

A State Estimation Scheme for Finite Quantum Systems

by

Thomas Baier

Submitted to

Central European University

Department of Mathematics and its Applications

In partial fulfillment of the requirements for the degree of Doctor of

Philosophy

Supervisor: Professor Dr. Dénes Petz

Budapest, Hungary

2009

Acknowledgments

First of all I indebted to my supervisor Dénes Petz for his helpful guidance and advice for the entirety of my graduate studies. I would also like to thank him for introducing me to the topic of quantum estimation theory and for many useful discussions (mathematical and otherwise). The mathematics Ph.D. program is a joint endeavor between the Rényi Institute of Mathematics and Central European University. I am grateful for both institutions and the excellent conditions they provided for my studies. Additionally I am much obliged to the Budapest University of Technology and Economics and the EU Research Training Network Quantum Probability and its Applications, which gave me the opportunity to start my doctoral studies in Hungary.

Abstract

The estimation of the state of a finite quantum system is studied. It is assumed that N identical copies of the unknown state are at hand and statistical data is obtained from different separate measurements. Complementary (or quasi-orthogonal) measurements, defined by an orthogonality relation, play a special role among separate measurements. Motivated by recent results on the existence of certain complementary subalgebras, the effect of complementarity of the measurements on the state estimation procedure is examined. For an unconstrained estimate optimality is shown for finite sample sizes, and for a constrained estimate, that always gives approvable results, optimality is shown in the limit of large N .

Table of Contents

Acknowledgments	ii
Abstract	iii
Introduction	1
1 Mathematical Description of Finite Quantum Systems	4
1.1 Quantum Probability Theory	4
1.1.1 The observable algebra	5
1.1.2 States and measurements	7
1.2 Complementarity	11
1.2.1 The traceless subspace	12
1.2.2 Complementary observables, algebras and measurements	13
1.2.3 Complete sets of complementary subalgebras	15
1.3 Composite Systems	20

2	State Estimation	22
2.1	Classical Parameter Estimation	22
2.2	Quantum State Estimation	28
2.2.1	State space and parameterization	30
2.2.2	Measurement	32
2.2.3	Construction of the estimate	35
2.2.4	Efficiency of the estimate	39
3	Complementarity and State Estimation	47
3.1	Preliminaries	48
3.1.1	Quality of an estimation scheme	49
3.1.2	Compared estimation schemes	50
3.2	Optimality of Quasi-orthogonal Measurements	51
3.2.1	Unconstrained estimate	51
3.2.2	Constrained estimate	55
3.2.3	Average information gain	60
3.2.4	Examples	62
3.3	Evaluation of von Neumann Measurements	69
	Conclusion	76
A	Some Integrals	79
A.1	Integration on the Probability Simplex	79
A.2	Integration on the Unitary Group	82
A.3	Summation	84
B	Sanov's Theorem for Unequal Sample Sizes	86

Introduction

The probably most widely known and challenging property of quantum theory is that it is non-deterministic. Even if complete knowledge about the state of a quantum system, acquired by determining a certain property of the system, is at hand, there are observable quantities that cannot be predicted with certainty from the obtained information. This inherent stochastic nature of quantum mechanics makes the determination of the unknown state of a quantum system a statistical problem and in general we can only estimate the unknown state of the system. The fundamental question in quantum state estimation is to find most efficient estimation procedure.

A scheme to estimate an unknown quantum state includes usually the following steps: Parameterization of the set of possible true states of the system, the selection of measurements to be performed on the given copies of the state and the construction of an estimate from the results of the measurements. The main challenge, however lies in the right choice of the measurements that are performed to obtain statistical data about the system. For a good estimate it is necessary to collect statistical data from measurements on several identical copies of the unknown state. In [28] separate measurements of observables with non-degenerate spectrum on the individual available copies of the state were studied. It was shown that in this setting complementary observables obtain asymptotically the optimal information gain. A similar result was obtained in [21], where it was shown that the

same holds true if the quality of the estimation scheme is measured by the average mean quadratic error matrix of the estimate. Nevertheless, for the full estimation of a quantum state a minimum number of different complementary observables is needed. A further result of [28] was the proof that in the case of quantum systems of dimension p^k , where p is a prime and $k \in \mathbb{N}$, sufficiently many complementary observables exists. The existence in the general case is a popular unsolved problem. From the relation of observables to Abelian subalgebras of the observable algebra, the concept of complementarity can be extended to subalgebras in general. Recently there were developments which showed existence and non-existence of the maximal number of complementary subalgebras in certain quantum systems [23, 18, 17] and their relevance in state estimation [22].

In the present work we study the role of complementary for measurement schemes that use general separate measurements (POVMs) on the given copies of a quantum system. We consider the following framework: The unknown state is one of the possible states of the system and eventual a priory knowledge is reflected by a prior probability distribution on the set of states, which is required to be invariant under unitary conjugation. We infer about the state by a number of different measurements, each of them performed separately on a certain number of identical copies of the state. From the obtained data an unconstrained point estimate for the state is obtained by linear inversion of the relation between the state and the measurement probabilities. From the unconstrained estimate, that may take values outside the set of states, a constrained estimate that always gives useful results can be constructed. The efficiency of such an estimation scheme is evaluated by the mean quadratic error matrix of the estimate averaged over the possible true states and different measurement schemes can be compared by the determinant of this matrix. In this setting a measurement scheme that applies complementary (or quasi-orthogonal) measurements can be compared to another scheme, if their measurements are related by unitaries. The main results of the thesis show that measurement schemes that consist of complementary

measurements are optimal with respect to the applied measure of efficiency. This result can be formulated for finite sample size for the unconstrained estimate and in an asymptotic setting for the constrained estimate. In many important cases a measurement scheme that uses complementary measurements can be connected to complementary subalgebras and their importance for state estimation is discussed.

The thesis is structured in the following way: Chapter 1 contains an introduction to the formalism of quantum probability theory. Complementarity of observables, subalgebras and general measurements is defined and the cases where existence of complete sets of complementary subalgebras is known are given. Chapter 2 gives an introduction to classical and quantum estimation theory. In particular the setting of the estimation scheme under study is explained and the different parts of the estimation schemes we apply in the thesis are discussed in detail. The measures we use to quantify the efficiency of an estimation scheme and additionally the measure used in [28] is introduced. In Chapter 3 the main results of the thesis are presented and their relation to the work in [28] is examined. The results are applied to the cases where existence of complementary subalgebras is known. Finally for a certain unitarily invariant prior distribution on the set of true states the mean quadratic error matrix and its average are explicitly evaluated in the case of a subclass of von Neumann measurements. The Conclusion gives again a summary of the results of the thesis.

Chapter 1

Mathematical Description of Finite Quantum Systems

1.1 Quantum Probability Theory

Quantum probability theory is a non-commutative generalization of classical probability theory. In classical probability theory events are described by elements of a σ -algebra, which has the algebraic structure of a Boolean lattice. In quantum probability theory algebras of operators on a Hilbert space are considered. The set of projections in such an algebra obeys a lattice structure as well and projections play the role of events in quantum probability. Probability measures are replaced by positive linear functionals on the algebra. Throughout the thesis we consider only finite dimensional quantum systems described by a subalgebra of the matrix algebra $M_n(\mathbb{C})$ on a Hilbert space with $\dim(\mathcal{H}) = n$. In the following we give an introduction to quantum probability in the finite dimensional setting.

1.1.1 The observable algebra

The finite quantum systems regarded in the thesis are described by subalgebras of the algebra $\mathcal{B}(\mathcal{H})$ of operators on a finite dimensional Hilbert space. $\mathcal{B}(\mathcal{H})$ satisfies the axioms of an unital $*$ -algebra where the $*$ -operation is given by taking adjoints. By fixing a basis in \mathcal{H} , the algebra $\mathcal{B}(\mathcal{H})$ can be identified with the algebra $M_n(\mathbb{C})$ of $n \times n$ matrices with complex entries.

Orthogonal projections can be characterized by the algebraic property $P = P^2 = P^*$ and they obey additionally the structure of a lattice. To examine the lattice structure we note that a projection $P \in \mathcal{B}(\mathcal{H})$ and the linear subspace of \mathcal{H} that is its range can be identified $P \equiv \text{Im}(P)$. The two equivalent conditions

$$\begin{aligned} (i) \quad P \leq Q & \text{ if } PQ = P \\ (ii) \quad P \leq Q & \text{ if } \text{Im}(P) \subset \text{Im}(Q) \end{aligned} \tag{1.1}$$

define a **partial ordering** on the set $\mathcal{P}(\mathcal{B})$ of projections in $\mathcal{B}(\mathcal{H})$. From condition (ii) it is easy to see that $P \geq Q$ and $Q \geq P$ implies $P = Q$ and that $P \geq Q$ and $Q \geq R$ implies $P \geq R$. A projection P is called a **minimal projection** if $Q \leq P$ implies either $Q = 0$ or $Q = P$.

From the partial ordering (1.1) two algebraic operations \vee and \wedge on $\mathcal{P}(\mathcal{B})$ can be defined: The intersection $\text{Im}(P) \cap \text{Im}(Q)$ is also a linear subspace of \mathcal{H} and it contains all linear subspaces that are smaller than $\text{Im}(P)$ and smaller than $\text{Im}(Q)$. As such the projection on $\text{Im}(P) \cap \text{Im}(Q)$ provides the greatest upper bound of the two projections P and Q and we denote it as $P \wedge Q$. By definition $\text{span}\{\text{Im}(P) \cup \text{Im}(Q)\}$ is the smallest linear subspace that contains $\text{Im}(P)$ and $\text{Im}(Q)$. Thus the projection onto $\text{span}\{\text{Im}(P) \cup \text{Im}(Q)\}$ is the smallest upper bound of the two projections P and Q and we denote it as $P \vee Q$.

With the operations \vee and \wedge the set $\mathcal{P}(\mathcal{B})$ forms a complete¹ **lattice** and \vee and \wedge are commutative and associative. The lattice $\mathcal{P}(\mathcal{B})$ has several further properties:

- The identity operator I is the largest element in $\mathcal{P}(\mathcal{B})$ with respect to the partial ordering (1.1) and it acts as the identity with respect to \wedge . Zero is the smallest in $\mathcal{P}(\mathcal{B})$ with respect to the partial ordering (1.1) and it acts as the zero element with respect to \vee .
- The projection $(I - P)$ fulfills $P \wedge (I - P) = 0$ and $P \vee (I - P) = I$. Therefore $(I - P)$ is called the complement of P and the lattice is called **complemented**.
- The lattice is not distributive, however **modular** if

$$P \leq Q \quad \Rightarrow \quad P \vee (Q \wedge R) = (P \vee Q) \wedge (P \vee R) = Q \wedge (P \vee R) \quad (1.2)$$

Additional to the full algebra $\mathcal{B}(\mathcal{H})$ we can consider $*$ -subalgebras $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ that contain the identity $I \in \mathcal{B}(\mathcal{H})$. A subalgebra \mathcal{A} induces a sublattice $\mathcal{P}(\mathcal{A}) \subset \mathcal{P}(\mathcal{B})$ as well. For proof it is sufficient to see that the largest upper bound of a set of projections $\{P_i : 1 \leq i \leq k\}$ is contained in the subalgebra \mathcal{A} . The largest upper bound $\sup\{P_i : 1 \leq i \leq k\}$ is given by the projection onto the span of the corresponding subspaces $Im(P_i)$. Since $Ker(P + Q) = Ker(P) \cap Ker(Q)$ for positive operators, $span\{Im(P_i) : 1 \leq i \leq k\}$ is identical with $Im(\sum P_i)$. Since $Q := \sum P_i$ is an element of the subalgebra \mathcal{A} , the projection onto $Im(Q)$, and thus the largest upper bound of the set $\{P_i : 1 \leq i \leq k\}$ is an element of \mathcal{A} as well. In this sense a subalgebra of $\mathcal{B}(\mathcal{H})$ can be considered a subsystem of $\mathcal{B}(\mathcal{H})$.

¹A lattice is called complete if for every subset $A \subset \mathcal{P}$ the greatest upper bound, $\sup(A)$, and the smallest lower bound, $\inf(A)$, exist.

A lattice that is not only modular but distributive is called **Boolean**². For the lattice $\mathcal{P}(\mathcal{A})$ of a subalgebra $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ this is the case if \mathcal{A} is commutative. Then it is generated by the minimal projections of \mathcal{A} , which satisfy $P_i P_j = 0$ and $\sum P_i = I$. In this case $\mathcal{P}(\mathcal{A})$ is isomorphic as a lattice to the σ -algebra of subsets of the index set $I = \{1, \dots, k\}$ we used to number the minimal projections in $\mathcal{P}(\mathcal{A})$.

1.1.2 States and measurements

In analogy to a probability measure in a classical probability space, a **state** on the algebra $\mathcal{B}(\mathcal{H})$ is defined as a linear functional $\phi : \mathcal{B}(\mathcal{H}) \mapsto \mathbb{C}$ such that

$$\phi(A) \geq 0 \quad \text{if } A \geq 0 \quad \text{and} \quad \phi(I) = 1 \quad (1.3)$$

The algebra $\mathcal{B}(\mathcal{H})$ can be equipped with the Hilbert-Schmidt inner product

$$\langle A, B \rangle = \frac{1}{n} \text{Tr}(A^* B) \quad (A, B \in \mathcal{B}(\mathcal{H})). \quad (1.4)$$

and with this inner product it is a Hilbert space itself. A state is an element of the dual of $\mathcal{B}(\mathcal{H})$ and it can be represented by a **density operator** $\rho \in \mathcal{B}(\mathcal{H})$ as $\phi(A) = \text{Tr}(\rho A)$.

By the properties imposed on a state the density operator ρ fulfills

$$\rho \geq 0 \quad \text{and} \quad \text{Tr } \rho = 1. \quad (1.5)$$

From (1.5) it follows that a density matrix is a self-adjoint operator with eigenvalues $\{\lambda_i : 1 \leq i \leq n, \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1\}$. A state on $\mathcal{B}(\mathcal{H})$ is called **pure** if the density operator is a rank one projection, i.e. it has a single non zero eigenvalue equal to one.

²A σ -algebra of sets is an example of a Boolean lattice.

An important role in quantum probability theory is played by **self-adjoint** operators $A = A^*$. By the spectral theorem a self-adjoint operator A has an orthogonal basis of eigenvectors and therefore a decomposition

$$A = \sum_i a_i P_i \quad (1.6)$$

where the $a_i \in \mathbb{R}$ are the eigenvalues of A and the P_i are orthogonal projections on the corresponding mutually orthogonal eigenspaces.

The eigenprojections of a self-adjoint operator fulfill $P_i P_j = 0$ and $\sum_i P_i = I$ and they generate an Abelian subalgebra $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$. The lattice $\mathcal{P}(\mathcal{A})$ is Boolean and a σ -additive map defined by $M_A : a_i \mapsto P_i$ gives a lattice isomorphism between the σ -algebra over the set of eigenvalues of A and the lattice $\mathcal{P}(\mathcal{A})$ of projections in \mathcal{A} . By linearity and the conditions (1.3) the state ϕ induces a probability measure on the set of eigenvalues of A through the map M_A . We get the probabilities

$$\text{Prob}(a_i) = \phi(P_i) = \text{Tr}(\rho P_i). \quad (1.7)$$

Thus the operator A can be associated with a random experiment that has the eigenvalues of A as outcomes which appear with the probabilities in (1.7). Then the expectation value of the outcomes is given by the value of the functional ϕ at A :

$$\langle A \rangle = \sum_i a_i \text{Prob}(a_i) = \phi(A) \quad (1.8)$$

In the description of a physical system self-adjoint operators are associated with observable quantities of the system and therefore called **observables**.

The concept of a random experiment in quantum theory can be extended to a more general form given by positive operator valued measures (**POVM**). A finite discrete POVM

is a σ -additive map from the σ -algebra over a finite set $\{a_1, a_2, \dots, a_d\}$ of outcomes into the set of positive operators on $\mathcal{B}(\mathcal{H})$. It can be defined by the assignment $a_i \rightarrow E_i$ such that $M := \{E_i : 1 \leq i \leq d\}$ is a set of $\#(M) = d$ positive operators $E_i \in \mathcal{B}(\mathcal{H})$ satisfying the conditions

$$E_i \geq 0 \quad \text{and} \quad \sum_{i=1}^d E_i = I. \quad (1.9)$$

Again, by linearity and the conditions (1.3) the state induces a probability measure on the set $\{a_1, a_2, \dots, a_d\}$ and $\phi(E_i)$ can be interpreted as the probability that the outcome a_i appears if the system is in the state ρ :

$$p_i := \text{Prob}(a_i) = \text{Tr}(\rho E_i). \quad (1.10)$$

Since in the following the values of the outcomes will not be important, we will describe a POVM simply by the set of operators $M := \{E_i : 1 \leq i \leq d\}$ and refer to it as a **measurement**. Performance of a measurement shall simply mean the performance of the random experiment associated with the POVM. If all operators in a POVM are projections, it corresponds to the measurement of an observable and commonly the notion **von Neumann measurement** is used.

Example 1.1.1 In the case of a qubit, i.e. $\dim(\mathcal{H}) = 2$, consider the operators

$$\begin{aligned} E_1 &= \frac{1}{4}I + \frac{1}{4\sqrt{3}}(\sigma_1 + \sigma_2 + \sigma_3) & E_2 &= \frac{1}{4}I + \frac{1}{4\sqrt{3}}(\sigma_1 - \sigma_2 - \sigma_3) \\ E_3 &= \frac{1}{4}I + \frac{1}{4\sqrt{3}}(-\sigma_1 + \sigma_2 - \sigma_3) & E_4 &= \frac{1}{4}I + \frac{1}{4\sqrt{3}}(-\sigma_1 - \sigma_2 + \sigma_3) \end{aligned} \quad (1.11)$$

where the σ_i are the Pauli matrices given in (1.15). The operators in $M := \{E_i : 1 \leq i \leq 4\}$ fulfill $\sum_{i=1}^4 E_i = I$. With the anticommutator $\{\sigma_i, \sigma_j\} := \sigma_i \sigma_j + \sigma_j \sigma_i = 0$ ($i \neq j$) of the Pauli matrices it follows that $E_i^2 = \frac{1}{2} E_i$ and therefore the E_i are subnormalized projections, thus positive and the set $M := \{E_i : 1 \leq i \leq 4\}$ forms a POVM. \diamond

As an important difference to classical theory, in quantum probability theory the performance of a measurement is accompanied by the so called reduction of the state. If the outcome a_i appears after a measurement is completed, this gain of information about the system needs to be reflected in a change of the state of the system. Therefore after a measurement in which the outcome a_i appeared, the state of the system is given by

$$\hat{\rho} = \frac{V_i \rho V_i^*}{\text{Tr}(V_i \rho V_i^*)} \quad (1.12)$$

where³ $E_i = V_i V_i^*$. In the case of a von Neumann measurement the operators V_i are projections and the repeated performance of the same measurement on the system will give the same result a_i with probability one.

To illustrate the difference in the measurement process in the classical and in the quantum case, one may assume that the actual outcome of a measurement is not recorded, but we know that the system is in one of the post-measurement states with the according probability. Then the state after the measurement can be described by the convex combination

$$\hat{\rho} = \sum_i \text{Prob}(a_i) \hat{\rho}_i \quad (1.13)$$

In this situation the measurement process can be considered as a state transformation (see e.g. [20]), i.e. a completely positive map $\mathcal{E} : \mathcal{B}(\mathcal{H}) \mapsto \mathcal{B}(\mathcal{H})$ that takes $\rho \mapsto \hat{\rho}$. The map \mathcal{E} can be written in the form $\mathcal{E}(\rho) = \sum_i V_i \rho V_i$ and in this context the V_i are called the Kraus operators of \mathcal{E} . The difference in observation of a classical and a quantum system becomes obvious: In the classical case \mathcal{E} is always the identity, observation does not influence the state of the system. In quantum mechanics \mathcal{E} is always different from the identity and the system is only left unchanged for certain states, e.g. for a von Neumann measurement

³ The actual form of the V_i depends on the physical realization of the POVM.

if ρ and the measured observable A commute. In this sense the order in which certain random variables are measured does not matter in the classical case. In the quantum case, due to the change of the state during a measurement, the outcome probabilities in measurements depend in general on preceding observations. In particular the outcome of the measurement of an observable B is only predictable with certainty from the outcome of a preceding measurement of an observable A if A and B commute.

1.2 Complementarity

By the non-commutativity of observations as described in the preceding section, knowledge obtained from a measurement of an observable A implies uncertainty in a subsequent measurement of observables that do not commute with A . For this reason in early quantum theory non-commuting observables were said to provide complementary information about a system. Motivated by this, two observables are called complementary if knowledge of one of them implies maximal uncertainty about the other. Complementarity of observables is related to an orthogonality relation between the Abelian subalgebras generated by their eigenprojections and from this geometric property complementarity of observables can be generalized to subalgebras [19] and POVMs .

Several information theoretic consequences of complementarity have been studied. Complementarity leads to extremal bounds in uncertainty relations and entropic uncertainty relations. An overview and applications can be found in [19]. In the context of state estimation the role of complementary observables was studied in [28]. We will examine the role of complementary measurements in Chapter 3.

1.2.1 The traceless subspace

Before we give a definition of complementarity, we examine the vector space structure of the algebra $\mathcal{B}(\mathcal{H})$. If $\dim(\mathcal{H}) = n$ we can choose an orthonormal basis in $\mathcal{B}(\mathcal{H})$ that consists of self-adjoint operators of the form $\{\sigma_i : 0 \leq i \leq n^2 - 1\}$, where $\sigma_0 \equiv I$ denotes the identity. The basis elements fulfill

$$\langle \sigma_i, \sigma_j \rangle = 0 \quad (i \neq j), \quad \langle \sigma_i, \sigma_i \rangle = 1. \quad (1.14)$$

By the form (1.4) of the Hilbert Schmidt inner product the orthogonality condition in (1.14) implies that $\mathcal{S} := \text{span}\{\sigma_i : 1 \leq i \leq n^2 - 1\}$ is the $(n^2 - 1)$ -dimensional linear subspace of traceless operators in $\mathcal{B}(\mathcal{H})$.

Example 1.2.1 In the case of $\dim(\mathcal{H}) = 2$ the **Pauli matrices**

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (1.15)$$

together with the identity form a standard basis of $M_2(\mathbb{C})$. They fulfill additionally the algebraic relations $\sigma_1\sigma_2 = -\sigma_2\sigma_1 = i\sigma_3$, $\sigma_2\sigma_3 = -\sigma_3\sigma_2 = i\sigma_1$ and $\sigma_3\sigma_1 = -\sigma_1\sigma_3 = i\sigma_2$. A similar basis for $\dim(\mathcal{H}) > 2$ is given by the so called Gell-Mann matrices (see e.g. [24]). \diamond

With respect to the above basis we can expand an arbitrary operator $A \in \mathcal{B}(\mathcal{H})$ as

$$A = \frac{\text{Tr } A}{n} I + \boldsymbol{\alpha} \quad \text{where} \quad \boldsymbol{\alpha} \in \mathcal{S} \quad (1.16)$$

and we call $\boldsymbol{\alpha}$ the **Bloch vector of the operator** A . This notation is in dependence on the notation of the case of a density operator in $\dim(\mathcal{H}) = 2$ as described in Chapter

2.2.1⁴. Of particular interest is the case of self-adjoint operators: By $\text{Tr}(A^*B) = \overline{\text{Tr}(B^*A)}$ the inner product $\langle A, B \rangle$ between self adjoint operators A and B takes always real values. Thus self adjoint operators form a real vector space over the basis $\{\sigma_i : 0 \leq i \leq n^2 - 1\}$ and α can be written as a vector with real components:

$$\alpha \in \mathbb{R}^{(n^2-1)} \quad \text{if } A = A^*.$$

Note that self adjoint operators do not form an algebra, in particular the product of two self adjoint operators is not self adjoint if the operators do not commute.

1.2.2 Complementary observables, algebras and measurements

As mentioned at the beginning of this chapter, complementarity arises from the fact that two non-commuting observables of a quantum system are not jointly measurable with arbitrary precision. As an extremal case of this, two observables A and B with eigenprojections $\{P_i : 1 \leq i \leq n\}$ and $\{Q_i : 1 \leq i \leq n\}$ and non-degenerate spectrum are called **complementary** if

$$\text{Tr}(P_i Q_j) = \frac{1}{n} \quad \forall i, j. \quad (1.17)$$

Complementarity has the following consequence for the measurement of the observables A and B on a quantum system: If the observable $A = \sum a_i P_i$ is determined in a measurement, which implies that the state after the measurement is $\hat{\rho} = P_i$ if the obtained value is a_i , a subsequent measurement of observable B will give any of the possible outcomes of B with probability $1/n$. In this sense exact knowledge of A implies complete uncertainty about B .

