-mass momentum of the atom. Therefore it appears as if the two components of the state separate in configuration space due to a constant force acting in the z-direction.

Now a measurement has to be made, i.e. we have to describe the registration of spots on the screen. The observable which corresponds to the measurement of the spots (called 'screen observable') shall be modeled by means of projection operators P_+ and P_- corresponding to the localization in the upper or lower half of the screen (cf. figure 2.1). The corresponding probabilities can then be expressed with respect to the incoming spin state $|\varphi\rangle$ as

$$\begin{aligned}\langle\psi_\tau|P_\pm \otimes \mathbb{1}|\psi_\tau\rangle &= |c_+|^2\langle\phi_+|P_\pm|\phi_+\rangle + |c_-|^2\langle\phi_-|P_\pm|\phi_-\rangle\\ &=: \langle\varphi|E_\pm|\varphi\rangle\end{aligned}$$

where the effects

$$E_\pm = \langle\phi_+|P_\pm|\phi_+\rangle|+\rangle\langle+| \;+\; \langle\phi_-|P_\pm|\phi_-\rangle|-\rangle\langle-| \tag{2.13}$$

constitute the unsharp spin observable actually measured in the experiment. One may immediately confirm that $E_+ + E_- = \mathbb{1}$; however, the effects (2.13) are no projections, i.e. $E_\pm^2 \neq E_\pm$, unless their eigenvalues $\langle\phi_+|P_\pm|\phi_+\rangle$ and $\langle\phi_+|P_\pm|\phi_+\rangle$ are 0 or 1. If the center of mass wave packets $|\phi_\pm\rangle$ were well separated and localized in the appropriate half planes, i.e. if $\langle\phi_\pm|P_\pm|\phi_\pm\rangle = 1$ and thus $\langle\phi_+|P_\mp|\phi_+\rangle = 0$, one would have recovered the familiar textbook description with $E_\pm$ coinciding just with the projections $|\pm\rangle\langle\pm|$. However due to the (unavoidable) spreading of wave packets this could only be achieved approximately and for special initial states $|\phi\rangle$.

[10] since there are no restrictions in the x- and y-directions the spatial part only depends on z.

A final note on more realistic descriptions of the experiment: It can be shown [Bus95] that any step towards a more realistic description of the experiment (such as realistic magnetic fields and proper screen observables, taking into account detector or screen efficiencies as well as non-instantaneous space localization measurements) will only increase the degree of unsharpness in the measured spin quantity, meaning that the measured observable $E_{\pm}$ resembles less and less the ideal projective case $P_{\pm}$.

Chapter 3

State Estimation

The question of how well the pure quantum state $|\psi\rangle$ of a physical system can be estimated is one of fundamental interest. It dates back to the early days of quantum mechanics and in particular to a handbook article by Wolfgang Pauli [Pau33]. Since the state contains all information available about a quantum system we can definitely calculate all probability distributions starting from this state. Now, the inverse question can be asked: Is it possible to use a set of probability distributions to reconstruct the quantum state in amplitude *and* phase?

It is possible if an *infinite* ensemble of identically prepared copies of a pure N-dimensional state $|\psi\rangle = \sum_i c_i|u_i\rangle = \sum_k d_k|v_k\rangle$ is at our disposal. Then, the probabilities of all outcomes

$$|c_i|^2 = |\langle u_i|\psi\rangle|^2\,, \quad |d_k|^2 = |\langle v_k|\psi\rangle|^2 \tag{3.1}$$

in two different bases $\{|u_i\rangle\}, \{|v_k\rangle\}$ can be measured with arbitrary accuracy. Equations (3.1) constitute a set of $2(N-1)$ independent (because of the two normalization constraints $\sum_i |c_i|^2 = \sum_k |d_k|^2 = 1$) experimental data needed to determine moduli and phases of $c_i = |c_i|e^{i\phi_i}$. The N moduli $|c_i|$ are given by the first set of data in (3.1); the $N-1$ relative phases (for example, $\phi_1 = 0$) then follow from

$$|d_k|^2 = \left|\sum_i \langle v_k|u_i\rangle |c_i| e^{i\phi_i}\right|^2 ,$$

a set of $N-1$ algebraic equations for the same number of unknowns (see also [Per93]).

In reality, however, only finite and usually small ensembles are available. This leads to the problem of optimal state estimation with limited physical resources. During the last couple of years, this problem attracted much interest in the context of quantum information processing, quantum cryptography and quantum computation [Nie01].

Common to all approaches is that they use the so-called *fidelity* as a figure of merit. It indicates how much on average the estimated state resembles the original (unknown) one. At the same time, it can describe the knowledge acquired about $|\psi\rangle$ through measurement. Maximizing the fidelity amounts to an optimization of the POVM used in the measurement scheme (how this optimization looks like for the qubit case will be seen in section 3.4). Before discussing fidelity in greater detail, we will first review some work that has been done in quantum state estimation theory to give some idea of the subject.

3.1 Schemes for Estimating a Pure Quantum State

Basically, two possible estimation schemes are conceivable: Either measurements on each ensemble member are made or one single measurement on the whole ensemble is carried out. Let us first discuss the latter possibility, which originated from a paper by Peres and Wootters [Per91].

There they asked the question: Is an ensemble of N identically prepared particles, viewed as an entity, more than the sum of its components? That is, could we learn more about the ensemble by performing a measurement on all constituent particles *together* than by performing separate measurements on each particle? Massar and Popescu [Mas95] described such optimal measurement procedures in the case of spin-$\frac{1}{2}$ particles. They furthermore proved the conjecture by Peres and Wootters and calculated the maximal fidelity $(N+1)/(N+2)$ obtainable with this procedure; as expected, the fidelity tends towards 1 (meaning the guess is exactly right) as N tends to infinity. These optimal measurements ('optimal' in the sense of yielding the most information about the state) are called 'non-local', referring to entangled systems exhibiting non-local Einstein-Podolsky-Rosen correlations. The operators characterizing those measurements do not factorize into components that act in the Hilbert spaces of individual particles only.

Following up on this, Derka *et al.* [Der98] presented a universal (i.e. always applicable, regardless of the physical system under study) algorithm for optimal quantum state estimation of an arbitrary finite dimensional pure system. In particular, they showed that finite[1] POVMs are sufficient for optimal state estimation. This result implied that an experimental realization of such measurements is – in principle – possible. Subsequently, Latorre *et al.* [Lat98] derived optimal POVMs to determine the pure state of a qubit with the *minimal* number of projectors when up to $N = 5$ copies of the unknown state are available. Vidal *et al.* [Vid99] and Acín *et al.* [Aci00] generalized these optimal and minimal measurements to mixed states and systems of arbitrary spin, respectively. However, the proposed strategies require the experimental implementation of rather intricate nonfactorizable operators for a simultaneous measurement on all N ensemble members. Additionally, it could be difficult to have all N quantum systems available at the same time. In short, the experimental implementation of such measurements seems, though feasible, quite involved. This suggests to consider 'local' measurements, i.e. separate measurements on each member of the ensemble.

The most general individual measurement procedures come under the name of LOCC (local operations and classical communication) schemes, as e.g. Bagan *et al.* [Bag02] pursue. In this framework, one allows a wide class of local operations for which, *depending on the outcome* of the local measurement performed on a copy, appropriate transformations can be applied on the subsequent copies of the state before measuring again. Now, any local (i.e. unitary) operation on an individual member of the ensemble may be viewed as a redefinition of the operator characterizing the measurement performed on that copy. Hence, one can equivalently change the measurement operators according to previous outcomes. Fischer *et al.* [Fis00] present such a 'self-learning' algorithm. There, a pure qubit state is estimated by a sequence of projective measurements. The algorithm is used (1) to update (via Bayes' theorem) the knowledge about the true state and (2) to choose the best projector for the next measurement. Numerical simulations show that one gets very close to the optimal upper limit (set by collective measurements) with small ensemble sizes $N \approx 40$. Hannemann *et al.* [Han02] gave an experimental realization of this algorithm using two hyperfine states of a single trapped $^{171}Yb^{+}$ ion as a qubit.

[1] Until then, the only solution to the problem of the best possible estimate of a state ρ, given by Holevo [Hol82], consisted of an *infinite* continuous set of operators.

3.2 Fidelity as Distance Measure for States

To estimate an unknown quantum state ρ, two things have to be done: (1) find an estimate ρ_s based on information gained about ρ through measurement (summarized in the index s) and (2) judge the quality of the given estimate by some measure of 'goodness'. In this section we will be concerned with this measure.

A variety of distance measures has been developed for quantum information theory (see, for example, [Fuc95] and [Nie01]), often in analogy to measures known from the classical theory. This analogy proved itself fruitful because quantum as well as classical information deal with probability distributions. A classical information source is modeled as a random variable, that is, as a probability distribution over some source alphabet (e.g. each character from an English text is seen as a random variable with its source alphabet being the Latin one). This characterization of information sources as probability distributions compelled the use of classical information-theoretic measures of distinguishability for probability distributions as the starting point for quantifying the same for quantum states.

Among those measures is the widely used Uhlmann fidelity, describing the *resemblance* between two mixed quantum states ρ and ρ_s:

$$f_s := \left(\mathrm{tr}[\sqrt{\rho^{1/2}\rho_s\rho^{1/2}}] \right)^2 . \tag{3.2}$$

f_s is not a metric ($\rho = \rho_s \Leftrightarrow f_s = 1$), but there exist measures derived from the fidelity being metrics. We shall not be concerned too much with the properties of f_s or the pros and cons to decide on one or the other distance measure (see again [Fuc95] and [Nie01] for exhaustive accounts). The reasons to choose fidelity in this work are twofold: first, it is a mathematically simple measure often used in the literature; second, it is a quite intuitive concept. To see this, we rewrite (3.2) for pure states $\rho = |\psi\rangle\langle\psi|$ and $\rho_s = |\psi_s\rangle\langle\psi_s|$: since $\rho^{1/2} = \sqrt{|\psi\rangle\langle\psi|} = \sqrt{|\psi\rangle\langle\psi|\psi\rangle\langle\psi|} = \sqrt{\rho^2} = \rho$,

$$\begin{aligned} f_s &= \left(\mathrm{tr}[\sqrt{\rho^{1/2}\rho_s\rho^{1/2}}] \right)^2 \\ &= \left(\mathrm{tr}[\sqrt{|\psi\rangle\langle\psi|\rho_s|\psi\rangle\langle\psi|}] \right)^2 \\ &= \left(\mathrm{tr}[\sqrt{|\langle\psi|\psi_s\rangle|^2|\psi\rangle\langle\psi|}] \right)^2 \\ &= \left(\mathrm{tr}[|\langle\psi|\psi_s\rangle|] \right)^2 \\ &= |\langle\psi|\psi_s\rangle|^2 , \end{aligned} \tag{3.3}$$

which is the well known overlap between two different pure states. Relation (3.3) shows, that among being symmetric in its arguments, f_s is confined to the unit interval $[0, 1]$ with $f_s = 1$ being the perfect guess ($|\psi_s\rangle = |\psi\rangle$) and $f_s = 0$ corresponding to $|\psi_s\rangle \perp |\psi\rangle$.

The fidelity defined so far depends on the outcomes of measurements and on the pre-measurement state $|\psi\rangle$. It is therefore advantageous to consider the *mean operation fidelity*, gained by averaging over all possible outcomes s of the measurement, performed on every possible pre-measurement state ($p_s = \langle\psi|E_s|\psi\rangle$ with the corresponding POVM $\{E_s\}$):

$$F = \int d\psi \; p(\psi) \sum_s p_s f_s \, . \tag{3.4}$$

Here, $d\psi$ is a measure on the space of pure states which is invariant under unitary transformations (see appendices of [Ban00] and [Sch94]) and the probability density $p(\psi)$ reflects the a priori knowledge about the pre-measurement state. The normalization of the integration measure is such that $\int d\psi p(\psi) = 1$. The term 'operation' in 'mean operation fidelity' refers to the measurement applied to the pure state $|\psi\rangle$ one wishes to estimate.

3.3 Disturbance vs. Information Gain in Quantum Operations

With the fidelity (3.4) at hand to judge the quality of an estimate for some unknown quantum state, we can answer the first part of our estimation problem: How to find an estimate if certain information about the state is available?

Usually the best possible estimate is desirable, which is tantamount to maximize the fidelity. Naturally, the more information one has the better the estimate will be. But: Every measurement is linked to an unavoidable disturbance, altering the initial state depending on the measurement's strength ("the more you get the more you wreck"). This gets important as soon as the measured (and estimated) state is needed in a subsequent task (e.g. in a quantum network), at best undisturbed. Now, the trade-off between information gain and disturbance is no longer negligible, especially if only *one* copy of the system is available (if there is an ensemble of identical copies to begin with, one copy can remain undisturbed and the trade-off becomes obsolete insofar as the untouched qubit is concerned).

Banaszek considered this quantum mechanical trade-off between information gain and state disturbance and provided an analytical description in terms of mean fidelities. His results shall be reviewed in this section, since they mark the starting point for my own line of investigation. In particular, some analytical techniques used in the derivation of the trade-off are important for my work. Therefore they shall be discussed in greater detail.

3.3.1 Fidelity Balance in Quantum Operations

In [Ban01a] the following problem is considered: Suppose we are given a single d-level particle in a completely unknown *pure* state $|\psi\rangle$. We want to make a guess about the quantum state of this particle, but at the same time we would like to alter the state as little as possible. Two fidelities can be associated with this procedure. The first one, denoted by F, describes how much the state after the operation resembles the original one. It is the mean operation fidelity introduced in the preceeding section. The second fidelity, denoted by G, characterizes the average quality of our guess. As pictured above, it is natural to expect a trade-off between these two quantities: The more information is extracted from the system, i.e. the larger G, the less the final state should resemble the initial one, hence the smaller F should be. What is the actual quantitative bound between F and G?

Two extreme cases are well known: If nothing is done to the particle we have $F = 1$, but then our guess about the state of the particle has to be random, which yields $G = 1/d$. On the other hand, the optimal estimation strategy for a *single* copy ([Bru99], [Aci00]) yields $G = 2/(d+1)$, but then the particle after the operation cannot provide any more information on the initial state; thus also $F = 2/(d+1)$. What does the constraint in between look like? Let us first derive F and G before going on to the actual trade-off.

Defining F and G

The most general strategy that can be applied to the particle has the form of a POVM $\{E_s\}$ with Kraus operators M_s, where $s = 1, \ldots, N$ labels all possible outcomes. Remember that N can (due to Neumarks theorem, cf. subsection 2.2.2) be arbitrary, that is N can be larger than d. The classical information gained from this measurement is given by the index s, which is subsequently used to estimate the initial state of the particle. Having obtained the outcome s with probability $\langle\psi|M_s^\dagger M_s|\psi\rangle$, the pre-measurement state is transformed into

$$|\psi\rangle \to \frac{M_s|\psi\rangle}{\sqrt{\langle\psi|M_s^\dagger M_s|\psi\rangle}}\,.$$

We shall measure the resemblance of the transformed state to the original one using the mean operation fidelity (3.4), that is, the squared modulus of the scalar product, averaged over all possible realizations of the experiment:

$$F = \int d\psi \sum_{s=1}^{N} |\langle\psi|M_s|\psi\rangle|^2\,. \tag{3.5}$$

Here, the uniform[2] probability distribution $p(\psi)$ has been absorbed into the canonical measure $d\psi$ over the space of pure states.

Given the outcome s of the operation, we can make a guess $|\psi_s\rangle$ what the original state was. The quality of this guess, assuming that the initial state was $|\psi\rangle$, can – in analogy to F – be quantified with the help of the overlap $|\langle\psi_s|\psi\rangle|^2$. The *mean estimation fidelity* G is thus given by averaging this expression over all outcomes s with the probability distribution $\langle\psi|M_s^\dagger M_s|\psi\rangle$, and by integrating over states $|\psi\rangle$:

$$G = \int d\psi \sum_{s=1}^{N} \langle\psi|M_s^\dagger M_s|\psi\rangle |\langle\psi_s|\psi\rangle|^2\,.$$

After evaluation of the integrals over $|\psi\rangle$ (see [Ban01a], [Ban00] and [Sch94] for details) we get

$$F = \frac{1}{d(d+1)}\left(d + \sum_{s=1}^{N} |\mathrm{tr}[M_s]|^2\right) \tag{3.6}$$

$$G = \frac{1}{d(d+1)}\left(d + \sum_{s=1}^{N} \langle\psi_s|M_s^\dagger M_s|\psi_s\rangle\right)\,. \tag{3.7}$$

Let me briefly direct attention to equation (3.7): This expression directly provides a recipe for *optimal* assignment of guesses $|\psi_s\rangle$ to outcomes of the operation: each of the components $\langle\psi_s|M_s^\dagger M_s|\psi_s\rangle$ in the sum over s is maximized if $|\psi_s\rangle$ is the eigenvector of $M_s^\dagger M_s$ corresponding to its maximum eigenvalue. This maximum-likelihood estimate for states, given a sample of size one (i.e. the measurement outcome s), will become important later on in section 6.2.

[2]Since nothing is a priori known about $|\psi\rangle$ it is reasonable to assume equipartition.

Minimal Measurements Optimize Trade-Off

Now, in order to relate the fidelities F and G to each other, it is helpful to consider the singular-value decomposition [Nie01] of the Kraus operators M_s,

$$M_s = V_s D_s W_s \,,$$

where V_s and W_s are unitary, and D_s is a semipositive definite diagonal matrix,

$$D_s = \sum_{i=0}^{d-1} \lambda_i^s |i\rangle\langle i| \,,$$

with the diagonal elements put in a decreasing order: $\lambda_0^s \geq \cdots \geq \lambda_{d-1}^s \geq 0$. We will first show that only the diagonal matrices D_s, i.e. the minimal measurement part of the corresponding Kraus operator M_s, are relevant to the trade-off (the following derivation is also valid if the polar decomposition for M_s us used, see below). Indeed, the modulus of the trace of the matrix M_s appearing in (3.6) is bounded by

$$\begin{aligned} |\mathrm{tr}[M_s]| &= \left| \sum_{i=0}^{d-1} \langle i|W_s V_s D_s|i\rangle \right| \\ &\leq \lambda_i^s \sum_{i=0}^{d-1} |\langle i|W_s V_s|i\rangle| \leq \sum_{i=0}^{d-1} \lambda_i^s \,, \end{aligned} \tag{3.8}$$

and moreover any quantum operation can easily be modified in such a way that the equality sign is reached. What needs to be done is to follow the operation M_s with an extra unitary transformation $W_s^\dagger V_s^\dagger$ depending on the outcome s. This corresponds to the modification of the Kraus operator according to $M_s \to W_s^\dagger V_s^\dagger M_s$, which makes each element of the operation a positive operator, i.e. a minimal measurement ($W_s^\dagger V_s^\dagger M_s = W_s^\dagger D_s W_s$ is a similarity transformation of D_s, preserving its positivity).

To clarify this point, let me rewrite the argument taking now the polar decomposition as – equivalent – basis for discussion. Accordingly, M_s can be written as $M_s = U_s|M_s|$. To compensate for the unitary part of M_s, known in the literature under the name of (unitary) *back-action* (see, for example [Wis95]), one only has to change the Hamiltonian evolution of the measured system with the unitary operation $U_s^{-1} = U_s^\dagger$, depending on the obtained measurement readout s. This feedback procedure avoids any additional disturbance of the state caused by the back-action term U_s. The change in Hamiltonian evolution can equally be accomplished by modifying the Kraus operators:

$$M_s = U_s|M_s| \to M_s' = U_s^\dagger M_s = |M_s| \,.$$

Completing the Trade-Off

Let us complete the derivation of the trade-off between F and G. As we are interested – given a *fixed* value of G – in the maximum value of F, we can assume with no loss of generality that

$$\begin{aligned} F &= \frac{1}{d(d+1)}(d+f) \\ G &= \frac{1}{d(d+1)}(d+g) \,, \end{aligned}$$

where f and g are given by (cf. (3.6), (3.7) and (3.8))

$$f = \sum_{s=1}^{N} \left(\sum_{i=0}^{d-1} \lambda_i^s \right)^2 , \qquad g = \sum_{s=1}^{N} (\lambda_0^s)^2 .$$

Note that f and g do not contain any unitary part, that is, we are exclusively dealing with minimal measurements. To relate f and g to each other, it is convenient to introduce vector notation which facilitates the use of vector inequalities (see Appendix A.1).