Complementarity is related to orthogonality in the traceless subspace \mathcal{S} (see [19]). If

⁴Note that unlike in the case of a density operator, when $\text{Tr}(\rho) = 1$, the vector α does not define the operator A completely.

we are given two operators A_1 and A_2 , their components in the traceless subspace \mathcal{S} are given by $(A_i - \frac{\text{Tr}(A_i)}{n} I)$ ($i = 1, 2$). If we look at the inner product

$$\begin{aligned} \frac{1}{n} \text{Tr} \left(\left(A_1 - \frac{\text{Tr}(A_1)}{n} I \right) \left(A_2 - \frac{\text{Tr}(A_2)}{n} I \right) \right) \\ = \frac{1}{n} \left(\text{Tr}(A_1 A_2) - \frac{2}{n} \text{Tr}(A_1) \text{Tr}(A_2) + \frac{1}{n} \text{Tr}(A_1) \text{Tr}(A_2) \right) \\ = \frac{1}{n} \left(\text{Tr}(A_1 A_2) - \frac{1}{n} \text{Tr}(A_1) \text{Tr}(A_2) \right) \end{aligned} \quad (1.18)$$

of their traceless components we find that they are orthogonal if and only if

$$\text{Tr}(A_1 A_2) = \frac{1}{n} \text{Tr}(A_1) \text{Tr}(A_2). \quad (1.19)$$

For rank 1 projections P_i and Q_j this is equal to condition (1.17). Thus complementarity of two observables is equivalent to the condition that the traceless components of the maximal Abelian subalgebras \mathcal{A}_1 and \mathcal{A}_2 generated by the eigenprojections of the of A_1 and A_2 are orthogonal. This gives raise to the generalization of complementarity to general subalgebras of $\mathcal{B}(\mathcal{H})$: Two subalgebras $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{B}(\mathcal{H})$ are called **complementary** or **quasi-orthogonal**, denoted as $\mathcal{A}_1 \perp_0 \mathcal{A}_2$, if they fulfill the equivalent conditions

$$\begin{aligned} (i) \quad & \mathcal{A}_1 \ominus \mathbb{C}I \perp \mathcal{A}_2 \ominus \mathbb{C}I \\ (ii) \quad & \text{Tr}(A_1 A_2) = \frac{1}{n} \text{Tr}(A_1) \text{Tr}(A_2) \quad \forall A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2. \end{aligned} \quad (1.20)$$

Since we required a subalgebra to contain the span of the identity, quasi-orthogonality is the maximum level of orthogonality two subalgebras can obtain.

From the viewpoint of quasi-orthogonality we can extend complementarity of observables to complementarity of POVMs. We call two measurements (POVMs) **quasi-orthogonal**, if the linear subspaces spanned by $M_1 = \{E_i : 1 \leq i \leq d_1\}$ and $M_2 = \{F_i : 1 \leq i \leq$

$d_2\}$ are quasi-orthogonal. This can be formulated by the following equivalent conditions:

$$\begin{aligned}
(i) \quad & \text{span}\{E_i\} \ominus \mathbb{C}I \perp \text{span}\{F_i\} \ominus \mathbb{C}I \\
(ii) \quad & \text{Tr}(E_i F_j) = \text{Tr}(E_i)\text{Tr}(F_j) \quad \forall i, j.
\end{aligned}
\tag{1.21}$$

While a von Neumann measurement is naturally related to a subalgebra, in the case of a POVM this is not necessarily the case. For example, the four operators $\{1/4(I \pm \sigma_x), 1/4(I \pm \sigma_y)\}$ form a POVM. By $[\sigma_x, \sigma_y] = i\sigma_z$ the algebra they generate is $M_2(\mathbb{C})$ while their span is only a subspace of $M_2(\mathbb{C})$.

1.2.3 Complete sets of complementary subalgebras

An important question for the state estimation problem discussed in the thesis is whether it is possible to find POVMs that are mutually quasi-orthogonal and at the same time span the whole algebra $\mathcal{B}(\mathcal{H})$. In the most relevant cases this is related to the question if there exist subalgebras $\mathcal{A}_i \subset M_n(\mathbb{C})$ such that $\mathcal{A}_i \perp_0 \mathcal{A}_j$ ($i \neq j$) and $\text{span}\{\mathcal{A}_i\} = M_n(\mathbb{C})$. An obvious condition for the \mathcal{A}_i is that the dimensions of their traceless components have to sum up to the dimension of traceless component of $M_n(\mathbb{C})$. If there is a set of complementary subalgebras such that this is the case we call it a **complete set**. Before we give a list of examples where the existence of a complete set of complementary subalgebras is known, let us describe the explicit structure of a $*$ -subalgebra \mathcal{A} of the matrix algebra $M_n(\mathbb{C})$:

By fixing a basis of the Hilbert space \mathcal{H} we can identify $\mathcal{B}(\mathcal{H})$ with the matrix algebra $M_n(\mathbb{C})$. The choice of a basis is arbitrary, nevertheless by the choice of a convenient basis the structure of the matrices in $M_n(\mathbb{C})$ representing the elements of a subalgebra $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ simplifies. A change between orthogonal bases of \mathcal{H} corresponds to a unitary operator $U \in U_{\mathbb{C}}(n)$ and the representation of \mathcal{A} in $M_n(\mathbb{C})$ for two different bases, denoted

here by \mathcal{A} and $\tilde{\mathcal{A}}$, is related by unitary conjugation:

$$\tilde{\mathcal{A}} = U\mathcal{A}U^* \quad (1.22)$$

where we define unitary conjugation of a set by unitary conjugation of its elements. The structure of subalgebras of $M_n(\mathbb{C})$ can be derived from results from non-commutative algebra (see e.g. [5]): First let us note that $M_n(\mathbb{C})$ is a central simple algebra, i.e. it has no proper ideal and the only elements that commute with all operators $A \in M_n(\mathbb{C})$ are scalar multiples of the identity. Conversely any finite dimensional simple algebra is isomorphic to a full matrix algebra. In the special case when \mathcal{A} is a central simple subalgebra of $M_n(\mathbb{C})$, i.e. \mathcal{A} is isomorphic to a full matrix algebra itself, by the centralizer theorem (see [5]) with an appropriate choice of basis

$$M_n(\mathbb{C}) \equiv \mathcal{A} \otimes \mathcal{A}' \quad (1.23)$$

where $\mathcal{A}' := \{A \in M_n(\mathbb{C}) : AB = BA \ \forall B \in \mathcal{A}\}$ is the commutant of \mathcal{A} in $M_n(\mathbb{C})$.

It can be shown that a $*$ -subalgebra \mathcal{A} is semisimple, therefore by the Artin Wedderburn structure theorem (see e.g. [5]) it is isomorphic to the direct sum of full matrix algebras:

$$\mathcal{A} \simeq \bigoplus_k M_{n_k}(\mathbb{C}) \quad (1.24)$$

We can choose a maximal family of minimal projections $\{P_1, P_2, \dots, P_{n_k}\}$ in the $M_{n_k}(\mathbb{C})$ such that $P_i P_j = 0 \ \forall i \neq j$ and $\sum_i P_i = I_{n_k}$, where I_{n_k} is the identity matrix in $M_{n_k}(\mathbb{C})$. The isomorphic images of the P_i in \mathcal{A} correlate with bases of the underlying Hilbert space \mathcal{H} . In this bases the algebra \mathcal{A} is represented by block diagonal matrices and $\mathcal{A} \equiv \bigoplus_k \mathcal{A}^{(k)}$, where $\mathcal{A}^{(k)} \simeq M_{n_k}(\mathbb{C})$. In this representation the algebra $\mathcal{A}^{(k)}$ is subalgebra of $M_{m_k}(\mathbb{C})$ isomorphic to a full matrix algebra, where m_k is the size of the block corresponding to $\mathcal{A}^{(k)}$. By (1.23), in an appropriate basis $\mathcal{A}^{(k)} \equiv M_{n_k}(\mathbb{C}) \otimes I_{d_k}$. Thus in an appropriate

basis the subalgebra \mathcal{A} has the form

$$\mathcal{A} \equiv \bigoplus_k M_{n_k}(\mathbb{C}) \otimes I_{d_k} \quad (1.25)$$

Let us discuss the following basic examples of $*$ -subalgebras of $M_n(\mathbb{C})$ that contain the identity matrix:

- An Abelian subalgebra \mathcal{A} of $M_n(\mathbb{C})$ is isomorphic to the direct product $\mathcal{A} \simeq \bigoplus_{k=1}^m \mathbb{C}$, where $m \leq n$, where the later is equivalent to the algebra of $m \times m$ diagonal matrices. As shown above, with the choice of an appropriate basis $\mathcal{A} \equiv \bigoplus_{k=1}^m \mathbb{C} \otimes I_{d_k}$ such that $\sum_k d_k = n$. \mathcal{A} is called **maximal Abelian** if $m = n$ and in this case the minimal projections of \mathcal{A} are of rank one and there are exactly n direct summands.
- If the dimension $n = kl$ is not prime, the algebra $M_n(\mathbb{C})$ is isomorphic to a tensor product algebra $M_n(\mathbb{C}) \simeq M_k(\mathbb{C}) \otimes M_l(\mathbb{C})$. Naturally, $M_k(\mathbb{C}) \otimes I_l$ (respectively $I_k \otimes M_l(\mathbb{C})$) are subalgebras of $M_n(\mathbb{C})$ isomorphic to the corresponding full matrix algebras.
- A subalgebra \mathcal{A} is called **homogenous** if all minimal projections in \mathcal{A} have the same rank. If two homogenous subalgebras \mathcal{A}_1 and \mathcal{A}_2 of $M_n(\mathbb{C})$ are isomorphic, we get in an appropriate basis by (1.25)

$$\bigoplus_k M_{n_k}(\mathbb{C}) \otimes I_{d_k} \equiv \mathcal{A}_1 \simeq U \mathcal{A}_2 U^* \equiv \bigoplus_k M_{n'_k}(\mathbb{C}) \otimes I_{d'_k} \quad (1.26)$$

with some unitary $U \in M_n(\mathbb{C})$. Since \mathcal{A}_1 and \mathcal{A}_2 were assumed to be isomorphic we have $n_k = n'_k$ and then homogeneity implies $d_k = d'_k = d$. Therefore \mathcal{A}_1 and \mathcal{A}_2 differ only by a choice of basis of \mathcal{H} and there is some $V \in U(n)$ such that $\mathcal{A}_1 = V \mathcal{A}_2 V^*$. In particular, a subalgebra isomorphic to a full matrix algebra, and especially $M_n(\mathbb{C})$, is homogenous.

Let us now return to complementary subalgebras. In the following we give a comprehensive list of the cases where the existence of a complete set of complementary subalgebras of $M_n(\mathbb{C})$ is known:

Example 1.2.2 Maximal Abelian subalgebras are generated by n minimal projections of $M_n(\mathbb{C})$. The dimension of the traceless subspace of a maximal Abelian subalgebra is $n - 1$. The dimension of the traceless subspace of $M_n(\mathbb{C})$ is $n^2 - 1$, thus an upper bound on the maximum number of complementary maximal Abelian subalgebras $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_m \subset M_n(\mathbb{C})$ such that $\mathcal{A}_i \perp_0 \mathcal{A}_j$ ($i \neq j$) is given by $m = n + 1$.

Existence of complementary maximal Abelian subalgebras is equivalent to existence of mutually unbiased bases of the Hilbert space \mathcal{H} : The minimal projections $\{P_i\}_{i=1}^n$ in some maximal Abelian subalgebra \mathcal{A}_1 and $\{Q_j\}_{j=1}^n$ in some maximal Abelian subalgebra \mathcal{A}_2 can be identified with orthonormal bases $\{e_i\}_{i=1}^n$ and $\{f_j\}_{j=1}^n$ of \mathcal{H} . Then the condition of quasi-orthogonality of \mathcal{A}_1 and \mathcal{A}_2 is equivalent to

$$|(e_i, f_j)| = \frac{1}{\sqrt{n}}. \quad (1.27)$$

Bases that fulfill this condition are called **mutually unbiased**. A complete set of $m = n + 1$ mutually unbiased bases respectively complementary maximal Abelian subalgebras in $M_n(\mathbb{C})$ is known to exist in the cases when $n = p^k$ is the power of a prime p [28]. In any other cases their existence is an open problem. \diamond

Example 1.2.3 Now consider a system of l qubits described by $\mathcal{B}(\mathcal{H}) = M_2^1(\mathbb{C}) \otimes M_2^2(\mathbb{C}) \otimes \dots \otimes M_2^l(\mathbb{C})$ with $n = \dim(\mathcal{H}) = 2^l$. We can construct an orthonormal basis of $M_{2^l}(\mathbb{C})$ from the tensor product

$$\boldsymbol{\sigma}^{(k)} = \sigma_{i_1} \otimes \sigma_{i_2} \otimes \dots \otimes \sigma_{i_l} \quad \text{where } (i_j = 0, 1, 2, 3) \quad (1.28)$$

of elements of a orthogonal basis of $M_2(\mathbb{C})$ with $\sigma_0 = I$. There are 2^{2l} such operators and by $Tr(A \otimes B) = Tr(A)Tr(B)$ they are traceless if at least one of the indices i_j is different from zero. The remaining element, let us denote it by $\sigma^{(0)}$, is the identity. Let us consider the case when the σ_i are the Pauli matrices given in (1.15). The eigenvalues of the Pauli matrices are ± 1 , thus each of the traceless operators in (1.28) has eigenvalues ± 1 and by induction on l it is easy to see that they have multiplicity $n/2$ and fulfill $(\sigma^{(k)})^2 = I$. Consequently the operators

$$P_{\pm}^{(k)} := \frac{1}{2}(I \pm \sigma^{(k)}) \quad (1 \leq k \leq n^2 - 1) \quad (1.29)$$

have eigenvalues 0 and 1 with multiplicity $n/2$, hence they are projections of rank $n/2$. They correspond to the measurement of observables with two distinct eigenvalues of multiplicity $n/2$. The algebras \mathcal{A}_k generated by $\{P_+^{(k)}, P_-^{(k)}\}$ are homogenous Abelian subalgebras and, by $P_+^{(k)} + P_-^{(k)} = I$, the traceless component of the \mathcal{A}_k is one dimensional. Since the $\sigma^{(k)}$ form an orthonormal basis of $\mathcal{B}(\mathcal{H})$, the \mathcal{A}_k form a complete set of $m = n^2 - 1$ complementary subalgebras. \diamond

Example 1.2.4 If the dimension of \mathcal{H} is given as $\dim(\mathcal{H}) = q^k$, the algebra $M_n(\mathbb{C})$ is isomorphic to the tensor product

$$M_{q^k}(\mathbb{C}) \simeq \bigotimes_{i=1}^k M_q(\mathbb{C}). \quad (1.30)$$

Then $M_n(\mathbb{C})$ contains full matrix subalgebras $\mathcal{A} \subset M_{q^k}(\mathbb{C})$ isomorphic to $M_q(\mathbb{C})$. The traceless subspace of the algebra \mathcal{A} has dimension $q^2 - 1$ thus an upper bound to the maximum number of such complementary subalgebras $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_m$ such that $\mathcal{A}_i \perp_0 \mathcal{A}_j$ ($i \neq j$) is given by $m = (q^{2k} - 1)/(q^2 - 1)$ and in this case they span the whole algebra. In [17] it was shown that if $q = p^l$ is a prime power with $p \geq 3$ there exists a complete set of such

subalgebras. ◇

Example 1.2.5 In the case of $M_4(\mathbb{C}) = M_2(\mathbb{C}) \otimes M_2(\mathbb{C})$, i.e. $p=2$ and $k=2$, it was shown in [23] that the upper bound of 5 quasi-orthogonal subalgebras $\mathcal{A} \simeq M_2(\mathbb{C})$ cannot be achieved. However it is possible to choose four quasi-orthogonal subalgebras $\mathcal{A}_i \simeq M_2(\mathbb{C})$ ($1 \leq i \leq 4$) and a maximal Abelian subalgebra \mathcal{A}_5 on the remaining orthogonal complement in \mathcal{S} [18]. ◇

Example 1.2.6 In the context of Example 1.2.4 when $q = p^l$ with a prime $p \geq 3$ the \mathcal{A}_i , as the isomorphic image of $M_q(\mathbb{C})$, obey a further decomposition into mutually quasi-orthogonal Abelian subalgebras. Since q was assumed to be a prime power, we can apply Example 1.2.2 to the algebras $M_q(\mathbb{C})$. In $M_q(\mathbb{C})$ we can find $m_M = q + 1$ complementary maximal Abelian subalgebras, and the decomposition of the algebras \mathcal{A}_i is obtained from the isomorphic images of this subalgebras. This subalgebras are not maximal Abelian in $M_{q^k}(\mathbb{C})$, from (1.25) we can see that their minimal projections are of rank $r = q^{(k-1)}$. Thus we obtain a decomposition of $M_{q^k}(\mathbb{C})$ into $m = m_M m_q$ homogenous Abelian subalgebras.

In [17] a detailed treatise of $M_3 \otimes M_3$ can be found. The dimension of the traceless subspace in this case is $n^2 - 1 = 80$. It has ten subalgebras $\mathcal{A}_1, \dots, \mathcal{A}_{10} \simeq M_3(\mathbb{C})$. In the algebras \mathcal{A}_i we can find complementary Abelian subalgebras $\mathcal{A}_i^{(k)}$ generated by the projections $\{P_i^{(k)}\}_{i=1}^3$ ($1 \leq i \leq 4$), where the $P_i^{(k)}$ are isomorphic images of minimal projections on $M_3(\mathbb{C})$ and have $rk(P_i^{(k)}) = 3$. ◇

1.3 Composite Systems

The composite of two quantum systems with Hilbert spaces \mathcal{H}_A and \mathcal{H}_B , is described by the algebra $\mathcal{B}(\mathcal{H}) = \mathcal{B}(\mathcal{H}_A) \otimes \mathcal{B}(\mathcal{H}_B)$ on the Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. Then $\mathcal{B}(\mathcal{H})$ contains $\mathcal{B}(\mathcal{H}_A) \otimes I_B$ and $I_A \otimes \mathcal{B}(\mathcal{H}_B)$ as subalgebras and these are associated with the

subsystems A and B. The state of the subsystem A is described by the restriction of the state as a linear functional on the algebra $\mathcal{B}(\mathcal{H})$ to the subalgebra $\mathcal{B}(\mathcal{H}_A) \otimes I_B \simeq \mathcal{B}(\mathcal{H}_A)$. In this case the system A can be described by the reduced density operator $\rho_A \in \mathcal{B}(\mathcal{H}_A)$ given by the partial trace:

$$\rho_A = \text{Tr}_B(\rho). \quad (1.31)$$

Any subalgebra \mathcal{A} of the composite system $\mathcal{B}(\mathcal{H})$ that is isomorphic to $\mathcal{B}(\mathcal{H}_A)$ can be identified with the subsystem A after an appropriate unitary conjugation such that

$$\mathcal{B}(\mathcal{H}_A) \otimes I_B \equiv W\mathcal{A}W^*. \quad (1.32)$$

For a physical system the unitary W can be realized through the evolution of the system under an appropriate time independent Hamilton operator H : The dynamic of the system is described by the von Neumann equation

$$\frac{d\rho}{dt} = -\frac{i}{\hbar}[H, \rho] \quad (1.33)$$

where the commutator is given by $[A, B] = AB - BA$. The solution to this equation is given by the unitary evolution

$$\rho(t) = e^{-iHt/\hbar}\rho(t=0)e^{iHt/\hbar} \quad (1.34)$$

and we can choose H and t such that $W = e^{-iHt/\hbar}$.

Chapter 2

State Estimation

In classical statistics, as well as in quantum theory, observations are not always predictable but they are distributed at random. In estimation theory one aims to guess the underlying distribution of a random observations from statistical data. In particular, we are interested in the situation when we can parameterize the possible true distributions by a some parameter θ from a set Θ and we try to find a single value $\hat{\theta}$, i.e. a point estimate, that gives a good approximation for the true value θ . In this chapter we give a short introduction to classical estimation theory followed by an introduction to the problem in the quantum case. We give a brief overview of results from quantum estimation theory that can be found in the literature. Furthermore we give a detailed description of the scheme we use for estimation of a quantum state in the thesis.

2.1 Classical Parameter Estimation

A statistical space $(X, \mathcal{S}, \mathcal{P})$ consists of a set X together with a σ -algebra \mathcal{S} of subsets of X and a family of probability distributions $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ that is parameterized

by a parameter $\boldsymbol{\theta}$ such that $(X, \mathcal{S}, P_{\boldsymbol{\theta}})$ are probability spaces. We will assume that the parameterization is identifiable, i.e. it is a one-to-one map between Θ and \mathcal{P} . We will consider $\Theta \subset \mathbb{R}^m$ and we will assume that \mathcal{P} is dominated by a finite measure μ , i.e. $P_{\boldsymbol{\theta}}$ is absolutely continuous with respect to μ for all values of the parameter $\boldsymbol{\theta}$. Furthermore let us consider in this section only the case when the probability distributions in \mathcal{P} obey either a probability density $f_{\boldsymbol{\theta}}(x)$ or a probability mass function $p_{\boldsymbol{\theta}}(i)$. In the following we discuss the discrete case. The formulas in the continuous case are obtained by replacing $p_{\boldsymbol{\theta}}(i)$ by $f_{\boldsymbol{\theta}}(x)$.

In the typical case in estimation theory we look at a sequence $\mathbf{X} = (X_1, X_2, \dots, X_N)$ of N i.i.d. random variables, called a sample of size N , where the X_k are distributed according to a distribution $P_{\boldsymbol{\theta}} \in \mathcal{P}$. We infer about the true value $\boldsymbol{\theta}$ from the realization of the random sequence \mathbf{X} . The outcome space of the sample is the Cartesian product $\times_{k=1}^N X$. Since the $X^{(k)}$ are independent and identically distributed, the probability mass function of \mathbf{X} is the product of the probability mass functions $p_{\boldsymbol{\theta}}$ of the $X^{(k)}$ and it is called **likelihood function**

$$L_{\boldsymbol{\theta}}(\mathbf{x}) := \prod_{i=1}^N p_{\boldsymbol{\theta}}(x_i).$$

A **statistic** T is a measurable function from the outcome space of the random sequences \mathbf{X} into the parameter space \mathbb{R}^m . It is called **sufficient** if

$$\text{Prob}(\mathbf{X} = \mathbf{x} | T(\mathbf{x}) = t) \quad \text{is independent of } \boldsymbol{\theta}. \quad (2.1)$$

Heuristically, in this case it contains all information about the parameter $\boldsymbol{\theta}$. By the Neyman-Fisher factorization theorem (see e.g. [13]) this is the case if

$$L_{\boldsymbol{\theta}}(\mathbf{x}) = g_{\boldsymbol{\theta}}(T(\mathbf{x}))h(\mathbf{x}). \quad (2.2)$$

A statistic is called **complete** if

$$E_{\theta}[g(T(X))] = 0 \quad \forall \theta \in \Theta \Rightarrow \text{Prob}_{\theta}(g(T(x)) = 0) = 0 \quad \forall \theta \in \Theta \quad (2.3)$$

for real valued measurable functions g .

Of particular interest is the case when the statistics T is used as an **estimate** of θ . Then it is natural to require some further properties of T such that the estimate gives values close to the true value of the parameter with high probability. An estimate is called **unbiased** if its expectation equals the true value of the parameter:

$$E_{\theta}[T] = \theta. \quad (2.4)$$

It is called **asymptotically unbiased** if we have a sequence of statistics T_N on samples of size N estimating θ such that $\lim_{N \rightarrow \infty} E_{\theta}[T_N] = \theta$. As a measure of efficiency of an estimate the **mean quadratic error matrix**

$$V_{\theta}(T) = E_{\theta}[(T - \theta)(T - \theta)^t] \quad (2.5)$$

is used, and if the estimate is unbiased, this is identical with its variance. An estimate T_1 is called more efficient than an estimate T_2 if

$$V_{\theta}(T_1) \leq V_{\theta}(T_2) \quad \forall \theta \in \Theta \quad (2.6)$$

where the inequality is to be understood in the matrix sense, i.e. $A \geq B$ if $A - B \geq 0$. An estimate with uniformly (i.e. for all θ) minimal mean quadratic error matrix does not always exist. Nevertheless, if there exists a sufficient and complete statistics T' , then by the Rao-Blackwell theorem (see e.g. [13]) there exists a unique unbiased estimate T of uniform

minimal variance and it is a function of \mathbf{T}' . For an unbiased estimate the following lower bound on the mean quadratic error matrix can be found: Suppose that the probability mass functions in \mathcal{P} have common support, $\text{supp}(p_{\boldsymbol{\theta}}) = C$ for all $\boldsymbol{\theta} \in \Theta$, where C is a bounded set in X . Further $\nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}$ exists and is finite on $\text{supp}(p_{\boldsymbol{\theta}})$, where $\nabla_{\boldsymbol{\theta}}$ denotes the gradient with respect to $\boldsymbol{\theta}$. Then the Cramer-Rao inequality (see e.g. [13])

$$V_{\boldsymbol{\theta}}(\mathbf{T}) \geq \frac{(\partial_{\boldsymbol{\theta}} \mathbf{E}_{\boldsymbol{\theta}}[\mathbf{T}])^2}{I(\boldsymbol{\theta})} \quad (2.7)$$

holds, where the matrix $I(\boldsymbol{\theta}) = \mathbf{E}_{\boldsymbol{\theta}}[\nabla_{\boldsymbol{\theta}} \log L_{\boldsymbol{\theta}}(\mathbf{X})][\nabla_{\boldsymbol{\theta}} \log L_{\boldsymbol{\theta}}(\mathbf{X})]^t$ is the Fisher information. From the observation that the Fisher information has the form of a variance and the observables X_k are independent, it is easy to see that it grows proportional to the sample size: $I_N = NI_1$ where $I_1(\boldsymbol{\theta}) = \mathbf{E}_{\boldsymbol{\theta}}[\nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(\mathbf{X})][\nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(\mathbf{X})]^t$. For the special case of an unbiased estimate additionally $(\partial_{\boldsymbol{\theta}} \mathbf{E}_{\boldsymbol{\theta}}[\mathbf{T}])^2 = 1$.