We define d real vectors $\mathbf{v}_i = (\lambda_i^1, \ldots, \lambda_i^N)$, where the index i runs from 0 to $d-1$. With this f and g can be rewritten:

$$f = \sum_{s=1}^{N} \left(\sum_{i=0}^{d-1} \lambda_i^s \right)^2 = \sum_{i,j=0}^{d-1} \sum_{s=1}^{N} \lambda_i^s \lambda_j^s = \sum_{i,j=0}^{d-1} \mathbf{v}_i \cdot \mathbf{v}_j \tag{3.9}$$

$$g = \sum_{s=1}^{N} (\lambda_0^s)^2 = |\mathbf{v}_0|^2 . \tag{3.10}$$

The trace of the completeness condition $\sum_s E_s = \mathbb{1}$ written in vector notation reads

$$\mathrm{tr}[\sum_{s=1}^{N} M_s^\dagger M_s] = \sum_{i=0}^{d-1} \sum_{s=1}^{N} (\lambda_i^s)^2 = \sum_{i=0}^{d-1} |\mathbf{v}_i|^2 = d . \tag{3.11}$$

Let us now suppose that the vector $\mathbf{v}_0$ is fixed. The mean estimation fidelity is then given by $G = (d + |\mathbf{v}_0|^2)/[d(d+1)]$. What is the maximum operation fidelity F that can be achieved with this constraint? The answer to this question is provided by applying the Schwarz inequality to (3.9):

$$f \leq \sum_{i,j=0}^{d-1} |\mathbf{v}_i| \cdot |\mathbf{v}_j| = \left(\sum_{i=0}^{d-1} |\mathbf{v}_i| \right)^2 = \left(\sqrt{g} + \sum_{i=1}^{d-1} |\mathbf{v}_i| \right)^2 . \tag{3.12}$$

In the last step we excluded from the sum over i the norm of the vector $|\mathbf{v}_0|$ which is fixed and equal to $\sqrt{g}$, cf. (3.10). The sum of the remaining vectors can be estimated using the inequality between the arithmetic and quadratic means,

$$\frac{1}{d-1} \sum_{i=1}^{d-1} |\mathbf{v}_i| \leq \sqrt{\frac{1}{d-1} \sum_{i=1}^{d-1} |\mathbf{v}_i|^2} = \sqrt{\frac{d-g}{d-1}} , \tag{3.13}$$

where we have evaluated the sum over i using (3.11). Inserting this bound into (3.12) we finally obtain the inequality

$$f \leq \left[\sqrt{g} + \sqrt{(d-1)(d-g)} \right]^2 , \tag{3.14}$$

which expressed in terms of F and G takes the form

$$\sqrt{F - \frac{1}{d+1}} \leq \sqrt{G - \frac{1}{d+1}} + \sqrt{(d-1)\left(\frac{2}{d+1} - G \right)} . \tag{3.15}$$

Example of an Optimal POVM

From the estimate (3.8) of the trace of M_s we saw that only minimal POVMs have a chance to optimize the trade-off between F and G. To be optimal, they have to make (3.14) an equality. One can therefore concentrate on either the semipositive matrices D_s or the minimal positive part $|M_s|$ and derive relations for the eigenvectors λ_i^s of the POVM. A more elegant way is to formulate the necessary and sufficient conditions leading to optimality in the vector notation.

A note on the ambiguous meaning of 'optimal': From now on we call measurements 'optimal' if they *saturate* the trade-off (3.14). There will be no confusion with the meaning 'optimal measurement' has in section 3.1 (where it denotes non-local measurements on ensembles) since we will restrict ourselves to single quantum systems or local measurements on ensembles.

The Schwarz inequality (3.12) becomes an equality if all the vectors $\mathbf{v}_0, \ldots, \mathbf{v}_{d-1}$ are collinear. Furthermore, the equality sign in eq. (3.13) holds if and only if $|\mathbf{v}_0| = \cdots = |\mathbf{v}_{d-1}|$. We can therefore choose $\mathbf{v}_0 = (\lambda_0^1, \ldots, \lambda_0^N) \in \mathbb{R}^N$ with $\lambda_0^s \geq 0$ (to ensure positivity of $\{E_s\}$), while

$$\mathbf{v}_1 = \mathbf{v}_2 = \cdots = \mathbf{v}_{d-1} = (\lambda_1^1, \ldots, \lambda_1^N)\,.$$

Given $\mathbf{v}_0 \| \mathbf{v}_1$ implies $\mathbf{v}_1 = a\mathbf{v}_0$ $(a > 0)$ or $\lambda_1^s = a\lambda_0^s$. From the trace of the completeness relation (3.11) and using g of eq. (3.10), we get:

$$\begin{aligned} d &= \sum_{i=0}^{d-1} |\mathbf{v}_i|^2 = |\mathbf{v}_0|^2 + (d-1)|\mathbf{v}_1|^2 \\ &= g\big[1 + (d-1)a^2\big] \end{aligned}$$

and hence

$$\lambda_1^s = \frac{1}{\sqrt{g}} \sqrt{\frac{d-g}{d-1}} \lambda_0^s\,.$$

With this proportionality[3] at hand, we make the ansatz $(|M_s| \equiv M_s)$

$$M_s = \lambda_0^s |s-1\rangle\langle s-1| + \lambda_1^s (\mathbb{1} - |s-1\rangle\langle s-1|) \tag{3.16}$$

where $\{|s-1\rangle\}$ denotes *any* orthonormal basis. Now $s = 1, \ldots, d$, that is, the number of measurement outcomes equals the dimensionality of the system. Choosing $\lambda_0^s = \sqrt{g/d}$, we arrive at the exemplary POVM given by Banaszek:

$$M_s = \sqrt{\frac{g}{d}} |s-1\rangle\langle s-1| + \sqrt{\frac{d-g}{d(d-1)}} (\mathbb{1} - |s-1\rangle\langle s-1|)\,. \tag{3.17}$$

The ansatz (3.16) restricts the class of POVMs to one with commuting effects. Konrad ([Kon03], p. 48) showed that POVMs with commuting effects correspond to weak measurements of sharp (hermitian) observables. It could not be shown in this thesis whether only POVMs describing smeared hermitian observables saturate the fidelity trade-off (3.14). It seems however, that the one-to-one correspondence[4] between the number of measurement outcomes N and dimensionality d of the system, characteristic for projective measurements, plays a role for optimal generalized measurements. Setting $g = d$ we recover a PVM with projectors $|s-1\rangle\langle s-1|$, $s = 1, \ldots, d$.

[3] Although $g = \sum_s (\lambda_0^s)^2$ depends on all λ_0^s, it has a *fixed* value as soon as we choose a certain $\mathbf{v}_0$.

[4] $N > d$ and $N < d$ stand for redundancy and loss of information, respectively. If $N < d$, information about several system states is conveyed with one outcome. If $N > d$, several outcomes contain (approximately) the same information.

Visualization in the FG-Plane

Before we see how optimal measurements for qubits look like, I want to visualize their position in the FG-plane. A simple transformation of eq. (3.15) shows that the region allowed by quantum mechanics for the fidelities F and G is bound by the fragment of an ellipse given by the equation

$$(F - F_0)^2 + d^2(G - G_0)^2 + 2(d-2)(F - F_0)(G - G_0) = \frac{d-1}{(d+1)^2} \tag{3.18}$$

with $F_0 = (d+2)/(2d+2)$ and $G_0 = 3/(2d+2)$. Figure 3.1 shows the region of all possible

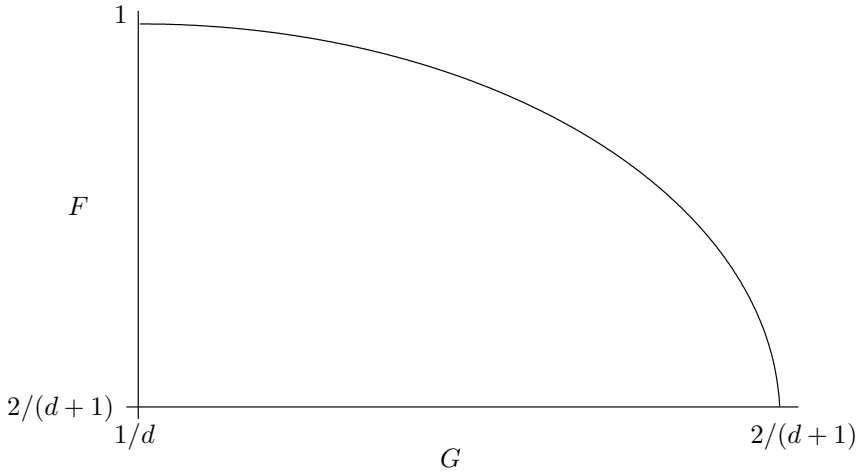


Figure 3.1: Rescaled bound for the operation fidelity F versus the estimation fidelity G, plotted for $d = 2$.

values of F and G for qubits, bound from above by (3.18). Each point corresponds to a certain POVM, with minimal Kraus operators lying on the boundary and non-minimal (that is, Kraus operators with non-trivial unitary part) POVM elements below that boundary. In particular, we can identify the two extreme cases mentioned at the beginning of our derivation. If nothing is done to the particle we are in the upper left corner, since $F = 1$ and $G = 1/d$. Projective measurements are located in the lower right corner: They disturb the particle's state maximally, $F = 2/(d+1)$, but at the same time the maximal amount of information is gained, hence $G = 2/(d+1)$.

3.3.2 Summary

To summarize the long account of [Ban01a]: We explicitly showed the derivation given by Banaszek of an analytical bound for the fidelities describing the quality of estimating the pure state of a single copy of a d-level particle, and the degree the unknown initial state is altered during this operation. No restrictions are imposed a priori on the used POVM. After averaging over all unknown states (with equal weights) it turns out that only minimal measurements have a chance to saturate the upper bound. This upper bound is subsequently obtained by applying certain inequalities to the terms whereby mean operation fidelity F and mean estimation fidelity G differ. Optimal POVMs saturating this upper bound seem to describe unsharp measurements of sharp observables.

Remarks

Banaszek *et al.* [Ban01b] generalized the fidelity trade-off (3.15) to the case of finite ensembles. With the increasing size of such an ensemble, one can extract more and more precise information about the prepared state. In the FG-plane, the optimal trade-off curve then stretches towards the upper right corner, with the limit $G = 1$ and $F = 1$ being reached as the ensemble size N goes to infinity (see figure 1 in [Ban01b]).

In relation (3.15) the pre-measurement state was the subject of inference. Equally, one can ask about the post-measurement state and give an estimate for it [Aud02a]. One finds that the mean estimation fidelity G_{post} for the post-measurement state is proportional to the mean operation fidelity G for the pre-measurement state:

$$G = \frac{1}{d+1}(1 + G_{\text{post}}) \,,$$

and thus (3.15) can be transcribed into a constraint between F and G_{post}. Furthermore, the best guesses for the pre- and post-measurement state are closely related to each other. If $|\chi^s_{\text{pre}}\rangle$ and $|\chi^s_{\text{post}}\rangle$ are the best pre- and post-measurement guesses given the outcome s, it turns out that

$$U_s|\chi^s_{\text{pre}}\rangle = |\chi^s_{\text{post}}\rangle \,.$$

In other words, the best guess for the post-measurement state is just the best guess for the pre-measurement state transported with the unitary part of the respective Kraus operator. Hence, for all minimal measurements the best estimates for the pre- and post-measurement state always agree if the ingoing state $|\psi\rangle$ is completely unknown. See [Aud02a] for details.

3.4 Optimizing Qubit Measurements

We now apply the trade-off (3.14) to a specific class of POVMs. The motivation to consider these measurements came from the desire to optimize the real-time visualization of a qubit's Rabi oscillations ([Aud01] and [Aud02c]), presented in chapter 5. The POVM used for this tracking procedure is the one we will optimize in this section. Unfortunately, it is first introduced in chapter 5. Hence, to avoid unnecessary repetition of facts, I ask the reader to scan the beginning of chapter 5 (especially subsection 5.2.1) to become familiar with these POVMs before continuing.

From now on, if not stated otherwise, $|\psi\rangle$ denotes a *pure qubit state*.

3.4.1 Constraint for a Single Measurement

To enhance the sequence of measurements (N-series) used for visualizing the evolution of $|c_1(t)|^2$ in real-time, one can optimize the individual measurement on which the sequence is based. This is a first step towards the optimization of a whole sequence, currently under investigation [Kle03]. Optimized N-series probably consist of adaptive single measurements. The sequences used so far [Aud01] only have fixed parameters. Despite this restriction, impressive results have been obtained (see, for example, figure 3 in [Aud01]). In this section, we will concentrate on the optimization of a *single* measurement because it can be treated in a complete analytical manner.

One N-series consists of N single measurements, each given by the set $\{M_s\}$ of Kraus operators ($0 \leq p_0, p_1 \leq 1$)

$$M_+ = \sqrt{p_0}|0\rangle\langle 0| + \sqrt{p_1}|1\rangle\langle 1|$$
$$M_- = \sqrt{1-p_0}|0\rangle\langle 0| + \sqrt{1-p_1}|1\rangle\langle 1|\,,$$

representing a weak measurement of some ordinary observable, e.g. energy or spin. Therefore it is of the same structure as example (3.17). Moreover, it is a non-adaptive scheme since p_0 and p_1 (equivalently $\bar{p}$ and Δp; we use these two parametrizations interchangeably) are fixed for the duration of the experiment, that is, all N-series. Optimizing this POVM is a straightforward application of result (3.14). We choose $p_1 \geq p_0$ without restriction of generality.

The optimality condition (3.14) reduces for $d = 2$ to

$$f = \left[\sqrt{g} + \sqrt{2-g}\right]^2 . \tag{3.19}$$

Calculating f and g yields

$$\begin{aligned} f &= \sum_{s=\pm} |\mathrm{tr}[M_s]|^2 \\ &= (\sqrt{p_0} + \sqrt{p_1})^2 + (\sqrt{1-p_0} + \sqrt{1-p_1})^2 \\ &= 2(1 + \sqrt{p_0 p_1} + \sqrt{(1-p_0)(1-p_1)})\,, \end{aligned}$$

and accordingly (remember that $E_s = M_s^\dagger M_s$)

$$\begin{aligned} g &= \sum_{s=\pm} \max\{\langle\psi_s|E_s|\psi_s\rangle\} \\ &= p_1 + 1 - p_0\,. \end{aligned}$$

With this, the fidelity balance (3.19) reads

$$\sqrt{p_0 p_1} + \sqrt{(1-p_0)(1-p_1)} = \sqrt{1-(p_0-p_1)^2}\,.$$

Above equation has two solutions: $p_0 = p_1$ ($\Leftrightarrow \Delta p = 0$) corresponds to the trivial POVM with $E_s \propto \mathbb{1}$ and is therefore excluded. Optimal single measurements for qubits are thus described by

$$p_0 + p_1 = 1 \quad \Leftrightarrow \quad \bar{p} = \frac{1}{2}\,. \tag{3.20}$$

Besides trivial measurements, projective measurements are optimal too. With $\bar{p} = 1/2$ fixed, Δp remains the only free parameter. What does that mean? Do these two parameters have an intuitive meaning? Let us return to the FG-plane to answer these questions.

3.4.2 Partitioning the FG-Plane

We have to parametrize F and G with $\bar{p}$ and Δp to see their working in the FG-plane. Starting from the results (3.6) and (3.7),

$$F = \frac{1}{3}\left(2 + \sqrt{\bar{p}^2 - \frac{1}{4}\Delta p^2} + \sqrt{(1-\bar{p})^2 - \frac{1}{4}\Delta p^2}\right) \tag{3.21}$$

$$G = \frac{1}{6}(3 + \Delta p)\,. \tag{3.22}$$

G in (3.22) measures the (average) resemblance of the guess with the unknown state, that is, our knowledge about it. The more knowledge we obtain about a state, the more it gets disturbed. Consequently, we can associate with Δp the parameter displaying the *strength*[5] of a measurement. No measurement (and therefore no information about the state) corresponds to $\Delta p = 0$ whereas projective measurements (and maximal information) have $\Delta p = 1$, cf. figure 3.1. Unfortunately, no such concrete meaning for $\bar{p}$ has been found.

Given Δp and $\bar{p}$ to construct any desired qubit POVM with Kraus operators $\{M_\pm\}$, where do they lie in the FG-plane? An answer to this question will illuminate the position of minimal measurements in the FG-plane and the meaning of $\bar{p}$.

To simplify our discussion, we plug (3.22) into (3.21) and obtain the following array of curves,

$$F(G,\bar{p}) = \frac{1}{3}\left(2 + \sqrt{\bar{p}^2 - 9(G-\frac{1}{2})^2} + \sqrt{(1-\bar{p})^2 - 9(G-\frac{1}{2})^2}\right), \tag{3.23}$$

mapping each *minimal* qubit POVM (characterized through Δp and $\bar{p}$) to a certain point in the FG-plane (the parameter Δp is implicitly fixed by specifying G), shown in figure 3.2. We see

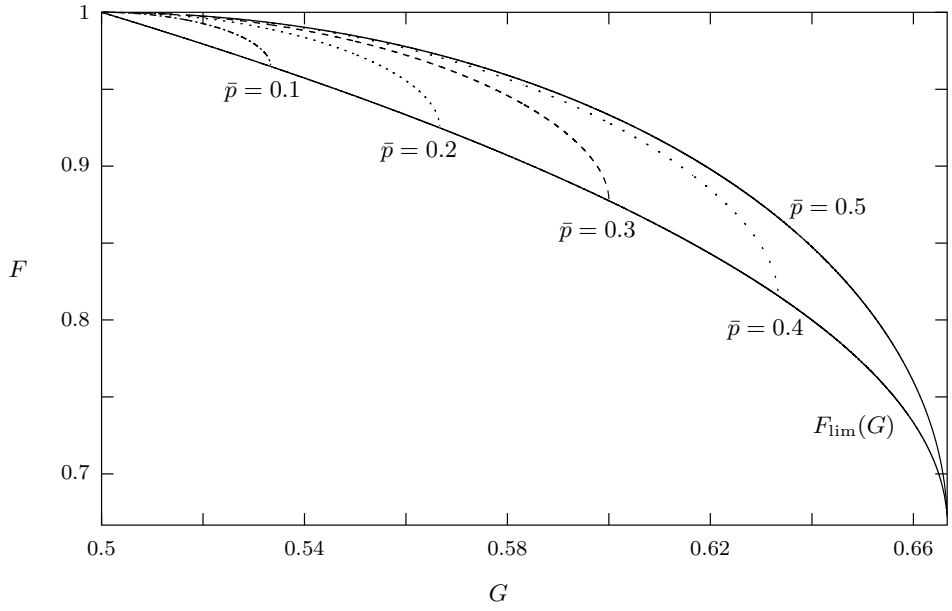


Figure 3.2: Different trade-off curves for qubit POVMs $\{E_\pm\}$. $\bar{p}$ parametrizes the different trade-off curves. $\bar{p} = 1/2$ corresponds to the optimal curve derived by Banaszek, cf. figure 3.1. $F_{\mathrm{limit}}(G)$ divides minimal from non-minimal measurements.

different trade-off curves, parametrized by $\bar{p}$. Except for $\bar{p} = 1/2$, to each curve (3.23) correspond two values of $\bar{p}$, because $F(G,\bar{p})$ is symmetric in $\bar{p}$, $F(G,\bar{p}) = F(G,1-\bar{p})$. Having decided on a certain value of $\bar{p}$ we can adjust the measurement strength with Δp thus moving along the specified curve, going from the left ($\Delta p = 0$) to the point marked by the curve $F_{\mathrm{lim}}(G)$.

[5] Another indication for the interpretation of Δp as the measurement strength comes from the level resolution time T_{lr}, see [Aud01] and [Aud02b].

Minimal vs. Non-Minimal

At this point, (3.23) is imaginary. No combination of values $0 \leq \bar{p} \leq 1$ and $0 \leq \Delta p \leq 1$ can break this 'barrier' set by $F_{\lim}(G)$. In other words, no minimal POVM $\{E_s\}$ can possibly lie below this limiting curve. Only non-minimal measurements can be found there. We thus narrowed down the region in the FG-plane where those measurements are found.

Let us specify this last statement more precisely. Setting one root term in (3.23) equal to zero and plugging the result $\bar{p} = 3(G - 1/2)$ back in yields

$$F_{\lim}(G) = \frac{1}{3}\left(2 + \sqrt{1 - 6(G - \frac{1}{2})}\right), \quad \frac{1}{2} \leq G \leq \frac{2}{3}.$$

With this finding, we can describe the distribution of minimal and non-minimal POVMs within the FG-plane, cf. figure 3.2: The absolute upper bound is set by the optimal curve $F(G, \bar{p} = 1/2)$, corresponding to all minimal POVMs fulfilling (3.20). No other minimal qubit POVM can improve that. Also minimal, but less effective, are all POVMs represented by $F(G, \bar{p})$, lying between $F(G, 1/2)$ and $F_{\lim}(G)$. The latter curve identifies the lower boundary for minimal measurements. Below this point, only non-minimal POVMs can be found.

But it is wrong to think that only *minimal* POVMs lie within the slice bordered by $F(G, 1/2)$ and $F_{\lim}(G)$.

Unitary Back-Action

Suppose we take some minimal POVM $\{E_s\}$. In the FG-plane one point corresponds to it, characterized by certain values of $\bar{p}$ and Δp or, equivalently, F and G. Now, we modify the Kraus operators $|M_s|$ by unitary operations U_s,

$$|M_s| \xrightarrow{U_s} M_s' = U_s|M_s|.$$

Formally, this back-action U_s renders the POVM non-minimal and our point within the slice should presumably move somewhere outside this region. But this is, in general, not true.

From (3.6) and (3.7) we know that

$$F = \frac{1}{d(d+1)}\left(d + \sum_{s=1}^{N} |\mathrm{tr}[M_s]|^2\right), \quad G = \frac{1}{d(d+1)}\left(d + \sum_{s=1}^{N} \langle\psi_s|E_s|\psi_s\rangle\right).$$

Now G only depends on the effects E_s and is therefore independent of U_s (cf. equations (2.10) and (2.11) on p. 10), whereas unitary parts influence the mean operation fidelity. They reduce the value of F as we saw from our derivation of the optimal trade-off, cf. estimate (3.8). Hence, our point within the slice of the FG-plane moves *vertically* downward by some amount δF, depending exclusively on the unitary back-action U_s. It now can happen that $\delta F = F - F'$ is such that the point (F', G), representing the POVM $\{E_s'\}$, lies again within the slice with some curve $F'(G, \bar{p}')$ going through it. Thus, certain non-minimal POVMs can be transformed into minimal ones by a suitable choice of $\bar{p}$, compensating for the unitary part U_s.

We saw that an appropriate choice of of $\bar{p}$ can 'compensate' certain types of unitary back-action, thus rendering it useless. Next, however, emphasis on these U_s has to be put, because they can work in favor of an enhanced mean operation fidelity.

Chapter 4

Partial Reversal of a Non-Unitary Operation by Unitary Back-Action

We showed in the last section that any non-minimal measurement cannot improve the trade-off between information gain and disturbance. The argument (3.8) is based on the assumption that nothing is known a priori about the quantum state $|\psi\rangle$, except that it is a pure state.

What happens if something is known about the state before measuring it? In particular, we are interested in the following question: *Given some a priori information about the initial pure quantum state, is it possible to improve the mean operation fidelity using non-minimal measurements?*

This question is not only of academic interest, but has practical significance in quantum computational tasks. Possessing the capability to partly reduce disturbances caused by non-unitary processes (e.g. unavoidable noise, imperfect wires, etc.) with unitary quantum gates could be a desirable feature of future quantum computers. At first sight, this seems unlikely, because of the inherent different nature of non-unitary and unitary operations. However, we will see that there exist unitary operations canceling specific non-unitary operations to some extent, thereby increasing the fidelity.

Why are we focusing on F and not on the trade-off between F and G? Certainly, to every mean operation fidelity F corresponds a mean estimation fidelity G, but we will not bother with it for two reasons: First, G depends only on effects, is therefore independent of unitary back-action terms. Second, a priori information influences the guess for $|\psi\rangle$, hence G depends on the knowledge we have about the state. However, incorporating this knowledge into our guess is a quite complicated issue we shall not be concerned with: The best guess, that is, the eigenvector of E_s to the highest eigenvalue, is only valid when nothing is known about $|\psi\rangle$ in advance; in case of a priori information, this estimation strategy does not have to be optimal.

We will deal with qubits throughout this chapter and the qubit POVM $\{E_\pm\}$ introduced in subsection 3.4.1. In particular, we choose $p_1 \geq p_0$ without restriction of generality.

4.1 A Trivial Example

Let me clarify the idea behind this cancellation procedure by a trivial example. Suppose we measure a *known* pure state $|\tilde{\psi}\rangle$ with a minimal Kraus operator $\sqrt{E_s}$, $|\tilde{\psi}_s\rangle = \sqrt{E_s}|\tilde{\psi}\rangle/\sqrt{p_s}$. Since the initial state and the operation are known, we can write down, given the orthonormal bases

$\{|\tilde{\psi}\rangle, |\tilde{\psi}^{\perp}\rangle\}$ and $\{|\tilde{\psi}_s\rangle, |\tilde{\psi}_s^{\perp}\rangle\}$, the unitary operation $U_s = |\tilde{\psi}\rangle\langle\tilde{\psi}_s| + |\tilde{\psi}^{\perp}\rangle\langle\tilde{\psi}_s^{\perp}|$ restoring the initial state. By construction, U_s is unitary. It maps the measured state and its orthogonal complement to their respective initial counterparts. Effectively, the action of a minimal measurement has been countermanded,

$$|\tilde{\psi}\rangle \xrightarrow{\sqrt{E_s}} |\tilde{\psi}_s\rangle \xrightarrow{U_s} |\tilde{\psi}\rangle\,.$$

Of course, no information has been obtained from $|\tilde{\psi}\rangle$ through measurement, since everything was known in advance. This fact makes it possible to construct such a unitary operation. If we do not know anything about $|\tilde{\psi}\rangle$ in the first place, this feat would – has to – be impossible. Otherwise information would be obtained without any disturbance. The state $|\tilde{\psi}\rangle$ could thus be cloned which is prohibited by the no-cloning theorem [Woo82].

4.2 Incorporating a Priori Information

Our task consists of two parts: (1) representing knowledge about the pre-measurement state within the fidelity picture, and (2) constructing some unitary back-action U_s given this information. In the following discussions, we will make use of the Bloch sphere. The reader not familiar with this picture is referred to Appendix B.

4.2.1 Confining the Integration Volume

From the definition (3.4) on page 17 we already know how to include a priori information into the mean operation fidelity,

$$F = \int d\psi\; p(\psi) \sum_{s=\pm} |\langle\psi|M_s|\psi\rangle|^2\,.$$

It is the probability distribution $p(\psi)$ which represents our knowledge about the pre-measurement state. If nothing is known in advance, it is reasonable to assume equipartition, that is, each state has the same probability $1/4\pi$.

Given a general probability distribution $p(\psi)$ over the surface of the Bloch sphere and a qubit POVM $\{E_s\}$, we can, in principle, calculate F. However, this will only be possible analytically for a few cases. We therefore restrict ourselves to a more special case of a priori information. We assume that the initial state lies – with equal probability – somewhere within a *well defined* region Ω, given by

$$\Omega := \{(\theta,\varphi) : \theta_1 \le \theta \le \theta_2 \wedge \varphi_1 \le \varphi \le \varphi_2 \text{ with } \theta_i \in [0,\pi], \varphi_i \in [0,2\pi]\}\,.$$

For $\theta_1 = 0$, $\theta_2 = \pi$ and $\varphi_1 = 0$, $\varphi_2 = 2\pi$ we recover the whole surface of the Bloch sphere. The 'sharp edges' of Ω are not necessarily unphysical. Take, for example, a polarizer used in some optical experiment. We want to make a statement about the polarization of the passed photon. But due to the polarizer's imperfect nature any adjusted direction is only known within some margin of error. We therefore know the polarization of the photon to be within a well defined region Ω (in general, equipartition of polarization angles within Ω is not given, but we assume it here for simplicity).

With this restriction, we denote the modified fidelity by

$$F_\Omega(M_s) := \frac{1}{N(\Omega)} \int_\Omega d\psi \sum_{s=\pm} |\langle\psi|M_s|\psi\rangle|^2\,. \tag{4.1}$$

The integration measure $d\psi$ reduces for qubits to a simple one in spherical coordinates,

$$\int_\Omega d\psi \;\rightarrow \int_{\varphi_1}^{\varphi_2} d\varphi \int_{\theta_1}^{\theta_2} d\theta \sin\theta \,.$$

Hence, normalization reads $N(\Omega) = (\varphi_2 - \varphi_1)(\cos\theta_1 - \cos\theta_2)$.

4.2.2 Constructing U_s

With Ω, a (simple) representation of our knowledge about the pre-measurement state is available. We now have to construct a unitary operation U_s. The example provided in the last subsection is a reasonable starting point for this.

In his dissertation ([Kon03], p. 63), Konrad showed that the Bloch vector of a qubit moves due to a minimal unsharp measurement on a great circle of the Bloch sphere towards the Bloch vector corresponding to the eigenvector of the Kraus operator $\sqrt{E_s}$ with the highest eigenvalue, cf. figure 4.1. With this picture the construction of U_s goes as follows: In a given region Ω, we

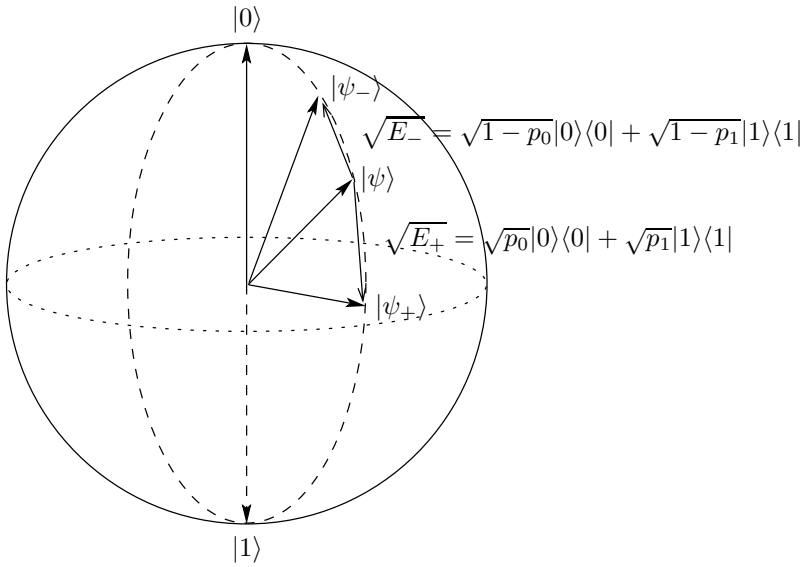


Figure 4.1: Shift of the Bloch vector of $|\psi\rangle$ due to a measurement with the POVM $\{E_\pm\}$ and results + and −. Here, $p_1 > p_0$. $|\psi\rangle$ moves on a great circle towards the eigenvector belonging to the largest eigenvalue of the relevant Kraus operator $\sqrt{E_\pm}$.

choose one specific state $|\tilde{\psi}\rangle$. A unitary operation can thus compensate the disturbance caused by $\sqrt{E_s}$. It is done by rotating the Bloch vector $\tilde{\boldsymbol{n}}$ of $|\tilde{\psi}\rangle$ about an axis lying in the xy-plane (cf. figure B.1 in Appendix B). A rotation about the z-axis is useless since $\sqrt{E_s}$ moves the Bloch vector $\tilde{\boldsymbol{n}}$ on a great circle towards one of the poles of the Bloch sphere. This *geometric* rotation on the Bloch sphere translates into a *rotation of states* in Hilbert space. Therefore U_s has the form $U_s = \exp\{\pm i/2(\alpha_s^x\sigma_x + \alpha_s^y\sigma_y)\}$. It can be further simplified if we rotate – after fixing $|\tilde{\psi}\rangle$ – the whole region Ω about the z-axis so that $\tilde{\boldsymbol{n}}$ lies in the yz-plane. This can always be done without restriction of generality. U_s thus corresponds to a rotation about the x-axis,

$$U_s = \exp\{\pm i\frac{\alpha_s}{2}\sigma_x\}\,. \tag{4.2}$$

The angle of rotation α_s is determined by

$$\cos\alpha_s := \tilde{\boldsymbol{n}} \cdot \tilde{\boldsymbol{n}}_s \,, \tag{4.3}$$

the scalar product between the *normed* Bloch vectors of $|\tilde{\psi}\rangle$ and $|\tilde{\psi}_s\rangle = \sqrt{E_s}|\tilde{\psi}\rangle/\sqrt{p_s}$. In (4.2) only $\alpha_s/2$ appears which is a leftover[1] of this translation from real space to Hilbert space: The scalar product (4.3) defines an angle between two real vectors, whereas (4.2) represents a rotation of states in Hilbert space about half that angle.

Additional care has to be taken with the sign, indicating the direction of rotation. A moment's thought shows that $U_+ = \exp\{-i\alpha_+/2\sigma_x\}$ rotates a state counterclockwise, so we can use it to compensate the action of $\sqrt{E_+}$ on $|\tilde{\psi}\rangle$ (therefore the subscript '+' in U_+). Accordingly, $U_- = \exp\{i\alpha_-/2\sigma_x\}$ compensates the action of $\sqrt{E_-}$ by rotating $\tilde{\boldsymbol{n}}_-$ clockwise.

So far, U_s compensates the action of the Kraus operator $\sqrt{E_s}$ on our chosen state. What happens to all other states within Ω? Since U_s is fixed, each state in Ω experiences the same rotation, that is, a movement along the great circle of $\tilde{\boldsymbol{n}}$, shown in figure 4.2.

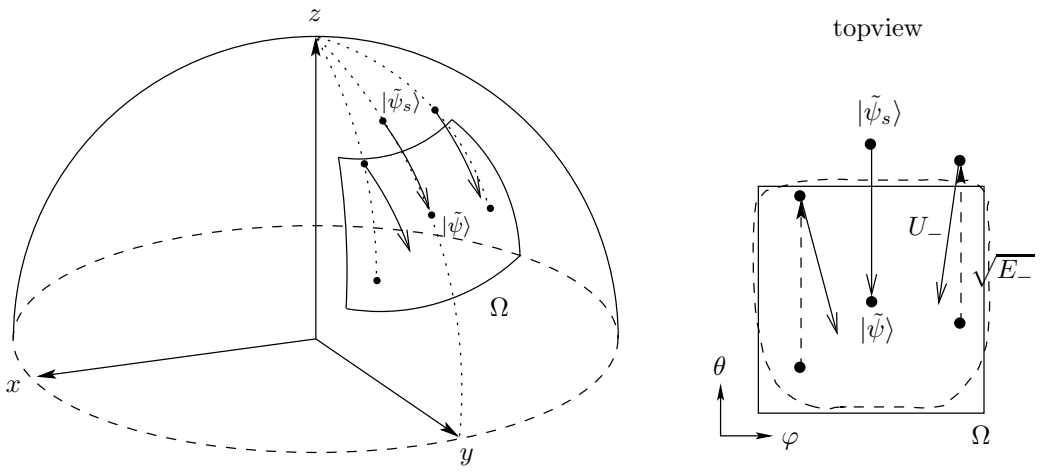


Figure 4.2: Action of U_- on $|\tilde{\psi}\rangle$ and two other states (dots) within Ω after a measurement with $\sqrt{E_-}$. U_- rotates each Bloch vector (only end points shown) along the great circle of the chosen state $\tilde{\boldsymbol{n}}$ for which U_- was constructed (left picture). This results in a deformation of Ω, due to the curved surface of the Bloch sphere, see topview on the right (deformation: dashed region).

In this picture the mean operation fidelity F_Ω is something like the mean geometric distance between measured and unmeasured points on the surface. It then becomes clear why a back-action term U_s has a chance to increase F_Ω. And why it does not work if the region extends over the whole Bloch sphere: The improvement gained, for example, in the right half of the sphere gets canceled in the left half, since the sign of U_s gets flipped and the resulting rotation moves all points in the *same* direction as $\sqrt{E_s}$ does, thus increasing the mean distance.

4.3 Partial Reversal – Preliminary Results

Having constructed U_s (4.2) for a given region Ω (after suitably rotating it so that $\tilde{\boldsymbol{n}}$ lies in the yz-plane), the fidelity can be calculated (see Appendix A.2 for the lengthy derivation). We start

[1] Actually, it is a property of the spinor group $SU(2)$, used for describing particles with spin.

discussing the results by stating them: I found that (4.1) can be written as a sum of two parts,

$$F_\Omega(U_s\sqrt{E_s}) = F_\Omega(\sqrt{E_s}) + \Delta F_\Omega\,. \tag{4.4}$$

The first term is the fidelity calculated for minimal measurements. I will call it 'minimal term' to distinguish it from the 'difference term' ΔF_Ω which comprises both operator parts, back-action U_s and minimal part $\sqrt{E_s}$. ΔF_Ω will decide whether U_s can improve on the fidelity or not, that is

$$F_\Omega(U_s\sqrt{E_s}) \stackrel{?}{\geq} F_\Omega(\sqrt{E_s})\,. \tag{4.5}$$

4.3.1 Minimal Term

The minimal term $F_\Omega(\sqrt{E_s})$ has the simple form

$$F_\Omega(\sqrt{E_s}) = \frac{1}{8}\left(3 - \frac{1}{3}f(\theta_1,\theta_2)\right) m(p_0,p_1) + \frac{1}{8}\left(5 + \frac{1}{3}f(\theta_1,\theta_2)\right)\,, \tag{4.6}$$

with $m(p_0,p_1) := \sqrt{p_0p_1} + \sqrt{(1-p_0)(1-p_1)}$ representing the minimal Kraus operators $\sqrt{E_s}$. In this parametrization, $m = 0$ stands for projective measurements, $m(p_0,p_0) = 1$ for no measurement at all ($E_s \propto \mathbb{1}$). Our knowledge Ω about the pre-measurement state is contained within the function $f(\theta_1,\theta_2)$,

$$f(\theta_1,\theta_2) := 1 + 2[\cos(2\theta_1) + \cos(2\theta_2)] + 4\cos\theta_1\cos\theta_2\,. \tag{4.7}$$

Before discussing this result further, we note that in the limit of no a priori knowledge, that is, $f(0,\pi) = 1$, we recover the result of Banaszek,

$$F_\Omega(\sqrt{E_s}) \equiv F(\sqrt{E_s}) = \frac{1}{3}m(p_0,p_1) + \frac{2}{3}\,.$$

Notice that we drop the subscript on F if the region Ω equals the whole Bloch sphere.

Interestingly, (4.6) is – by virtue of (4.7) – *independent* of any phase information. How does that come about? Remember that every minimal qubit POVM moves states along a great circle on the surface of the Bloch sphere (cf. figure 4.1). Therefore any averaging over phases, corresponding to a movement along lines of latitude, does not contribute to (4.6).

It is $-3 \leq f(\theta_1,\theta_2) \leq 9$, which implies that $F_\Omega(\sqrt{E_s})$ is not necessarily larger than $F(\sqrt{E_s})$. In other words: For minimal measurements, information about the pre-measurement state can increase as well as decrease $F_\Omega(\sqrt{E_s})$ over the value of $F(\sqrt{E_s})$. This somehow seems counter-intuitive, since information about something usually helps in deciding on different alternatives. The point is, that one is not *using* the given information for a decision.

Figure 4.3 shows an example of how different a priori knowledge can change the value of $F_\Omega(\sqrt{E_s})$ in comparison to $F(\sqrt{E_s})$. Both plots display the relation between the two fidelities for different regions $\Delta\theta = \theta_2 - \theta_1$. In the left plot $\Delta\theta$ is centered around the meridian ($\theta = \pi/2$) of the Bloch sphere. In this case, $\Delta\theta = 0$ corresponds to the meridian circle; therefore $F < 1$ since no phase information is available (one does not know the position of $|\psi\rangle$ on the circle). $\Delta\theta = \pi$ denotes the full Bloch sphere, hence $F_\Omega = F$. In the right plot $\Delta\theta$ is centered around the 'north pole', that is, the state $|0\rangle$. Now $F_\Omega = 1$ for $\Delta\theta = 0$ because the poles are point regions. $\Delta\theta = \pi$ corresponds to the upper half of the Bloch sphere. This shows that it crucially depends on the given region (a priori information) whether F_Ω or F has a higher value.

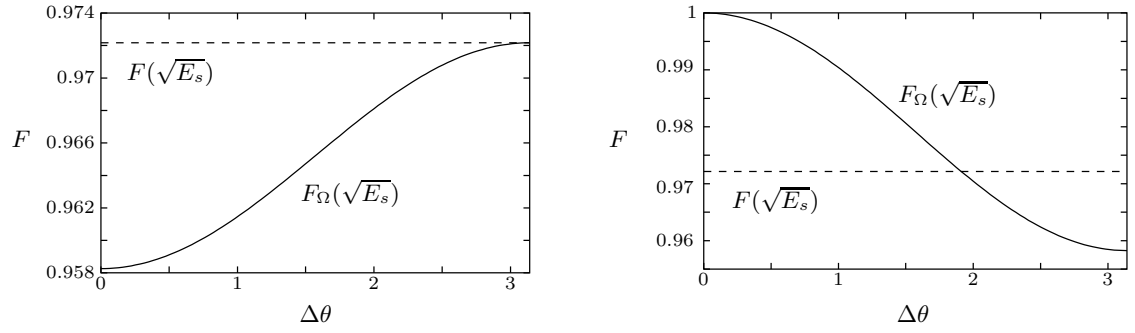


Figure 4.3: Comparison of fidelities $F_\Omega(\sqrt{E_s})$ and $F(\sqrt{E_s})$ for different a priori information $\Delta\theta = \theta_2 - \theta_1$. Left: $\Delta\theta$ is centered around the equator ($\theta = \pi/2$) of the Bloch sphere; thus $\Delta\theta = \pi$ denotes the full sphere. Right: $\Delta\theta$ is centered around the state $|0\rangle$ ($\theta = 0$). Hence $\Delta\theta = \pi$ denotes the upper half of the Bloch sphere. $p_0 = 0.3$ and $p_1 = 0.7$ were chosen for both plots.

Another interesting feature of the dependency of F_Ω on the chosen region $\Delta\theta$ comes from the zeros of $f(\theta_1, \theta_2)$. The only non-trivial[2] region for which $f(\theta_1, \theta_2) = 0$, is given by

$$\theta_2 = \theta_1 + \frac{2}{3}\pi\,, \quad 0 \le \theta_1 \le \frac{\pi}{3}\,. \tag{4.8}$$

For all $\Delta\theta$ fulfilling (4.8), the fidelity takes on the same value $F_\Omega = (3m(p_0, p_1) + 5)/8$. These '$\Delta\theta$-belts' thus describe (from the fidelity point of view) the same information about the pre-measurement state.

However, the minimal part F_Ω is only half of the story, the important part in (4.4) being the difference term ΔF_Ω.

4.3.2 Difference Term

The difference term ΔF_Ω is given by

$$\begin{aligned}\Delta F_\Omega = \frac{1}{N(\Omega)} \int_\Omega d\psi \Bigg\{ &\sum_{s=\pm} \left(|\langle\psi|\sigma_x \sqrt{E_s}|\psi\rangle|^2 - |\langle\psi|\sqrt{E_s}|\psi\rangle|^2 \right) \sin^2 \frac{\alpha_s}{2} \\ &+ \frac{1}{2}\langle\psi|\sigma_y|\psi\rangle \Big[(\sqrt{p_1} - \sqrt{p_0})\langle\psi|\sqrt{E_+}|\psi\rangle \sin\alpha_+ + (\sqrt{1-p_0} - \sqrt{1-p_1})\langle\psi|\sqrt{E_-}|\psi\rangle \sin\alpha_- \Big] \Bigg\}\,.\end{aligned} \tag{4.9}$$

To decide whether inequality (4.5) is true for some regions Ω, we need to take a closer look at the two terms in (4.9).

For the following discussion we assume that Ω is restricted to the right half of the Bloch sphere, that is, $0 \le \theta \le \pi$ and $0 \le \varphi \le \pi$ (cf. figure B.1 in Appendix B, p. 67). Furthermore, $p_1 \ge p_0$ without restriction of generality.

[2] That is, no point regions or circles. $f(\pi/2, \pi/2) = -3$ corresponds to the meridian circle, $f(0,0) = f(\pi,\pi) = 9$ correspond to points $|0\rangle$ and $|1\rangle$, respectively.