An important case are so called **exponential families** \mathcal{P} whose elements have likelihood functions

$$L_{\boldsymbol{\theta}}(\mathbf{x}) = \exp\left(\sum_{i=1}^n \eta_i(\boldsymbol{\theta}) \mathbf{T}_i(\mathbf{x}) - B(\boldsymbol{\theta})\right) h(\mathbf{x}) \quad (2.8)$$

with respect to some common measure. Here η_i and \mathbf{T}_i are real valued functions and the η_i are called **natural parameters** of the family. An exponential family is called of **full rank** if the parameter set $\Theta \subset \mathbb{R}^m$ contains a open set of \mathbb{R}^m . By the Neyman-Fisher theorem $\mathbf{T} = (\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_n)$ is a sufficient statistics and if additionally the family is of full rank, \mathbf{T} is complete. In this case \mathbf{T} attains the Cramer-Rao bound, and in particular if \mathbf{T} is unbiased it is of uniform minimal variance (see e.g. [13]).

Finally a sequence of estimates \mathbf{T}_N is called **weakly consistent**, if $\mathbf{T}_N \rightarrow \boldsymbol{\theta}$ in probability. If the mean quadratic error matrix of a sequence of asymptotically unbiased estimates \mathbf{T}_N goes to zero asymptotically, this implies weak consistency of \mathbf{T}_N . This follows if we apply Chebyshev's inequality $\text{Prob}(|X - \mathbf{E}[X]| \geq a) \leq \text{Var}(X)/a^2$ for a real valued random

variable \mathbf{X} to the components of \mathbb{T}_N .

Example 2.1.1 As an example consider a statistical space over a finite set $X = \{1, 2, \dots, m\}$ with the family \mathcal{P} of all discrete probability distributions with mass function p_θ , $\text{supp}(p_\theta) = X$. Let us denote $\mathbf{p} := (p_1, p_2, \dots, p_m)$ as the vector containing the probability $p_i := \text{Prob}(i)$ that i appears. Then \mathcal{P} is the interior of the probability simplex

$$\Omega_m = \left\{ \mathbf{p} = (p_1, p_2, \dots, p_m) : p_i \geq 0, \sum p_i = 1 \right\} \subset \mathbb{R}^m. \quad (2.9)$$

Due to the condition $\sum_i p_i = 1$ we can parameterize \mathcal{P} by the first $(m - 1)$ components of \mathbf{p} and the parameter set is given by

$$\Theta_m = \left\{ \boldsymbol{\theta} = (p_1, p_2, \dots, p_{m-1}) : p_i > 0, \sum p_i < 1 \right\} \subset \mathbb{R}^{m-1}. \quad (2.10)$$

Given a sample of length N , i.e. a sequence of N i.i.d. random variables $X_k \sim p_\theta$ ($1 \leq k \leq N$), let us denote the number of occurrences of the value i in a outcome sequence by n_i . The likelihood functions $L_\theta(\mathbf{x})$ of the sample form an exponential family (see e.g. [13]):

$$L_\theta(\mathbf{X} = (x_1, x_2, \dots, x_N)) = \exp \left(\sum_{i=1}^{m-1} \frac{n_i}{N} \log \frac{p_i^N}{p_m^N} - \log p_m^N \right) I_\Omega(n_1, n_2, \dots, n_m) \quad (2.11)$$

where $I_\Omega(n_1, n_2, \dots, n_m)$ is the indicator function on the set $\{n_i \in \mathbb{N} : \sum_{i=1}^m n_i = N\}$. This family is dominated by the counting measure and it is identifiable. The natural parameter is given by

$$\boldsymbol{\eta} = \left(\log \frac{p_1^N}{p_m^N}, \log \frac{p_2^N}{p_m^N}, \dots, \log \frac{p_{m-1}^N}{p_m^N} \right) \in \mathbb{R}^{m-1}. \quad (2.12)$$

Since Θ contains an open set of \mathbb{R}^{m-1} , it is an exponential family of full rank. Thus $\boldsymbol{\nu}(x_1, x_2, \dots, x_N) := (\frac{n_1}{N}, \frac{n_2}{N}, \dots, \frac{n_{m-1}}{N})$ is a complete and sufficient statistics and it is called

the empirical distribution of the X_k . The components $\nu_i := \frac{n_i}{N}$ of $\boldsymbol{\nu}$ are called **relative frequencies** of the outcomes i . The vector $\boldsymbol{\nu}$ is by (2.19) an unbiased estimate for $\boldsymbol{\theta}$ and it is the the uniform minimum variance unbiased estimate on the interior of the probability simplex (2.9) as discussed in the previous section.

The numbers n_i of the occurrences of the outcome i in a realization of the sample are random variables as well, nevertheless by the condition $\sum_i n_i = N$ they are not independent. In the following we derive the marginal distributions, expectations and variances of the n_i : The number of realizations of \mathbf{X} that contain i exactly n_i times is given by the multinomial coefficient

$$\binom{N}{n_1 \ n_2 \ \dots \ n_m} = \frac{N!}{n_1! n_2! \dots n_m!} \quad (2.13)$$

and thus the n_i follow a **multinomial** or **polynomial distribution** with probability mass function

$$p_{\boldsymbol{\theta}}(n_1, n_2, \dots, n_m) = \binom{N}{n_1 \ n_2 \ \dots \ n_m} p_1^{n_1} p_2^{n_2} \dots p_m^{n_m} I_{\Omega}(n_1, n_2, \dots, n_m). \quad (2.14)$$

Let us define the Bernoulli variables $B_i^{(k)} \sim (p_i, 1 - p_i)$ on the sample space defined by

$$B_i^{(k)} = \begin{cases} 1 & \text{if } X_k = i \\ 0 & \text{if } X_k \neq i \end{cases} \quad (2.15)$$

Then we can write the n_i as the sum of i.i.d. random variables: $n_i = \sum_{k=1}^N B_i^{(k)}$. It is immediate by considering the multinomial distribution for the sequence of random variables $B_i^{(k)}$ instead of X_k that the marginal distributions of (2.14) for the n_i are given by the binomial distribution

$$p_{\boldsymbol{\theta}}(n_i) = \binom{N}{n_i} p_i^{n_i} (1 - p_i)^{(N - n_i)}. \quad (2.16)$$

The binomial distribution has expectation $\mathbb{E}[n_i] = Np_i$ and variance $\text{Var}(n_i) = N(p_i - p_i^2)$.

To calculate the covariances we use

$$n_i n_j = \left(\sum_{k=1}^n \mathbb{B}_i^{(k)} \right) \left(\sum_{l=1}^N \mathbb{B}_j^{(l)} \right) = \left(\sum_{k=1}^N \mathbb{B}_i^{(k)} \mathbb{B}_j^{(k)} \right) + \left(\sum_{k \neq l}^N \mathbb{B}_i^{(k)} \mathbb{B}_j^{(l)} \right) \quad (2.17)$$

The first sum on the right hand side of this equation is always zero by the definition of the $\mathbb{B}_i^{(k)}$. Since $k \neq l$, in the second sum we have the products of independent random variables. Thus

$$\begin{aligned} \text{Cov}(n_i, n_j) &= \mathbb{E}[n_i n_j] - \mathbb{E}[n_i] \mathbb{E}[n_j] = \sum_{k \neq l}^N \mathbb{E}[\mathbb{B}_i^{(k)}] \mathbb{E}[\mathbb{B}_j^{(l)}] - \mathbb{E}[n_i] \mathbb{E}[n_j] \\ &= (N^2 - N) p_i p_j - N^2 p_i p_j = -N p_i p_j. \end{aligned} \quad (2.18)$$

Altogether we get

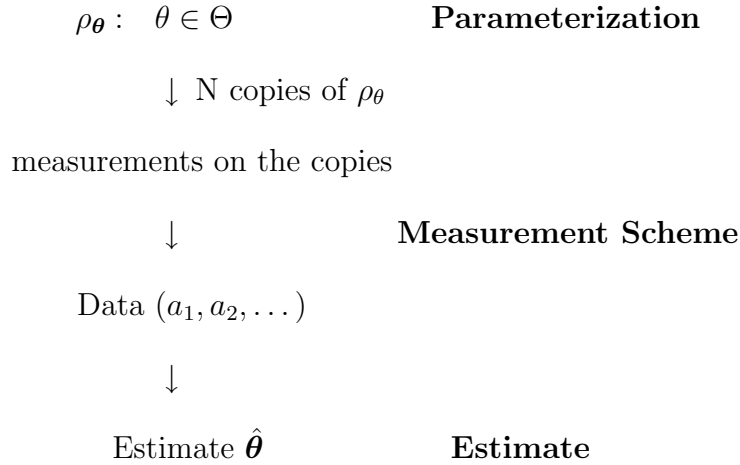
$$\begin{aligned} \mathbb{E}[n_i] &= Np_i, \\ \text{Var}(n_i) &= N(p_i - p_i^2), \\ \text{Cov}(n_i, n_j) &= -N p_i p_j \quad (i \neq j). \end{aligned} \quad (2.19)$$

◇

2.2 Quantum State Estimation

In quantum statistics we consider a family $\mathcal{R} = \{\rho_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ of density operators parameterized by a parameter from a set $\Theta \subset \mathbb{R}^k$. As in the classical case the aim of estimation theory is to find an estimate of the parameter $\boldsymbol{\theta}$ from the outcome of observations of the system. For this purpose we consider a sample of N identical copies of an unknown state $\rho_{\boldsymbol{\theta}} \in \mathcal{B}(\mathcal{H})$ of the quantum system. In order to obtain experimental data we need to specify

a set of measurements we perform on the sample. From the statistical data obtained from the outcome of the measurements we form an estimate of the true state of the system. Thus a quantum estimation scheme consist usually from the following steps:



To compare the quality of an estimation scheme a measure of quality of the estimate needs to be chosen. Additionally to the influence factors discussed for a classical estimate the efficiency of a quantum estimation scheme depends on the chosen set of measurements.

Quantum state estimation is an active field of research. Let us mention in the following some results from this field: One of the main questions is the optimal choice of the measurement in an estimation scheme. The measurements on the sample can be categorized by their correlations into collective, separable and separate measurements¹. For the problem discussed in the thesis the restriction to separate measurement is adequate, however, it is known that in some situations a collective measurement on the sample is superior to separable measurements [16, 6], and that separable, but correlated measurements in the form of adaptive measurement schemes can perform better than comparable separated

¹ A collective measurement is a POVM on the composite system of the available copies of the state. If the operators in a measurement can be written in the form $E_i = \sum_j E_{ij}^1 \otimes E_{ij}^2 \otimes \dots \otimes E_{ij}^N$ with $E_{ij}^k \geq 0$ ($1 \leq k \leq N$), a measurement is called separable. Separate measurements are defined in (2.24).

measurements [6, 11].

For the qubit case ($\dim(\mathcal{H}) = 2$) several estimation schemes using separate measurements can be found in the literature. Besides the so called standard scheme using the spin observables S_i (see example 2.2.2), a minimal estimation scheme with the single POVM of example 1.1.1 [11], schemes estimating the orientation and length of the Bloch vector [1] and even continuous POVM [27] have been studied. For the construction of an estimate from the statistical data, the procedure of inverting the linear relationship between the state and the outcome probabilities in a measurement, as used in the present work, is considered standard [21, 1, 11]. Alternatively the concepts of maximum likelihood estimation [10] as well as Bayesian estimation [1, 15] have been studied in the quantum setting.

As measures for the quality of the estimation scheme we use the determinant of the mean quadratic error matrix, also known as generalized variance (see e.g [12]). This measure was used in [21]. In [28] the average information gain was used, which is asymptotically related to the determinant of the mean quadratic error matrix. Commonly also the trace of the mean quadratic error matrix [11, 27] or the quantum fidelity between the estimate and the true state is used [1]. Let us also mention that different bounds on the mean quadratic error matrix similar to the classical Cramer-Rao bound have been derived in the quantum case [9, 6, 7] and asymptotical attainability was shown for some of them [6, 7, 1].

In the following we discuss the setting of the state estimation schemes considered in this thesis. In particular, the individual steps of the estimation scheme are described in detail. .

2.2.1 State space and parameterization

Recall that a **density operator** or **state** $\rho \in \mathcal{B}(\mathcal{H})$ is defined by the conditions $\rho \geq 0$ and $\text{Tr } \rho = 1$. By the condition of unit trace the component of ρ in the span of the identity is

constant and the state is entirely described by its component in the traceless subspace \mathcal{S} :

$$\rho = \frac{1}{n}I + \boldsymbol{\theta}, \quad \text{where} \quad \boldsymbol{\theta} \in \mathcal{S}. \quad (2.20)$$

We will call the Bloch vector $\boldsymbol{\theta}$ of ρ the **state vector**² of the system and we will denote by $\mathcal{T} \subset \mathcal{S}$ the set of state vectors. As mentioned in section 1.2.1 with the choice of a selfadjoint basis in $\mathcal{B}(\mathcal{H})$, the state vector $\boldsymbol{\theta}$ can be written as a vector with real components. Thus it can serve as a parameterization of the state and $\mathcal{T} \equiv \Theta \subset \mathbb{R}^{n^2-1}$ is the set of parameters.

Example 2.2.1 In the qubit case, i.e. $\dim(\mathcal{H}) = 2$, the **Pauli matrices**

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (2.21)$$

together with the identity form a standard basis of $M_2(\mathbb{C})$. The state vector $\boldsymbol{\theta}$ can be written as a vector in \mathbb{R}^3 and up to a normalization constant it is the well-known **Bloch vector** of the state. The set of state vectors is the Bloch ball $\mathcal{T} = \{\boldsymbol{\theta} : \boldsymbol{\theta} \in \mathbb{R}^3, \|\boldsymbol{\theta}\|^2 \leq (n-1)/n^2\}$ (see below). \diamond

Due to the positivity condition imposed on the states, the structure of the set \mathcal{T} is difficult to describe in the case $\dim(\mathcal{H}) > 2$. In the following we summarize some of the properties of \mathcal{T} : The set \mathcal{T} is a convex set and by choosing a basis of \mathcal{H} in which a given ρ is diagonal, it is easy to see that ρ is the convex combination of rank one projections. Thus the extremal points of this set are the pure states. Since a self-adjoint operator is non-positive if it has some negative eigenvalues, it follows also that the boundary of \mathcal{T} is given by states that have at least one eigenvalue equal to zero. This are the states that have

² Also the notion generalized Bloch vector or coherence vector are common

non-invertible density matrices. Conversely invertible states are in the interior of \mathcal{T} . We can conclude the following bounds on the set of states: The eigenvalues λ_i of ρ can be written in the form $\lambda_i = 1/n + \Delta_i$ where Δ_i are the eigenvalues of $\boldsymbol{\theta}$ and sum up to zero. With $\|\boldsymbol{\theta}\|^2 = \langle \boldsymbol{\theta}, \boldsymbol{\theta} \rangle = 1/n \sum_i \Delta_i^2$ it follows that

$$\|\boldsymbol{\theta}\|^2 = \frac{n \text{Tr}(\rho^2) - 1}{n^2}. \quad (2.22)$$

For a pure state $\text{Tr}(\rho^2) = 1$, thus the Bloch vector of a pure state has length $\|\boldsymbol{\theta}\|^2 = (n-1)/n^2$. In general $\text{Tr}(\rho^2)$ is smaller than one, thus all state vectors are contained in a ball of radius $\sqrt{n-1}/n$ in \mathbb{R}^{n^2-1} . On the other hand, if all Δ_i are smaller than $n/2$ in absolute value, ρ will be positive. This is ensured if $\sum_i \Delta_i^2 \leq 1/n^2$, thus a ball of radius $\|\boldsymbol{\theta}\| \leq n^{-3/2}$ is surely contained in \mathcal{T} .

In general the pure states form only a subset of a sphere in \mathbb{R}^{n^2-1} . For a description of this subset we note that the unitary group $U(n)$ acts on the set of states by conjugation: $(U, \rho) \mapsto U\rho U^*$ where $U \in U(n)$. This action defines also an action of $U(n)$ on the set \mathcal{T} by $(U, \boldsymbol{\theta}) \mapsto U\boldsymbol{\theta}U^*$. The set \mathcal{T} can be described by the orbit manifolds under this action of $U(n)$. Especially, the action is transitive on the set $\{\boldsymbol{\theta} \in \mathcal{T} : |\boldsymbol{\theta}|^2 = (n-1)/n^2\}$ of pure state vectors and this set is described by the orbit manifold of pure state vectors. It is homeomorphic to a sphere in \mathbb{R}^{n^2-1} only in the case of $\dim(\mathcal{H}) = 2$ [26]. In this case it is the Bloch ball defined in Example 2.2.2.

2.2.2 Measurement

To obtain experimental data from a sample of quantum states we need to perform a measurement. A sample of N identical copies of a quantum state is described by a density operator

$$\rho_{\boldsymbol{\theta}}^{\otimes N} := \bigotimes_{k=1}^N \rho_{\boldsymbol{\theta}} \in \mathcal{B}(\mathcal{H}^{\otimes N}). \quad (2.23)$$

where $\mathcal{H}^{\otimes N} := \otimes_{i=1}^N \mathcal{H}$. A measurement is described by a set of positive operators in $\mathcal{B}(\mathcal{H}^{\otimes N})$ that form a partition of the identity. In the thesis we will consider only so called separate measurements. They correspond to the situation when each copy of the sample is measured separately without any correlations between this measurements. This is the case if the operators $E_i \in M$ in a measurement M are of the form

$$E_i = E_i^1 \otimes E_i^2 \otimes \cdots \otimes E_i^N \quad (2.24)$$

By $\text{Tr}(\rho^{\otimes N} E_i) = \prod_{k=1}^N \text{Tr}(\rho E_i^k)$ we may simply forget about the tensor product structure and instead of a measurement of the composite system we may consider \mathcal{M} as a collection of measurements $M^{(k)}$ defined on the individual systems $\mathcal{B}(\mathcal{H})$. Thus we call a set $\mathcal{M} := \{M^{(k)} : 1 \leq k \leq m\}$ of m different measurements a **measurement scheme**³. In an estimation scheme we perform the measurement $M^{(k)}$ on $N^{(k)}$ copies of the state, where $\sum_{k=1}^m N^{(k)} = N$. A measurement $M^{(k)}$ and the family \mathcal{R} of states induce a classical statistical space $(X^{(k)}, \mathcal{P}^{(k)})_{M^{(k)}}$ on the outcome set $X^{(k)}$ of the measurement. Therefore from a measurement scheme we get a collection of m different classical statistical spaces $(X^{(k)}, \mathcal{P}^{(k)})_{M^{(k)}}$.

For our calculations we will use the expansion of the operators E_i in the basis $\{I, \sigma_i : 1 \leq i \leq n^2 - 1\}$:

$$E_i = \frac{\text{Tr} E_i}{n} I + \mathbf{u}_i \quad (2.25)$$

In the following we will use \mathbf{u}_i synonymously to E_i , if there is no ambiguity⁴. Condition (1.9) implies that the components of E_i in the traceless subspace \mathcal{S} sum up to zero.

³ We collect only different measurements in \mathcal{M} , so if the same measurement $M^{(k)}$ is performed on several copies of ρ , it appears only once in \mathcal{M} . Therefore in general m is smaller than N .

⁴ For example speaking about a measurement $M^{(k)}$ we may sometimes refer to the set $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d\}$ instead of $\{E_1, E_2, \dots, E_d\}$ if it is clear from the context what are the E_i .

Therefore

$$\sum_{i=1}^d \mathbf{u}_i = 0 \quad \text{and} \quad \sum_i \text{Tr } E_i = n \quad (2.26)$$

is equivalent to the second condition in (1.9). If the vectors

$$\left\{ \mathbf{u}_i^{(k)} = E_i^{(k)} - \frac{\text{Tr } E_i^{(k)}}{n} I : E_i^{(k)} \in M^{(k)}, 1 \leq k \leq m \right\} \quad (2.27)$$

linearly span the whole space \mathcal{S} , we call the measurement scheme **informationally complete**. This notion is commonly also used for POVMs [25] which span the whole algebra $\mathcal{B}(\mathcal{H})$. From (2.29) it is easy to see that informational completeness is a necessary condition if we want to estimate the parameter $\boldsymbol{\theta}$: The measurement probabilities will not depend on the component of $\boldsymbol{\theta}$ in the orthogonal complement of the subspace spanned by the operators in the measurement and we obtain no information about this component of $\boldsymbol{\theta}$.

Example 2.2.2 For a quantum system with $\dim(\mathcal{H}) = 2$ the measurement of a non-degenerate observable A_i consists of two minimal projections $M_{A_i} = \{1/2 I \pm \mathbf{u}_i\}$. This two projections are characterized completely by the Bloch vectors $\mathbf{u}_i \in \mathcal{T}$. An informational complete measurement scheme using such observables consists of A_i ($1 \leq i \leq 3$) such that the vectors \mathbf{u}_i are linearly independent in \mathbb{R}^3 . As a special case let us consider the spin observables $S_i := \sigma_i$ ($1 \leq i \leq 3$), where the σ_i are the Pauli matrices. They have eigenprojections $M_{S_i} = \{1/2(I \pm \sigma_i)\}$ and they are mutually complementary observables on $M_2(\mathbb{C})$. Since the Pauli matrices form a basis of \mathcal{S} , the measurement scheme $\mathcal{M} = \{M_{S_1}, M_{S_2}, M_{S_3}\}$ defines an informationally complete measurement scheme. \diamond

Example 2.2.3 The POVM described in example (1.1.1) provides an informational complete measurement scheme for a quantum system with $\dim(\mathcal{H}) = 2$ consisting of a single POVM. \diamond

In general, the vectors $\mathbf{u}_i^{(k)}$ that belong to some measurement $M^{(k)}$ will not span the whole space \mathcal{S} but only a subspace $\mathcal{S}^{(k)} = \text{span}\{\mathbf{u}_i^{(k)} : 1 \leq i \leq d^{(k)}\}$. Quasi-orthogonality of two measurements $M^{(k)}$ and $M^{(l)}$, as defined in section 1.2.2, corresponds to the case when the subspaces $\mathcal{S}^{(k)}$ and $\mathcal{S}^{(l)}$ are orthogonal in \mathcal{S} with respect to the inner product (1.4).

2.2.3 Construction of the estimate

Suppose now we are given an ensemble of N identical copies of a quantum system in the state $\boldsymbol{\theta}$ which is not known to us. To estimate the state we apply the following strategy: We choose an informationally complete measurement scheme \mathcal{M} consisting of elements $M^{(k)}$, $1 \leq k \leq m$, which have cardinality $\#(M^{(k)}) = d^{(k)}$ and we require

$$\sum_{k=1}^m (d^{(k)} - 1) = n^2 - 1. \quad (2.28)$$

Furthermore we divide the ensemble into subensembles of size $N^{(k)}$. On the individual copies in k th subensemble we perform the measurement $M^{(k)}$. We will find that there is a linear relation between the probabilities $p_i^{(k)}$ of the measurement outcomes and the state vector $\boldsymbol{\theta}$. We use this relation to construct an estimate $\hat{\boldsymbol{\theta}}$ of the state vector from a classical estimate $\boldsymbol{\nu}$ of the probabilities.