Term 2

The second term in (4.9) reads

$$\frac{1}{2}\langle\psi|\sigma_y|\psi\rangle\Big[(\sqrt{p_1}-\sqrt{p_0})\langle\psi|\sqrt{E_+}|\psi\rangle\sin\alpha_+ + (\sqrt{1-p_0}-\sqrt{1-p_1})\langle\psi|\sqrt{E_-}|\psi\rangle\sin\alpha_-\Big]\,. \quad (4.10)$$

$\langle\psi|\sigma_y|\psi\rangle =: y$ equals the y-component of the Bloch vector of $|\psi\rangle$. It is $y \geq 0$ because of our choice of Ω (if $|\psi\rangle$ lies in the xz-plane, $y = 0$). Going on to the bracket-term $[\cdot]$ we note that $\sqrt{p_1}-\sqrt{p_0} \geq 0$ and $\sqrt{1-p_0}-\sqrt{1-p_1} \geq 0$. $\sqrt{E_s}$ are positive operators, so by definition (this is independent of our choice of Ω) $\langle\psi|\sqrt{E_s}|\psi\rangle \geq 0$. The two rotation angles α_s are confined to the interval $0 \leq \alpha_s < \pi$, and thus $\sin\alpha_s \geq 0$ ($\alpha_s = 0$ would mean that $|\tilde{\psi}\rangle$ lies in one of the poles of the Bloch sphere. $\alpha_s \approx \pi$ calls for a projective measurement, transporting, for example, $|\tilde{\psi}\rangle \approx |1\rangle$ to the antipodal state $|0\rangle$).

Given our assumptions, (4.10) is greater or equal to zero. The bracket term $[\cdot]$ is hereby greater or equal to zero, independent of Ω.

Term 1

For both[3] parts of the sum

$$\sum_{s=\pm}\Big(|\langle\psi|\sigma_x\sqrt{E_s}|\psi\rangle|^2 - |\langle\psi|\sqrt{E_s}|\psi\rangle|^2\Big)\sin^2\frac{\alpha_s}{2} \quad (4.11)$$

to be positive, we need to decide whether $|\langle\psi|\sigma_x\sqrt{E_s}|\psi\rangle|^2 - |\langle\psi|\sqrt{E_s}|\psi\rangle|^2 \geq 0$ or not. To this end we rewrite the terms in the difference as scalar products which can be interpreted geometrically. Out of that interpretation will then arise physical conditions. The following argumentation is valid for either $s = +$ or $s = -$.

We start by defining the pure states

$$|\psi'\rangle := \sqrt{E_s}|\psi\rangle/\sqrt{p_s} \quad \text{and} \quad |\psi''\rangle := \sigma_x|\psi'\rangle\,,$$

having (since σ_x is unitary) norm equal to one, $\langle\psi'|\psi'\rangle = \langle\psi''|\psi''\rangle = 1$. Thus,

$$|\langle\psi|\sigma_x\sqrt{E_s}|\psi\rangle|^2 - |\langle\psi|\sqrt{E_s}|\psi\rangle|^2 = p_s\Big(|\langle\psi|\psi''\rangle|^2 - |\langle\psi|\psi'\rangle|^2\Big)\,. \quad (4.12)$$

Subsequently, we omit the factor $p_s > 0$ since it is irrelevant for deciding if (4.11) can be greater than zero for certain POVMs and regions Ω. For a geometrical representation we rewrite (4.12) using Bloch vectors: With (cf. Appendix B, p. 67)

$$\rho' = \frac{1}{2}(\mathbb{1} + \boldsymbol{n}'\cdot\boldsymbol{\sigma})\,, \quad \rho'' = \frac{1}{2}(\mathbb{1} + \boldsymbol{n}''\cdot\boldsymbol{\sigma})$$

and the identity $\mathrm{tr}[\rho\rho'] = |\langle\psi|\psi'\rangle|^2$, (4.12) reads

$$\begin{aligned}|\langle\psi|\psi''\rangle|^2 - |\langle\psi|\psi'\rangle|^2 &= \mathrm{tr}[\rho\rho''] - \mathrm{tr}[\rho\rho'] \\ &= \frac{1}{2}(\boldsymbol{n}\cdot\boldsymbol{n}'' - \boldsymbol{n}\cdot\boldsymbol{n}')\,. \end{aligned} \quad (4.13)$$

[3] To decide whether one part of the sum is larger than the other is a far more difficult question to answer.

Utilizing $\sigma_i\sigma_j = \delta_{ij}\mathbb{1} + \sum_k \epsilon_{ijk}\sigma_k$, $(k = x, y, z)$, the Bloch vectors $\boldsymbol{n}'$ and $\boldsymbol{n}''$ evaluate to

$$\boldsymbol{n}' = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} , \quad \boldsymbol{n}'' = \begin{pmatrix} x' \\ -y' \\ -z' \end{pmatrix} . \tag{4.14}$$

Because their components in the $\boldsymbol{e}_x$-direction are equal, they cancel in the difference (4.13) and we can restrict ourselves to the 2-dimensional subspace $\mathcal{V} := \mathrm{span}\{\boldsymbol{e}_y, \boldsymbol{e}_z\}$ ($\boldsymbol{e}_i$ denotes Cartesian unit vectors in $\mathbb{R}^3$). If we further define by $|\boldsymbol{n}'| := |\boldsymbol{n}'|_{\mathcal{V}}$ a 2-norm on $\mathcal{V}$, (4.13) can be written

$$\frac{1}{2}(\boldsymbol{n}\cdot\boldsymbol{n}'' - \boldsymbol{n}\cdot\boldsymbol{n}') = \frac{1}{2}|\boldsymbol{n}|\cdot|\boldsymbol{n}'|(\cos\gamma - \cos\beta) .$$

Consequently,

$$|\langle\psi|\sigma_x\sqrt{E_s}|\psi\rangle|^2 - |\langle\psi|\sqrt{E_s}|\psi\rangle|^2 \geq 0$$

reduces to the simple trigonometric relation

$$\cos\gamma - \cos\beta \geq 0 \tag{4.15}$$

we need to fulfill so that $\Delta F_\Omega \geq 0$.

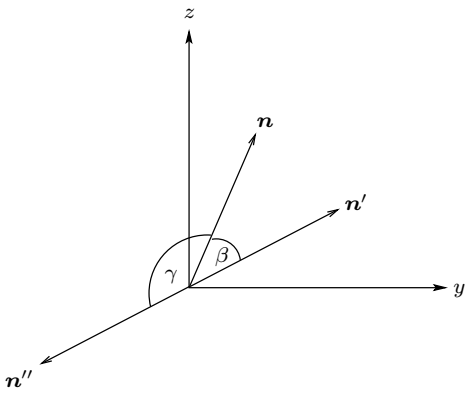


Figure 4.4: Relative positions of Bloch vectors $\boldsymbol{n}$, $\boldsymbol{n}'$ and $\boldsymbol{n}''$ in the yz-plane spanned by $\mathcal{V}$. Because of (4.14), $\gamma + \beta = \pi$ always holds. Here, $\boldsymbol{n}'$ results from $\boldsymbol{n}$ by measuring it with $\sqrt{E_+}$ (identify $(y = 0, z = 1)$ with the north pole of the Bloch sphere).

Figure 4.4 pictures the relative positions of states $|\psi\rangle$, $|\psi'\rangle$ and $|\psi''\rangle$ through their respective Bloch vectors in $\mathcal{V}$. Due to relations (4.14), $\gamma + \beta = \pi$ is satisfied for all states within Ω. Thus, to comply (4.15),

$$\frac{\pi}{2} \leq \beta < \pi . \tag{4.16}$$

That is a demand on the strength of the POVM. Remember, that $\boldsymbol{n}'$ emerges from $\boldsymbol{n}$ by measuring it. The angle by which $\boldsymbol{n}$ has to be moved is now restricted to relatively strong measurements, characterized by the constraint (4.16).

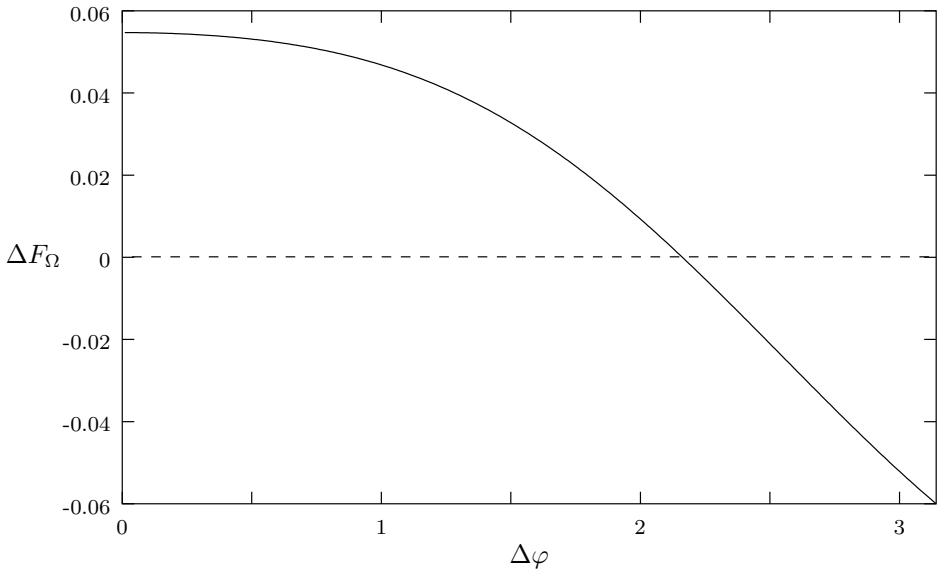


Figure 4.5: ΔF_Ω as a function of Ω, represented by $\Delta\varphi$. It is $\Delta\theta = \pi$ and $\Delta\varphi$ is symmetric with respect to the yz-plane, that is, $\varphi = \pi/2$. Thus, $\Delta\varphi = \pi/2$ corresponds to the right half of the Bloch sphere, whereas the full sphere has $\Delta\varphi = \pi$. Interestingly, $\Delta F_\Omega > 0$ for $\Delta\varphi > \pi/2$. The dashed zero line is shown for convenience. It is $\tilde{\theta} = \pi/2$, $p_0 = 0.2$ and $p_1 = 0.8$.

4.3.3 First Results

The preceeding discussion suggested that $\Delta F_\Omega \geq 0$ if Ω is restricted to the right half of the Bloch sphere, that is, $0 \leq \theta, \varphi \leq \pi$. Indeed, this conjecture is correct as figure 4.5 shows. For a fixed measurement $U_s\sqrt{E_s}$ ($\tilde{\theta}$ denotes the location of $|\tilde{\psi}\rangle$ used for constructing U_s), ΔF_Ω is plotted as a function of Ω, represented by the phase interval $\Delta\varphi$; since $\Delta\theta = \theta_2 - \theta_1 = \pi$, a characterization of Ω by $\Delta\varphi$ is sufficient. $\Delta\varphi$ is symmetric with respect to the yz-plane, that is, the great circle $\varphi = \pi/2$. As expected, $\Delta F_\Omega > 0$ if Ω lies within the right half of the Bloch sphere. Surprisingly, the positivity of ΔF_Ω extends beyond that: now, $\langle\psi|\sigma_y|\psi\rangle < 0$ and the resulting negative term (4.10) is overcompensated by (4.11), up to the point $\Delta\varphi \approx 2.1$.

We saw in figure 4.3 that knowledge about the pre-measurement state does not necessarily imply a higher fidelity. This leads us to the question: Do combinations of U_s and Ω exist so that $F_\Omega(U_s\sqrt{E_s}) > F(\sqrt{E_s})$? Figure 4.6 reveals the positive answer. The constant fidelity $F(\sqrt{E_s})$ is compared with $F_\Omega(U_s\sqrt{E_s})$ for $\tilde{\theta} = \pi/2$ and measurement $p_0 = 0.2$, $p_1 = 0.8$. Again we see that the noticeable (several percent compared to $F(\sqrt{E_s})$) increase in fidelity also holds for regions Ω extending over the right half of the Bloch sphere.

To complete this short compilation of results, let us take a closer look at how the measurement parameters influence the fidelity difference ΔF_Ω. The left plot of figure 4.7 displays the dependency of ΔF_Ω on the measurement strength Δp for optimal measurements $\bar{p} = 1/2$, $\tilde{\theta} = \pi/2$ and Ω being the right half of the Bloch sphere. ΔF_Ω increases with Δp, a result which is consistent with requirement (4.16), as stronger measurements are likely to yield $\Delta F_\Omega > 0$. Interesting is the almost quadratic progression. However, this could also be due to the special choice of parameters p_0, p_1 and $\tilde{\theta}$.

The latter is subject of the right plot in figure 4.7. Here, the dependency of ΔF_Ω on the

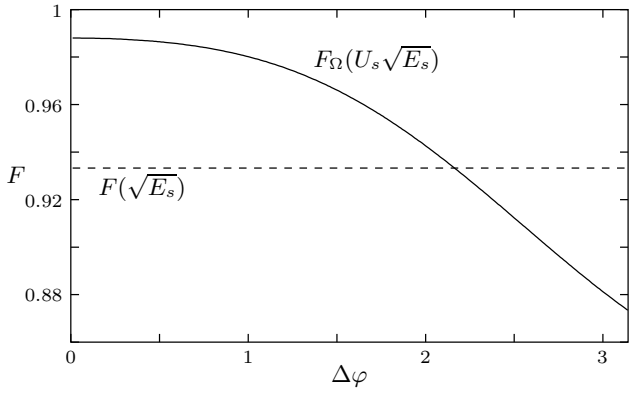


Figure 4.6: Comparison of $F_\Omega(U_s\sqrt{E_s})$ with constant $F(\sqrt{E_s})$ for different Ω, expressed by $\Delta\varphi$ since $\Delta\theta = \pi$. Notice the absolute gain in fidelity due to U_s and Ω. As in figure 4.5, $F_\Omega(U_s\sqrt{E_s})$ drops below $F(\sqrt{E_s})$ at $\Delta\varphi > \pi/2$. It is $\tilde{\theta} = \pi/2$ and $p_0 = 0.2$, $p_1 = 0.8$.

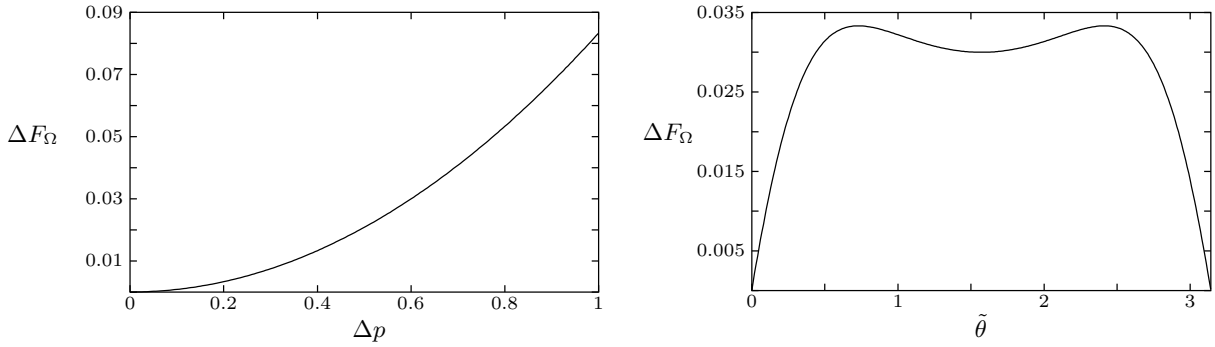


Figure 4.7: Left: Dependency of ΔF_Ω on the measurement strength for minimal measurements with $\bar{p} = 1/2$. Right: Influence of the position of $|\tilde{\psi}\rangle$ on ΔF_Ω. The peculiar two-maxima structure is probably due to a characteristic trait of the qubit POVM $\{E_\pm\}$; see also figure 2.4 on page 65 of [Kon03]. It is $p_0 = 0.2$, $p_1 = 0.8$ and Ω equals the right half of the Bloch sphere.

position $\tilde{\theta}$ of our chosen state $|\tilde{\psi}\rangle$ is investigated. The prominent feature is the two-maxima structure. This could be due to the special POVM we used in our calculations, see also figure 2.4 on page 65 of [Kon03]. The structure gets unsymmetrical if the measurements become non-optimal, that is, $\bar{p} \neq 1/2$. However, numerical examples indicate that a change of Δp and Ω do not destroy this feature.

The results found here allow for a more thorough investigation. Ideas for further work are given in the outlook, see chapter 7.

Chapter 5

Intermezzo: Real-Time Visualization of Rabi-Oscillations using Weak Measurements

As the title already indicates this chapter is intended as a transition from questions related to state estimation towards parameter estimation discussed in chapter 6. So far we discussed quantum state estimation and the optimal trade-off between information gain and disturbance. From this compromise we derived optimal measurements for qubits and investigated the influence a unitary back-action – in combination with information about the pre-measurement state – has on the fidelity.

The motivation to optimize qubit POVMs $\{E_\pm\}$ in section 3.4 stems from the desire to improve the real-time visualization of a qubit's Rabi oscillations ([Aud01], [Aud02b] and [Aud02c]). Besides measurements, the algorithm of how to find a 'best guess' for the variable used to characterize the evolution of the monitored state can be subject to optimization. This will be the content of chapter 6, where we discuss – in addition to the one described in section 5.3 – two different estimation schemes for the characteristic parameter $|c_1(t)|^2$.

We will first outline the overall scheme (section 5.1) before reviewing [Aud01] and, in detail, the key ingredients for our discussion of parameter estimation in the next chapter: namely the POVM used for tracking (section 5.2) and the algorithm for estimating $|c_1(t)|^2$ (section 5.3). Section 5.4 introduces the mean square deviation $\overline{\mathcal{D}}$ as the criterion defining the quality of a guess.

Throughout this and the following chapter we apply terminology from probability theory used in statistical inference of parameters. See also Appendix C for a compilation of definitions as well as [Lar83] and [Kre70] for thorough accounts.

5.1 Real-Time Monitoring in a Nutshell

Under the influence of a periodic driving potential, the resulting motion of a normalized qubit state (Schrödinger picture)

$$|\psi\rangle = c_0(t)|0\rangle + c_1(t)|1\rangle \tag{5.1}$$

involves oscillations of the probabilities $|c_0(t)|^2$ and $|c_1(t)|^2$, called Rabi oscillations. Usually they are measured using projective measurements. For this purpose an ensemble is prepared

in the initial state and a projection measurement is carried out at time t_0 on each ensemble member thus leading to the determination of $|c_1(t_0)|^2$. Repeating the procedure for different times result in the Rabi oscillations of $|c_1(t)|^2$.

In [Aud01] and [Aud02c] Audretsch *et al.* considered the situation where only *one* qubit is available and the objective is to visualize its Rabi dynamics in real-time. For this purpose the special class of unsharp measurements is employed.

To monitor the evolution of $|c_1(t)|^2$, the qubit is subjected to a sequence of unsharp measurements with a POVM $\{E_\pm\}$ (how unsharp a measurement has to be shall not be of concern to us here). Successive measurements are separated by a time $\tau > 0$ during which the system evolves freely. The duration of a measurement $\delta\tau$ is assumed to be much smaller than τ, thus $\delta\tau = 0$ is justified. The obtained sequence (e.g. $-++-++---+---++-\cdots$) of measurement readouts is then split into M blocks of N, called N-series. From each N-series the relative frequency $r = N_+/N$ of positive results $+$ is used to calculate a 'best guess' $g_r(t_m)$ for $|c_1(t_m)|^2$ at time t_m $(m = 1, \ldots, M)$, denoting the end of the respective N-series (for example $t_1 = N\tau$ identifies the end of the first N-series). The sequence $g_r(t_1), \ldots, g_r(t_M)$ represents the approximate visualization of the qubit's Rabi oscillations $|c_1(t)|^2$.

5.2 Single Measurement and N-Series

In classical physics it is possible to track the evolution of an individual system without disturbing it. In quantum mechanics, this can only be achieved approximately. Projective measurements are not suited for this purpose since they severely alter the system's dynamics. In fact, a sequence of projection measurements results in the well-known Zeno effect, a suppression of the *original* dynamics by measurement (the modified Zeno dynamics is almost perfectly traceable). In our case, Quantum Non-Demolition[1] (QND) measurements are no alternative either since they require for non-trivial dynamics (i.e. a Hamiltonian $H \neq \mathbb{1}$) an observable with continuous spectrum [Per89]. We are thus forced to resort to generalized measurements.

5.2.1 Motivating the Single Measurement $\{E_\pm\}$

These POVMs have to fulfill two complementary conditions:

1. The measurement must not disturb the system too much, so that its evolution remains close to the undisturbed case.

2. The coupling of the measurement apparatus to the qubit should nevertheless be so strong that the readout shows accurately enough the modified dynamics of the system (including the disturbance by the measurement).

What kind of POVM complies with these two demands? The qubit (5.1) is given in the orthonormed basis states $\{|0\rangle, |1\rangle\}$ which are eigenstates of a sharp observable A; this could be, for example, energy or spin. It is therefore advisable to consider minimal measurements which are diagonal in that basis: We specify the measurement to have only two outcomes or pointer

[1] The characteristic trait of QND measurements is represented by the fact, that the system under study is at all times in an eigenstate of the measured observable. Hence the term 'non-demolition'.

readings, denoted by + and −, with Kraus operators

$$M_+ := u_0^+|0\rangle\langle 0| + u_1^+|1\rangle\langle 1|$$
$$M_- := u_0^-|0\rangle\langle 0| + u_1^-|1\rangle\langle 1|\,.$$

We further require that $u_{0,1}^+$ and $u_{0,1}^-$ be positive. An overall phase factor $\exp\{i\varphi_\pm\}$ is omitted without loss of generality.

Why do we choose a minimal POVM which commutes with the sharp observable A of the system?

Since the measurement scheme contains consecutive measurements, the post-measurement state should not be altered qualitatively, i.e., a superposition of states $|0\rangle$ and $|1\rangle$ should, though modified, remain a superposition. We therefore require diagonality of the Kraus operators in the $\{|0\rangle, |1\rangle\}$-basis.

The minimality (and therefore the positivity of $u_{0,1}^+$ and $u_{0,1}^-$) of $M_\pm$ ensures that there is no unnecessary disturbance of the state (5.1). What does 'unnecessary' mean? Suppose we take a non-minimal Kraus operator $M_\pm$. It can be split due to the polar-decomposition theorem into a minimal part $|M_\pm|$ and a unitary back-action term $U_\pm$. The minimal part is, because of its relation to the effects $E_\pm = |M_\pm|^2$, connected to the acquisition of information. The unitary evolution $U_\pm$ represents an additional Hamiltonian evolution not leading to any increase of information about the system (see, for example, [Wis95]). It should therefore be set as $U_\pm = \mathbb{1}$ (neglecting a global phase factor).

Why $U_\pm \neq \mathbb{1}$ is indeed obstructive can best be seen in the Bloch sphere picture: For the non-minimal Kraus operators $M_\pm$ to be diagonalizable (to comply condition 1. on page 42), they have to be normal, that is, $[M_\pm^\dagger, M_\pm] = 0$, which amounts to $[|M_\pm|^2, U_\pm] = 0$. Hence, $U_\pm$ and $|M_\pm|$ have to be diagonal in the same basis. Now if $U_\pm$ is diagonal with respect to $\{|0\rangle, |1\rangle\}$ it can be written as $U_\pm = \exp\{-i\theta_\pm\sigma_z\}$ for some angles $\theta_\pm$ and σ_z as the Pauli operator. The resonant Rabi oscillations that our qubit (5.1) performs are represented by a rotation of the normalized Bloch vector in the yz-plane (given that the initial state lies in this plane). A minimal measurement induces a change of the state *within* the yz-plane, cf. figure 4.1 on page 31. $U_\pm$ on the other hand rotates the Bloch vector of $|\psi\rangle$ about the z-axis thereby *leaving* the yz-plane. As a consequence, $U_\pm$ cannot compensate the inevitable influence of $|M_\pm|$. Moreover, in a sequence of measurements, $U_\pm$ heavily disturbs the original Rabi oscillation one is trying to visualize. Therefore we have to choose $U_\pm = \mathbb{1}$ in order to disturb the Rabi oscillations by the action of $M_\pm$ as little as possible. In real experiments, any non-trivial $U_\pm$ can be compensated by means of feedback (cf. Wiseman [Wis95]).

Having motivated the POVM $\{E_\pm\}$, we introduce

$$p_0 := (u_0^+)^2 = 1 - (u_0^-)^2\,, \quad p_1 := (u_1^+)^2 = 1 - (u_1^-)^2\,,$$

and the parameters

$$\bar{p} := \frac{1}{2}(p_0 + p_1)\,, \quad \Delta p := p_1 - p_0$$

used to characterize the single measurement. Because $\{E_\pm\}$ constitutes a POVM, we have $0 \leq p_0, p_1 \leq 1$. Assuming further $p_1 \geq p_0$, $\bar{p}$ and Δp are likewise confined to the unit interval. With this parametrization at hand, the POVM writes

$$E_+ := p_0|0\rangle\langle 0| + p_1|1\rangle\langle 1|$$
$$E_- := (1-p_0)|0\rangle\langle 0| + (1-p_1)|1\rangle\langle 1|\,. \tag{5.2}$$

Accordingly, the probabilities read $p_\pm = \langle\psi|E_\pm|\psi\rangle$, with $p_+ + p_- = 1$. In the limiting case $p_0 = 0$ and $p_1 = 1$ a projection valued measure of the sharp observable A is obtained. Thus $\Delta p = 1$ corresponds to a projective measurement of A. In the opposite limit $\Delta p \ll 1$ the effects are nearly proportional to the identity operator and the probabilities $p_\pm$ of the outcomes become nearly independent of the initial state. This property justifies to call the POVM $\{E_\pm\}$ a smeared/unsharp version of the sharp hermitian observable A.

5.2.2 One N-Series

With the POVM (5.2) at hand, we now perform on the initial state $|\psi_0\rangle$ a sequence of N unsharp measurements with the *same parameters* in an *immediate succession*. This is called an 'N-series'. Using the same set of parameters (p_0, p_1) for the whole series necessarily renders the scheme non-adaptive. The phrase 'immediate succession' makes the whole sequence unphysical, since any real measurements are separated in time by $\tau > 0$. However, it turns out (cf. Appendix B of [Aud01]) that the results derived from this assumption $\tau = 0$ can be applied to the case of N-series with dynamical evolution between two single measurements if the number of measurements N in one N-series fulfills[2]

$$(N-1)^2 \ll \frac{\max\{u_0^+, u_1^-\}}{2\max\{|u_1^+ - u_0^+|, |u_1^- - u_0^-|\}} \frac{T_R}{\pi\tau} \,. \tag{5.3}$$

Here it was assumed that the matrix elements of the operation M_+ satisfy $u_0^+ > u_1^+$. Otherwise the indices 0 and 1 have to be permutated. T_R denotes the Rabi period.

In the following we assume that N is chosen according to (5.3), that is, we do not have to worry about any Hamiltonian evolution of the system between two measurements. Since the Kraus operators M_+ and M_- commute, the final state after the N-series does not depend on the order of $+$ and $-$ results. For any particular sequence of results with a total number N_+ of results $+$ the normalized state $|\psi\rangle$ is transformed according to $|\psi\rangle \to M(N_+, N)|\psi\rangle$ with

$$M(N_+, N) = M_+^{N_+} M_-^{N-N_+} \,.$$

In view of the 'best guess' defined in the next section, we restrict ourselves to the information that the total number of $+$ results is N_+, regardless of when they occurred in the sequence. That is, we are only interested in the relative frequency $r := N_+/N$ of positive results. r will be the measurement outcome attributed to one N-series. Thus, by interpreting the N-series conceptually as one single measurement, we have to work out the effects related to this measurement.

The conditional probability $p(N_+|N, |c_1|^2)$ that N_+ positive results (given N and $|c_1|^2$) are measured in one N-series is $\binom{N}{N_+}$ times the probability that a particular ordered sequence of N_+ positive and $N - N_+$ negative results is obtained:

$$p(N_+|N, |c_1|^2) = \langle\psi|E(N_+, N)|\psi\rangle \,,$$

[2]With dynamical evolution $U = \exp\{-iH\tau\}$ ($\hbar = 1$) in between two measurements, the $M_\pm$ no longer commutate with U. This residual commutator accumulates the longer one N-series gets. Hence restriction (5.3).

where the effect $E(N_+, N)$ is given by

$$\begin{aligned} E(N_+, N) &= \binom{N}{N_+} M^\dagger(N_+, N) M(N_+, N) \\ &= \binom{N}{N_+} \left[p_0^{N_+} (1-p_0)^{N-N_+} |0\rangle\langle 0| + p_1^{N_+} (1-p_1)^{N-N_+} |1\rangle\langle 1| \right] . \end{aligned} \tag{5.4}$$

The corresponding conditional probability is a linear combination of binomial distributions,

$$p(N_+ \big| N, |c_1|^2) = \binom{N}{N_+} [\alpha_0 + (\alpha_1 - \alpha_0)|c_1|^2] . \tag{5.5}$$

Here we introduced the abbreviation $\alpha_i := p_i^{N_+}(1-p_i)^{N-N_+}$ $(i = 0, 1)$ and used $|c_0|^2 = 1 - |c_1|^2$. Notice, that we only care about measurement statistics (5.5) and do not pay attention to the state after one N-series.

5.3 Best Guess for $|c_1|^2$

So far we motivated a special POVM $\{E_\pm\}$, representing a smeared version of the sharp observable A, suited for the purpose of tracking the evolution of $|c_1(t)|^2$ in real-time. To this end, sequences of unsharp measurements were defined and effects for one such N-series derived. In doing so we conceived the series conceptually as one single measurement with POVM $\{E(N_+, N)\}$ and relative frequency $r = N_+/N$ as measurement outcome.

Now, to approximately monitor the evolution of $|c_1(t)|^2$, several N-series at different times t_m $(m = 1, \ldots, M)$ are necessary, each of them yielding a best guess we are now going to motivate.

Imagine we have an *ensemble* of initial states $|\psi_0\rangle$ and we make measurements on this ensemble with the POVM (5.4). Then, the statistical expectation value of the relative frequency of positive results $r = N_+/N$ is given by

$$\begin{aligned} \langle r \rangle &:= \sum_{N_+=0}^{N} \frac{N_+}{N} p(N_+ \big| N, |c_1|^2) \\ &= p_0 |c_0|^2 + p_1 |c_1|^2 . \end{aligned} \tag{5.6}$$

Based on (5.6) it is easy to relate $|c_1|^2$ of the initial state with the expectation value $\langle r \rangle$ of an N-series of unsharp measurements starting with $|\psi_0\rangle$,

$$|c_1|^2 = \frac{\langle r \rangle - p_0}{\Delta p} . \tag{5.7}$$

The important point here is that $|c_1|^2$ and $\langle r \rangle$ refer to an ensemble represented by the state $|\psi_0\rangle$.

If only *one* realization of $|\psi_0\rangle$ is available, only one N-series can be measured and only a best guess g_r for the quantity $|c_1|^2$ can be obtained. Equation (5.7) suggests to choose[3] it as

$$g_r := \frac{r - p_0}{\Delta p} . \tag{5.8}$$

[3] In [Aud01], g_r is denoted by G_1. Here, the capital letter G could be mistaken for the mean estimation fidelity. In this thesis, every guess will be denoted by g, with a subscript referring to the method of derivation.

By virtue of (5.7), g_r is an unbiased estimator for the unknown value of $|c_1|^2$. This property represented the main motivation for the definition of g_r by equation (5.8). For $r < p_0$, the guess (5.8) becomes negative. This seems unphysical at first, but it is the unbiasedness which guarantees that g_r fluctuates (in the long run) around the true value of $|c_1|^2$.

Although (5.8) was constructed in order to meet the unbiasedness criterion, I found an alternative motivation of g_r based on ensembles (the following subsection does not contain an alternative *derivation* of g_r, because g_r was not derived in the first place but taken from an analogy).

5.3.1 Motivating g_r From Ensembles Using Maximum Likelihood

Suppose we are given an ensemble of N *identical* qubits $|\psi\rangle = c_0|0\rangle + c_1|1\rangle$. On each member of the ensemble a measurement is made with the same POVM $\{E_\pm\}$ (5.2). The aim is to calculate an estimate for $|c_1|^2$ given the sequence of N measurement results (e.g. $++--++-+--$ $---+-\cdots$). To this end, the maximum likelihood method is employed.

The maximum likelihood method for generating estimators of unknown parameters is a very powerful technique, giving – in general – 'good' estimators. It can be described as follows: Suppose we draw a sample $\{x_1, \ldots, x_N\}$ from the random variable X, whose probability function $p_X(u)$ depends on an unknown parameter u. The maximum likelihood estimator $\tilde{u}$ is then defined as the value which maximizes the so-called likelihood function $l(u) = p_X(x_1)\cdots p_X(x_N)$, the joint probability function of the sample, evaluated at the observed sample values $x_1, \ldots, x_N$.

With the POVM $\{E_\pm\}$, the probabilities to get $+$ or $-$ in a measurement are $p_+ = \langle E_+\rangle = p_0|c_0|^2 + p_1|c_1|^2 = p_0 + \Delta p|c_1|^2$ and $p_- = 1 - p_+$. The estimator $|\tilde{c}_1|^2$ maximizes the likelihood function

$$\begin{aligned} l(|c_1|^2) &= \binom{N}{N_+} p_+^{N_+}(1-p_+)^{N-N_+} \\ &= \binom{N}{N_+} (p_0 + \Delta p|c_1|^2)^{N_+}(1 - p_0 - \Delta p|c_1|^2)^{N-N_+}\,. \end{aligned} \tag{5.9}$$

Since we are not interested in the order of results $+$ and $-$, the binomial factor $\binom{N}{N_+}$ has to be included. Maximizing (5.9) requires the first derivative[4] to vanish,

$$\begin{aligned} \frac{\partial l}{\partial |c_1|^2} = \Delta p (p_0 + \Delta p|c_1|^2)^{N_+-1}(1-p_0-\Delta p|c_1|^2)^{N-N_+-1} \\ \cdot \left[N_+(1-p_0-\Delta p|c_1|^2) - (N-N_+)(p_0+\Delta p|c_1|^2)\right] \stackrel{!}{=} 0\,. \end{aligned} \tag{5.10}$$

This equation has only one physically reasonable solution: $\Delta p = 0$ can be discarded at once, just as $|\tilde{c}_1|^2 = -p_0/\Delta p$ and $|\tilde{c}_1|^2 = (1-p_0)/\Delta p$, derived from setting the second and third product term in (5.10) equal to zero. Both estimators solely depend on the parameters characterizing the POVM. Hence they are constant. In addition, $-p_0/\Delta p < 0$ renders this estimator unphysical. We are thus left with the fourth term in (5.10), which gives

$$|\tilde{c}_1|^2_{\mathrm{ML}} = \frac{N_+ - Np_0}{N\Delta p} = \frac{r - p_0}{\Delta p} = g_r\,. \tag{5.11}$$

[4] In general, $l(|c_1|^2, p_0, \Delta p, N_+, N)$ is a function of several variables. In this special case, everything but $|c_1|^2$ is fixed and we could have replaced the partial derivative by $d/d|c_1|^2$.

This peculiar result merits further discussion. Originally, g_r was constructed as an unbiased estimator, starting from the results of one N-series. That is, a sequence of measurements is conducted on *one and only one* state. The derivation of (5.11) started from an ensemble, that is, N separate measurements are carried out on an *ensemble of identical* states. Why do both methods yield the same estimator?

Obviously, these two conceptually different procedures agree on the level of estimators. However, in the ensemble approach, independency of measurements is guaranteed in contrast to the N-series. So both methods are not only conceptually but physically different. This fact is expressed in the mean square deviation $\overline{\mathcal{D}}$ (see section 5.4 for the definition):

$$\lim_{N\to\infty} \overline{\mathcal{D}}(g_r) = \overline{\mathcal{D}}_{\text{projective}} \tag{5.12}$$

$$\lim_{N\to\infty} \overline{\mathcal{D}}(|\tilde{c}_1|^2_{\text{ML}}) = 0\,. \tag{5.13}$$

In case of an infinite ensemble, an infinite amount of information is available and $\overline{\mathcal{D}}(|\tilde{c}_1|^2_{\text{ML}})$ has to go to zero ($|c_1|^2$ is known exactly). For the N-series, infinite N means infinitely many unsharp measurements. That is equivalent ([Kon03], p. 58) to one projective measurement. Hence, the lower limit $\overline{\mathcal{D}}_{\text{projective}}$, derived in the forthcoming section.

To sum up: Although two physically different measurement schemes yield the same guess for $|c_1|^2$, their dissimilarity shows up on the level of mean square deviations.

5.4 Mean Square Deviation as Measure of Quality

Experimental physicists often use the standard deviation in their statistical analysis of measured data. This stochastic concept quantifies the variation of a measured quantity X by the average quadratic distance of all obtained values from the *measured average* $\langle X\rangle$,

$$\sigma^2(X) = \langle (X - \langle X\rangle)^2\rangle\,.$$

The average $\langle\cdot\rangle$ is hereby taken with respect to the probability distribution p_X underlying X.

In the real-time visualization of Rabi oscillations, we are interested in an estimate g preferably close to the actual value of $|c_1|^2$ for which a measured average no longer exists. However, we can define – in analogy to σ^2 – the mean square deviation $\mathcal{D}$,

$$\begin{aligned}\mathcal{D}(g) &:= \langle (g - |c_1|^2)^2\rangle \\ &= \sum_{N_+=0}^{N} p(N_+\big|N, |c_1|^2)(g - |c_1|^2)^2\,.\end{aligned}$$

Although this measure is well defined it still depends on $|c_1|^2$, which is – in general – unknown to us. If, however, some information about $|c_1|^2$ is available, $\mathcal{D}$ indicates quantitative differences between estimators.

At large, $\mathcal{D}$ leads us to a strange situation – to evaluate our guess about $|c_1|^2$, we need to know $|c_1|^2$! A way out of this dilemma is not to look at the mean square deviation for a particular $|c_1|^2$ but at the deviation averaged additionally over all possible $|c_1|^2$ with the probability distribution $p(|c_1|^2)$. For reasons to become clear in section 6.3, the 'prior' $p(|c_1|^2)$ expresses our a priori knowledge about $|c_1|^2$. Since $0 \le |c_1|^2 \le 1$ is the only information we have, it is reasonable to

assume equipartition in the unit interval, hence $p(|c_1|^2) = 1/(1-0)$. With this, $\mathcal{D}$ averaged over all states[5] $|c_1|^2$ reads

$$\begin{aligned}\overline{\mathcal{D}}(g) &:= \int_0^1 d|c_1|^2\, \mathcal{D}(g) \\ &= \int_0^1 d|c_1|^2 \sum_{N_+=0}^{N} p(N_+|N, |c_1|^2)(g - |c_1|^2)^2 \,. \end{aligned} \tag{5.14}$$

We will use the term *mean square deviation* synonymously for $\mathcal{D}$ and $\overline{\mathcal{D}}$. In case of ambiguities, the respective definition is given.

5.4.1 Calculating $\overline{\mathcal{D}}(g_r)$

We start by calculating $\mathcal{D}(g_r)$:

$$\begin{aligned}\mathcal{D}(g_r) &= \langle (g_r - |c_1|^2)^2 \rangle \\ &= \left[\left(\frac{p_1}{\Delta p} + |c_1|^2 \right) - \frac{\langle r \rangle}{\Delta p} \right]^2 + \frac{1}{\Delta p^2} \left[\langle r^2 \rangle - \langle r \rangle^2 \right] \\ &= \left[\langle g_r \rangle - |c_1|^2 \right]^2 + \frac{\sigma^2(r)}{\Delta p^2} \,. \end{aligned}$$

Since g_r is unbiased, the first bracket vanishes and $\mathcal{D}(g_r)$ becomes proportional to the variance of $r = N_+/N$. Calculating this variance yields

$$\mathcal{D}(g_r) = |c_0|^2 |c_1|^2 + \frac{1}{N} \frac{|c_0|^2 p_0(1-p_0) + |c_1|^2 p_1(1-p_1)}{\Delta p^2} \,.$$

In case of a single measurement, $N = 1$, $\mathcal{D}(g_r)$ assumes its minimal value $|c_0|^2|c_1|^2$ for projective measurements, $\Delta p = 1$. For all other choices of parameters N, p_0 and p_1, there is a larger deviation. This again shows that these measurements are unsharp, because they provide less reliable information about $|c_1|^2$. For a given experimental setup, i.e., fixed Kraus operators, $\mathcal{D}(g_r)$ decreases with an increasing number of repetitions N so that it is favorable to refer to the N-series.

Using the parametrization $\bar{p} = (p_0 + p_1)/2$ and $\Delta p = p_1 - p_0$, $\overline{\mathcal{D}}(g_r)$ is given by

$$\overline{\mathcal{D}}(g_r) = \frac{1}{6} + \frac{1}{N\Delta p^2} \left(\bar{p}(1-\bar{p}) - \frac{\Delta p^2}{4} \right) \,. \tag{5.15}$$

Proving (5.12) and (5.13)

We are now able to prove the qualitative difference between the two estimators g_r and $|\tilde{c}_1|^2_{\mathrm{ML}}$, expressed by their limiting behavior for $N \to \infty$, (5.12) and (5.13). From (5.15) we see that

$$\lim_{N\to\infty} \overline{\mathcal{D}}(g_r) = \frac{1}{6} =: \overline{\mathcal{D}}_{\text{projective}} \,.$$

[5] This equivalence $\int d|c_1|^2 \equiv \int d\psi$ is only valid if one assumes equal distributions for azimuth and polar angles.

To calculate $\overline{\mathcal{D}}$ for the maximum likelihood estimate $|\tilde{c}_1|^2_{\mathrm{ML}}$, we have to use $l(|c_1|^2)$ as probability distribution instead of $p(N_+\big|N, |c_1|^2)$:

$$\begin{aligned}\overline{\mathcal{D}}(|\tilde{c}_1|^2_{\mathrm{ML}}) &= \langle(|\tilde{c}_1|^2_{\mathrm{ML}} - |c_1|^2)^2\rangle \\ &= \int_0^1 d|c_1|^2 \sum_{N_+=0}^{N} l(|c_1|^2)(|\tilde{c}_1|^4_{\mathrm{ML}} - 2|\tilde{c}_1|^2_{\mathrm{ML}}|c_1|^2 + |c_1|^4) \\ &= \int_0^1 d|c_1|^2 \, \frac{p_+(1-p_+)}{N\Delta p^2} \,.\end{aligned}$$

The integral as well as the integrand are bound for non-trivial measurements (i.e. $\Delta p \neq 0$), so it suffices to look at the latter:

$$\lim_{N\to\infty} \overline{\mathcal{D}}(|\tilde{c}_1|^2_{\mathrm{ML}}) = 0 \,.$$

Chapter 6

Parameter Estimation

Chapter 3 dealt with the estimation of pure quantum states and the corresponding optimal measurements. In particular, POVMs describing unsharp observables were of interest, since they are used in the real-time visualization of Rabi dynamics, reviewed in chapter 5. There, a qubit's state evolution $|\psi(t)\rangle = c_0(t)|0\rangle + c_1(t)|1\rangle$ is *indirectly* monitored by $|c_1(t)|^2$, respectively the sequence $g_r(t_m)$, $m = 1, \ldots, M$.