The probability of an individual outcome to appear in a particular measurement is

$$p_i^{(k)} = \text{Tr} \rho E_i^{(k)} = \frac{\text{Tr} E_i^{(k)}}{n} + n \langle \mathbf{u}_i^{(k)}, \boldsymbol{\theta} \rangle \quad (1 \leq k \leq m, 1 \leq i \leq d^{(k)} - 1) \quad (2.29)$$

and

$$p_{d^{(k)}}^{(k)} = 1 - \sum_{j=1}^{d^{(k)}-1} p_j^{(k)} \quad (1 \leq k \leq m). \quad (2.30)$$

Thus the probabilities $p_i^{(k)}$ and the state vector $\boldsymbol{\theta}$ are related by the $(n^2 - 1)$ linear equations (2.29). Let us, with the abbreviation $r_i^{(k)} := \text{Tr} E_i^{(k)}$, introduce the $(n^2 - 1) \times (n^2 - 1)$ matrix

T and the vectors \mathbf{p} , \mathbf{r} as

$$\begin{aligned}
 T &= \begin{bmatrix} T^{(1)} \\ \vdots \\ T^{(m)} \end{bmatrix}, \quad T^{(k)} = \begin{bmatrix} (\mathbf{u}_1^{(k)})^t \\ \vdots \\ (\mathbf{u}_{d^{(k)}-1}^{(k)})^t \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} \mathbf{p}^{(1)} \\ \vdots \\ \mathbf{p}^{(m)} \end{bmatrix}, \quad \mathbf{p}^{(k)} = \begin{bmatrix} p_1^{(k)} \\ \vdots \\ p_{d^{(k)}-1}^{(k)} \end{bmatrix}, \\
 \mathbf{r} &= \begin{bmatrix} \mathbf{r}^{(1)} \\ \vdots \\ \mathbf{r}^{(m)} \end{bmatrix}, \quad \mathbf{r}^{(k)} = \begin{bmatrix} r_1^{(k)} \\ \vdots \\ r_{d^{(k)}-1}^{(k)} \end{bmatrix}
 \end{aligned} \tag{2.31}$$

where $(\mathbf{u}_i^{(k)})^t$ denotes the transposed of the vector $\mathbf{u}_i^{(k)}$. Using this notation we can represent the system (2.29) in a matrix form

$$n \cdot T\boldsymbol{\theta} = \mathbf{p} - \frac{1}{n} \mathbf{r}. \tag{2.32}$$

Since we chose the measurement scheme to be informationally complete and we required (2.28), the rows $\mathbf{u}_i^{(k)}$ of $T^{(k)}$ form a set of linear independent vectors in the subspace $\mathcal{S}^{(k)}$ and the matrix T is quadratic and of full rank. Thus we can invert equation (2.32) to

$$\boldsymbol{\theta} = \frac{1}{n} T^{-1} \left(\mathbf{p} - \frac{1}{n} \mathbf{r} \right). \tag{2.33}$$

Given an estimate $\boldsymbol{\nu}$ for the classical probabilities in \mathbf{p} from the measured data we obtain an **unconstrained estimate** $\hat{\boldsymbol{\theta}}$ as

$$\hat{\boldsymbol{\theta}} = \frac{1}{n} T^{-1} \left(\boldsymbol{\nu} - \frac{1}{n} \mathbf{r} \right). \tag{2.34}$$

The vectors $\mathbf{p}^{(k)}$ parameterize discrete probability distributions as discussed in Chapter 2.1.1. They take values from the parameter set $\Theta_{d^{(k)}} := \{\mathbf{p}^{(k)} : p_i^{(k)} \geq 0, \sum_{i=1}^{d^{(k)}-1} p_i^{(k)} \leq 1\}$

of a probability simplex as defined in (2.10). Thus (2.32) and (2.33) define an affine transformation that maps the set of state vectors $\mathcal{T} \subset \mathcal{S}$ into a subset $\mathcal{T}_\Theta \subset \times_{k=1}^m \Theta_{d^{(k)}}$ of the Cartesian product of the parameter sets $\Theta_{d^{(k)}}$. The notion *unconstrained* results from the fact, that the estimate $\boldsymbol{\nu}$ of the vector \boldsymbol{p} may fall outside the image \mathcal{T}_Θ of \mathcal{T} under the above transformation (2.32) and thus it can happen that $\hat{\boldsymbol{\theta}}$ is not a state.

We can obtain a **constrained estimate** $\hat{\boldsymbol{\theta}}_c$ from $\hat{\boldsymbol{\theta}}$ that fulfills the constraint $\hat{\boldsymbol{\theta}}_c \in \mathcal{T}$ by adding a appropriate correction

$$\hat{\boldsymbol{\theta}}_c(\boldsymbol{\nu}) = \hat{\boldsymbol{\theta}}(\boldsymbol{\nu}) + \Delta(\boldsymbol{\nu}) \quad \text{such that} \quad \hat{\boldsymbol{\theta}}_c \in \mathcal{T} \quad \text{and} \quad \Delta(\boldsymbol{\nu}) = 0 \quad \text{if} \quad \hat{\boldsymbol{\theta}}(\boldsymbol{\nu}) \in \mathcal{T} \quad (2.35)$$

One possibility to correct the unconstrained estimate, proposed in [21], is to take the state in \mathcal{T} closest to the unconstrained estimate in the Hilbert Schmidt norm:

$$\hat{\boldsymbol{\theta}}_c := \operatorname{argmin}_{\omega \in \mathcal{T}} \operatorname{Tr} (\hat{\boldsymbol{\theta}} - \omega)^2. \quad (2.36)$$

The minimum in (2.36) can be calculated in the eigenbasis of $\hat{\boldsymbol{\theta}}$ by the following algorithm: If $\hat{\boldsymbol{\theta}}$ falls outside the set \mathcal{T} the estimated density operator $\hat{\rho} = 1/n I + \hat{\boldsymbol{\theta}}$ violates the positivity condition imposed on states and it will have negative eigenvalues. If $\lambda_1, \lambda_2; \dots, \lambda_k \leq 0$ denote the non positive eigenvalues of $\hat{\rho}$, we replace the eigenvalues λ_i by

$$\tilde{\lambda}_i = 0 \quad \text{if} \quad 1 \leq i \leq k \quad \text{and} \quad \tilde{\lambda}_i = \lambda_i + \sum_{j=1}^k \frac{\lambda_j}{n-k} \quad \text{if} \quad k < i < n. \quad (2.37)$$

If we repeat this procedure until $\tilde{\lambda}_i \geq 0$ for all i , the minimizer of (2.36) is obtained by $\tilde{\rho} = 1/n I + \hat{\boldsymbol{\theta}}_c = \operatorname{Diag}(\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_n)$ in the eigenbasis of $\hat{\boldsymbol{\theta}}$. An other possibility to correct the unconstrained estimate is to consider the cases when $\hat{\boldsymbol{\theta}} \notin \mathcal{T}$ as cases of error and use a trivial correction $\Delta(\boldsymbol{\nu}) = -\hat{\boldsymbol{\theta}}$ if $\hat{\boldsymbol{\theta}} \notin \mathcal{T}$.

Nevertheless if the true state $\boldsymbol{\theta}$ is not on the boundary of \mathcal{T} , which is the case if it has

strictly positive eigenvalues, the probability of the unconstrained estimate to fall outside the set \mathcal{T} goes exponentially to zero with increasing $N_{min} := \min_k \{N^{(k)}\}$. As a slight modification of Sanov's theorem the rate of convergence can be bounded as (see appendix)

$$\limsup_{N_{min} \rightarrow \infty} \frac{1}{N_{min}} \log (\text{Prob}_{\theta}(\hat{\theta} \notin \mathcal{T})) \leq -D(\boldsymbol{\nu}^* \|\mathbf{p}) \leq \liminf_{N_{min} \rightarrow \infty} \frac{1}{N_{max}} \log (\text{Prob}_{\theta}(\hat{\theta} \notin \mathcal{T})) \quad (2.38)$$

where $N_{max} = \max_k \{N^{(k)}\}$,

$$D(\boldsymbol{\nu} \|\mathbf{p}) = \sum_k D(\boldsymbol{\nu}^{(k)} \|\mathbf{p}^{(k)})$$

is the relative entropy $D(\boldsymbol{\nu}^{(k)} \|\mathbf{p}^{(k)}) = \sum_{i=1}^{d^{(k)}} \nu_i^{(k)} (\log \nu_i^{(k)} - \log p_i^{(k)})$ and

$$\boldsymbol{\nu}^* = \text{argmin}_{\boldsymbol{\nu} \notin \mathcal{T}_{\Omega}} D(\boldsymbol{\nu} \|\mathbf{p})$$

is the $\boldsymbol{\nu} \notin \mathcal{T}$ closest to \mathbf{p} in relative entropy. Note that the rate of convergence depends on the true state θ as well as on the chosen measurements scheme \mathcal{M} .

As a consequence of (2.38) the constrained estimate is **asymptotically unbiased** if we use an unbiased estimate, $\text{E}_{\theta}[\boldsymbol{\nu}] = \mathbf{p}$, for the classical probabilities and the true state is in the interior of \mathcal{T} ⁵:

$$\lim_{N_{min} \rightarrow \infty} \text{E}_{\theta}[\hat{\theta}_c] = \frac{1}{n} T^{-1} (\text{E}_{\theta}[\boldsymbol{\nu}] - \frac{1}{n} \mathbf{r}) + \lim_{N_{min} \rightarrow \infty} \text{E}_{\theta}[\Delta] = \theta \quad \forall \theta \in \text{Int}(\mathcal{T}) \quad (2.39)$$

where the last equality follows from the fact that Δ is bounded and $\text{Prob}(\Delta(\boldsymbol{\nu}) > 0) \rightarrow 0$

⁵ This implies $D(\boldsymbol{\nu}^* \|\mathbf{p}) > 0$.

as $N_{min} \rightarrow \infty$.

The most natural example of an estimate $\boldsymbol{\nu}$ are the relative frequencies discussed in Example 2.1.1:

$$\boldsymbol{\nu} = \begin{bmatrix} \boldsymbol{\nu}^{(1)} \\ \vdots \\ \boldsymbol{\nu}^{(m)} \end{bmatrix}, \quad \boldsymbol{\nu}^{(k)} = \begin{bmatrix} n_1^{(k)}/N^{(k)} \\ \vdots \\ n_{d^{(k)}-1}^{(k)}/N^{(k)} \end{bmatrix}, \quad (2.40)$$

where $n_i^{(k)}$ is the number of times a specific outcome appears among the observed outcomes of the measurement $M^{(k)}$. If the true state is in the interior of \mathcal{T} , the $p_i^{(k)}$ will be in the interior of the sets $\Theta_{d^{(k)}}$. This corresponds to the situation discussed in Chapter 2.1.1 and the relative frequencies are the minimum variance unbiased estimate for the probability vector \boldsymbol{p} if $\boldsymbol{\theta} \in \text{Int}(\mathcal{T})$.

2.2.4 Efficiency of the estimate

Mean quadratic error matrix

Given the true state $\boldsymbol{\theta}$, we describe the error of an estimate $\hat{\boldsymbol{\theta}}$ by the **mean quadratic error matrix**

$$V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \mathbb{E}_{\boldsymbol{\theta}}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^t]. \quad (2.41)$$

The unconstrained estimate $\hat{\boldsymbol{\theta}}$ (2.34) as a random vector is a linear transformation of the random vector $\boldsymbol{\nu}$ and in this case the matrix $V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ depends on the mean quadratic error matrix $W_{\boldsymbol{\theta}}(\boldsymbol{\nu}) = \mathbb{E}_{\boldsymbol{\theta}}[(\boldsymbol{\nu} - \boldsymbol{p})(\boldsymbol{\nu} - \boldsymbol{p})^t]$ of the classical estimate $\boldsymbol{\nu}$ by

$$V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \frac{1}{n^2} T^{-1} W_{\boldsymbol{\theta}}(\boldsymbol{\nu}) (T^{-1})^t. \quad (2.42)$$

Here both $V_{\boldsymbol{\theta}}$ and $W_{\boldsymbol{\theta}}$ depend by (2.29) as well on the state $\boldsymbol{\theta}$ as on the particular measurement scheme \mathcal{M} .

Since the individual measurements are statistically independent, W_{θ} is a block diagonal matrix where the blocks $W_{\theta}^{(k)}$ are related to the measurements $M^{(k)}$:

$$W_{\theta} = \begin{pmatrix} W^{(1)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & W^{(m)} \end{pmatrix} \quad (2.43)$$

If we use the relative frequencies (2.40) as a classical estimate, the matrix elements of $W^{(k)}$ are given as (see Chapter 2.1.1)

$$W_{(i,j)}^{(k)} = \frac{1}{N^{(k)}} \left(\delta_{ij} p_i^{(k)} - p_i^{(k)} p_j^{(k)} \right) \quad (1 \leq k \leq m; 1 \leq i, j \leq d^{(k)} - 1). \quad (2.44)$$

The mean quadratic error matrix of the constrained estimate is given by

$$V_{\theta}(\hat{\theta} + \Delta) = V_{\theta}(\hat{\theta}) + E_{\theta}[\Delta \Delta^t] + E_{\theta}[\Delta(\theta - \hat{\theta})^t + (\theta - \hat{\theta})\Delta^t] \quad (2.45)$$

Since $\hat{\theta}$ and Δ are bounded and $\Delta = 0$ if $\hat{\theta} \notin \mathcal{T}$, by (2.38) the mean quadratic error matrix of the constrained estimate $\hat{\theta}_c$ approaches the one of the unconstrained $\hat{\theta}$ exponentially fast on the interior $Int(\mathcal{T})$. Using the formula $E_{\theta}[X] = \text{Prob}(A)E_{\theta}[X|A] + \text{Prob}(\bar{A})E_{\theta}[X|\bar{A}]$ for a random variable X we get

$$\|V_{\theta}(\hat{\theta}_c) - V_{\theta}(\hat{\theta})\| \leq \text{Prob}_{\theta}(\hat{\theta} \notin \mathcal{T}) M \quad (2.46)$$

where M is a constant. If estimate ν is (weakly) consistent, i.e. $\lim_{n \rightarrow \infty} W_{\theta}(\nu) = 0$, it follows from (2.42) that the same holds for $\hat{\theta}$ as well as for $\hat{\theta}_c$.

Information gain

An other way, used in [28], to quantify the quality of an estimation scheme can be obtained from an information theoretic point of view: For this purpose we define the differential entropy of a random variables \mathbf{X} with density $f(\mathbf{X})$ and support $\text{supp}(f) = S$ as $h(\mathbf{X}) = -\int_S f(\mathbf{x}) \ln f(\mathbf{x}) d\mathbf{x}$. The entropy can be interpreted as a measure of uncertainty of a random variable (see e.g. [4]). Assume we know that the true states $\boldsymbol{\theta}$ are distributed according to a prior probability measure μ on \mathcal{T} with density $f_0(\mathbf{x})$, $\text{supp}(f_0) = \mathcal{T}$. Then we can define the **information gain** \mathcal{I} in an estimation scheme by the difference of the entropy of the prior distribution, related to the uncertainty about the true state before the estimation, and the entropy of the estimate $\hat{\boldsymbol{\theta}}$ given $\boldsymbol{\theta}$, related to the uncertainty after inferring about the state:

$$\mathcal{I}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = h(\boldsymbol{\theta}) - h(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}) \quad (2.47)$$

where $h(\boldsymbol{\theta}) = -\int_{\mathcal{T}} f_0(\mathbf{x}) \ln f_0(\mathbf{x}) d\mathbf{x}$ is the entropy of the prior distribution and $h(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}) = \int_{\mathcal{T}} f_0(\mathbf{x}) h(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta} = \mathbf{x}) d\mathbf{x}$ is the conditional entropy of $\hat{\boldsymbol{\theta}}$ given $\boldsymbol{\theta}$. Let us remark that this definition requires an estimate $\hat{\boldsymbol{\theta}}$ that has an absolutely continuous distribution. This is not the case for the estimates we discussed in the previous section. Therefore either a modified definition of the information gain or a modified estimate has to be used. This problem is discussed in the remark at the end of the section. According to [28], for a uniform prior distribution the information gain is equal to the mutual information $\mathcal{I}(\hat{\boldsymbol{\theta}} : \boldsymbol{\theta}) = h(\boldsymbol{\theta}) - h(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$. The mutual information measures the information that $\hat{\boldsymbol{\theta}}$ contains about $\boldsymbol{\theta}$ [14].

The information gain can be related to the mean quadratic error matrix in the limit of large sample sizes: For simplicity let us consider the relative frequencies $\boldsymbol{\nu}$ as an estimate for \mathbf{p} and let us discuss the case of equal sample sizes, i.e. $N_{min} = N_{max}$ only. By the central limit theorem, for large N the distribution of the relative frequency vector $\boldsymbol{\nu}_N$ (as

the sum of the appearances of the specific outcomes) given $\mathbf{p} = p$ converges to a normal distribution:

$$\frac{1}{\sqrt{N}}(\boldsymbol{\nu}_N | \mathbf{p} = p) \xrightarrow{D} \mathcal{N}(p, \widehat{W}) \quad (2.48)$$

where D denotes convergence in distribution and $\widehat{W} := W_{N^{(k)}=1}$ is the mean quadratic error matrix of $\boldsymbol{\nu}$ for $N^{(k)} = 1$. The density function of the normal distribution $\mathcal{N}(0, V)$ for a r -dimensional random variable is given by

$$f_{\mathcal{N}}(\mathbf{x}) = \frac{1}{(2\pi)^{r/2} \text{Det}(V)^{1/2}} e^{-1/2 \langle \mathbf{x}, V^{-1} \mathbf{x} \rangle}. \quad (2.49)$$

Since the map T is a linear transformation, the distribution of the estimates $\hat{\boldsymbol{\theta}}_N$ given the true state $\boldsymbol{\theta} = \theta$ converge also to a normal distribution:

$$\frac{1}{\sqrt{N}}(\hat{\boldsymbol{\theta}}_N | \boldsymbol{\theta} = \theta) \xrightarrow{D} \mathcal{N}(\theta, V_{\boldsymbol{\theta}}). \quad (2.50)$$

For a normal distributed r -dimensional random variable $\mathbf{X} \sim \mathcal{N}(0, V)$ the differential entropy depends on the covariance matrix V only and is given by

$$h(\mathbf{X}) = \frac{1}{2} \ln((2\pi e)^r \text{Det}(V)). \quad (2.51)$$

For a sequence of random variables \mathbf{X}_N that converge to $\mathbf{X} \sim \mathcal{N}(0, V)$, their entropies converge to $h(\mathbf{X})$ if $h(\mathbf{X}_N)$ is finite for some N [2]. If the $\hat{\boldsymbol{\theta}}_N$ fulfill this condition⁶, the conditional entropy tends to

$$h(\hat{\boldsymbol{\theta}} | \boldsymbol{\theta} = \theta) \rightarrow \frac{1}{2} \ln V_{\boldsymbol{\theta}} + \frac{1}{2} \ln(2\pi e)^{n^2-1} - \frac{1}{2} \ln N \quad (2.52)$$

⁶See the remark at the end of the section.

in the limit of large N , where we used the formula $h(a\mathbf{X}) = h(\mathbf{X}) + \ln |a|$. The uncertainty of a normal distributed random variable can be illustrated by defining the **uncertainty volume** of the distribution $\mathcal{N}(0, V)$ as the volume of the set for which the density $f_{\mathcal{N}}$ of $\mathcal{N}(0, V)$ exceeds $1/e$ times its maximum value [28]

$$V_e = \text{Vol}\left(\left\{x : f_{\mathcal{N}}(x) \geq \frac{1}{e} (2\pi)^{n^2-1} |V|^{-1/2}\right\}\right). \quad (2.53)$$

This region corresponds to the ellipsoid enclosed by the hypersurface on which the argument of the exponential function in (2.49) is constant one. For the limiting distribution of the estimate $\hat{\boldsymbol{\theta}}$ this is given as

$$\left\langle (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \frac{1}{n^2} T^* W^{-1} T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right\rangle = 1. \quad (2.54)$$

This ellipsoid is the image of a $n^2 - 1$ dimensional sphere of radius n^2 under the linear map

$$(T^* W^{-1} T)^{-1/2} = (T^{-1} W (T^*)^{-1})^{1/2} \quad (2.55)$$

which is well defined as here the covariance matrix is positive definite. Denoting the volume enclosed in this sphere by V_0 , we get the uncertainty volume to be

$$V_e = V_0 \text{Det} \left([T^{-1} W (T^*)^{-1}]^{1/2} \right) = V_0 \frac{\text{Det}(W)^{1/2}}{\text{Det}(T)} = V_0 \frac{\prod_k \text{Det}(W^{(k)})^{1/2}}{\text{Det}(T)}. \quad (2.56)$$

In the last step we used that the error matrix W is a block diagonal matrix and thus its determinant is the product of the determinants of the blocks $W^{(k)}$.

In the case that the true states of the system are distributed uniformly with respect to the Lebesgue measure on \mathcal{T} , the entropy $h(\boldsymbol{\theta})$ of the prior distribution is the logarithm of the volume $V_0 := \text{Vol}(\mathcal{T})$ and for large N the entropy of $\hat{\boldsymbol{\theta}}$ given a certain value θ of $\boldsymbol{\theta}$

tends to

$$h(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta} = \theta) \rightarrow \ln\left(\frac{V_e}{V_0}\right) + \frac{1}{2}\ln(2\pi e)^{n^2-1} - \ln V_0 - \frac{1}{2}\ln N \quad (2.57)$$

From the average of (2.57) over the possible true states with respect to the prior distribution we get for the information gain

$$\mathcal{I}(\hat{\boldsymbol{\theta}}) \rightarrow -\left\langle \ln\left(\frac{V_e}{V_0}\right) \right\rangle - \frac{1}{2}\ln(2\pi e)^{n^2-1} + \ln V_0 + \frac{1}{2}\ln N \quad (2.58)$$

This expression was, up to a normalization constant, used in [28] to define the information gain. Rewritten in terms of the mean quadratic error matrix of this reads as

$$\mathcal{I} \rightarrow -\frac{1}{2}\sum_k \langle \ln(\text{Det}(W^{(k)})) \rangle + \ln(\text{Det}(T)) - \frac{1}{2}\ln(2\pi e)^{n^2-1} + \ln(V_0 V_0) + \frac{1}{2}\ln N \quad (2.59)$$

Remark 1 Let us remark that the definition (2.47) is not proper in the case when the distribution of $\hat{\boldsymbol{\theta}}_N$ is not absolutely continuous with respect to the prior distribution. In particular, this is the case when $\hat{\boldsymbol{\theta}}_N$ is a point estimate based on the relative frequencies $\nu_N^{(k)}$, which are discrete random variables⁷ that take values in the set $\{i/N : 0 \leq i \leq N\}$.

There are two ways to resolve this problem:

Instead of (2.47) in [28] the right hand side of (2.57) was used to define the information gain. The uncertainty volume for a discrete distribution can be defined as

$$V_e = \text{Vol}\left(\left\{x : \text{Prob}(x) \geq \frac{1}{e} \max_X(\text{Prob}(x))\right\}\right) \quad (2.60)$$

⁷ The differential entropy of a discrete random variable X is sometimes defined to be $-\infty$, however in this case, while X converges to a continuous normal distributed random variable \tilde{X} , the entropy $h(X)$ does not converge to $h(\tilde{X})$ [2], nor does its entropy.

and with this definition convergence of the distribution function leads to (2.57) and (2.59).

Alternatively, instead of using a point estimate we can construct an estimate $\hat{\boldsymbol{\theta}}_N^{\text{ac}}$ that obeys a density function. It can be constructed in a similar way as a histogram in the case of a one dimensional random variable: For a given N we partition the interval $[-1/(2N), 1 + 1/(2N)]$ into intervals $[i - 1/(2N), i + 1/(2N)]$ ($0 \leq i \leq N$). Each of this intervals contains exactly one possible value of the relative frequencies $\nu^{(k)}$. This induces a set $C \supset \mathcal{T}$ in \mathbb{R}^{n^2-1} that contains the possible estimates $\hat{\boldsymbol{\theta}}_N$ and a partition of C into cells C_k of volume Δ such that each cell contains exactly one of the possible values of $\hat{\boldsymbol{\theta}}_N$. Then an estimate $\hat{\boldsymbol{\theta}}_N^{\text{ac}}$ is constructed by putting the weight equal to the probability $\text{Prob}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_i^{(k)})$ to the cell C_k if C_k contains $\hat{\boldsymbol{\theta}}_N(\nu_i^{(k)})$. Then $\hat{\boldsymbol{\theta}}_N^{\text{ac}}$ has the density

$$f(\hat{\boldsymbol{\theta}}_N^{\text{ac}} = \boldsymbol{x}) = \text{Prob}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_N : \hat{\boldsymbol{\theta}}_N \in C_k \text{ and } \boldsymbol{x} \in C_k) \Delta^{-1} \quad (2.61)$$

In this case $\hat{\boldsymbol{\theta}}_N$ is a quantization of the random variable $\boldsymbol{\theta}_N^{\text{ac}}$. Under a quantization of a random variable \mathbf{X} we understand the following (see e.g. [4], chapter 9.3): Let $f(x)$ be the density of \mathbf{X} with $\text{supp}(f)$ contained in some compact set C . Then we partition C of \mathbf{X} into N cells C_k of Volume Δ and there is an x_i in each cell such that

$$p_i := f(x_i) \Delta = \int_{C_k} f(x) dx \quad (2.62)$$

and we define the quantized random variable \mathbf{X}^Δ as

$$\mathbf{X}^\Delta = x_i \quad \text{if } \mathbf{X} \in C_k \quad (2.63)$$

and zero else. Then the limit of the entropy $H(\mathbf{X}^\Delta) = -\sum_i p_i \ln p_i$ as Δ is given by

$$H(\mathbf{X}^\Delta) + \ln \Delta \rightarrow h(\mathbf{X}) \quad (2.64)$$

The random variable $\hat{\boldsymbol{\theta}}_N^{\text{ac}}$ converges to a normal distributed random variable as its quantization does. Its differential entropy $h(\hat{\boldsymbol{\theta}}_N^{\text{ac}}|\boldsymbol{\theta} = \theta)$ given the true state θ is finite for all N and it converges to the entropy of the limiting normal distribution.

Chapter 3

Complementarity and State Estimation

In this chapter we study complementarity in the context of state estimation. In the case of an estimation scheme $\mathcal{M} = \{M^{(k)} : 1 \leq i \leq m\}$ using different separated measurements $M^{(k)}$ we can compare the situation when the $M^{(k)}$ are pairwise quasi-orthogonal measurements to the case when they are not. In [28] this problem was studied for measurement schemes that use observables with non-degenerate spectrum. It was shown that the complementary observables obtain asymptotically the maximum information gain. In the present work we show that also for general measurements, as defined in Chapter 1.2.2, quasi-orthogonality leads to maximal efficiency among comparable state estimation schemes. As an efficiency measure of the estimation scheme we average the mean quadratic error matrix of the estimate with respect to a prior distribution of the unknown state. We compare the performance of measurement schemes by means of the determinant of this average. In case of the unconstrained estimate the obtained result applies for finite sample sizes while for the constrained estimate quasi-orthogonal measurements are shown to be asymptotic

optimal. Furthermore, we discuss the relation to the result in [28] and show that it can be extended to general quasi-orthogonal measurements as well. In Chapter 3.2.4 the cases where the existence of quasi-orthogonal measurements is known, mainly in the context of quasi-orthogonal subalgebras of $\mathcal{B}(\mathcal{H})$, are discussed in detail.