Motivated by this work [Aud01], the question arose whether there exist better ways for processing information obtained from one N-series into a guess for $|c_1|^2$. Sections 6.2 and 6.3 deal with this question. As laid down with the fidelity F in chapter 3, a measure is needed defining the quality of a guess. This will be the purely classical statistical measure $\overline{\mathcal{D}}$ already introduced in section 5.4. Although our system under study (qubit performing Rabi oscillations) is quantum mechanical, processing measured data only requires classical probability theory. The focus is therefore shifted from questions of state inference and optimization of measurements, to questions of parameter estimation. A comparison of all developed guesses in section 6.4 concludes this chapter.

6.1 Stating the Problem

For a given POVM $\{E(N_+, N)\}$ (5.4) the sequence of results S (e.g. $S = \{++--+-++-\cdots\}$) of one N-series is governed by the probability distribution (5.5),

$$p(N_+|N, |c_1|^2) = \binom{N}{N_+}[\alpha_0 + (\alpha_1 - \alpha_0)|c_1|^2]\,. \tag{6.1}$$

It gives the probability that N_+ times out of N the result reads $+$, regardless of its occurrence within the sequence; remember, that $\alpha_i = p_i^{N_+}(1-p_i)^{N-N_+}$ $(i = 0, 1)$ is assumed to be constant during the N-series. In addition, $p_1 > p_0$ is chosen without loss of generality.

In writing down (6.1) we conceived the whole N-series as one single measurement with the number N_+ as measurement outcome. Unfortunately, we are not able to calculate $p(N_+|N, |c_1|^2)$ unless the value of $|c_1|^2$ is known. So far, the only information available is contained in the measured sample. Thus, our problem is to find an estimate g for $|c_1|^2$, given the information of the sample S. For this kind of estimation problem, probability theory provides a set of tools; see, for example, chapter 7 of [Lar83] and chapter 12 of [Kre70].

Notice that we are entirely dealing with classical probability theory. The fact that (6.1) derives from a quantum mechanical process is of no interest here.

6.2 Revisiting Banaszek: Maximum Likelihood Estimator

Our first approach (besides the estimate g_r presented in section 5.3) to the aforementioned estimation problem employs the maximum likelihood method as a tool. Banaszek [Ban01a] already used it implicitly for samples of size one in his optimal assignment of guesses for $|\psi\rangle$, cf. subsection 3.3.1 on page 18. To see this, we restate the argument: Maximization of the mean estimation fidelity

$$G = \frac{1}{d(d+1)}\left(d + \sum_s \langle\psi_s|E_s|\psi_s\rangle\right)$$

yields as best guess for $|\psi\rangle$ the eigenvector of E_s to its highest eigenvalue. In effect, the probability distribution $\langle\psi_s|E_s|\psi_s\rangle$ with the unknown parameter 'ψ_s' is maximized for the sample consisting of the measurement result s. For example, taking $E_+ = p_0|0\rangle\langle 0| + p_1|1\rangle\langle 1|$ as effect ($p_1 > p_0$) and $|\psi_+\rangle = c_0^+|0\rangle + c_1^+|1\rangle$ for the guess, we get

$$\langle\psi_+|E_+|\psi_+\rangle = p_0 + \Delta p|c_1^+|^2 =: p(+\big||c_1^+|^2)\,.$$

$p(+\big||c_1^+|^2)$ is a straight line, with its maximum $p(+\big|1) = p_1$ at the boundary. So the best guess for $|\psi\rangle$ is $|c_1^+|^2 = 1 \Leftrightarrow |\psi_+\rangle = |1\rangle$. With the same line of reasoning, $|\psi_-\rangle = |0\rangle$. Applying this result to the problem of estimating $|c_1|^2$ given the sample S is now straightforward.

6.2.1 Estimator for N-Series

In case of an N-series the sample of size one consists of the number N_+ and the probability $p(N_+\big|N, |c_1|^2)$ itself is the likelihood function we need to maximize. Since

$$\frac{\partial}{\partial|c_1|^2}p(N_+\big|N, |c_1|^2) = \binom{N}{N_+}(\alpha_1 - \alpha_0) \neq 0\,,$$

the global maximum of $p(N_+\big|N, |c_1|^2)$ again lies at the boundary. Looking at (6.1) we see that $|c_1|^2 = 1$ maximizes $p(N_+\big|N, |c_1|^2)$ if $\alpha_1 > \alpha_0$. Otherwise, $|c_1|^2 = 0$ is the best choice. Thus, we found a maximum likelihood estimate g_{ML} for $|c_1|^2$, given the sample information S:

$$g_{\text{ML}} := \begin{cases} 1 & \text{if } \alpha_1 > \alpha_0\,, \\ 0 & \text{if } \alpha_1 < \alpha_0\,. \end{cases} \tag{6.2}$$

The condition $\alpha_1 = \alpha_0$ implies $E(N_+, N) \propto \mathbb{1}$, that is, no measurement at all.

At first sight, (6.2) does not look like a very accurate guess for a variable which varies continuously within the unit interval. We will see, however, that when compared to other estimators via $\overline{\mathcal{D}}$, it is favorable to refer to g_{ML} rather than to g_r.

6.2.2 Mean Square Deviation $\overline{\mathcal{D}}(g_{\text{ML}})$

From the definition for $\mathcal{D}$,

$$\begin{aligned}\mathcal{D}(g_{\text{ML}}) &= \langle(g_{\text{ML}} - |c_1|^2)^2\rangle \\ &= \langle g_{\text{ML}}^2\rangle - 2\langle g_{\text{ML}}\rangle|c_1|^2 + |c_1|^4\,.\end{aligned}$$

Since the average $\langle g_{\rm ML}\rangle$ is taken over all possible measurement outcomes N_+, we need to include the maximum likelihood estimate for every sample N_+. Therefore,

$$\langle g_{\rm ML}\rangle = \sum_{N_+\in I} 1\cdot p(N_+\big|N,|c_1|^2) + \sum_{N_+\in J} 0\cdot p(N_+\big|N,|c_1|^2) = \langle g_{\rm ML}^2\rangle\,, \tag{6.3}$$

with the index sets $I := \{N_+ : \alpha_1 > \alpha_0\}$ and $J := \{N_+ : \alpha_1 < \alpha_0\}$ reflecting the estimation algorithm: The best guess $g_{\rm ML} = 1$ maximizes the corresponding probability; for example, if there is only one measurement, $N = 1$, (6.3) simplifies to $\langle g_{\rm ML}\rangle = 1\cdot p_+ + 0\cdot p_-$.

With $\langle g_{\rm ML}\rangle$ given, $\overline{\mathcal{D}}(g_{\rm ML})$ reads

$$\overline{\mathcal{D}}(g_{\rm ML}) = \int_0^1 d|c_1|^2\left[(1-2|c_1|^2)\langle g_{\rm ML}\rangle + |c_1|^4\right]\,. \tag{6.4}$$

Because the summation in (6.3) is performed over the index set I, $\overline{\mathcal{D}}(g_{\rm ML})$ cannot be evaluated in closed form for arbitrary p_0 and p_1. There exists, however, a closed solution for the special case $\bar p = 1/2$: The condition $\alpha_1 > \alpha_0$ transforms into

$$N_+ > \frac{N}{1+f(p_0,p_1)}$$

with $f(p_0,p_1) := [\ln p_1 - \ln p_0]/[\ln(1-p_0) - \ln(1-p_1)]$. If $\bar p = 1/2$ then $f(p_0, 1-p_0) = 1$ and the average (6.3) reads[1]

$$\begin{aligned}\langle g_{\rm ML}\rangle &= \sum_{N_+=[N/2]+1}^{N} p(N_+\big|N,|c_1|^2)\\ &= \frac{2^{-N}N!(1-\Delta p^2)^{-[\frac{N}{2}]}}{(\Delta p^2-1)\Gamma(N-[\frac{N}{2}])\Gamma(2+[\frac{N}{2}])}\\ &\quad\cdot\Big\{(|c_1|^2-1)(1-\Delta p)^{2+2[\frac{N}{2}]}(1+\Delta p)^N\,{}_2\mathcal{F}_1(1,1-N+[\frac{N}{2}],2+[\frac{N}{2}],\frac{\Delta p-1}{\Delta p+1})\\ &\qquad -|c_1|^2(1+\Delta p)^{2+2[\frac{N}{2}]}(1-\Delta p)^N\,{}_2\mathcal{F}_1(1,1-N+[\frac{N}{2}],2+[\frac{N}{2}],\frac{\Delta p+1}{\Delta p-1})\Big\}\,.\end{aligned} \tag{6.5}$$

The functions ${}_2\mathcal{F}_1(a,b,c,z)$ are solutions of the hypergeometric differential equation $z(1-z)y'' + [c-(a+b+1)z]y' - aby = 0$, given in their integral form by

$${}_2\mathcal{F}_1(a,b,c,z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 dt\, t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\,.$$

Although analytical, (6.5) is quite clumsy and useless for general calculations. For a comparison of different estimators, we will always refer to special values of N, for which $\overline{\mathcal{D}}(g_{\rm ML})$ is easily calculated.

[1]The Gauß-bracket $[N/2]$ denotes the greatest integer smaller or equal to $N/2$; for example, $[15/2] = 7$. $\Gamma(c)$ is the usual gamma function.

6.3 Incorporating a Priori Information: Bayes' Theorem

Both estimation procedures presented so far only use information contained in the measured sample S to calculate the respective estimators g_r and g_{ML}. These approaches ignore information about $|c_1|^2$, eventually existing before any measurement is performed. Including this additional prior information about the unknown parameter yields a new estimator we are going to derive in this section. The scheme is based on Bayes' theorem for conditional probabilities and consists of three steps:

1. Conversion of the prior information about $|c_1|^2$ into a probability distribution.
2. Calculation of the a posteriori distribution on the basis of the sample information S and the prior distribution via Bayes' theorem.
3. Calculation of the estimator from the a posteriori distribution.

The next two subsections illustrate this method in general, before we apply it to the problem of one N-series.

6.3.1 Prior Information and a Posteriori Distribution

Suppose we are given a random variable X whose probability distribution p depends on an unknown parameter θ. 'Unknown' means that we know the functional form of p but not the concrete value of the parameter θ. Take, for example, a Poisson distribution $p = \exp(-\mu)\mu^x/x!$. Its functional form (Poissonian) is known but the mean μ is unknown to us and shall be estimated with the help of an experiment.

We denote the distribution p as a conditional probability,

$$p(x\big|\theta)\,,$$

to express the fact that it can only be calculated if the value of θ is known. In applying this notation we interpret θ as value of a random variable Θ. Associated with it is a certain probability distribution $p(\theta)$, containing our knowledge about θ. For the moment we have to assume that the *prior* $p(\theta)$ is given to us. Converting some given information about θ into a prior distribution (point one on the above list) depends on the specific problem. We come back to this in subsection 6.3.3.

With the distribution $p(x\big|\theta)$ and the prior $p(\theta)$ known, an experiment is performed, providing additional information about θ in the form of a sample x^*. This knowledge is used to update the probability distribution $p(\theta)$ for θ via Bayes' theorem,

$$p(\theta) \to p(\theta\big|x^*) = p(\theta)\frac{p(x^*\big|\theta)}{p(x^*)}\,. \tag{6.6}$$

$p(x^*) = \int d\theta p(x^* \cap \theta)$ denotes the *marginal*. (6.6) is a 'dynamical' law for probabilities: It relates the prior distribution $p(\theta)$, together with the obtained sample information $p(x^*\big|\theta)$, to the *a posteriori* distribution $p(\theta\big|x^*)$, reflecting our improved knowledge about θ.

With the posterior distribution $p(\theta\big|x^*)$ now at hand, we can tackle the third problem on our list: constructing a 'best' estimator for θ.

6.3.2 Best Guess, Bayesian Estimator

What is the 'best' estimate for θ, given the posterior distribution $p(\theta\big|x^*)$? There can be no 'right' answer to this question, because the word 'best' involves value judgements and the problem is therefore one of decision theory. What we need is a criterion, defining the quality of an estimate.

Every estimate $\tilde{\theta}$ we make for θ will have an error and elementary parameter estimation theory provides several error functions quantifying the error made with a guess. Without discussing the pros and cons of different error functions, we take – in analogy to our definition of $\overline{\mathcal{D}}$, eq. (5.14) – the estimate that minimizes the expected square of the error,

$$\begin{aligned}\langle(\theta-\tilde{\theta})^2\rangle &= \langle\theta^2\rangle - 2\tilde{\theta}\langle\theta\rangle + \tilde{\theta}^2 \\ &= (\tilde{\theta}-\langle\theta\rangle)^2 + (\langle\theta^2\rangle - \langle\theta\rangle^2)\,. \end{aligned} \tag{6.7}$$

The average, however, is taken over the a posteriori distribution $p(\theta\big|x^*)$ and not over the a priori distribution $p(x^*\big|\theta)$ as is done in $\overline{\mathcal{D}}$. The reason to choose the a posteriori instead of the a priori distribution lies in the fact that the latter ignores prior information about θ while the former takes it into account and is therefore what we want.

From (6.7) we see that the choice

$$\tilde{\theta} = \langle\theta\rangle = \int d\theta\; \theta p(\theta\big|x^*) \tag{6.8}$$

always minimizes the expected square error and the minimum achievable value in (6.7) is the variance of θ. (6.8) is called *Bayesian estimator*.

6.3.3 Implementation to N-Series, Mean Square Deviation $\overline{\mathcal{D}}(g_{\mathbf{Bay}})$

The implementation of (6.6) and (6.8) to one N-series is forthright: Bayes' theorem reads

$$p(|c_1|^2\big|N_+) = p(|c_1|^2)\frac{p(N_+\big|N,|c_1|^2)}{p(N_+)}\,, \tag{6.9}$$

with the a priori distribution $p(N_+\big|N,|c_1|^2)$ given by (6.1), and the marginal

$$\begin{aligned} p(N_+) &= \int_0^1 d|c_1|^2\; p(N_+\big|N,|c_1|^2) \\ &= \frac{1}{2}\binom{N}{N_+}[\alpha_0+\alpha_1]\,. \end{aligned}$$

$0 \le |c_1|^2 \le 1$ is the only a priori information we have about $|c_1|^2$. Since we do not know its distribution within the unit interval it is reasonable to assume equipartition. Hence, the prior writes

$$p(|c_1|^2) = \begin{cases} \frac{1}{1-0} = 1 & 0 \le |c_1|^2 \le 1\,, \\ 0 & \text{else.} \end{cases}$$

For the Bayesian estimator $g_{\rm Bay}$ we get

$$\begin{aligned} g_{\rm Bay} &= \int_0^1 d|c_1|^2\ |c_1|^2 p(|c_1|^2\big|N_+) \\ &= \int_0^1 d|c_1|^2\ 2|c_1|^2 \frac{\alpha_0 + (\alpha_1 - \alpha_0)|c_1|^2}{\alpha_0 + \alpha_1} \\ &= \frac{\alpha_0 + 2\alpha_1}{3(\alpha_0 + \alpha_1)}\,. \end{aligned} \tag{6.10}$$

Mean Square Deviation

To compare (6.10) with g_r and $g_{\rm ML}$ we need the mean square deviation $\overline{\mathcal{D}}$, defined in section 5.4, eq. (5.14). There, the average in $\overline{\mathcal{D}}$ is taken – besides states $|c_1|^2$ – over the distribution $p(N_+\big|N, |c_1|^2)$, whereas $g_{\rm Bay}$ is derived from minimizing the mean square deviation (6.7), averaged over the a posteriori distribution $p(|c_1|^2\big|N_+)$ for $|c_1|^2$. So, why should $g_{\rm Bay}$ be subjected to $\overline{\mathcal{D}}$ when it was derived from a different mean square deviation?

First of all, we need consistency if all three guesses are to be compared with each other. Second, $\overline{\mathcal{D}}$ and (6.7) are not really different. In writing down (6.7)

$$\begin{aligned} \overline{\mathcal{D}}_{\rm post}(g_{\rm Bay}) &:= \langle (g_{\rm Bay} - |c_1|^2)^2 \rangle \\ &= \int_0^1 d|c_1|^2\ p(|c_1|^2\big|N_+)(g_{\rm Bay} - |c_1|^2)^2\,, \end{aligned}$$

we see that this quantity still depends on the measurement outcome N_+. $\overline{\mathcal{D}}(g_{\rm Bay})$ on the other hand is independent of N_+. To get rid of this dependency, $\overline{\mathcal{D}}_{\rm post}(g_{\rm Bay})$ has to be averaged over all measurement outcomes N_+ with the probability distribution $p(N_+)$. But this leads – because of Bayes' theorem (6.9) and the trivial prior $p(|c_1|^2) = 1$ – exactly to our definition $\overline{\mathcal{D}}(g_{\rm Bay})$,

$$\begin{aligned} \sum_{N_+=0}^{N} p(N_+)\overline{\mathcal{D}}_{\rm post}(g_{\rm Bay}) &= \int_0^1 d|c_1|^2 \sum_{N_+=0}^{N} p(|c_1|^2\big|N_+)p(N_+)(g_{\rm Bay} - |c_1|^2)^2 \\ &= \int_0^1 d|c_1|^2 \sum_{N_+=0}^{N} p(N_+\big|N, |c_1|^2)(g_{\rm Bay} - |c_1|^2)^2 \\ &= \overline{\mathcal{D}}(g_{\rm Bay})\,. \end{aligned} \tag{6.11}$$

Thus, a comparison of $g_{\rm Bay}$, g_r and $g_{\rm ML}$ with $\overline{\mathcal{D}}$ is well suited.

6.4 Comparing all Schemes

With all three guesses g_r, $g_{\rm ML}$ and $g_{\rm Bay}$ established, we are in a position to compare them with the help of the mean square deviation $\overline{\mathcal{D}}$. This comparison will make up the first half of this section. The second half is devoted to a discussion of $\mathcal{D}$, that is, we discard the averaging over states. This procedure will further illuminate the behavior of our guesses if more information about $|c_1|^2$ is available than 'equipartition in the unit interval'.

For the remainder of this discussion we set $\bar{p} = 1/2$. $\overline{\mathcal{D}}(g_{\rm ML})$ is then given in closed form, cf. equations (6.4) and (6.5). The only motive to choose $\bar{p} = 1/2$ is convenience for further

calculations. Nevertheless, there seem to exist interconnections to the fidelity balance and optimal measurements, discussed in section 3.4, page 25. However, this topic shall not be discussed here. We furthermore compare the mean square deviations for specific N, since a closed solution for $\overline{\mathcal{D}}(g_{\mathrm{Bay}})$ does not exist. The values of N are chosen as to fulfill inequality (5.3). This ensures that all guesses are applicable to the experimental situation, where measurements are not performed in an immediate succession.

Discussing $\overline{\mathcal{D}}$

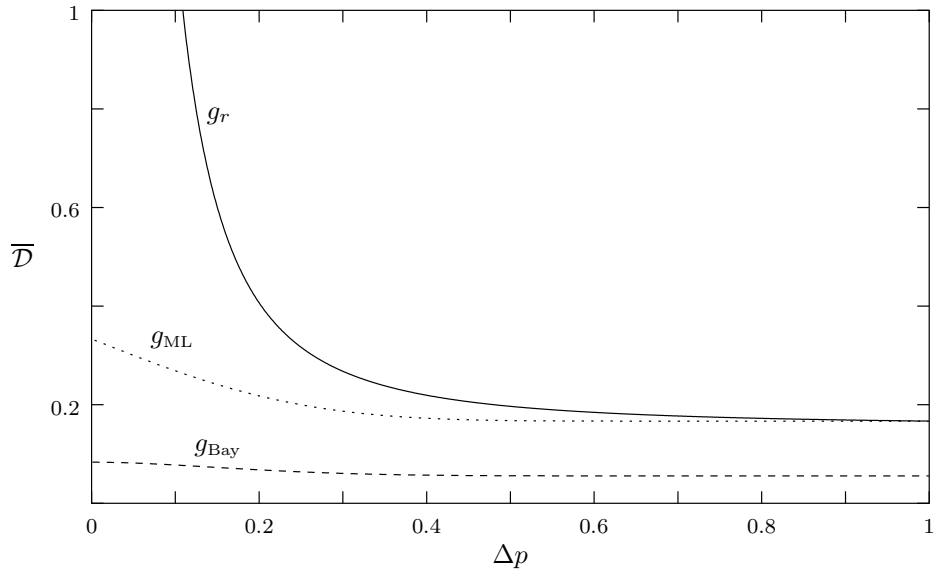


Figure 6.1: Mean square deviation $\overline{\mathcal{D}}$ vs. measurement strength Δp for one N-series with $N = 25$ measurements. In contrast to $\overline{\mathcal{D}}(g_r)$, $\overline{\mathcal{D}}(g_{\mathrm{ML}})$ and $\overline{\mathcal{D}}(g_{\mathrm{Bay}})$ are confined to the interval $[1/6, 1/3]$ and $[1/18, 1/12]$, respectively. It is $\bar{p} = 1/2$.

Figure 6.1 shows the mean square deviation $\overline{\mathcal{D}}$ plotted against the measurement strength Δp for one N-series with $N = 25$ measurements. Without looking at details of particular curve shapes, one notices that

$$\overline{\mathcal{D}}(g_{\mathrm{Bay}}) < \overline{\mathcal{D}}(g_{\mathrm{ML}}) \leq \overline{\mathcal{D}}(g_r)\,. \tag{6.12}$$

This is the main result. Explicit calculations of $\overline{\mathcal{D}}$ for $N \leq 4$ confirm relation (6.12), further supported by numerical calculations with Mathematica up to $N = 80$ (with N this high, condition (5.3) is no longer fulfilled for the entire range of Δp).