In the end of the chapter, we give an explicit calculation of the average mean quadratic error matrix and its determinant in the case of von Neumann measurements related to homogenous Abelian subalgebras and a specific prior distribution of the true states. From the result the performance of all such measurement schemes can be compared.

3.1 Preliminaries

In the following we consider the situation when the unknown state originates from the whole set \mathcal{T} of possible states, i.e. $\mathcal{R} = \mathcal{T}$. As described in Chapter 2.2.1 we parameterize this set by the Bloch vector of the state. We will additionally assume that the possible states are distributed with respect to a prior probability measure μ on \mathcal{T} . The main condition we impose on the measure μ is that it is **unitarily invariant**:

$$\int_A f(\rho) d\mu(\rho) = \int_{U^*AU^*} f(U\rho U^*) d\mu(\rho) \quad \text{and} \quad \int_{\mathcal{T}} d\mu(\rho) = 1 \quad (3.1)$$

for integrable functions f on \mathcal{T} and measurable subsets $A \subset \mathcal{T}$. We denote the average with respect to μ by $\langle \cdot \rangle$. Since unitary conjugation corresponds to an orthogonal transformation of the state vectors $\boldsymbol{\theta}$, an example for such a measure is the normalized Lebesgue measure on $\mathcal{T} \subset \mathbb{R}^{(n^2-1)}$. Another example is given in Chapter 3.3.

3.1.1 Quality of an estimation scheme

In Chapter 2.2.4 we discussed the mean quadratic error matrix as measure of efficiency of an estimate given a certain true state ρ . To evaluate the quality of an estimation scheme we use the average of this measure with respect to the prior distribution μ on the set \mathcal{T} .

If we use the mean quadratic error matrix of the estimate given the true state, the average mean quadratic error matrix is given as

$$\begin{aligned} \langle V_{\theta} \rangle &= \int_{\mathcal{T}} V_{\theta} d\mu(\theta) \\ &= \frac{1}{n^2} T^{-1} \int_{\mathcal{T}} \begin{pmatrix} W^{(1)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & W^{(m)} \end{pmatrix} d\mu(\theta) (T^{-1})^t \end{aligned} \quad (3.2)$$

where the $W^{(k)}$ are the mean quadratic error matrices of the classical estimates $\nu^{(k)}$ and T was defined in (2.31). Since for two different measurement schemes the average mean quadratic error matrices are not necessarily comparable by positive semi-definiteness. Therefore we use, analog to [21],

$$\text{Det}(\langle V_{\theta} \rangle) = \frac{1}{n^2} \text{Det}(\langle W_{\theta} \rangle) \text{Det}(T)^{-2} \quad (3.3)$$

to compare the quality of different measurement schemes.

This quantity is similar to the generalized variance, which is defined as the determinant of the mean quadratic error matrix. It provides single number that can be used to compare different measurement schemes. The generalized variance can be interpreted as a volume related to the scatter of a random variable (see e.g. [12]). Thus a smaller value of the generalized variance (3.12) means better efficiency of the estimation scheme.

3.1.2 Compared estimation schemes

If we want to compare measurement schemes with regard to quasi-orthogonality of the measurements they use, we need to restrict us to certain families within this is meaningful. For instance in the case of a qubit, from the view point of complementarity there is no sense in comparing a measurement scheme using a single POVM, as described in Example 1.1.1, to the standard scheme that uses the measurement of the three complementary spin observables (see Example 2.2.2). Since complementarity of two measurements is a geometric property between the subspaces their measurement operators span, we will compare schemes which differ only in the orientation of this subspaces. For this purpose let us consider unitarily conjugated measurements $M^{(k)}$ and $\hat{M}^{(k)}$

$$\hat{M}^{(k)} = UM^{(k)}U^* = \{UE_i^{(k)}U^* : E_i^{(k)} \in M^{(k)}\} \quad (3.4)$$

where $U \in U_n(\mathbb{C})$. Unitary conjugation of the operators $E_i^{(k)}$ results in an orthogonal transformation their Bloch vectors $\mathbf{u}_i^{(k)}$ (2.25). Thus unitary conjugation of the measurement $M^{(k)}$ results in a rotation of the subspace $\mathcal{S}^{(k)}$, while the geometrical configuration of the vectors $\mathbf{u}_i^{(k)}$ and $\hat{\mathbf{u}}_i^{(k)} = U\mathbf{u}_i^{(k)}U^*$ remains the same.

Example 3.1.1 As an example of two unitarily conjugated measurements consider two observables A and B with non-degenerate spectrum of eigenvalues. The measurements they define are the sets of projections $\{P_i^{(A)} : 1 \leq i \leq n\}$ and $\{P_i^{(B)} : 1 \leq i \leq n\}$ on the one dimensional spans of their eigenvectors. The eigenbasis of A is related to the one of B by some unitary transformation U , and $\{P_i^{(B)} : 1 \leq i \leq n\}$ can be obtained by unitary conjugation of the set $\{P_i^{(A)} : 1 \leq i \leq n\}$ by U . \diamond

We will compare measurement schemes within the following families: Given an informational complete measurement scheme $\mathcal{M} = \{M^{(1)}, M^{(2)} \dots, M^{(m)}\}$ we define the family

$\mathcal{U}(\mathcal{M})$ as the set of all measurement schemes whose elements are unitarily conjugated to the elements of \mathcal{M} :

$$\mathcal{U}(\mathcal{M}) = \{ \mathcal{M}' = \{ U_1 M^{(1)} U_1^*, \dots, U_m M^{(m)} U_m^* \} : \mathcal{M}' \text{ informationally complete and } U^{(k)} \in U_n(\mathbb{C}) \} \quad (3.5)$$

This means that the subspaces $\mathcal{S}^{(k)}$ spanned by the individual measurements $M^{(k)}$ are rotated in \mathcal{S} . In the next section we will show that if such a family contains a measurement scheme of quasi-orthogonal measurements, it is the scheme with the highest efficiency.

3.2 Optimality of Quasi-orthogonal Measurements

In this section we compare estimation schemes as described in Chapter 2 within the families defined in the previous section. We discuss the unconstrained and constrained estimate and compare the estimation schemes by means of the determinant of the average mean quadratic error matrix as well as the average information gain.

3.2.1 Unconstrained estimate

The unconstrained estimate of a state was defined in (2.34) as

$$\hat{\boldsymbol{\theta}} = \frac{1}{n} T^{-1} \left(\boldsymbol{\nu} - \frac{1}{n} \mathbf{r} \right). \quad (3.6)$$

where $\boldsymbol{\nu}$ was an estimate for the classical measurement probabilities. For an estimation scheme using a measurement scheme using m different separate measurements on a finite number of N identical quantum states as described in Chapter 2.2.2, we obtain the following theorem :

Theorem 1 *Let the possible true states of the system be distributed according to a unitarily*

invariant probability measure μ on \mathcal{T} and let $\mathcal{M}_0 = \{M_0^{(1)}, M_0^{(2)}, \dots, M_0^{(m)}\}$ be a measurement scheme consisting of pairwise quasi-orthogonal measurements $M_0^{(k)}$, $1 \leq k \leq m$. Let $\mathcal{M}_1 = \{U_1 M_0^{(1)} U_1^*, U_2 M_0^{(2)} U_2^* \dots, U_m M_0^{(m)} U_m^*\} \in \mathcal{U}(\mathcal{M}_0)$ be another measurement scheme obtained by unitaries from \mathcal{M}_0 . Then

$$\text{Det}(\langle V^{\mathcal{M}_0}(\hat{\theta}) \rangle) \leq \text{Det}(\langle V^{\mathcal{M}_1}(\hat{\theta}) \rangle)$$

Thus \mathcal{M}_0 is optimal within the family $\mathcal{U}(\mathcal{M}_0)$.

Proof: For the proof we first show that the average mean quadratic error matrix $\langle W_{\theta} \rangle$ of the classical estimate is the same for both \mathcal{M}_0 and \mathcal{M}_1 and only the matrix T , as defined in (2.31), depends on the particular choice of the measurement scheme. To see this we notice, that the blocks $W_{\theta}^{(k)}$ are functions of the probabilities $p_i^{(k)}$ of the k th measurement only, and as such they are functions of ρ and $M^{(k)}$:

$$W_{\theta}^{(k)} = W^{(k)}(p_1^{(k)}, p_2^{(k)}, \dots, p_{d^{(k)}-1}^{(k)}) = W^{(k)}(\rho, M^{(k)}). \quad (3.7)$$

By our assumption every measurement $M^{(k)} := U_k M_0^{(k)} U_k^*$ in \mathcal{M}_1 is unitarily conjugated to some $M_0^{(k)}$. Consequently an element $E_i^{(k)} \in M^{(k)}$ is unitarily conjugated to the element $E_{0,i}^{(k)} \in M_0^{(k)}$. From (1.10) and the cyclic invariance of the trace we get

$$p_i^{(k)} = \text{Tr}(\rho E_i^{(k)}) = \text{Tr}(\rho (U_k E_{0,i}^{(k)} U_k^*)) = \text{Tr}((U_k^* \rho U_k) E_{0,i}^{(k)}). \quad (3.8)$$

This allows us to move the unitarily conjugation between the arguments of (3.7) and we get the relation

$$W^{(k)}(\rho, M^{(k)}) = W^{(k)}(\rho, U_k M_0^{(k)} U_k^*) = W^{(k)}(U_k^* \rho U_k, M_0^{(k)}). \quad (3.9)$$

By our assumption on the measure μ substituting $U_k^* \rho U_k \rightarrow \rho$ will not change the value of the integrals of the blocks $W^{(k)}$:

$$\begin{aligned} \langle W_{\boldsymbol{\theta}}^{(k)} \rangle &= \int_{\mathcal{T}} W^{(k)}(\rho, M^{(k)}) d\mu(\rho) \\ &= \int_{\mathcal{T}} W^{(k)}((U_k^* \rho U_k, M_0^{(k)}) d\mu(\rho) = \int_{\mathcal{T}} W^{(k)}(\rho, M_0^{(k)}) d\mu(\rho) \dots \end{aligned} \quad (3.10)$$

Thus for \mathcal{M}_0 and \mathcal{M}_1 the average mean quadratic error matrix $\langle W_{\boldsymbol{\theta}} \rangle$ of the estimate ν of the classical probabilities is identical and we get for the average mean quadratic error matrix of $\hat{\boldsymbol{\theta}}$:

$$\langle V_{\boldsymbol{\theta}}^{\mathcal{M}_i} \rangle = \frac{1}{n^2} T_{\mathcal{M}_i}^{-1} \langle W_{\boldsymbol{\theta}} \rangle (T_{\mathcal{M}_i}^{-1})^t = \frac{1}{n^2} T_{\mathcal{M}_i}^{-1} \begin{pmatrix} A^{(1)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & A^{(m)} \end{pmatrix} (T_{\mathcal{M}_i}^{-1})^t \quad (3.11)$$

for $i = 0, 1$ with some constant matrices $A^{(k)} := \langle W_{\boldsymbol{\theta}}^{(k)} \rangle$.

Thus for all measurement schemes from the family $\mathcal{U}(\mathcal{M}_0)$ the mean quadratic error matrix $\langle V_{\boldsymbol{\theta}} \rangle$ depends on the transformation matrix T only. By

$$\text{Det}(\langle V_{\boldsymbol{\theta}} \rangle) = \frac{1}{n^2} \text{Det}(\langle W_{\boldsymbol{\theta}} \rangle) \text{Det}(T)^{-2} \quad (3.12)$$

it is enough to show that for all measurement schemes $\mathcal{M} \in \mathcal{U}(\mathcal{M}_0)$ the $\text{Det}(T)$ is maximal for \mathcal{M}_0 . $\text{Det}(T)$ corresponds to the volume of the parallelepiped $\text{Par}\{\mathbf{u}_i^{(k)} : 1 \leq i \leq (d^{(k)} - 1), 1 \leq k \leq m\}$ spanned by the row vectors in T . To calculate this volume for a given T we apply the Gram-Schmidt orthogonalization method (without normalization) on its row vectors: We can perform the orthogonalization first in the subspaces $\mathcal{S}^{(k)}$ spanned by the vectors $\{\mathbf{u}_i^{(k)} : 1 \leq i \leq (d^{(k)} - 1)\}$ independently. We start with $\mathbf{u}_1^{(k)} =: \tilde{\mathbf{u}}_1^{(k)}$ and

obtain recursively

$$\tilde{\mathbf{u}}_i^{(k)} = \mathbf{u}_i^{(k)} - P_{i-1} \mathbf{u}_i^{(k)} \quad (3.13)$$

where P_{i-1} is the orthogonal projection on $\text{span}\{\tilde{\mathbf{u}}_1^{(k)}, \tilde{\mathbf{u}}_2^{(k)}, \dots, \tilde{\mathbf{u}}_{i-1}^{(k)}\}$ in $\mathcal{B}(\mathcal{H})$. The volume of the parallelepiped spanned by the resulting vectors remains invariant in this procedure since (3.13) are elementary row operations on T . If the $\mathcal{S}^{(k)}$ are already orthogonal, orthogonalization in the subspaces $\mathcal{S}^{(k)}$ is sufficient and

$$\begin{aligned} \text{Det}(T) &= \text{Vol}(\text{Par}\{\mathbf{u}_i^{(k)} : 1 \leq i \leq (d^{(k)} - 1), 1 \leq k \leq m\}) \\ &= \text{Vol}(\text{Par}\{\tilde{\mathbf{u}}_i^{(k)} : 1 \leq i \leq (d^{(k)} - 1), 1 \leq k \leq m\}) \\ &= \prod_{k=1}^m \prod_{i=1}^{d^{(k)}-1} \|\tilde{\mathbf{u}}_i^{(k)}\|. \end{aligned} \quad (3.14)$$

Otherwise we have to continue the orthogonalization procedure. In this case first note that for measurements $M^{(k)}$ and $\hat{M}^{(k)}$ unitarily conjugated by some U the Gram-Schmidt procedure gives a geometrically equal result in the subspaces $\mathcal{S}^{(k)}$ and $\hat{\mathcal{S}}^{(k)}$. Orthogonalization and unitary conjugation can be interchanged and the following diagram commutes:

$$\begin{array}{ccc} M^{(k)} : & \{\mathbf{u}_1^{(k)}, \mathbf{u}_2^{(k)}, \dots, \mathbf{u}_{d^{(k)}-1}^{(k)}\} & \xrightarrow{G-S} & \{\tilde{\mathbf{u}}_1^{(k)}, \tilde{\mathbf{u}}_2^{(k)}, \dots, \tilde{\mathbf{u}}_{d^{(k)}-1}^{(k)}\} \\ & \downarrow U & & \downarrow U \\ \hat{M}^{(k)} : & U\{\mathbf{u}_1^{(k)}, \mathbf{u}_2^{(k)}, \dots, \mathbf{u}_{d^{(k)}-1}^{(k)}\}U^* & \xrightarrow{G-S} & U\{\tilde{\mathbf{u}}_1^{(k)}, \tilde{\mathbf{u}}_2^{(k)}, \dots, \tilde{\mathbf{u}}_{d^{(k)}-1}^{(k)}\}U^* \end{array} \quad (3.15)$$

In particular, since unitary conjugation by U results in a orthogonal rotation of the Bloch vectors $\mathbf{u}_i^{(k)}$, we have $\|\tilde{\mathbf{u}}_i^{(k)}\| = \|U\tilde{\mathbf{u}}_i^{(k)}U^*\|$.

If we need to continue the orthogonalization procedure, this will decrease the length of the vectors $\tilde{\mathbf{u}}_i^{(k)}$, and by (3.14) this results in a smaller volume of $\text{Par}\{\mathbf{u}_i^{(k)} : 1 \leq i \leq (d^{(k)} - 1), 1 \leq k \leq m\}$. Thus T has maximal determinant, if (and actually only if) the

$\mathcal{S}^{(k)}$ are orthogonal. □

3.2.2 Constrained estimate

Since the unconstrained estimate may fall outside the set of states, a constrained estimate was constructed from the classical estimate $\boldsymbol{\nu}$ in (2.35) by adding a correction $\Delta(\boldsymbol{\nu})$ as

$$\hat{\boldsymbol{\theta}}_c(\boldsymbol{\nu}) = \hat{\boldsymbol{\theta}}(\boldsymbol{\nu}) + \Delta(\boldsymbol{\nu}) \quad \text{such that} \quad \hat{\boldsymbol{\theta}}_c \in \mathcal{T} \quad \text{and} \quad \Delta(\boldsymbol{\nu}) = 0 \quad \text{if} \quad \hat{\boldsymbol{\theta}}(\boldsymbol{\nu}) \in \mathcal{T}. \quad (3.16)$$

For the constrained estimate $\hat{\boldsymbol{\theta}}_c$ already the explicit calculation of the mean quadratic error matrix is much more complicated since $\hat{\boldsymbol{\theta}}_c$ is not a simple linear transformation of the classical estimate $\boldsymbol{\nu}$ anymore. Nevertheless, we know that for large sample sizes $\hat{\boldsymbol{\theta}}_c$ is with high probability equal to $\hat{\boldsymbol{\theta}}$. Therefore the optimality of quasi-orthogonal measurement schemes can be shown for the constrained estimate in the limit when the number of available copies of the unknown state tends to infinity. This is stated by the following

Theorem 2 *Let the possible true states of the system be distributed according to a unitarily invariant probability measure μ on \mathcal{T} with $\mu(\partial\mathcal{T}) = 0$. Let the measurement scheme $\mathcal{M}_0 = \{M_0^{(1)}, M_0^{(2)} \dots, M_0^{(m)}\}$ consist of pairwise quasi-orthogonal measurements $M^{(k)}$, $1 \leq k \leq m$, and let $\mathcal{M}_1 = \{U_1 M_0^{(1)} U_1^*, U_2 M_0^{(2)} U_2^* \dots, U_m M_0^{(m)} U_m^*\} \in \mathcal{U}(\mathcal{M}_0)$ be another measurement scheme obtained by unitaries from \mathcal{M}_0 . If $\hat{\boldsymbol{\theta}}_c \in \mathcal{T}$ is an estimate of the form (3.16), then the inequality*

$$\text{Det}(\langle V^{\mathcal{M}_0}(\hat{\boldsymbol{\theta}}_c) \rangle) \leq \text{Det}(\langle V^{\mathcal{M}_1}(\hat{\boldsymbol{\theta}}_c) \rangle)$$

holds if $N_{\min} = \min\{N^{(k)}\}$ is large enough.

Proof: Since the estimates of $\boldsymbol{\theta}$ we constructed in the previous chapter are consistent for any informationally complete measurement scheme, in all cases the mean quadratic

error matrix will asymptotically go to zero. Thus for the proof of the theorem we show that for the two measurement schemes \mathcal{M}_0 and \mathcal{M}_1 the error of the constrained estimates approaches the error of the unconstrained ones faster than the difference between the errors of the unconstrained estimators of the two schemes goes to zero. For this it is sufficient to show

$$\frac{|\text{Det}\langle V_{\theta}(\hat{\theta}^{\mathcal{M}_i}) \rangle - \text{Det}\langle V_{\theta}(\hat{\theta}_c^{\mathcal{M}_i}) \rangle|}{|\text{Det}\langle V_{\theta}(\hat{\theta}^{\mathcal{M}_0}) \rangle - \text{Det}\langle V_{\theta}(\hat{\theta}^{\mathcal{M}_1}) \rangle|} \rightarrow 0 \quad \text{for both } i = 0, 1 \quad (3.17)$$

as $N_{\min} \rightarrow \infty$. To ensure that the denominator is different from zero, we need to exclude the case when the average mean quadratic error matrix of \mathcal{M}_1 has the same determinant as the one of \mathcal{M}_0 . By the proof of Theorem 1 this implies that \mathcal{M}_1 consists of pairwise quasi-orthogonal measurements as well and we can find a single unitary U such that $\mathcal{M}_1 = U\mathcal{M}_0U^*$. In this case, however, we obtain equality in Theorem 2.

For the proof of (3.17) let us consider closed sets \mathcal{T}_ϵ that are contained in the interior of \mathcal{T} and have probability $(1 - \epsilon)$ with respect to the prior distribution μ :

$$\mathcal{T}_\epsilon \subset \text{Int}(\mathcal{T}), \quad \mathcal{T}_\epsilon \text{ closed}, \quad \mu(\mathcal{T}_\epsilon) = 1 - \epsilon. \quad (3.18)$$

This is possible since we assumed that $\mu(\partial\mathcal{T}) = 0$. Then we can evaluate the nominator in (3.17) under the condition that the true state is from the set \mathcal{T}_ϵ and we show first that

$$\frac{|\text{Det}\langle V_{\theta}(\hat{\theta}^{\mathcal{M}_i}) | \mathcal{T}_\epsilon \rangle - \text{Det}\langle V_{\theta}(\hat{\theta}_c^{\mathcal{M}_i}) | \mathcal{T}_\epsilon \rangle|}{|\text{Det}\langle V_{\theta}(\hat{\theta}^{\mathcal{M}_1}) \rangle - \text{Det}\langle V_{\theta}(\hat{\theta}^{\mathcal{M}_2}) \rangle|} \rightarrow 0 \quad \text{for both } i = 0, 1 \quad (3.19)$$

as $N_{\min} \rightarrow \infty$. By Weyl's Perturbation theorem (see e.g. [3]) for self-adjoint matrices A and B of the same size, the maximum difference between the eigenvalues $\lambda_i^\downarrow(A)$ of A , ordered in descending order, and the $\lambda_i^\downarrow(B)$ of B is smaller than the operator norm¹ of their

¹ For a self-adjoint operator A the operator norm is defined as $\|A\| = \max\{|\lambda_i(A)| : \lambda_i(A) \text{ is an eigenvalue of } A\}$. Let us also note that in finite dimensions all norms are equivalent.

difference

$$\max_j |\lambda_j^\downarrow(A) - \lambda_j^\downarrow(B)| \leq \|A - B\| \quad (3.20)$$

Since the determinant of a positive² matrix is the product of the eigenvalues, $\text{Det}(A) = \prod_j \lambda_j^\downarrow(A)$, the difference of the determinants of A and B is bounded by

$$\text{Det}(A) - \text{Det}(B) \leq \|A - B\| c_1 + \|A - B\|^2 c_2 + \dots + \|A - B\|^n c_n \quad (3.21)$$

with coefficients c_i depending on the spectrum of A .

We can apply the above to the average mean quadratic error matrices of the constrained and unconstrained estimate. Since additionally $\|\langle V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c^{\mathcal{M}_i}) | \mathcal{T}_\epsilon \rangle - \langle V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_i}) | \mathcal{T}_\epsilon \rangle\| \leq \langle \|V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c^{\mathcal{M}_i}) - V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_i})\| | \mathcal{T}_\epsilon \rangle$ holds³, the following is sufficient for (3.19):

$$\frac{\langle \|V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c^{\mathcal{M}_i}) - V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_i})\| | \mathcal{T}_\epsilon \rangle}{|\text{Det}\langle V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_1}) \rangle - \text{Det}\langle V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_2}) \rangle|} \rightarrow 0 \quad \text{for both } i = 0, 1 \quad (3.22)$$

as $N_{\min} \rightarrow \infty$. To show (3.22) we look at the asymptotic behavior of the denominator and nominator:

Since the Fisher information grows proportional to the sample size, by the Cramer-Rao inequality (2.7) the elements of the mean quadratic error matrices of the optimal classical estimate tend to zero proportional to the inverse of the sample sizes $N^{(k)}$: $W^{(k)} = \frac{1}{N^{(k)}} W_{N^{(k)}=1}^{(k)} =: \frac{1}{N^{(k)}} \widehat{W}^{(k)}$. Consequently, taking into account (3.12), we get for two different

²Let us assume that A and B are positive, since this is the case for the mean quadratic error matrix. Note that a positive operator is self-adjoint.

³ The norm is a convex function and therefore this follows from Jensen's inequality

measurement schemes \mathcal{M}_0 and \mathcal{M}_1

$$\begin{aligned}
& \left| \text{Det}\langle V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_1}) \rangle - \text{Det}\langle V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_2}) \rangle \right| \\
&= \frac{1}{n^2} \left| \text{Det}(T_{\mathcal{M}_1})^{-2} \prod_{k=1}^m \text{Det}\langle W_{\mathcal{M}_1}^{(k)} \rangle - \text{Det}(T_{\mathcal{M}_2})^{-2} \prod_{k=1}^m \text{Det}\langle W_{\mathcal{M}_2}^{(k)} \rangle \right| \\
&= \frac{1}{n^2} \left| \text{Det}(T_{\mathcal{M}_1})^{-2} \prod_{k=1}^m \text{Det}\langle \frac{1}{N^{(k)}} \widehat{W}_{\mathcal{M}_1}^{(k)} \rangle - \text{Det}(T_{\mathcal{M}_2})^{-2} \prod_{k=1}^m \text{Det}\langle \frac{1}{N^{(k)}} \widehat{W}_{\mathcal{M}_2}^{(k)} \rangle \right| \quad (3.23) \\
&= \frac{C_1}{\prod_{k=1}^m N^{(k)(d^{(k)}-1)}}
\end{aligned}$$

with some constant C_1 depending on the choice of \mathcal{M}_1 and \mathcal{M}_2 . For the last equality recall that the $W^{(k)}$ are $(d^{(k)} - 1) \times (d^{(k)} - 1)$ matrices and therefore $\text{Det}(\frac{1}{N^{(k)}} \widehat{W}^{(k)}) = \frac{1}{N^{(k)(d^{(k)}-1)}} \text{Det}(\widehat{W}^{(k)})$.