We see that $\overline{\mathcal{D}}(g_r)$ goes off to infinity for Δp going to zero, in contrast to the other two guesses. This behavior is evident from $\overline{\mathcal{D}}(g_r) \propto \Delta p^{-2}$ (cf. equation (5.15)), but can also be inferred from the structure of g_r itself. Remember, that $g_r = (r - p_0)/\Delta p$ can take on arbitrary large positive and negative values if Δp is taken to be sufficiently small. We therefore expect the mean square of all fluctuations of g_r around $|c_1|^2$ to become quite large. In fact, they approach infinity as the measurement strength goes to zero. In this limit, $g_r \to \pm\infty$, which means no reasonable conclusion about $|c_1|^2$ can be made on the basis of the relative frequency $r = N_+/N$.

For fairly strong measurements, $\overline{\mathcal{D}}(g_r)$ clings to the horizontal asymptote $\overline{\mathcal{D}}_{\text{projective}} = 1/6$. This behavior gets more pronounced the higher N, that is, $\overline{\mathcal{D}}(g_r)$ drops faster to $\overline{\mathcal{D}}_{\text{projective}}$ for smaller Δp. In the limit of infinitely many measurements,

$$\lim_{N\to\infty} \overline{\mathcal{D}}(g_r) = \overline{\mathcal{D}}_{\text{projective}} = \frac{1}{6}\,, \tag{6.13}$$

independent of Δp. This expresses the fact that an infinite series of unsharp measurements is equivalent to one projective measurement. See p. 58 of [Kon03] for a proof of this statement.

The maximum likelihood guess g_{ML} displays the same asymptotic behavior (6.13) for large Δp and high N. In contrast to $\overline{\mathcal{D}}(g_r)$, $\overline{\mathcal{D}}(g_{\text{ML}})$ is bounded above by

$$\lim_{\Delta p\to 0} \overline{\mathcal{D}}(g_{\text{ML}}) = \frac{1}{3}\,,$$

verified numerically for $N \leq 30$. This restriction of $\overline{\mathcal{D}}(g_{\text{ML}})$ over the whole range of measurements can again be made clear by looking at g_{ML} itself. It is either 0 or 1, so fluctuations are bound to this interval, which implies a finite mean square deviation. Although g_{ML} seems a very crude estimate in comparison to g_r, we see that $\overline{\mathcal{D}}(g_{\text{ML}})$ *always* lies below $\overline{\mathcal{D}}(g_r)$, a somehow unexpected result.

The Bayesian estimator $\overline{\mathcal{D}}(g_{\text{Bay}})$ even lies below the maximum likelihood deviation, a consequence of the incorporated a priori information about $|c_1|^2$. Striking feature of $\overline{\mathcal{D}}(g_{\text{Bay}})$ is its small absolute value, lying in between 1/18 for projective measurements and 1/12 for $\Delta p = 0$ (analytically as well as numerically verified for $N \leq 30$). In analogy to the finite deviation of g_{ML}, these small deviations can be attributed to the special behavior of g_{Bay}. A moment's thought shows that g_{Bay} is restricted to the interval

$$\frac{1}{3} \leq g_{\text{Bay}} \leq \frac{2}{3}\,. \tag{6.14}$$

This accounts for its low mean square deviation. In fact, g_{Bay} is the best possible estimate one can make, taking $\overline{\mathcal{D}}$ as the criterion. Recall, that we got g_{Bay} by minimizing the mean square error, averaged over the a posteriori distribution $p(|c_1|^2\big|N_+)$ for $|c_1|^2$. Now, because of (6.11), the minimum of $\overline{\mathcal{D}}_{\text{post}}(g_{\text{Bay}})$ is also the minimum for our criterion $\overline{\mathcal{D}}$: Assume we want to find the guess g minimizing $\overline{\mathcal{D}}$, defined in (5.14). $\overline{\mathcal{D}}(g)$ then reads

$$\begin{aligned}\overline{\mathcal{D}}(g) &= \sum_{N_+=0}^{N} \int_0^1 d|c_1|^2 p(N_+\big|N, |c_1|^2)(g - |c_1|^2)^2 \\ &= \sum_{N_+=0}^{N} \binom{N}{N_+} \left[\frac{1}{2}(\alpha_0+\alpha_1)g^2 + \frac{1}{3}(\alpha_0+2\alpha_1)g + \frac{1}{3}\alpha_0 + \frac{1}{4}(\alpha_1-\alpha_0)\right]\,.\end{aligned}$$

Solving for $\partial\overline{\mathcal{D}}(g)/\partial g = 0$ yields $g = g_{\text{Bay}}$. Hence, there exists no estimate having a lower mean square deviation $\overline{\mathcal{D}}$ than g_{Bay}.

Remark: The POVMs corresponding to $g_{\text{Bay}} = 1/3$ respectively $g_{\text{Bay}} = 2/3$ describe phase damping, cf. [Nie01], page 384. For example, if $g_{\text{Bay}} = 2/3$, then $p_0 = 0$ and $E_+ = p_1|1\rangle\langle 1|$, $E_- = |0\rangle\langle 0| + (1-p_1)|1\rangle\langle 1|$.

Is $\overline{\mathcal{D}}$ an Appropriate Measure?

Although g_{Bay} has the lowest possible deviation, it is an unbiased estimator, as can be seen from relation (6.14) (take, for example, $|c_1|^2 = 0.8$). This poses the question whether $\overline{\mathcal{D}}$ is an appropriate measure to define the quality of a guess, if used in the context of real-time visualization of dynamics. In this respect the mean square deviation competes with criteria normally used in parameter estimation theory (cf. Appendix C.2). g_r is – by construction – an unbiased guess, which ensures the correct estimate (true value of $|c_1|^2$) in the limit of infinitely many ensembles (of finite size). Nevertheless, $\overline{\mathcal{D}}(g_r)$ lies above $\overline{\mathcal{D}}(g_{\text{ML}})$ and $\overline{\mathcal{D}}(g_{\text{Bay}})$. g_{ML} does not seem to be unbiased, although this could not be verified analytically.

So far, a conclusive answer to what criterion is best used cannot be given. The classical measures unbiasedness and efficiency are not applicable to the mean square deviation, since no measured average for $|c_1|^2$ exists. Consistency seems a reasonable criterion for a guess: The error made with a guess should go to zero if the ensemble size goes to infinity. Then, an infinite amount of information would be available and the guess would be exact.

$\overline{\mathcal{D}}$ defined in (5.14) was developed in close analogy to the classical measure variance. It thus represents a first step towards a measure suited to analyze the error made with a guess obtained from an N-series. See [Kle03] for a thorough discussion.

Discussing $\mathcal{D}$

Should we find ourselves in the situation where more information about $|c_1|^2$ is available to us than 'equipartition within the unit interval', the average over states is unnecessary. Then $\mathcal{D}$ is a sufficient measure to make statements about different guesses and their dependence on $|c_1|^2$.

Figure 6.2 shows the mean square deviation $\mathcal{D}$ versus measurement strength Δp for $N = 4$ measurements and different states $|c_1|^2$. The four plots display $\mathcal{D}(g)$ for the eigenstates $|0\rangle$ and $|1\rangle$ ($|c_1|^2 = 0$ and $|c_1|^2 = 1$, respectively) as well as two superpositions $|\psi\rangle = c_0|0\rangle + c_1|1\rangle$ ($|c_1|^2 = 0.3$ and $|c_1|^2 = 0.7$). Except for the lower left plot, axes labels have been suppressed for better readability.

In contrast to figure 6.1, statements about the quality of a particular guess now depend on the measurement strength as well as the state one is presupposing. For $|c_1|^2$ sufficiently away from the corresponding eigenstates, we qualitatively recover the main result (6.12), now stated with $\mathcal{D}$ instead of $\overline{\mathcal{D}}$:

$$\mathcal{D}(g_{\text{Bay}}) < \mathcal{D}(g_{\text{ML}}) \leq \mathcal{D}(g_r)\,.$$

This is not surprising. Because g_{Bay} lies within the interval $[1/3, 2/3]$, the estimate never lies far away from the actual value of $|c_1|^2$. On the other hand, the maximum likelihood estimate is either 0 or 1, so on average, g_{ML} lies further away from $|c_1|^2$ than g_{Bay}. Again, g_r has the highest deviation $\mathcal{D}$, regardless of the applied measurement strength.

The situation changes if $|c_1|^2$ lies close to either one of the respective eigenstates $|0\rangle$ or $|1\rangle$. For strong measurements, g_{ML} and g_r outrun the Bayesian estimate, because of its confinement to the interval $[1/3, 2/3]$. The maximum likelihood estimate is now almost perfectly well suited to follow the evolution of $|c_1|^2$, as well as g_r, whose deviation $\mathcal{D}(g_r)$ drops to zero for projective measurements, and $|c_1|^2 = 0$ or $|c_1|^2 = 1$.

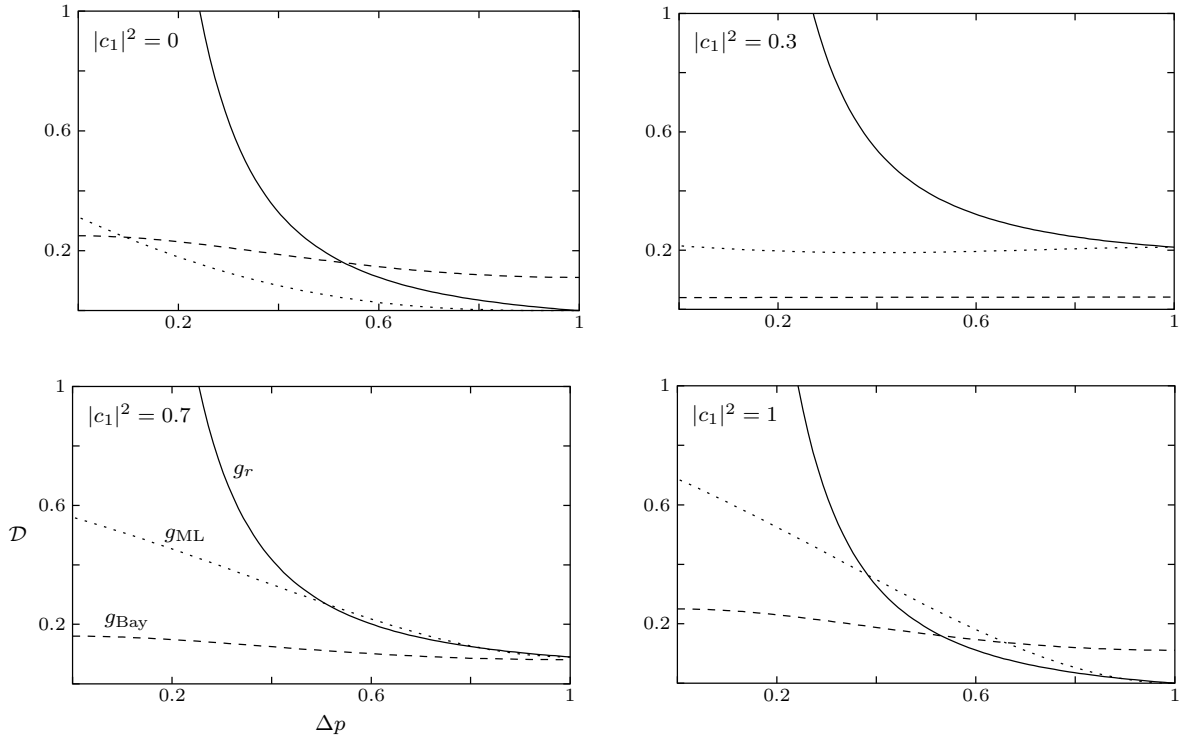


Figure 6.2: Mean square deviation $\mathcal{D}$ vs. measurement strength Δp for one N-series with $N = 4$ measurements and different values of $|c_1|^2$. In all plots g_r is displayed as straight line, g_{ML} as dotted line, g_{Bay} as dashed line. Labeling of axes has been suppressed except for the lower left plot to enhance legibility. For $|c_1|^2$ near an eigenstate and strong measurements, $\mathcal{D}(g_r)$ and $\mathcal{D}(g_{\mathrm{ML}})$ surpass the Bayesian estimate, whose $\mathcal{D}(g_{\mathrm{Bay}})$ is better otherwise. It is $\bar{p} = 1/2$.

Chapter 7

Summary and Outlook

7.1 Summary

Several aspects of state and parameter estimation in quantum theory have been studied in this work using the language of generalized measurements (POVM measurements).

In the first part, comprising chapters 3 and 4, the delicate balance between information gained from a measurement and thereby induced state disturbance was investigated. For one measurement (mapping pure states to pure states) on a single system, the resulting trade-off [Ban01a] in terms of mean operation fidelities F and G was investigated for qubits and a special class of qubit POVMs. This special class consists of measurements with commuting effects. Such unsharp measurements can be interpreted in terms of ordinary observables, as, for example, energy or spin. The optimal balance between information gain and disturbance was determined to be a simple constraint, restricting the parameters characterizing these unsharp measurements.

A clear visualization of the trade-off is provided by the FG-plane. Here, the trade-off could be expressed in terms of parameters characterizing the unsharp qubit POVM. This parametrization revealed a simple structure, separating minimal from non-minimal measurements (Kraus operators of non-minimal measurements have a non-trivial unitary back-action term in their polar decomposition). Non-minimal measurements were found to deteriorate the optimal balance noticeably.

If nothing is known a priori about the pure qubit state to be estimated, a unitary back-action term, i.e. non-minimal measurements, can not improve on the mean operation fidelity F. If, however, a priori information about the qubit is available, then – as was shown in chapter 4 – a suitably chosen back-action improves on the mean operation fidelity, calculated for minimal measurements and a priori information. The latter fidelity, in turn, can be larger or smaller than the mean operation fidelity calculated if no a priori information about the qubit exists; that is, information not necessarily implies a higher fidelity. Interestingly, the mean operation fidelity calculated for a priori information and non-minimal measurements, containing this special back-action, can be higher than the mean operation fidelity calculated if there is no a priori information.

In the second part, comprising chapters 5 and 6, focus is shifted to the estimation of parameters characterizing dynamics of a qubit. Building upon previous work on the real-time visualization of a qubit's Rabi oscillations which employs sequences of unsharp measurements

(N-series), different estimation procedures were devised and compared. The originally proposed estimator [Aud01] was constructed to fulfill unbiasedness, a criterion used in classical estimation theory. The second estimator (guess) developed uses the maximum likelihood method. It discards, however, eventually available pre-measurement information about the qubit. The third guess (Bayesian estimator) takes this fact into account by means of Bayes' theorem.

To compare all guesses, a mean square deviation was constructed to define a measure of quality. With this measure at hand, it was shown that for one N-series, the Bayesian estimator is superior to the maximum likelihood – as well as the original estimator. The latter, in turn, is inferior (i.e. has a larger mean square deviation) to the maximum likelihood estimator.

7.2 Outlook

The increase in fidelity by certain unitary back-actions and pre-measurement information about the qubit state is an interesting and unexpected effect, requiring further investigation. In this connection, the distinguishability of two non-orthogonal pure states, quantifiable e.g. by the error probability, poses an interesting problem. It would be worthwhile to check whether a reduction of the error probability is possible by means of the special non-minimal measurements introduced in chapter 4.

Without information about the pre-measurement state, the fidelity trade-off between F and G is optimized for the special class of minimal measurements with commuting effects. The question arises if these unsharp measurements are the only class saturating the optimal upper bound.

In chapter 6 the mean square deviation was used to compare all three estimators obtained from the measurement results of one N-series. By definition, the measurements constituting one N-series were carried out in an immediate succession. That is, no dynamical evolution of the state in between two measurements could take place. If dynamical processes are taken into account, the question poses itself whether the mean square deviation is an appropriate measure any longer. Numerical simulations of the estimators derived in this work are currently made to elucidate things.

Appendix A

Calculations in Chapters 3 and 4

A.1 Inequalities Used for Trade-Off

In deriving the fidelity balance (3.15) between F and G we made use of two vector inequalities from linear algebra. I placed the omitted details in this appendix in order not to interrupt the flux of the presentation.

Cauchy-Schwarz Inequality

The Cauchy-Schwarz inequality states that for two arbitrary vectors $\mathbf{x}$,$\mathbf{y} \in \mathbb{R}^N$ the inequality

$$\mathbf{x} \cdot \mathbf{y} \leq |\mathbf{x}| \cdot |\mathbf{y}| \tag{A.1}$$

always holds, with equality being reached if and only if $\mathbf{x} = a\mathbf{y}$ ($a > 0$). $|\cdot|$ denotes the usual 2-norm. The proof of (A.1) can be found in every introductory textbook on linear algebra. Here, I only prove the equality sign: if $\mathbf{x} = a\mathbf{y}$ with $a > 0$, we get

$$\begin{aligned} \text{l.h.s} &= \mathbf{x} \cdot \mathbf{y} = a\mathbf{y} \cdot \mathbf{y} = a\mathbf{y}^2 \\ \text{r.h.s} &= |\mathbf{x}| \cdot |\mathbf{y}| = a|\mathbf{y}| \cdot |\mathbf{y}| = a\mathbf{y}^2 \,. \end{aligned}$$

Inequality Between Quadratic and Arithmetic Mean

For a sequence of real numbers $a_i \in \mathbb{R}$ and $N \in \mathbb{N}$, the arithmetic mean m_A and the quadratic mean m_Q are defined through

$$m_A := \frac{1}{N}\sum_{i=1}^{N} a_i \quad \text{and} \quad m_Q := \left(\frac{1}{N}\sum_{i=1}^{N} a_i^2\right)^{1/2} ,$$

with the inequality $m_A \leq m_Q \; \forall a_i \in \mathbb{R}$ holding between them. Equality is reached if and only if $a_1 = \cdots = a_N$. To see this, we evaluate both sides of the inequality assuming $a_i = a \; \forall i$:

$$\begin{aligned} \text{l.h.s} &= \frac{1}{N}\sum_{i=1}^{N} a = \frac{1}{N} Na = a \\ \text{r.h.s} &= \left(\frac{1}{N}\sum_{i=1}^{N} a^2\right)^{1/2} = \left(\frac{1}{N} Na^2\right)^{1/2} = a \,. \end{aligned}$$

A.2 Calculation of $F_\Omega(U_s\sqrt{E_s})$

A calculation becomes more error-prone the longer it gets. I therefore give the explicit derivation of $F_\Omega(U_s\sqrt{E_s})$ in section 4.3 (p. 32), so that the reader should be able to reconstruct every step of it.

Suppose we transformed our knowledge about the pre-measurement state $|\psi\rangle$ into a region Ω on the surface of the Bloch sphere. After choosing $|\tilde{\psi}\rangle$, we rotate Ω via the unitary transformation $U_z = \exp\{i\xi\sigma_z\}$, such that $|\tilde{\psi}\rangle$ lies in the yz-plane. Requiring the action of $\sqrt{E_s}$ on $|\tilde{\psi}\rangle$ to be fully compensated by $U_s = \exp\{\pm i\alpha_s/2\sigma_x\}$ determines the two angles α_s by which every state within Ω will be rotated. α_s are given by (4.3), the scalar product between the normed Bloch vectors of $|\psi\rangle$ and $|\tilde{\psi}\rangle$.

We start with the definition (4.1):

$$F_\Omega(U_s\sqrt{E_s}) = \frac{1}{N(\Omega)}\int_\Omega d\psi \sum_{s=\pm} |\langle\psi|U_s\sqrt{E_s}|\psi\rangle|^2 \,.$$

Evaluation of the integral is not necessary for a qualitative analysis, so it suffices to calculate the integrand,

$$\sum_{s=\pm} |\langle\psi|U_s\sqrt{E_s}|\psi\rangle|^2 = |\langle\psi|\exp\{-i\frac{\alpha_+}{2}\sigma_x\}\sqrt{E_+}|\psi\rangle|^2 + |\langle\psi|\exp\{i\frac{\alpha_-}{2}\sigma_x\}\sqrt{E_-}|\psi\rangle|^2 \,.$$

Both terms are calculated in the same fashion, so we focus on the term with α_+. Using

$$\exp\{\pm i\frac{\alpha_s}{2}\sigma_x\} = \cos\frac{\alpha_s}{2}\mathbb{1} \pm i\sin\frac{\alpha_s}{2}\sigma_x \,,$$

we get

$$\begin{aligned}
|\langle\psi|\exp\{-i\frac{\alpha_+}{2}\sigma_x\}\sqrt{E_+}|\psi\rangle|^2 &= |\langle\psi|\big(\cos\frac{\alpha_+}{2}\mathbb{1} - i\sin\frac{\alpha_+}{2}\sigma_x\big)\sqrt{E_+}|\psi\rangle|^2 \\
&= |\cos\frac{\alpha_+}{2}\langle\psi|\sqrt{E_+}|\psi\rangle - i\sin\frac{\alpha_+}{2}\langle\psi|\sigma_x\sqrt{E_+}|\psi\rangle|^2 \\
&= \cos^2\frac{\alpha_+}{2}|\langle\psi|\sqrt{E_+}|\psi\rangle|^2 + \sin^2\frac{\alpha_+}{2}|\langle\psi|\sigma_x\sqrt{E_+}|\psi\rangle|^2 \\
&\quad + \frac{i}{2}\sin\alpha_+\langle\psi|\sqrt{E_+}|\psi\rangle\Big(\langle\psi|\sigma_x\sqrt{E_+}|\psi\rangle^* - \langle\psi|\sigma_x\sqrt{E_+}|\psi\rangle\Big) \,.
\end{aligned} \tag{A.2}$$

In the last step we used positivity[1] of $\sqrt{E_+}$ and $2\sin(\alpha_+/2)\cos(\alpha_+/2) = \sin\alpha_+$. The Pauli matrix $\sigma_x = |0\rangle\langle 1| + |1\rangle\langle 0|$ is hermitian, $\langle\psi|\sigma_x\sqrt{E_+}|\psi\rangle^* = \langle\psi|(\sigma_x\sqrt{E_+})^\dagger|\psi\rangle = \langle\psi|\sigma_x\sqrt{E_+}|\psi\rangle$, and (A.2) writes

$$\begin{aligned}
|\langle\psi|\exp\{-i\frac{\alpha_+}{2}\sigma_x\}\sqrt{E_+}|\psi\rangle|^2 &= \cos^2\frac{\alpha_+}{2}|\langle\psi|\sqrt{E_+}|\psi\rangle|^2 + \sin^2\frac{\alpha_+}{2}|\langle\psi|\sigma_x\sqrt{E_+}|\psi\rangle|^2 \\
&\quad + \frac{i}{2}\sin\alpha_+\langle\psi|\sqrt{E_+}|\psi\rangle\langle\psi|[\sqrt{E_+},\sigma_x]|\psi\rangle \,.
\end{aligned}$$

[1] The modulus in $|\langle\psi|\sqrt{E_+}|\psi\rangle|^2$ can be omitted. We will leave it for optical homogeneity, though.