In calculating the mean of the differences of the quadratic error matrices of in the nominator of (3.22) only values of $\boldsymbol{\nu}$ contribute where $\hat{\boldsymbol{\theta}}_c$ is different from $\hat{\boldsymbol{\theta}}$. Thus with the formula $\text{E}_{\boldsymbol{\theta}}[\mathbf{X}] = \text{Prob}(A)\text{E}_{\boldsymbol{\theta}}[\mathbf{X}|A] + \text{Prob}(\bar{A})\text{E}_{\boldsymbol{\theta}}[\mathbf{X}|\bar{A}]$ we get for the denominator

$$\begin{aligned}
\|V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c) - V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})\| &= \|\text{E}_{\boldsymbol{\theta}}[\Delta\Delta^t + \Delta(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^t + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\Delta^t]\| \\
&= \text{Prob}(\hat{\boldsymbol{\theta}} \neq \hat{\boldsymbol{\theta}}_c) \|\text{E}_{\boldsymbol{\theta}}[\Delta\Delta^t + \Delta(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^t + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\Delta^t | \hat{\boldsymbol{\theta}} \neq \hat{\boldsymbol{\theta}}_c]\| \quad (3.24) \\
&\leq \text{Prob}(\hat{\boldsymbol{\theta}} \notin \mathcal{T}) C_2
\end{aligned}$$

where we chose some constant and finite C_2 such that this holds for both \mathcal{M}_1 and \mathcal{M}_2 . According to Sanov's theorem (2.38) the right hand side vanishes exponentially fast with growing N_{\min} if $\boldsymbol{\theta} \in \text{Int}(\mathcal{T})$. Consequently, in the limit of large N we can bound the

quotient of the left hand sides of (3.24) and (3.23) by

$$\begin{aligned}
& \limsup_{N_{\min} \rightarrow \infty} \frac{\|V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c) - V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})\|}{|\text{Det}\langle V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_1}) \rangle - \text{Det}\langle V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_2}) \rangle|} \\
& \leq C_3 \limsup_{N_{\min} \rightarrow \infty} \prod_{k=1}^m N^{(k)(d^{(k)}-1)} \text{Prob}(\hat{\boldsymbol{\theta}} \notin \mathcal{T}) \\
& \leq C_3 \limsup_{N_{\min} \rightarrow \infty} \prod_{k=1}^m N^{(k)(d^{(k)}-1)} \exp(-N_{\min}(D(\boldsymbol{\nu}^* \|\mathbf{p}) - \delta)).
\end{aligned} \tag{3.25}$$

for all $\delta > 0$ and a constant C_3 . On \mathcal{T}_ϵ we have $D(\boldsymbol{\nu}^* \|\mathbf{p}) \geq D_\epsilon > 0$. If additionally the size of the largest sample does not grow in N_{\min} exponentially, i.e. $N_{\max} \leq N_{\min}^r$ for some $r \in \mathbb{N}$, the limit in the last line equals zero for all $\boldsymbol{\theta} \in \mathcal{T}_\epsilon$ and

$$\frac{\|V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c) - V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})\|}{|\text{Det}\langle V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_1}) \rangle - \text{Det}\langle V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_2}) \rangle|} \rightarrow 0 \tag{3.26}$$

as $N_{\min} \rightarrow \infty$ on \mathcal{T}_ϵ . Since $D_\epsilon > 0$ convergence in (3.26) is uniform on \mathcal{T}_ϵ and (3.22) holds for the average of the expression in (3.26) and for all $\epsilon > 0$.

To complete the proof of the original inequality (3.17), we note that the contribution of the set $\mathcal{T} \setminus \mathcal{T}_\epsilon$ to the average of the mean quadratic error matrix is small: Again by the formula $E_\theta[\mathbf{X}] = \text{Prob}(A)E_\theta[\mathbf{X}|A] + \text{Prob}(\bar{A})E_\theta[\mathbf{X}|\bar{A}]$ we have

$$\begin{aligned}
\langle \|V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c^{\mathcal{M}_i}) - V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_i})\| \rangle &= \langle \|V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c^{\mathcal{M}_i}) - V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_i})\| | \mathcal{T}_\epsilon \rangle \\
&+ \epsilon \left(\langle \|V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c^{\mathcal{M}_i}) - V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_i})\| | \mathcal{T} \setminus \mathcal{T}_\epsilon \rangle - \langle \|V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c^{\mathcal{M}_i}) - V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_i})\| | \mathcal{T}_\epsilon \rangle \right)
\end{aligned} \tag{3.27}$$

and therefore

$$\frac{\langle \|V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c^{\mathcal{M}_i}) - V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_i})\| \rangle}{\langle \|V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c^{\mathcal{M}_i}) - V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_i})\| | \mathcal{T}_\epsilon \rangle} = 1 + \epsilon \left(\frac{\overbrace{\langle \|V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c^{\mathcal{M}_i}) - V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_i})\| | \mathcal{T} \setminus \mathcal{T}_\epsilon \rangle}^*}{\langle \|V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c^{\mathcal{M}_i}) - V_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^{\mathcal{M}_i})\| | \mathcal{T}_\epsilon \rangle} - 1 \right) \tag{3.28}$$

By the Cramer-Rao inequality the mean quadratic error matrix is proportional to the inverse of the sample size for the unconstrained estimate as well as for the constrained estimate. Thus the nominator and denominator of the term (*) vanish equally fast and this term can be bounded by a constant C_4 for all N . Thus

$$\begin{aligned} \frac{\langle \|V_{\theta}(\hat{\theta}_c^{\mathcal{M}_i}) - V_{\theta}(\hat{\theta}^{\mathcal{M}_i})\| \rangle}{|\text{Det}\langle V_{\theta}(\hat{\theta}^{\mathcal{M}_1}) \rangle - \text{Det}\langle V_{\theta}(\hat{\theta}^{\mathcal{M}_2}) \rangle|} &= \frac{\langle \|V_{\theta}(\hat{\theta}_c^{\mathcal{M}_i}) - V_{\theta}(\hat{\theta}^{\mathcal{M}_i})\| | \mathcal{T}_{\epsilon} \rangle}{|\text{Det}\langle V_{\theta}(\hat{\theta}^{\mathcal{M}_1}) \rangle - \text{Det}\langle V_{\theta}(\hat{\theta}^{\mathcal{M}_2}) \rangle|} \\ &+ \epsilon (C_4 - 1) \frac{\langle \|V_{\theta}(\hat{\theta}_c^{\mathcal{M}_i}) - V_{\theta}(\hat{\theta}^{\mathcal{M}_i})\| | \mathcal{T}_{\epsilon} \rangle}{|\text{Det}\langle V_{\theta}(\hat{\theta}^{\mathcal{M}_1}) \rangle - \text{Det}\langle V_{\theta}(\hat{\theta}^{\mathcal{M}_2}) \rangle|} \end{aligned} \quad (3.29)$$

and taking the limit of N in the last equation gives the desired result. \square

3.2.3 Average information gain

In the previous section we studied the efficiency of measurement schemes based on the mean quadratic error matrix. As an alternative measure of the efficiency of an estimation scheme the information gain $\mathcal{I}(\hat{\theta})$ was discussed in Chapter 2.2.4. In [28] it was shown that among estimation schemes using observables with non-degenerate spectrum $\mathcal{I}(\hat{\theta})$ is maximal in the case when they are complementary. As discussed in Chapter 2.2.4 the information gain is asymptotically related to the mean quadratic error matrix of $\hat{\theta}$. Therefore we can extend this result to general measurements (POVMs) as well and for the unconstrained estimate we get the following

Theorem 3 *Let the possible true states of the system be distributed according to a unitarily invariant probability measure μ on \mathcal{T} and let $\mathcal{M}_0 = \{M_0^{(1)}, M_0^{(2)}, \dots, M_0^{(m)}\}$ be a measurement scheme consisting of pairwise quasi-orthogonal measurements $M_0^{(k)}$, $1 \leq k \leq m$. Let $\mathcal{M}_1 = \{U_1 M_0^{(1)} U_1^*, U_2 M_0^{(2)} U_2^*, \dots, U_m M_0^{(m)} U_m^*\} \in \mathcal{U}(\mathcal{M}_0)$ be another measurement scheme*

obtained by unitaries from \mathcal{M}_0 . Then

$$\lim_{N \rightarrow \infty} I^{\mathcal{M}_0}(\hat{\boldsymbol{\theta}}) - \ln N \geq \lim_{N \rightarrow \infty} I^{\mathcal{M}_1}(\hat{\boldsymbol{\theta}}) - \ln N$$

and \mathcal{M}_0 is asymptotically optimal within the family $\mathcal{U}(\mathcal{M}_0)$.

The proof follows closely the reasoning of Theorem 1. In (2.52) it was shown that in the limit of large N the average information gain tends to

$$\mathcal{I}(\hat{\boldsymbol{\theta}}) \rightarrow -\frac{1}{2} \sum_k \langle \ln(\text{Det}(W^{(k)})) \rangle + \ln(\text{Det}(T)) + \frac{1}{2} \ln N - \frac{1}{2} \ln(2\pi e)^{n^2-1} + h(\boldsymbol{\theta}) \quad (3.30)$$

The average of the information gain depends on the averages of the individual blocks of the mean quadratic error matrix W of the classical estimate $\boldsymbol{\nu}$. In the same way as in the proof of Theorem 1 it follows from (3.9), analog to (3.10), that for a unitarily invariant measure μ

$$\int_{\mathcal{T}} \ln(W^{(k)}(\rho, M^{(k)})) d\mu(\rho) = \int_{U_k \mathcal{T} U_k^*} \ln W^{(k)}\left(\left(\rho, U_k M_0^{(k)} U_k^*\right)\right) d\mu(\rho) \quad (3.31)$$

and the averages $\langle \ln(\text{Det}(W^{(k)})) \rangle$ are constant within the families $\mathcal{U}(\mathcal{M}_0)$. Then the theorem follows from the maximality of the determinant of T for \mathcal{M}_0 . \square

In case of the unconstrained estimate we draw our attention again to the modifications of the remark in Chapter 2.2.4. In case we define the information gain \mathcal{I} through the uncertainty volume (2.60), it is immediate that the previous theorem holds equally true for the constrained estimate since $\text{Prob}(\hat{\boldsymbol{\theta}}_c \neq \hat{\boldsymbol{\theta}})$ tends to zero.

3.2.4 Examples

Let us discuss some examples where the existence of quasi-orthogonal measurements is known. These examples are mostly related to the cases where the existence of a complete set of complementary subalgebras is known. The relation of measurements to subalgebras $\mathcal{A} \subset M_n(\mathbb{C})$ can be understood in the following sense: Similar to the definition of measurement schemes on the whole algebra $M_n(\mathbb{C})$ in Chapter 2.2.2, we can define them on a subalgebra $\mathcal{A} \subset M_n(\mathbb{C})$ only:

$$\mathcal{M}_{\mathcal{A}} := \{M^{(k)} : 1 \leq k \leq m, M^{(k)} \subset \mathcal{A}\} \quad (3.32)$$

On the other hand the operators $\{E_i^{(k)} : E_i^{(k)} \in M^{(k)}, M^{(k)} \in \mathcal{M}_{\mathcal{A}}\}$ contained in some $\mathcal{M}_{\mathcal{A}}$ generate a certain subalgebra \mathcal{A} of $M_n(\mathbb{C})$. If the operators used in the measurements in $\mathcal{M}_{\mathcal{A}}$ at the same time span and generate the algebra \mathcal{A} we can speak about a relation between the subalgebra \mathcal{A} and the measurement scheme $\mathcal{M}_{\mathcal{A}}$. Then $\mathcal{M}_{\mathcal{A}}$ is also informationally complete on the subalgebra \mathcal{A} . Furthermore we can join measurement schemes on different subalgebras \mathcal{A}_i as $\mathcal{M} = \cup_i \mathcal{M}_{\mathcal{A}_i}$ in order to get a measurement scheme on the whole algebra $\mathcal{B}(\mathcal{H})$. This situation applies to the most common cases: A measurement scheme using von Neumann measurements is related to a collection of Abelian subalgebras of $\mathcal{B}(\mathcal{H})$. A measurement scheme built up from measurements on subsystems is related to a collection of matrix subalgebras of $\mathcal{B}(\mathcal{H})$.

A measurement scheme $\mathcal{M}_{\mathcal{A}}$ defined on a certain subalgebra \mathcal{A} naturally transfers to a measurement scheme on a subalgebra $U\mathcal{A}U^*$ by taking the unitary conjugate of the measurement operators in $\mathcal{M}_{\mathcal{A}}$ as $\mathcal{M}_{U\mathcal{A}U^*} := U\mathcal{M}_{\mathcal{A}}U^*$. To apply the above theorems in this setting we compare estimation schemes that are related to subalgebras isomorphic to each other by unitary conjugation. More precisely, suppose there exists a complete set of complementary subalgebras \mathcal{A}_i ($1 \leq i \leq m$) and measurement schemes $\mathcal{M}_{\mathcal{A}_i}$ on them such

that $\mathcal{M}_0 = \cup_{i=1}^m \mathcal{M}_{\mathcal{A}_i}$ is informational complete. Then we consider the family

$$\mathcal{U}(\mathcal{M}_0) := \left\{ \mathcal{M} = \bigcup_{i=1}^m U_i \mathcal{M}_{\mathcal{A}_i} U_i^* : \mathcal{M} \text{ informational complete and } U^{(i)} \in U_n(\mathbb{C}) \right\}.$$

Each of the measurement schemes⁴ \mathcal{M} in $\mathcal{U}(\mathcal{M}_0)$ is related to a set of subalgebras $\{U_k \mathcal{A}_i U_k^* : 1 \leq i \leq m\}$. The family $\mathcal{U}(\mathcal{M}_0)$ is a subset of the families we defined in (3.5) as we conjugate all the measurements in $\mathcal{M}_{\mathcal{A}_i}$ with the same unitary $U^{(i)}$. We can apply the results from the previous sections: For any particular choice of measurement schemes $\mathcal{M}_{\mathcal{A}_i}$ the estimation scheme is optimal within the family $\mathcal{U}(\mathcal{M}_0)$ in the sense of the above theorems if the measurements in \mathcal{M} are related to the complementary subalgebras \mathcal{A}_i .

The cases where existence of a complete set of quasi-orthogonal subalgebras is known was discussed in Chapter 1.2.3. In the following we will discuss this examples in the context of state estimation schemes:

First we will look at commutative subalgebras: A von Neumann measurement $M^{(k)} = \{P_1^{(k)}, \dots, P_d^{(k)}\}$ of an observable A_k is related to the Abelian subalgebra \mathcal{A}_k generated by its eigenprojections. Clearly $M^{(k)}$ is informationally complete on the subalgebra \mathcal{A}_k . For simplicity we shall restrict ourselves to examples of homogenous Abelian algebras generated by projections of rank $rk(P_i^{(k)}) = r$. This corresponds to the measurement of an observable A_k with $\frac{n}{r}$ distinct eigenvalues of multiplicity r . The projections $P_i^{(k)}$ generate a homogenous subalgebra \mathcal{A}_k which is isomorphic to the algebra \mathcal{A}_0 of $n \times n$ diagonal matrices with entries of multiplicity r by unitary conjugation:

$$\mathcal{A}_k = U_k \mathcal{A}_0 U_k^* \tag{3.33}$$

⁴ note again that $U \mathcal{M}_{\mathcal{A}} U^*$ is related to the algebra $U \mathcal{A} U^*$.

for some unitary operators $U^{(k)}$. In this setting we need $m = \frac{r(n^2-1)}{(n-r)}$ such measurements each of them consisting of $d = \frac{n}{r}$ positive operators in order to get an informational complete measurement scheme on $\mathcal{B}(\mathcal{H})$. We get the two extremal cases:

Example 3.2.1 The measurement of observables A_k with a non degenerate spectrum of eigenvalues is related to maximal Abelian subalgebras \mathcal{A}_k of $M_n(\mathbb{C})$. In this case the $P_i^{(k)}$ are minimal projections of $\mathcal{B}(\mathcal{H})$ and in our measurement schemes we have $m = n + 1$ measurements $M^{(k)}$, each of which consist of $d^{(k)} = n$ elements of the form

$$P_i^{(k)} = \frac{1}{n}I + \mathbf{u}_i^{(k)} \quad (1 \leq i \leq n)$$

and the matrix T takes the form

$$T = \begin{bmatrix} T^{(1)} \\ \vdots \\ T^{(n+1)} \end{bmatrix}, \quad T^{(k)} = \begin{bmatrix} \mathbf{u}_1^{(k) t} \\ \vdots \\ \mathbf{u}_{n-1}^{(k) t} \end{bmatrix}, \quad (3.34)$$

It is known in the cases where n is a prime power that a complete set of complementary maximal Abelian subalgebras exists [28]. \diamond

Example 3.2.2 In the case of $n = 2^j$, $j \in \mathbb{N}$, we can consider subalgebras generated by projections of rank $n/2$, which corresponds to the measurement of observables with two distinct eigenvalues of multiplicity $n/2$. Then every measurement gives information about only one parameter and thus we consider measurement schemes \mathcal{M} that consist of $m = n^2 - 1$ measurements $M^{(k)} = \{P_+^{(k)}, P_-^{(k)}\}$ where

$$P_{\pm}^{(k)} = \frac{1}{2}I \pm \mathbf{u}^{(k)} \quad (3.35)$$

The matrix T takes the form

$$T = \begin{bmatrix} \mathbf{u}^{(1) t} \\ \vdots \\ \mathbf{u}^{(n^2-1) t} \end{bmatrix} \quad (3.36)$$

◇

In the case of von Neumann measurements associated with homogenous subalgebras, all the measurements used are unitarily conjugate to a single $M_0 = \{P_i^{(0)} : 1 \leq i \leq m\}$ where $m = n/r$ and M_0 is the measurement associated with the algebra \mathcal{A}_0 in (3.33). Within the families

$$\mathcal{U} = \left\{ \mathcal{M} = \{U_i M_0 U_i^* : 1 \leq i \leq m\} : \mathcal{M} \text{ informational complete, } U^{(i)} \in U_n(\mathbb{C}) \right\}$$

the case when the subalgebras related to the $U_i M_0 U_i^*$ are complementary is optimal under the condition that they exists.

Example 3.2.3 If we replace in Example 3.2.2 the Bloch vectors $\mathbf{u}^{(k)}$ of the measurement operators by Bloch vectors such that the eigenvalues of the $\mathbf{u}^{(k)}$ are smaller than $1/2$ we can always obtain an informational complete set of quasi-orthogonal measurements. The measurement operators

$$E_{\pm}^{(k)} = \frac{1}{2}I \pm \mathbf{u}_i^{(k)} \quad (1 \leq i \leq n^2 - 1)$$

are always positive for $\|\mathbf{u}^{(k)}\| \leq 1/4n$ and we can find quasi-orthogonal measurements by choosing the vectors $\mathbf{u}^{(k)}$ orthogonal. ◇

In the following we present examples that are related to measurement schemes on non-commutative subalgebras corresponding to physical subsystems:

Let us consider the composite of k quantum systems with Hilbert spaces \mathcal{H}_i with

$\dim(\mathcal{H}_i) = n_i$. It is described by the tensor product algebra

$$\mathcal{B}(\mathcal{H}) = \bigotimes_{i=1}^k \mathcal{B}(\mathcal{H}_i) \equiv \bigotimes_{i=1}^k M_{n_i}(\mathbb{C}) \quad (3.37)$$

in a suitable chosen basis of $\mathcal{H} = \otimes_i \mathcal{H}_i$. Let us consider subalgebras $\mathcal{A} \simeq \mathcal{B}(\mathcal{H}_i)$ that are isomorphic to the algebra of one of the subsystems. A measurement $M \subset \mathcal{A}$ contained in the subalgebra \mathcal{A} can be experimentally realized by a measurement on the physical subsystem $\mathcal{B}(\mathcal{H}_i)$ in the following way: We can choose a unitary W such that $W\mathcal{A}W^* \equiv \mathcal{B}(\mathcal{H}_i) \otimes I$ and W can be implemented by the time evolution of the system under some suitable Hamiltonian H as $W^{-1} = e^{iHt/\hbar}$. Under this time evolution a given state $\rho \in \mathcal{B}(\mathcal{H})$ evolves to $W^{-1}\rho(W^{-1})^*$ and performing the conjugate measurement WMW^* , contained in $\mathcal{B}(\mathcal{H}_i)$, on $W^{-1}\rho(W^{-1})^*$ is equivalent to performing M on ρ . Suppose there is a set of subalgebras $\{\mathcal{A}_i : \mathcal{A}_i \simeq \mathcal{B}(\mathcal{H}_{j_i}) \text{ for some } j_i, 1 \leq i \leq m\}$ such that $\text{span}\{\mathcal{A}_i : 1 \leq i \leq m\} = \mathcal{B}(\mathcal{H})$. Then a measurement scheme \mathcal{M} on the composite system can be constructed as the union of informational complete schemes $\mathcal{M}_{\mathcal{A}_i}$ on the subalgebras \mathcal{A}_i where the measurements in $\mathcal{M}_{\mathcal{A}_i}$ are performed as described above with some W_i such that $W_i\mathcal{A}_iW_i^* \equiv \mathcal{B}(\mathcal{H}_{j_i}) \otimes I$. The measurement scheme \mathcal{M} corresponds to a collection of measurement schemes on the physical subsystems $\mathcal{B}(\mathcal{H}_{j_i})$ and unitary time evolutions $W_i = e^{iH_it_i/\hbar}$. In case there exists a complete set of complementary subalgebras $\mathcal{A}_i \simeq \mathcal{B}(\mathcal{H}_{j_i})$ for some j_i , we can apply the results of the previous sections to the families

$$\mathcal{U}(\mathcal{M}_0) = \left\{ \mathcal{M} = \cup_{i=1}^m U_i \mathcal{M}_{\mathcal{A}_i} U_i^* : \mathcal{M} \text{ informational complete, } U^{(i)} \in U_n(\mathbb{C}) \right\}$$

for some $\mathcal{M}_0 = \cup_{i=1}^m \mathcal{M}_{\mathcal{A}_i}$ with informationally complete $\mathcal{M}_{\mathcal{A}_i}$ on \mathcal{A}_i . For any particular form of the $\mathcal{M}_{\mathcal{A}_i}$ it is optimal if we choose the measurement scheme \mathcal{M}_0 associated with the complementary subalgebras \mathcal{A}_i .