With the commutator $[\sqrt{E_+},\sigma_x] = i(\sqrt{p_0}-\sqrt{p_1})\sigma_y$, we finally arrive at

$$|\langle\psi|\exp\{-i\frac{\alpha_+}{2}\sigma_x\}\sqrt{E_+}|\psi\rangle|^2 = \cos^2\frac{\alpha_+}{2}|\langle\psi|\sqrt{E_+}|\psi\rangle|^2 + \sin^2\frac{\alpha_+}{2}|\langle\psi|\sigma_x\sqrt{E_+}|\psi\rangle|^2 + \frac{1}{2}(\sqrt{p_1}-\sqrt{p_0})\langle\psi|\sqrt{E_+}|\psi\rangle\langle\psi|\sigma_y|\psi\rangle\sin\alpha_+\,. \tag{A.3}$$

Similarly, the second term with α_- is evaluated along the same line. One gets

$$|\langle\psi|\exp\{i\frac{\alpha_-}{2}\sigma_x\}\sqrt{E_-}|\psi\rangle|^2 = \cos^2\frac{\alpha_-}{2}|\langle\psi|\sqrt{E_-}|\psi\rangle|^2 + \sin^2\frac{\alpha_-}{2}|\langle\psi|\sigma_x\sqrt{E_-}|\psi\rangle|^2 + \frac{1}{2}(\sqrt{1-p_0}-\sqrt{1-p_1})\langle\psi|\sqrt{E_-}|\psi\rangle\langle\psi|\sigma_y|\psi\rangle\sin\alpha_-\,. \tag{A.4}$$

With (A.3) and (A.4), $F_\Omega(U_s\sqrt{E_s})$ reads

$$F_\Omega(U_s\sqrt{E_s}) = \frac{1}{N(\Omega)}\int_\Omega d\psi\left\{\sum_{s=\pm}|\langle\psi|\sqrt{E_s}|\psi\rangle|^2\cos^2\frac{\alpha_s}{2} + |\langle\psi|\sigma_x\sqrt{E_s}|\psi\rangle|^2\sin^2\frac{\alpha_s}{2} + \frac{1}{2}\langle\psi|\sigma_y|\psi\rangle\Big[(\sqrt{p_1}-\sqrt{p_0})\langle\psi|\sqrt{E_+}|\psi\rangle\sin\alpha_+ + (\sqrt{1-p_0}-\sqrt{1-p_1})\langle\psi|\sqrt{E_-}|\psi\rangle\sin\alpha_-\Big]\right\}.$$

To connect $F_\Omega(\sqrt{E_s})$ to $F_\Omega(U_s\sqrt{E_s})$, the final substitution $\cos^2(\alpha_s/2) = 1-\sin^2(\alpha_s/2)$ is made which leads to the result (4.4) with minimal term (4.6) and difference term (4.9).

Appendix B

Bloch Sphere Representation

The bit is the fundamental concept in classical computation and classical communication. Quantum computation and quantum communication are built upon an analogous concept, the *quantum bit*, or *qubit* for short. In addition to their classical counterparts, qubits can be in superpositions of states $|0\rangle$ and $|1\rangle$

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle\,, \tag{B.1}$$

making them much more powerful when it comes to certain computational tasks. Because of normalization $|\alpha|^2 + |\beta|^2 = 1$, eq. (B.1) may be rewritten (a global phase factor $e^{i\gamma}$ is omitted since it has no observable effects)

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle\,, \tag{B.2}$$

where $0 \leq \theta \leq \pi$ and $0 \leq \varphi \leq 2\pi$ are azimuthal and polar angles, respectively. θ and φ uniquely define a point on the surface of the three-dimensional unit sphere, also called *Bloch sphere* (cf. figure B.1). It provides a useful means of visualizing the pure state of a single qubit, and often serves as an excellent testbed for ideas about quantum computation and quantum information. Many operations on single qubits are neatly described within the Bloch sphere picture. However, it must be kept in mind that this intuition is limited because there is no generalization of the Bloch sphere known for multiple qubits.

The *general* description of states in the Bloch sphere picture makes use of an operator basis spanned by the three Pauli matrices σ_i, $i = \{x, y, z\}$ and the unit matrix $\mathbb{1}$. Every operator A in two-dimensional Hilbert space $\mathcal{H}_2$ can be decomposed according to

$$A = \frac{1}{2}\mathrm{tr}[A]\mathbb{1} + \frac{1}{2}\sum_i \mathrm{tr}[A\sigma_i]\sigma_i\,.$$

This way, we can write any qubit density matrix ρ in the form

$$\rho = \frac{1}{2}\left(\mathbb{1} + \boldsymbol{n}\cdot\boldsymbol{\sigma}\right)\,,$$

with $\boldsymbol{\sigma} := (\sigma_x, \sigma_y, \sigma_z)$; $\boldsymbol{n} = \sum_i \mathrm{tr}[\rho\sigma_i]$ is the so-called *Bloch vector*. It uniquely defines a point

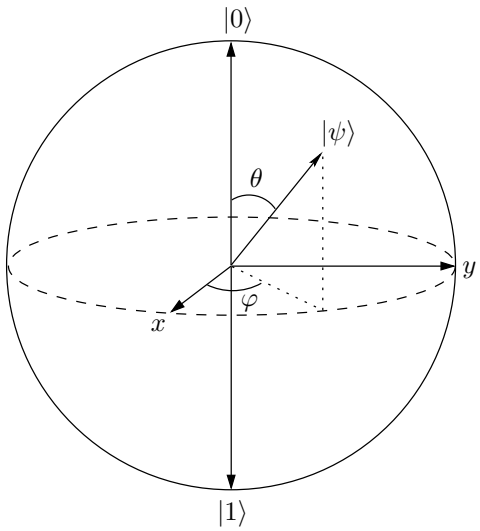


Figure B.1: Bloch sphere representation of a qubit. Pure states lie on the surface, mixed states within the sphere.

within the Bloch sphere as can be seen by calculating $\mathrm{tr}[\rho^2]$:

$$\begin{aligned}\mathrm{tr}[\rho^2] &= \frac{1}{4}\mathrm{tr}[(\mathbb{1} + \boldsymbol{n}\cdot\boldsymbol{\sigma})(\mathbb{1} + \boldsymbol{n}\cdot\boldsymbol{\sigma})] \\ &= \frac{1}{4}\mathrm{tr}[\mathbb{1} + 2\boldsymbol{n}\cdot\boldsymbol{\sigma} + \sum_{i,j} n_i n_j \sigma_i \sigma_j] \\ &= \frac{1}{2}(1 + n^2)\,;\end{aligned}$$

here, we used $\sigma_i\sigma_j = \delta_{ij}\mathbb{1} + \sum_k \epsilon_{ijk}\sigma_k$ in the last step. For pure states, $\mathrm{tr}[\rho^2] = 1$, so $|\boldsymbol{n}| = n \leq 1$. Since $\rho = \frac{1}{2}\mathbb{1}$ denotes the total mixture, n is bounded between zero[1] and one, $0 \leq n \leq 1$. Consequently, mixed states lie within the Bloch sphere whereas pure states populate its surface. This can also be seen by noting that in the latter case $\boldsymbol{n} = \sum_i \mathrm{tr}[\rho\sigma_i] = \langle\psi|\boldsymbol{\sigma}|\psi\rangle$. Inserting (B.2) we get

$$\boldsymbol{n} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)\,,$$

which is just the unit vector in spherical coordinates. Consequently, every pure state is completely characterized by azimuth θ and polar angle φ; additionally, mixed states require the length n.

[1] because the σ_i are linearly independent, every component of the Bloch vector has to vanish.

Appendix C

Notes on Probability Theory

Basics of probability theory can be found in every introductory textbook on the subject, e.g. [Lar83] and [Kre70]. Therefore this appendix should not be viewed as a thorough introduction but rather as a convenience to the reader. We assume a certain familiarity with elementary set theory, since it provides a clear and exact language for stating and solving probability problems.

Part one reviews some basic definitions and results given in [Nie01], whereas part two provides definitions of quality measures for estimators.

C.1 Basics

The basic notion of probability theory is that of a *random variable*. A random variable X may take one of a number of values, x, with probability $p_X(x)$. We use upper case to denote the random variable, and x to denote the values that random variable may take. All possible values are elements of some *sample space* Ω which can be continuous or discrete. In this thesis we are only concerned with discrete sample spaces, i.e. any random variable can take values from a finite (or countably infinite) set of values.

A probability function p_X of some random variable X assigns a real number to any element or subset of the sample space of X. From intuition it is clear what properties such a probability function must have: probabilities always have to be positive and add up to one, since one event always happens in any experiment. The third property a probability measure should satisfy is that of additivity: if two events are *independent* (i.e. every outcome of an experiment depends in no way on any other outcome), their respective probabilities of occurrence should add up. A strict formulation of these intuitive rules in abstract probability theory was first given by *Kolmogorov* [Kol33]:

Definition C.1 (probability axioms). *A probability function p_X of a random variable X is a real-valued set function defined on the class of all subsets Σ of the sample space Ω and has to fulfill the following three rules:*

(i) $p_X(\Omega) = 1$

(ii) $p_X(\Sigma) \geq 0 \quad$ *for all* $\Sigma \subset \Omega$

(iii) $p_X(\Sigma_1 \cup \Sigma_2 \cup \cdots) = p_X(\Sigma_1) + p_X(\Sigma_2) + \cdots \quad$ *if* $\Sigma_i \cap \Sigma_j = \emptyset \quad \forall\, i \neq j.$

In experiments, one often has to consider two random variables X and Y (if no ambiguities arise, the subscripts on p are dropped in the following). The *conditional probability* that $Y = y$ given that $X = x$ is defined by

$$p(y|x) := \frac{p(x \cap y)}{p(x)}\,, \tag{C.1}$$

where $p(x \cap y)$ is the probability that $X = x$ *and* $Y = y$. When $p(x) = 0$ the convention $p(y|x) = 0$ is made. For independent random variables X and Y, $p(x \cap y) = p(x)p(y)$ for all x and y; additionally, one has $p(y|x) = p(y)$ for all x and y.

Bayes' theorem relates the conditional probabilities for Y given X to those for X given Y,

$$p(x|y) = p(y|x)\frac{p(x)}{p(y)}\,; \tag{C.2}$$

it is a direct consequence of (C.1), since $p(x \cap y) = p(y \cap x)$. The probability $p(y)$ appearing in (C.2) is often re-expressed using the law of *total probability*. It states that if X and Y are two random variables, then the probabilities for Y can be expressed in terms of the probabilities for X, and the conditional probabilities for Y given X,

$$p(y) = \sum_x p(y|x)p(x)\,,$$

where the sum is over all values x of the respective sample space. The *expectation*, *average* or *mean* of a random variable X is defined by

$$\langle X \rangle := \sum_x p(x)x\,.$$

$\langle\cdot\rangle$ is a linear function; for independent X and Y, $\langle XY \rangle = \langle X \rangle \langle Y \rangle$. The *variance* of a random variable X is defined by the expression

$$\sigma^2(X) := \langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2\,.$$

Its square root $\sigma(X) := \sqrt{\sigma^2(X)}$, called *standard deviation*, is a measure of the spread of a random variable about the average.

C.2 Properties of Estimators

One problem of statistical inference is that of estimating one or more unknown parameters u of some probability distribution $p_X(u)$ for the random variable X. Here we shall not be concerned with methods of *how to find* a guess (called 'estimator') $\tilde{u}(x_1, \ldots, x_N)$ given a sample of size N. Instead we will review three criteria judging the quality of a given estimator $\tilde{u}$.

Definition C.2 (unbiasedness). *An estimator $\tilde{u}$ of an unknown parameter u is* unbiased *if* $\langle \tilde{u} \rangle = u$.

Thus, an unbiased estimator is a random variable whose expected value is the parameter being estimated; if we were to take samples of size N repeatedly and for each compute the observed value of $\tilde{u}$ (the estimate for that sample outcome), then the average of these observed values would be u, the parameter being estimated.

The property of unbiasedness ensures that the estimator $\tilde{u}$ yields the true value of the parameter u in the limit of infinite ensembles. However, this statement does not imply that the estimator has very high probability of lying close to the unknown parameter for any given ensemble. Suppose we have two unbiased estimators $\tilde{u}_1$, $\tilde{u}_2$ for the same sample. Clearly, we would take the estimator who has the higher probability of lying close to the parameter. This leads us to consider the variances of unbiased estimators:

Definition C.3 (efficiency). *If $\tilde{u}_1$ and $\tilde{u}_2$ are both unbiased estimators of u for the same sample, then $\tilde{u}_1$ is* more efficient *than $\tilde{u}_2$ if*

$$\sigma^2(\tilde{u}_1) < \sigma^2(\tilde{u}_2)\,.$$

There is one further property which estimators may possess, that of consistency:

Definition C.4 (consistency). *$\tilde{u}$ is a* consistent *estimator of u if*

$$\lim_{N\to\infty}[p(\tilde{u}-u)<\epsilon]=1 \quad \textit{for any } \epsilon>0$$

(N is the sample size).

Note first of all that an estimator $\tilde{u}$ is consistent if a sequence of probabilities (evaluated from the probability distribution of $\tilde{u}$) converges to one as the sample size increases. Thus consistency, strictly speaking, has to do only with the limiting behavior of an estimator as the sample size increases without limit and does not imply that the observed value of $\tilde{u}$ is necessarily close to u for any specific sample size N (cf. efficiency). This is not the same as unbiasedness, where only different ensembles of *fixed* size N are considered.

There is a theorem (cf. page 233 of [Lar83]) which gives a relatively easy check on whether a particular estimator is consistent. It says that if $\tilde{u}$ is a random variable whose value gets closer and closer to u as N increases, and whose variance shrinks to zero as N increases, it will be a consistent estimator of u.

Bibliography

[Aci00] A. Acin, J. I. Latorre and P. Pascual. *Optimal generalized quantum measurements for arbitrary spin systems.* Phys. Rev. A **61** (2000) 022113.

[Aud01] J. Audretsch, T. Konrad and A. Scherer. *A sequence of unsharp measurements enabling a real-time visualization of a quantum oscillation.* Phys. Rev. A **63** (2001) 052102.

[Aud02a] J. Audretsch, L. Diósi and T. Konrad. *Estimating the post-measurement state.* quant-ph (2002) 0210205.

[Aud02b] J. Audretsch, L. Diósi and T. Konrad. *Evolution of a qubit under the influence of a sucession of weak measurements with unitary feedback.* Phys. Rev. A **66** (2002) 022310.

[Aud02c] J. Audretsch, T. Konrad and A. Scherer. *Quantum-optical weak measurements can visualize photon dynamics in real time.* Phys. Rev. A **65** (2002) 033814.

[Bag02] E. Bagan, M. Baig and R. Muñoz-Tapia. *Optimal scheme for estimating a pure qubit state via local measurements.* quant-ph:0205026v2 (2002).

[Ban00] K. Banaszek. *Optimal quantum teleportation with an arbitrary pure state.* Phys. Rev. A **62** (2000) 024301.

[Ban01a] K. Banaszek. *Fidelity Balance in Quantum Operations.* Phys. Rev. Lett. **86** (2001) 1366.

[Ban01b] K. Banaszek and I. Devetak. *Fidelity trade-off for finite ensembles of identically prepared qubits.* Phys. Rev. A **64** (2001) 052307.

[Bru99] D. Bruß and C. Macchiavello. *Optimal state estimation for d-dimensional quantum systems.* Physics Letters A **253** (1999) 249.

[Bus91] P. Busch, P. J. Lahti and P. Mittelstaedt. *The Quantum Theory of Measurement.* Springer-Verlag Berlin (1991).

[Bus95] P. Busch, M. Grabowski and P. J. Lahti. *Operational Quantum Physics.* Lecture Notes in Physics. Springer-Verlag (1995).

[Der98] 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** (1998) 1571.

[Fis00] D. G. Fischer, S. H. Kienle and M. Freyberger. *Quantum-state estimation by self-learning measurements.* Phys. Rev. A **61** (2000) 032306.

[Fuc95] Christopher A. Fuchs. *Distinguishability and Accessible Information in Quantum Theory.* Dissertation Center for Advanced Studies, Department of Physics and Astronomy, University of New Mexico Albuquerque, NM 87131-1156 (1995).

[Han02] Th. Hannemann, D. Reiss, Ch. Balzer, W. Neuhauser, P. E. Toschek and Ch. Wunderlich. *Self-learning estimation of quantum states.* Phys. Rev. A **65** (2002) 050303.

[Hol82] A. S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory.* North-Holland Amsterdam (1982).

[Kle03] F. E. Klee. *Sequential Measurements.* Diplomarbeit Universität Konstanz (2003). In preparation.

[Kol33] A. Kolmogoroff. *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer-Verlag Berlin (1933).

[Kon03] T. Konrad. *Less is More: On the Theory and Application of Weak and Unsharp Measurements in Quantum Mechanics.* Dissertation Universität Konstanz, März (2003).

[Kra83] K. Kraus. *States, Effects and Operations.* Springer-Verlag Berlin (1983).

[Kre70] E. Kreyszig. *Statistische Methoden und ihre Anwendungen.* Vandenhoek & Ruprecht Göttingen, Germany (1970).

[Lar83] H. J. Larson. *Introduction to Probability Theory and Statistical Inference.* J. Wiley & Sons third edition (1983).

[Lat98] J. I. Latorre, P. Pascual and R. Tarrach. *Minimal Optimal Generalized Quantum Measurements.* Phys. Rev. Lett. **81** (1998) 1351.

[Mas95] S. Massar and S.Popescu. *Optimal Extraction of Information from Finite Quantum Ensembles.* Phys. Rev. Lett. **74** (1995) 1259.

[Nie01] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press (2001).

[Pau33] W. Pauli. *Die allgemeinen Prinzipien der Wellenmechanik.* Springer-Verlag Berlin (1933).

[Per89] A. Peres. *Quantum limited detectors for weak classical signals.* Phys. Rev. D **39** (1989) 2943.

[Per91] A. Peres and W. K. Wootters. *Optimal Detection of Quantum Information.* Phys. Rev. Lett. **66** (1991) 1119.

[Per93] A. Peres. *Quantum Theory: Concepts and Methods.* Kluwer Academic Publishers (1993).

[Sch94] R. Schack, G. M. D'Ariano and C. M. Caves. *Hypersensitivity to perturbation in the quantum kicked top.* Phys. Rev. E **50** (1994) 972.

[Vid99] G. Vidal, J. I. Latorre, P. Pascual and R. Tarrach. *Optimal minimal measurements of mixed states.* Phys. Rev. A **60** (1999) 126.

[vN32] J. von Neumann. *Mathematische Grundlagen der Quantenmechanik.* Springer-Verlag Berlin (1932).

[Wis95] H. M. Wiseman. *Using feedback to eliminate back-action in quantum measurements.* Phys. Rev. A **51** (1995) 2459.

[Woo82] W. K. Wootters and W. H. Zurek. *A single quantum cannot be cloned.* Nature **299** (1982) 802.

Acknowledgements

After all is said and done, it is a pleasure for me to thank all those who contributed to this effort in manifold ways.

First of all, I have to thank Prof. Audretsch and his amiable little group for making it possible to write this thesis in a cordial and stimulating environment. I am most indebted to Thomas Konrad, who never refused a question of mine and shared his profound insights into quantum measurement theory with me, despite his other time-consuming duties. Thanks also to my roommate Felix Klee, for numerous interesting discussions about various serious and not so serious physical problems, and the help with minor and not so minor computer and software (Mathematica) problems.

I am thankful to Anja (special accolade for overcoming her 'Theo-Phobie' and recklessly working her way through chapters 3, 4 and 5), Björn, Andreas and Thomas for proofreading and constructive comments.

Hugs and kisses to my girlfriend Irena, who not only read the whole draft but has always been there for me, especially if things went not the way I intended them to go.

I am grateful to all my friends: Anja, Andreas, Andreas (no typo), Björn, Hans, Harald, Irena, Matthias, Matthias (again, no typo), Konrad and Uli. Furthermore Ester, Mike, Karsten, Steffi, EP&N and everyone else I forgot to mention. Life besides physics would not be the same without you guys (and gals, of course).

To my family: Thank you for your unconditional support, to keep me grounded at every instant, and for giving me the freedom to pursue my goals in life.