Example 3.2.4 In the case of k copies of a r level system with Hilbert space \mathcal{H}_A , $\dim(\mathcal{H}_A) = r$, the algebra of the composite system is given by

$$\mathcal{B}(\mathcal{H}) \simeq M_{r^k}(\mathbb{C}) \simeq \bigotimes_{i=1}^k M_r(\mathbb{C})$$

In the case when $r = p^l$ is the power of prime $p \geq 3$ there exists a complete set of $m = (r^{2k} - 1)/(r^2 - 1)$ complementary $\mathcal{A}_i \simeq M_r(\mathbb{C})$. We can construct informationally complete $\mathcal{M}_{\mathcal{A}_i}$ on \mathcal{A}_i from a single informationally complete measurement scheme $\mathcal{M}_{M_r(\mathbb{C})}$ on $M_r(\mathbb{C})$ and unitaries W_i such that $W_i \mathcal{A}_i W_i^* \equiv M_r(\mathbb{C}) \otimes I$. In other words in the estimation scheme we perform the same measurements from $\mathcal{M}_{M_r(\mathbb{C})}$ on one of the subsystem after exposing the composite system to different time evolutions W_i . Within the families

$$\mathcal{U} = \left\{ \mathcal{M} = \bigcup_{i=1}^m W_i \mathcal{M}_{M_r(\mathbb{C})} W_i^* : \mathcal{M} \text{ informational complete, } W_i \in U_n(\mathbb{C}) \right\} \quad (3.38)$$

The choice of the W_i is optimal in the sense of Chapter 3 if the subalgebras $\mathcal{A}_i = W_i^{-1}(M_r(\mathbb{C}) \otimes I)(W_i^{-1})^*$ are complementary. \diamond

Example 3.2.5 In the case of two qubits, i.e. $r=2$ and $k=2$ in the previous example, it was shown in [23] that the upper bound of 5 quasi-orthogonal subalgebras $\mathcal{A} \simeq M_2(\mathbb{C})$ can not be achieved. However it is possible to choose four quasi-orthogonal subalgebras $\mathcal{A}_i \simeq M_2(\mathbb{C})$ ($1 \leq i \leq 4$) and the remaining orthogonal complement in \mathcal{S} together with the identity forms a maximal Abelian subalgebra \mathcal{A}_5 of $M_2(\mathbb{C}) \otimes M_2(\mathbb{C})$ [18]. Thus we can construct a measurement scheme on $\mathcal{B}(\mathcal{H})$ from an informationally complete measurement scheme $\mathcal{M}_{M_2(\mathbb{C})}$ on the subsystem $M_2(\mathbb{C}) \otimes I$ together with unitaries W_i such that $W_i \mathcal{A}_i W_i^* \equiv M_2(\mathbb{C}) \otimes I$ for $1 \leq i \leq 4$ and the measurement M_A of a non-degenerate observables on $M_2(\mathbb{C}) \otimes M_2(\mathbb{C})$ related to \mathcal{A}_5 . This measurement scheme is optimal within

the family

$$\mathcal{U} = \left\{ \mathcal{M} = \cup_{i=1}^4 W_i \mathcal{M}_{M_2} W_i^* \cup W_5 \{M_A\} W_5^* : \text{informational complete, } W_i \in U_4(\mathbb{C}) \right\} \quad (3.39)$$

◇

As a last example we specify additionally the measurement schemes on the subalgebras:

Example 3.2.6 In the context of the previous two examples, we can further specify the form of the measurement schemes $\mathcal{M}_{M_r(\mathbb{C})}$ on the subsystem $\mathcal{B}(\mathcal{H}_A)$. Since r was assumed to be a prime power, there exists a complete set $\{\mathcal{A}_i, 1 \leq i \leq r+1\}$ of complementary maximal Abelian subalgebras of the subsystem $\mathcal{B}(\mathcal{H}_A)$, or equivalently, a complete set of $(r+1)$ complementary observables on the subsystem \mathcal{H}_A . Together with optimal choice of the W_i (where $1 \leq i \leq \frac{r^{2k}-1}{r^2-1}$) that obtained complementary subalgebras of $\mathcal{B}(\mathcal{H})$ isomorphic to $\mathcal{B}(\mathcal{H}_A)$ we get a complete set of complementary subalgebras $\mathcal{A}_{ik} = W_i(\mathcal{A}_k \otimes I)W_i^*$. Note that this are not maximal Abelian subalgebras of $\mathcal{B}(\mathcal{H})$. The \mathcal{A}_{ik} are generated by the isomorphic image of minimal projections in the algebras $M_r(\mathbb{C}) \otimes I$. Since I denotes here the r^{k-1} dimensional identity, this projections are not minimal projections in $M_{r^k}(\mathbb{C})$ but they are of rank $r^{(k-1)}$. Thus the subalgebras \mathcal{A}_{ik} are homogenous Abelian subalgebras and the correspond to von Neumann measurements $M_{\mathcal{A}_{ik}}$ of r rank $r^{(k-1)}$ projections. We obtain a measurement scheme $\mathcal{M} = \{M_{\mathcal{A}_{ik}} : 1 \leq i \leq m\}$ with $m = (r+1) \frac{r^{2k}-1}{r^2-1} = \frac{r^{2k}-1}{r-1}$ measurements. Let us remark, that alternatively to the procedure of measuring non-degenerate observables on the subsystem after exposing the system to some time evolution, the measurement $M_{\mathcal{A}_{ik}}$ can be performed directly on the composite system $\mathcal{B}(\mathcal{H})$ as well.

◇

In all the examples above we can use the relative frequencies ν as a classical estimate

for the probabilities and we estimate the state by (2.34)

$$\hat{\boldsymbol{\theta}} = \frac{1}{n} T^{-1} (\boldsymbol{\nu} - \frac{1}{n} \mathbf{r}) \quad (3.40)$$

The mean quadratic error matrix is given by

$$V_{\boldsymbol{\theta}} = \frac{1}{n^2} T^{-1} W (T^{-1})^* \quad (3.41)$$

where W is block-diagonal with blocks of the form

$$W^{(k)} = \frac{1}{N^{(k)}} \begin{pmatrix} p_1^{(k)} - p_1^{(k)2} & \cdots & p_1^{(k)} p_{\frac{n}{r}-1}^{(k)} \\ \vdots & \ddots & \vdots \\ p_{\frac{n}{r}-1}^{(k)} p_1^{(k)} & \cdots & p_{\frac{n}{r}-1}^{(k)} - p_{\frac{n}{r}-1}^{(k)2} \end{pmatrix} \quad (3.42)$$

3.3 Evaluation of von Neumann Measurements

In this section we return to measurement schemes that use von Neumann measurements associated with homogenous Abelian subalgebras of $\mathcal{B}(\mathcal{H})$. Such measurements correspond to the measurement of observables with eigenvalues of multiplicity r and the measurement operators are projections of rank r . As discussed in Chapter 3.2.4 a measurement scheme consists of $\frac{r(n^2-1)}{(n-r)}$ measurements of such observables. From the results of Chapter 3.2 for a fixed r it is optimal to choose this observables complementary. In this section we evaluate the average mean quadratic error matrix and its determinant for the complementary case. In the calculations we choose a specific prior distribution for which the boundary of the set of states has measure zero. The result allows us to show that measurement schemes using observables with non-degenerate spectrum have maximal efficiency.

To evaluate the average we first specify the measure μ we will use on the set \mathcal{T} of states. Therefore note that we can generate the set \mathcal{T} from diagonal matrices by unitary

conjugation. Let us denote $\Omega = \{\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_n) : \lambda_i \geq 0, \sum \lambda_i = 1\}$ as the set of all diagonal matrices with an n-dimensional probability vector on the diagonal. On the Cartesian product $U_{\mathbb{C}}(n) \times \Omega$ we can define a function

$$g : U_{\mathbb{C}}(n) \times \Omega \rightarrow \mathcal{T} : \quad g(U, \Lambda) = U\Lambda U^* = U\Lambda U^* = \rho \quad (3.43)$$

$U_{\mathbb{C}}(n) \times \Omega$ can be equipped with the normalized product measure of the Haar measure on $U_{\mathbb{C}}(n)$ and the Lebesgue measure on Ω . The push-forward measure μ by g , denoted as μg^{-1} , is a unitarily invariant measure on \mathcal{T} :

$$\begin{aligned} \int_{\mathcal{T}} f(\rho) d\mu g^{-1}(\rho) &= \int_{U_{\mathbb{C}}(n) \times \Omega} f(g(U, \Lambda)) d\mu(U) \times d\mu(\Lambda) \\ &\stackrel{*}{=} \int_{U_{\mathbb{C}}(n) \times \Omega} f(g(VU, \Lambda)) d\mu(U) \times d\mu(\Lambda) = \int_{\mathcal{T}} f(V\rho V^*) d\mu g^{-1}(\rho) \end{aligned} \quad (3.44)$$

where the equality (*) represents the invariance of the Haar measure and holds for all $V \in U_{\mathbb{C}}(n)$. Let us remark that this measure is different from the normalized Lebesgue measure on $\mathcal{T} \subset \mathbb{R}^{(n^2-1)}$ which requires a different measure on Ω .

In order to evaluate the integral of (3.42) we first calculate the explicit form of the probabilities

$$p_i^{(k)} = \text{Tr}(\rho P_i^{(k)}) = \text{Tr}(U\Lambda U^* P_i^{(k)}) = \text{Tr}(\Lambda U^* P_i^{(k)} U) \quad (3.45)$$

related to the projection $P_i^{(k)}$ (we use the notation of Examples 3.2.1 and 3.2.2). By the invariance condition (3.44) of the Haar measure we can assume that the $P_i^{(k)}$ are diagonal in the same basis as Λ and of block diagonal form

$$P_i^{(k)} = \begin{bmatrix} \boxed{0} & & & & \\ & \dots & & & \\ & & \boxed{I_r} & & 0 \\ & & & \dots & \\ 0 & & & & \boxed{0} \end{bmatrix} \quad (3.46)$$

where each block is of size $r \times r$ such that the i th block contains the r -dimensional identity and all other blocks are zero. The matrix elements of the product $U^* P_i^{(k)} U$ are given by

$$(U^* P_i^{(k)} U)_{pq} = \sum_{s,t=1}^n \bar{U}_{sp} (P_i^{(k)})_{st} U_{tq} \quad (3.47)$$

In the basis where Λ is diagonal left multiplication of a matrix by Λ multiplies the p th row with λ_p . Evaluating the trace gives

$$\begin{aligned} \text{Tr} (\Lambda U^* P_i^{(k)} U) &= \text{Tr} \left(\left[\sum_{s,t=1}^n \lambda_p \bar{U}_{sp} (P_i^{(k)})_{st} U_{tq} \right]_{p,q=1}^n \right) \\ &= \sum_{p,s,t}^n \lambda_p \bar{U}_{sp} U_{tp} (P_i^{(k)})_{st} = \sum_{p,s}^n \lambda_p |U_{sp}|^2 (P_i^{(k)})_{ss} \end{aligned} \quad (3.48)$$

where in the last step we used the diagonal form (3.46) of the $P_i^{(k)}$. Especially in the last sum, $(P_i^{(k)})_{ss} \neq 0$ only for $(i-1)r + 1 \leq s \leq ir$. Since changing the index i is the same than permuting the diagonal blocks of $P_i^{(k)}$ in (3.46) by some unitary conjugation, it is enough to consider $i = 1, 2$ and we get

$$p_i^{(k)} = p_1^{(k)} = \sum_{i=1}^n \sum_{k=1}^r \lambda_i |U_{ki}|^2 \quad (3.49)$$

$$p_i^{(k)} p_j^{(k)} = p_1^{(k)} p_2^{(k)} = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^r \sum_{l=r+1}^{2r} \lambda_i \lambda_j |U_{ki}|^2 |U_{lj}|^2 \quad (i \neq j) \quad (3.50)$$

All diagonal elements of the blocks $\langle W^{(k)} \rangle$ will be identical after taking the average as well

as all off diagonal elements and we get

$$\langle W^{(k)} \rangle = \frac{1}{N^{(k)}} \begin{pmatrix} a & b & \cdots & b \\ b & a & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \cdots & b & a \end{pmatrix} \quad (3.51)$$

Furthermore all the blocks of $\langle W \rangle$ will be the same since also all measurements we use are unitarily conjugated to each other. Thus it is enough if we calculate the averages $a = \langle p_1^{(k)} - p_1^{(k)2} \rangle$ and $b = \langle -p_1^{(k)} p_2^{(k)} \rangle$. We get

$$\begin{aligned} a &= \int_{\mathcal{T}} p_1^{(k)} - p_1^{(k)2} d\mu g^{-1}(\rho) = \\ &= \sum_{i=1}^n \sum_{k=1}^r \int_{\Omega} \lambda_i d\mu(\Lambda) \int_{U_{\mathbb{C}(n)}} |U_{ki}|^2 d\mu(U) - \sum_{i,j=1}^n \sum_{k,l=1}^r \int_{\Omega} \lambda_i \lambda_j d\mu(\Lambda) \int_{U_{\mathbb{C}(n)}} |U_{ki}|^2 |U_{lj}|^2 d\mu(U) \end{aligned}$$

for the diagonal elements and

$$b = - \int_{\mathcal{T}} p_1^{(k)} p_2^{(k)} d\mu g^{-1}(\rho) = - \sum_{i,j=1}^n \sum_{k=1}^r \sum_{l=r+1}^{2r} \int_{\Omega} \lambda_i \lambda_j d\mu(\Lambda) \int_{U_{\mathbb{C}(n)}} |U_{ki}|^2 |U_{lj}|^2 d\mu(U)$$

for off diagonal elements. The appearing integrals and sums are evaluated in appendix A and we get the result

$$a = \int_{\mathcal{T}} p_1 - p_1^2 d\mu g^{-1}(\rho) = r \frac{(n+2)}{(n+1)^2} - r^2 \frac{(n+2)}{n(n+1)^2} \quad (3.52)$$

$$b = - \int_{\mathcal{T}} p_1 p_2 d\mu g^{-1}(\rho) = -r^2 \frac{(n+2)}{n(n+1)^2} \quad (3.53)$$

To evaluate the determinant of $\langle V \rangle$ we need to calculate the determinant of $\langle W \rangle$ and T .

We will use the following formula for our calculations: The determinant of a $k \times k$ matrix B with elements $B_{ij} = \delta_{ij}(a - b) + b$ is obtained by

$$\text{Det}(B) = \begin{vmatrix} a & b & \cdots & b \\ b & a & \cdots & b \\ \vdots & & \ddots & \vdots \\ b & b & \cdots & a \end{vmatrix} = (a - b)^{(k-1)} (a + (k - 1) b) \quad (3.54)$$

This formula can be derived by carrying B into upper triangular form by first subtracting the last line each other ones and then canceling the off-diagonal elements in the last line:

$$\text{Det}(B) = \text{Det} \begin{bmatrix} a-b & 0 & \cdots & \cdots & b-a \\ 0 & a-b & 0 & \cdots & b-a \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & b-a \\ b & b & \cdots & b & a \end{bmatrix} = \text{Det} \begin{bmatrix} a-b & 0 & \cdots & \cdots & b-a \\ 0 & a-b & 0 & \cdots & b-a \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & b-a \\ 0 & 0 & \cdots & 0 & a+(n-1)b \end{bmatrix} \quad (3.55)$$

Since the matrix on the right hand side is upper triangular, the determinant is the product of the diagonal entries and we obtain (3.54).

In our case additionally the elements are the form $a = a_0 + b$ and (3.54) simplifies to $\text{Det}(A) = a_0^{k-1}(a_0 + kb)$. Together with (3.52) and (3.53) we obtain

$$\text{Det}(\langle W \rangle) = \left[\prod_{k=1}^m \frac{1}{N^{(k)}} \right]^{(n/r-1)} \left[\frac{(n+2)}{(n+1)^2} \right]^{(n/r-1)m} n^{-m} r^{n/r m} \quad (3.56)$$

To calculate the determinant of T we recall the geometrical configuration of the row vectors of T . The inner product between two such vectors is given by

$$\langle \mathbf{u}_i^{(k)}, \mathbf{u}_j^{(l)} \rangle = \frac{1}{n} \text{Tr}(P_i^{(k)} P_j^{(l)}) - \frac{r^2}{n^2} \quad (3.57)$$

Thus the vectors have length $|\mathbf{u}_i^{(k)}|^2 = \frac{r}{n} (1 - \frac{r}{n})$ and the inner product between any two

of them is $\langle \mathbf{u}_i^{(k)}, \mathbf{u}_j^{(k)} \rangle = -\frac{r^2}{n^2}$. The product TT^* contains the inner products between the vectors $\mathbf{u}_i^{(k)}$ as elements. For the optimal measurement we can assume the subspaces $\mathcal{S}^{(k)}$ to be orthogonal and T becomes block-diagonal with

$$(T^{(k)}(T^{(k)})^*)_{ij} = \langle \mathbf{u}_i^{(k)}, \mathbf{u}_j^{(k)} \rangle = \delta_{ij} \frac{r}{n} - \frac{r^2}{n^2} \quad (3.58)$$

The blocks are of size $(n/(r-1)) \times (n/(r-1))$ and again we can use formula (3.54) to evaluate the determinant

$$\begin{aligned} \text{Det}(T^{(k)}(T^{(k)})^*) &= \left(\frac{r}{n}\right)^{(n/r-2)} \left(\frac{r}{n} - \left(\frac{n}{r} - 1\right) \frac{r^2}{n^2}\right) \\ &= \left(\frac{r}{n}\right)^{(n/r-2)} \left(\frac{nr - (n-r)r}{n^2}\right) = \left(\frac{r}{n}\right)^{n/r} \end{aligned} \quad (3.59)$$

Since the values (3.58) of the inner products are identical for all k , all blocks have the same determinant and their product becomes

$$\text{Det}(T)^2 = \text{Det}(TT^*) = n^{-n/r} m r^{n/r} m. \quad (3.60)$$

Putting (3.56) and (3.60) together we get

$$\text{Det}(\langle V_{\theta} \rangle) = \frac{1}{n^{2(n^2-1)}} \frac{\text{Det}(A)}{\text{Det}(T)^2} = \frac{1}{n^{2(n^2-1)}} \left[\prod_{k=1}^m \frac{1}{N^{(k)}} \right]^{(n/r-1)} \left[\frac{n(n+2)}{(n+1)^2} \right]^{(n^2-1)} \quad (3.61)$$

where $\sum_k N^{(k)} = N$. For fixed r this expression becomes minimal if we choose all $N^{(k)} = \frac{N}{m}$.

Then

$$\text{Det}(\langle V_{\theta} \rangle) = \left[\frac{m}{N} \right]^{(n^2-1)} \left[\frac{(n+2)}{n(n+1)^2} \right]^{(n^2-1)} \quad (3.62)$$

which is minimal if we choose $r = 1$. The case $r = 1$ corresponds to the measurement of observables with non-degenerate spectrum and measurement schemes consisting of such

observables are optimal in the above setting.

Conclusion

In the thesis we considered the problem of estimating an unknown quantum state by separate measurements on identical copies of the state. The following setting was considered for the state estimation problem: It was assumed that the unknown state originates from the set \mathcal{T} of all possible states of the system and a priori knowledge is reflected by a prior probability measure on \mathcal{T} . For the estimation of the state N identical copies of the state are at hand and the following estimation scheme is carried out: To obtain statistical data a set of measurements (or a measurement scheme) $\mathcal{M} = \{M^{(1)}, M^{(2)}, \dots, M^{(m)}\}$ is chosen. The ensemble of the given states is divided into m subensembles and the measurements $M^{(k)}$ are performed separately on the copies of the subensembles. To form a point estimate for the unknown state the set \mathcal{T} is parameterized by the generalized Bloch vector $\boldsymbol{\theta}$. An unconstrained estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained from the empirical distribution (or relative frequencies) of the measurement outcomes by linear inversion of the relation between the measurement probabilities and the Bloch vector. Since the unconstrained estimate may take values outside the set of states, additionally an estimate $\hat{\boldsymbol{\theta}}_c$ constrained onto the set \mathcal{T} is considered. The efficiency of an estimation scheme is evaluated by the average of the mean quadratic error matrix, where the average is taken with respect to the prior distribution on the true states. To compare different measurement schemes the determinant of this matrix is used.

The thesis studied the role of complementary measurements in the above state estimation problem. The main result showed that if a measurement scheme $\mathcal{M}_0 = \{M_0^{(1)}, M_0^{(2)}, \dots, M_0^{(m)}\}$ consists of complementary measurements, it performs better than any measurement scheme of the form $\mathcal{M}' = \{U_1 M^{(1)} U_1^*, U_2 M^{(2)} U_2^*, \dots, U_m M^{(m)} U_m^*\}$ obtained with unitaries $U_k \in U_{\mathbb{C}}(n)$ ($1 \leq k \leq m$). In other words, \mathcal{M}_0 is optimal within the family $\mathcal{U}(\mathcal{M}_0)$ formed by the \mathcal{M}' . This was stated in

- Theorem 1 for the unconstrained estimate: The necessary condition for the theorem is that the prior probability measure is invariant under unitary conjugation: $\mu(\rho_\theta) = \mu(U \rho_\theta U^*)$ ($U \in U_{\mathbb{C}}(n)$). The theorem holds for finite numbers of available copies of the unknown state.
- Theorem 2 for the constrained estimate: The main conditions for the theorem are that the prior probability measure is unitarily invariant and vanishes on the boundary of the set of states: $\mu(\partial\mathcal{T}) = 0$. Furthermore it is assumed that the constrained estimate differs from the unconstrained only if the unconstrained estimate falls outside the set \mathcal{T} . The theorem is shown in the asymptotic case when the number of available states tends to infinity.

Additionally in Theorem 3 the relation of the results to the work in [28], where a similar optimality result for complementary observables with non-degenerate spectrum with respect to the information gain was shown, is discussed.

By means of examples where the existence of complete sets of complementary subalgebras is known, the relation of the families $\mathcal{U}(\mathcal{M}_0)$ to subalgebras is discussed: A subsystem of a quantum system is described by a subalgebra $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ and in general a measurement $M^{(k)}$ is contained in a subsystem. If we are given two subsystem \mathcal{A} and $U\mathcal{A}U^*$ related by unitary conjugation, a measurement $M^{(k)} \subset \mathcal{A}$ performed on the subsystem \mathcal{A} given the true state ρ is equivalent to performing $UM^{(k)}U^* \subset U\mathcal{A}U^*$ on the subsystem $U\mathcal{A}U^*$ given

the true state $U\rho U^*$. By the results shown in the thesis, if the true states are unitarily invariant distributed, it is on average optimal to choose the subsystems complementary: Measurements $M^{(k)}$ performed on complementary subsystems $\mathcal{A}^{(k)}$ perform better than equivalent measurements $UM^{(k)}U^*$ on the subsystems $U\mathcal{A}^{(k)}U^*$.

Finally an example of a unitary invariant prior distribution on \mathcal{T} is constructed from the Haar measure on $U_{\mathbb{C}}(n)$ and the Lebesgue measure on the probability simplex. For this prior distribution the determinant of the average mean quadratic error matrix was evaluated explicitly for measurement schemes using von Neumann measurements associated with homogenous Abelian subalgebras generated by projections of the same rank r . It was shown that the case of maximal Abelian subalgebras, which is related to measurements of observables with non-degenerate spectrum, is optimal.

Appendix A

Some Integrals

A.1 Integration on the Probability Simplex

The set of discrete probability distributions on a set of n different outcomes can be parameterized as the $n - 1$ dimensional simplex

$$\Omega_{n-1} = \{\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_n) : \lambda_i \geq 0, \sum \lambda_i = 1\} \subset \mathbb{R}^n. \quad (\text{A.1})$$

which is a $n - 1$ dimensional manifold in \mathbb{R}^n . We change to the cumulative probabilities by the coordinate transformation

$$A : \quad x_i = \sum_{k=1}^i \lambda_k \quad \text{and} \quad A^{-1} : \quad \lambda_i = x_i - x_{i-1} \quad (\text{A.2})$$

Since we have $\lambda_n = 1$ we get a chart $\phi : \mathbb{R}^{n-1} \supset U \rightarrow \mathbb{R}^n$ for Ω_{n-1} from the set $U = \{\mathbf{x} : 0 \leq x_1 \leq x_2 \leq \dots \leq x_{n-1} \leq 1\} \subset \mathbb{R}^{n-1}$ to \mathbb{R}^n by .

$$\phi(\mathbf{x}) = A^{-1} \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} \quad (\text{A.3})$$

Integration of a function $f : \Omega \rightarrow \mathbb{R}$ is given by

$$\int_{\Omega} f d\Omega = \int_U f(\phi(\mathbf{x})) \sqrt{\text{Det } G_n(\mathbf{x})} dx_1 \dots dx_n \quad (\text{A.4})$$

where $G_n = (D\phi)^t D\phi$ is the measure tensor of ϕ and $\text{Det } G_n$ is the Gram's determinant.

The derivative of ϕ and the measure tensor are given by

$$D\phi = \left. \begin{pmatrix} \overbrace{\begin{matrix} 1 & 0 & \dots & 0 \\ -1 & 1 & 0 & \vdots \\ 0 & -1 & 1 & \ddots \\ \vdots & \ddots & \ddots & \ddots \\ \vdots & & 0 & -1 & 1 \\ 0 & \dots & \dots & 0 & -1 \end{matrix}}^{(n-1)} \\ \end{pmatrix} \right\} n, \quad G_n = \left. \begin{pmatrix} \overbrace{\begin{matrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 2 & -1 \\ 0 & \dots & \dots & -1 & 2 \end{matrix}}^{(n-1)} \\ \end{pmatrix} \right\} (n-1)$$

By expanding the determinant along the first row we can get the recursive formula $\text{Det } G_n = 2 \text{Det } G_{(n-1)} - \text{Det } G_{(n-2)}$. By induction on n it is easy to see from this formula that $\text{Det } G = n$.

We will use the following integral formula

$$\int_0^a \int_0^{x_k} \dots \int_0^{x_3} \int_0^{x_2} dx_1 dx_2 \dots dx_{k-1} dx_k = \frac{1}{k!} a^k. \quad (\text{A.5})$$

With (A.5) the hyperarea of Ω_n becomes

$$\int_{\Omega} d\Omega = \int_U \sqrt{n} dx_1 \dots dx_{n-1} \quad (\text{A.6})$$

$$= \int_0^1 \int_0^{x_{n-1}} \dots \int_0^{x_3} \int_0^{x_2} \sqrt{n} dx_1 dx_2 \dots dx_{n-2} dx_{n-1} = \frac{\sqrt{n}}{(n-1)!}. \quad (\text{A.7})$$

Integrating λ_n on Ω

$$\int_{\Omega} \lambda_n d\Omega = \int_{\mathcal{U}} (1 - x_{n-1}) \sqrt{n} dx_1 \dots dx_{n-1} \quad (\text{A.8})$$

$$= \int_0^1 (1 - x_{n-1}) \underbrace{\int_0^{x_{n-1}} \dots \int_0^{x_3} \int_0^{x_2} \sqrt{n} dx_1 dx_2 \dots dx_{n-2}}_{n-2 \text{ terms}} dx_{n-1} \quad (\text{A.9})$$

$$= \int_0^1 (1 - x_{n-1}) x_{n-1}^{(n-2)} \frac{\sqrt{n}}{(n-2)!} dx_{n-1} \quad (\text{A.10})$$

$$= \int_0^1 x_{n-1}^{(n-1)} \frac{\sqrt{n}}{(n-1)!} dx_{n-1} = \frac{\sqrt{n}}{(n)!} \quad (\text{A.11})$$

where the last line follows from integration by parts. Similar, integration of λ_n^2 on Ω

$$\int_{\Omega} \lambda_n^2 d\Omega = \int_{\mathcal{U}} (1 - x_{n-1})^2 \sqrt{n} dx_1 \dots dx_{n-1} \quad (\text{A.12})$$

$$= \int_0^1 (1 - x_{n-1})^2 x_{n-1}^{(n-2)} \frac{\sqrt{n}}{(n-2)!} dx_{n-1} \quad (\text{A.13})$$

$$= \int_0^1 2x_{n-1}^n \frac{\sqrt{n}}{n!} dx_{n-1} = \frac{\sqrt{n}}{(n+1)!} \quad (\text{A.14})$$

Integration of mixed terms $\lambda_n \lambda_{n-1}$ on Ω becomes

$$\begin{aligned} \int_{\Omega} \lambda_n \lambda_{n-1} d\Omega &= \int_{\mathcal{U}} (1 - x_{n-1})(x_{n-1} - x_{n-2}) \sqrt{n} dx_1 \dots dx_{n-1} \\ &= \int_0^1 \int_0^{x_{n-1}} (1 - x_{n-1})(x_{n-1} - x_{n-2}) \underbrace{\int_0^{x_{n-2}} \dots \int_0^{x_2} \sqrt{n} dx_1 \dots dx_{n-3}}_{n-3 \text{ terms}} dx_{n-2} dx_{n-1} \\ &= \int_0^1 \int_0^{x_{n-1}} (1 - x_{n-1})(x_{n-1} - x_{n-2}) x_{n-2}^{(n-3)} \frac{\sqrt{n}}{(n-3)!} dx_{n-1} \\ &= \frac{\sqrt{n}}{(n+1)!} \end{aligned}$$

where the last step follows after a lengthy calculation with integrals of the above type.

Normalizing the measure on Ω and using the symmetry of the indices finally results in the

integrals

$$\int_{\Omega} \lambda_i \, d\Omega = \frac{1}{n} \quad (\text{A.15})$$

$$\int_{\Omega} \lambda_i^2 \, d\Omega = \frac{2}{n(n+1)} \quad (\text{A.16})$$

$$\int_{\Omega} \lambda_i \lambda_j \, d\Omega = \frac{1}{n(n+1)} \quad i \neq j. \quad (\text{A.17})$$

A.2 Integration on the Unitary Group

Consider some $U = (U_{ij})_{i,j=1}^n$ in the complex unitary group $U_{\mathbb{C}}(n)$, and denote by μ the normalized Haar measure on $U_{\mathbb{C}}(n)$, then:

$$\int_{U_{\mathbb{C}}(n)} |U_{ki}|^2 \, d\mu(U) = \frac{1}{n} \quad (\text{A.18})$$

$$\int_{U_{\mathbb{C}}(n)} |U_{ki}|^4 \, d\mu(U) = \frac{2}{n(n+1)} \quad (\text{A.19})$$

$$\int_{U_{\mathbb{C}}(n)} |U_{ki}|^2 |U_{li}|^2 \, d\mu(U) = \frac{1}{n(n+1)} \quad k \neq l \quad (\text{A.20})$$

$$\int_{U_{\mathbb{C}}(n)} |U_{ki}|^2 |U_{lj}|^2 \, d\mu(U) = \frac{1}{(n-1)(n+1)} \quad i \neq j, k \neq l \quad (\text{A.21})$$

This moments can be calculated without explicit formulation of the Haar measure and we will replicate the proof given in [8]:

The first step in the proof is to notice that the elements of a matrix can permuted by some unitaries V and W such that $U_{\pi(i)\sigma(j)} = (VUW)_{ij}$ where π and σ are permutations. Then for a measurable function f the invariance condition of the Haar measure gives

$$\int f(U_{\pi(i)\sigma(j)}, U_{\pi(k)\sigma(l)}) \, d\mu(U) = \int f(U_{ij}, U_{kl}) \, d\mu(U) \quad (\text{A.22})$$

and it is mostly enough to consider the indices $i, j, k, l = 1, 2$.

The first equation (A.18) follows simply from $\sum_{i=1}^n |U_{ij}|^2 = 1$ and the equality of the integrals $\int |U_{ij}|^2 d\mu(U)$ for all $1 \leq i, j \leq n$ as mentioned in (A.22).

For the next step we superpose the elements U_{11} and U_{12} with help of the unitary matrix

$$V = \left(\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \oplus I_{n-2} \right) \quad (\text{A.23})$$

Using again the invariance condition of the Haar measure we get the identity

$$\begin{aligned} \int |U_{11}|^4 d\mu(U) &= \int |(VU)_{11}|^4 d\mu(U) = \int |U_{11} \cos \theta + U_{21} \sin \theta|^4 d\mu(U) \\ &= \int \left(|U_{11}|^2 \cos^2 \theta + |U_{21}|^2 \sin^2 \theta + (U_{11}U_{21}^* + U_{21}U_{11}^*) \cos \theta \sin \theta \right)^2 d\mu(U) \\ &= \int |U_{11}|^4 \cos^4 \theta + |U_{21}|^4 \sin^4 \theta + 4|U_{11}|^2 |U_{21}|^2 \cos^2 \theta \sin^2 \theta \\ &\quad + (U_{11}U_{21}^* + U_{11}^*U_{21}) (|U_{11}|^2 \cos^3 \theta \sin \theta + |U_{21}|^2 \sin^3 \theta \cos \theta) \\ &\quad + ((U_{11}U_{21}^*)^2 + (U_{11}^*U_{21})^2) \sin^2 \theta \cos^2 \theta d\mu(U) \end{aligned} \quad (\text{A.24})$$

By multiplying U with the unitary matrix $\text{diag}(e^{i\phi}, e^{2i\phi}, 1, \dots, 1)$ we get for terms that contain products of the form $U_{11}U_{21}^*$ or their complex conjugate the identity $\int U_{11}U_{21}^* \dots d\mu(U) = e^{-i\phi} \int U_{11}U_{21}^* \dots d\mu(U)$ for all ϕ . Therefore the terms in the last two lines of (A.24) vanish.

Setting additionally $\theta = \pi/4$ together with $\int |U_{11}|^4 d\mu(U) = \int |U_{12}|^4 d\mu(U)$ gives

$$\int |U_{11}|^4 d\mu(U) = 2 \int |U_{11}|^2 |U_{21}|^2 d\mu(U) = 2 \int |U_{ij}|^2 |U_{kj}|^2 d\mu(U) \quad (\text{A.25})$$

Using the normality condition on the rows of a unitary we get $\sum_{i,k} |U_{ij}|^2 |U_{kj}|^2 = 1$. Split-

ting up the sum gives together with (A.25)

$$\begin{aligned}
\int d\mu(U) &= \int \left(\sum_{i=k} |U_{ij}|^4 + \sum_{i \neq k} |U_{ij}|^2 |U_{kj}|^2 \right) d\mu(U) \\
&= n \int |U_{11}|^4 d\mu(U) + \frac{n(n-1)}{2} \int |U_{11}|^4 d\mu(U) \\
&= \frac{n(n+1)}{2} \int |U_{11}|^4 d\mu(U)
\end{aligned} \tag{A.26}$$

which proofs (A.19) and (A.20). Finally equation (A.21) follows in a similar way from $\sum_{i,k} |U_{ij}|^2 |U_{kl}|^2 = 1$ using (A.20).

A.3 Summation

To obtain (3.52) and (3.53) we need to calculate the sums in (3.49) and (3.50):

$$p_1^{(k)} = \sum_{i=1}^n \sum_{k=1}^r \int_{\Omega} \lambda_i d\mu(\Lambda) \int_{U_{\mathbb{C}(n)}} |U_{ki}|^2 d\mu(U) = n r \frac{1}{n} \frac{1}{n} = \frac{r}{n} \tag{A.27}$$

The square of the probability

$$\begin{aligned}
p_1^{(k)^2} &= \sum_{i=1}^n \sum_{k=1}^r \int_{\Omega} \lambda_i d\mu(\Lambda) \int_{U_{\mathbb{C}(n)}} |U_{ki}|^2 d\mu(U) \sum_{j=1}^n \sum_{l=1}^r \int_{\Omega} \lambda_j d\mu(\Lambda) \int_{U_{\mathbb{C}(n)}} |U_{lj}|^2 d\mu(U) \\
&= \sum_{i,j=1}^n \int_{\Omega} \lambda_i \lambda_j d\mu(\Lambda) \sum_{k,l=1}^r \int_{U_{\mathbb{C}(n)}} |U_{ki}|^2 |U_{lj}|^2 d\mu(U) \\
&= \sum_{i=j}^n \int_{\Omega} \lambda_i^2 d\mu_{\Lambda} \left(\sum_{k=l}^r \int_{U_{\mathbb{C}(n)}} |U_{ki}|^2 |U_{ki}|^2 d\mu_U + \sum_{k \neq l}^r \int_{U_{\mathbb{C}(n)}} |U_{ki}|^2 |U_{li}|^2 d\mu_U \right) \\
&\quad + \sum_{i \neq j}^n \int_{\Omega} \lambda_i \lambda_j d\mu_{\Lambda} \left(\sum_{k=l}^r \int_{U_{\mathbb{C}(n)}} |U_{ki}|^2 |U_{kj}|^2 d\mu_U + \sum_{k \neq l}^r \int_{U_{\mathbb{C}(n)}} |U_{ki}|^2 |U_{lj}|^2 d\mu_U \right)
\end{aligned} \tag{A.28}$$

For pairs (i, j) with $1 \leq i, j \leq k$ there are k pairs where $i = j$ and $n(n-1)$ pairs where $i \neq j$, we get

$$\begin{aligned}
p_1^{(k)^2} &= n \frac{2}{n(n+1)} \left(r \frac{2}{n(n+1)} + r(r-1) \frac{1}{n(n+1)} \right) \\
&\quad + n(n-1) \frac{1}{n(n+1)} \left(r \frac{1}{n(n+1)} + r(r-1) \frac{1}{(n-1)(n+1)} \right) \\
&= \frac{2}{n(n+1)^2} (r+r^2) + \frac{1}{n(n+1)^2} (nr^2-r) = \frac{r}{n(n+1)^2} + \frac{r^2(n+2)}{n(n+1)^2} \\
&= \frac{r}{n} \left(\frac{1+r(n+2)}{(n+1)^2} \right)
\end{aligned} \tag{A.29}$$

and (A.27) minus (A.29) gives the desired result:

$$a = \int_{\mathcal{T}} p_1 - p_1^2 d\mu g^{-1}(\rho) = r \frac{(n+2)}{(n+1)^2} - r^2 \frac{(n+2)}{n(n+1)^2} \tag{A.30}$$

For the off diagonal elements (3.53) we get:

$$\begin{aligned}
p_1^{(k)} p_2^{(k)} &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^r \sum_{l=r+1}^{2r} \int_{\Omega} \lambda_i \lambda_j d\mu(\Lambda) \int_{U_{\mathbb{C}(n)}} |U_{ki}|^2 |U_{lj}|^2 d\mu(U) \\
&= \sum_{i=j=1}^n \int_{\Omega} \lambda_i^2 d\mu(\Lambda) \sum_{k=1}^r \sum_{l=r+1}^{2r} \int_{U_{\mathbb{C}(n)}} |U_{ki}|^2 |U_{li}|^2 d\mu(U) \\
&\quad + \sum_{i \neq j}^n \int_{\Omega} \lambda_i \lambda_j d\mu(\Lambda) \sum_{k=1}^r \sum_{l=r+1}^{2r} \int_{U_{\mathbb{C}(n)}} |U_{ki}|^2 |U_{lj}|^2 d\mu(U) \\
&= n \frac{2}{n(n+1)} r^2 \frac{1}{n(n+1)} + n(n-1) \frac{1}{n(n+1)} r^2 \frac{1}{(n-1)(n+1)} \\
&= r^2 \frac{n+2}{n(n+1)^2}
\end{aligned} \tag{A.31}$$

Appendix B

Sanov's Theorem for Unequal Sample Sizes

We consider the problem of drawing samples of sizes $N^{(k)}$ from m different discrete probability distributions $\mathbf{p}^{(k)} \in \Omega_{d^{(k)}} (1 \leq k \leq m)$ and derive bounds to the probability that the relative frequency vector (or the empirical distributions) $\boldsymbol{\nu}$ of the samples fall into some set $A \subset \times_{k=1}^m \Omega_{d^{(k)}}$.

Theorem 4 *Let $X^{(k)} \sim \mathbf{p}^{(k)}$ be independent discrete random variables and the vectors \mathbf{p} and $\boldsymbol{\nu}$ defined as in (2.31). Let $A \subset \times_{k=1}^m \Omega_{d^{(k)}}$ be a set that is the closure of its interior. Then*

$$\limsup_{N_{\min} \rightarrow \infty} \frac{1}{N_{\min}} \log (\text{Prob}(\boldsymbol{\nu} \notin A)) \leq -D(\boldsymbol{\nu}^* \parallel \mathbf{p}) \leq \liminf_{N_{\min} \rightarrow \infty} \frac{1}{N_{\max}} \log (\text{Prob}(\boldsymbol{\nu} \notin A)) \quad (\text{B.1})$$

where $N_{\min} = \min_k \{N^{(k)}\}$, $N_{\max} = \max_k \{N^{(k)}\}$ and

$$D(\boldsymbol{\nu} \parallel \mathbf{p}) = \sum_k D(\boldsymbol{\nu}^{(k)} \parallel \mathbf{p}^{(k)})$$

is the relative entropy $D(\boldsymbol{\nu}^{(k)} \parallel \mathbf{p}^{(k)}) = \sum_i \nu_i^{(k)} (\log \nu_i^{(k)} - \log p_i^{(k)})$ and

$$\boldsymbol{\nu}^* = \operatorname{argmin}_{\boldsymbol{\nu} \in \mathcal{T}_\Omega} D(\boldsymbol{\nu} \parallel \mathbf{p})$$

is the $\boldsymbol{\nu}$ closest to \mathbf{p} in relative entropy.

The proof of (B.1) follows similar ideas as the proof of the usual version of Sanov's theorem (see e.g. [4]): Let us introduce the notation of types or empirical distributions as the set of probability vectors with rational components $\mathcal{P}_d = \Omega_d \cap \mathbb{Q}^d$. Let us denote the set of distributions where all probabilities have the common denominator N as \mathcal{P}_d^N . In the context of our estimation problem, we defined a map from the set of states into the Cartesian product of probability simplexes $\mathbf{p} \in \times_{k=1}^m \Omega_{d^{(k)}}$, while our estimates $\boldsymbol{\nu} \in \times_{k=1}^m \mathcal{P}_{d^{(k)}}^{N^{(k)}}$ (Here we defined the vector \mathbf{p} (respectively $\boldsymbol{\nu}$) similar to (2.31)).

For a sample of size $N^{(k)}$ drawn from a distribution $\mathbf{p}^{(k)}$, the probability to get $\boldsymbol{\nu}^{(k)} \in \mathcal{P}_{d^{(k)}}^{N^{(k)}}$ for the empirical distribution of the sample is given by

$$\frac{1}{(N^{(k)} + 1)^{d^{(k)}}} 2^{-N^{(k)} D(\boldsymbol{\nu}^{(k)} \parallel \mathbf{p}^{(k)})} \leq \operatorname{Prob}(\boldsymbol{\nu}^{(k)}) \leq 2^{-N^{(k)} D(\boldsymbol{\nu}^{(k)} \parallel \mathbf{p}^{(k)})} \quad (\text{B.2})$$

The maximum number of different empirical distributions with denominator $N^{(k)}$ is bounded by

$$\#(\mathcal{P}_{d^{(k)}}^{N^{(k)}}) \leq (N + 1)^{d^{(k)}} \quad (\text{B.3})$$

The bounds corresponding to (B.2) and (B.3) for $\boldsymbol{\nu}$ and $\times_k \mathcal{P}_{d^{(k)}}^{N^{(k)}}$ follow from

$$\text{Prob}(\boldsymbol{\nu}) = \prod_k \text{Prob}(\boldsymbol{\nu}^{(k)}) \quad \text{and} \quad \# \left(\times_k \mathcal{P}_{d^{(k)}}^{N^{(k)}} \right) = \prod_k \#(\mathcal{P}_{d^{(k)}}^{N^{(k)}}) \quad (\text{B.4})$$

In the following we derive bounds on the probability for $\boldsymbol{\nu} \in \times_{k=1}^m \mathcal{P}_{d^{(k)}}^{N^{(k)}}$ to fall into a well behaved set $A \subset \times_{k=1}^m \Omega_{d^{(k)}(k)}$ if we take samples of size $N^{(k)}$ from the distributions $\boldsymbol{p}^{(k)}$. An upper bound can be derived from the upper bounds on the probability of $\boldsymbol{\nu}$ and the number of different $\boldsymbol{\nu}$ in A :

$$\begin{aligned} \text{Prob} \left(\times_k \mathcal{P}_{d^{(k)}}^{N^{(k)}} \cap A \right) &= \sum_{\boldsymbol{\nu} \in \times_k \mathcal{P}_{d^{(k)}}^{N^{(k)}} \cap A} \text{Prob}(\boldsymbol{\nu}) \\ &\leq \sum_{\boldsymbol{\nu} \in \times_k \mathcal{P}_{d^{(k)}}^{N^{(k)}} \cap A} \prod_k 2^{-N^{(k)} D(\boldsymbol{\nu}^{(k)} \| \boldsymbol{p}^{(k)})} \\ &\leq \sum_{\boldsymbol{\nu} \in \times_k \mathcal{P}_{d^{(k)}}^{N^{(k)}} \cap A} 2^{-\min_{\boldsymbol{\nu} \in A} \sum_k N^{(k)} D(\boldsymbol{\nu}^{(k)} \| \boldsymbol{p}^{(k)})} \\ &\stackrel{*}{\leq} \sum_{\boldsymbol{\nu} \in \times_k \mathcal{P}_{d^{(k)}}^{N^{(k)}} \cap A} 2^{-\min_{\boldsymbol{\nu} \in A} \sum_k N_{\min} D(\boldsymbol{\nu}^{(k)} \| \boldsymbol{p}^{(k)})} \\ &\leq \sum_{\boldsymbol{\nu} \in \times_k \mathcal{P}_{d^{(k)}}^{N^{(k)}} \cap A} 2^{-N_{\min} D(\boldsymbol{\nu}^* \| \boldsymbol{p})} \\ &\leq \left(\prod_{k=1}^m (N^{(k)} + 1)^{d^{(k)}} \right) 2^{-N_{\min} D(\boldsymbol{\nu}^* \| \boldsymbol{p})} \end{aligned} \quad (\text{B.5})$$

Therefore we get the limit

$$\begin{aligned} \limsup_{N_{\min} \rightarrow \infty} \frac{1}{N_{\min}} \log \text{Prob} \left(\times_k \mathcal{P}_{d^{(k)}}^{N^{(k)}} \cap A \right) &\leq \\ &\lim_{N_{\min} \rightarrow \infty} \sum_k d^{(k)} \frac{\log(N^{(k)} + 1)}{N_{\min}} - D(\boldsymbol{\nu}^* \| \boldsymbol{p}) \end{aligned} \quad (\text{B.6})$$

Since the bound (*) is weak if the $N^{(k)}$ differ a lot, we get don't get convergence of the right hand side in (B.6) if N_{\max} grows exponentially in N_{\min} . However, we can establish further bounds: For a sequences of samples sizes with $N_{\min} \rightarrow \infty$ consider a subsequence $\{N_i^{(k)}\}_{k=1}^m$ such that $\min_k \{N_i^{(k)}\} < \min_k \{N_{i+1}^{(k)}\}$. Analog to (B.5)

$$\left(\prod (N^{(k)} + 1)^{d^{(k)}} \right) 2^{-\sum_k N^{(k)} D(\boldsymbol{\nu}^{*(k)} \| \boldsymbol{p}^{(k)})} \geq \text{Prob} \left(\times_k \mathcal{P}_{d^{(k)}}^{N^{(k)}} \cap A \right) \quad (\text{B.7})$$

where $\boldsymbol{\nu}_i^* = \text{argmin}_{\boldsymbol{\nu} \in A} \left(-\sum_k N_i^{(k)} D(\boldsymbol{\nu}^{(k)} \| \boldsymbol{p}^{(k)}) \right)$. Since the \mathbb{Q} is dense in \mathbb{R} we can find a sequence of $\boldsymbol{\nu}'_i \in \times_k \mathcal{P}_{d^{(k)}}^{\min_k \{N_i^{(k)}\}} \cap A$ with $\|\boldsymbol{\nu}'_i - \boldsymbol{\nu}_i^*\| \rightarrow 0$ if $i \rightarrow \infty$. We obtain a lower bound for the case when all sample sizes are equal, $N_i^{(k)} = \min_k \{N_i^{(k)}\}$ for all k , by the probability of the single $\boldsymbol{\nu}_i$:

$$\text{Prob} \left(\times_k \mathcal{P}_{d^{(k)}}^{\min_k \{N_i^{(k)}\}} \cap A \right) \geq \frac{1}{(N_{\min} + 1)^{md^{(k)}}} 2^{-N_{\min} D(\boldsymbol{\nu}_i \| \boldsymbol{p})} \quad (\text{B.8})$$

Already if the difference between N_{\max} and N_{\min} grows linearly, for big enough N_{\min} the right hand side of (B.8) grows atop of the left hand side of equation (B.7)¹ and the convergence of $\text{Prob} \left(\times_k \mathcal{P}_{d^{(k)}}^{\min_k \{N_i^{(k)}\}} \right)$ follows from ordinary Sanov's theorem.

On the other hand, if $\boldsymbol{\nu}^*$ is an accumulation point of the set A we can find a sequence of $\boldsymbol{\nu}_i \rightarrow \boldsymbol{\nu}^*$ in A and the probability of $\boldsymbol{\nu}$ to fall in A can be bounded by the probability of

¹ it may happen that $D(\boldsymbol{\nu}_i^{*(k)} \| \boldsymbol{p}^{(k)}) = 0$ for N_{\max} , however in this case one may consider N'_{\max} as the maximum of the $N_i^{(k)}$ with nonzero $D(\boldsymbol{\nu}^{(k)} \| \boldsymbol{p}^{(k)})$. In the case if N_{\max} does not grow linearly in N_{\min} but N'_{\max} does, (B.5) still holds.

ν_i :

$$\begin{aligned}
\text{Prob} \left(\times_k \mathcal{P}_{d^{(k)}}^{N^{(k)}} \cap A \right) &\geq \text{Prob}(\nu_i) \\
&\geq \frac{1}{\prod_k (N^{(k)} + 1)^{d^{(k)}}} 2^{-\sum_k N^{(k)} D(\nu_i \| \mathbf{p})} \\
&\geq \frac{1}{\prod_k (N_{\max} + 1)^{d^{(k)}}} 2^{-\sum_k N_{\max} D(\nu_i \| \mathbf{p})} \\
&\geq \frac{1}{(N_{\max} + 1)^{m d_{\max}^{(k)}}} 2^{N_{\max} D(\nu_i \| \mathbf{p})}
\end{aligned} \tag{B.9}$$

which gives the second inequality in (B.1) by

$$\begin{aligned}
\liminf_{N_{\min} \rightarrow \infty} \frac{1}{N_{\max}} \log \text{Prob} \left(\times_k \mathcal{P}_{d^{(k)}}^{N^{(k)}} \cap A \right) &\geq \\
&\lim_{N_{\min} \rightarrow \infty} m d_{\max}^{(k)} \frac{\log(N_{\max} + 1)}{N_{\max}} - D(\nu^* \| \mathbf{p}) \tag{B.10}
\end{aligned}$$

Bibliography

- [1] E. Bagan, M. A. Ballester, R. D. Gill, A. Monras, and R. Muñoz-Tapia. Optimal full estimation of qubit mixed states. *Phys. Rev. A*, 73:032301, 2006.
- [2] A. R. Barron. Entropy and the central limit theorem. *Ann. Probab.*, 14(1):336–342, 1986.
- [3] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.
- [4] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley Series in Telecommunications. John Wiley & Sons Inc., New York, 1991. A Wiley-Interscience Publication.
- [5] B. Farb and R. K. Dennis. *Noncommutative algebra*, volume 144 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1993.
- [6] R. D. Gill and S. Massar. State estimation for large ensembles. *Phys. Rev. A*, 61(4):042312, Mar 2000.
- [7] M. Hayashi and K. Matsumoto. Statistical model with measurement degree of freedom and quantum physics. In *Asymptotic Theory of Quantum Statistical Inference M. Hayashi eds., Ch. 13*, World Scientific 2005.
- [8] F. Hiai and D. Petz. *The semicircle law, free random variables and entropy*, volume 77 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2000.
- [9] A. S. Holevo. *Probabilistic and statistical aspects of quantum theory*, volume 1 of *North-Holland Series in Statistics and Probability*. North-Holland Publishing Co., Amsterdam, 1982. Translated from the Russian by the author.
- [10] Z. Hradil, J. Řeháček, J. Fiurášek, and M. Ježek. Maximum-likelihood methods in quantum mechanics. In *Quantum state estimation*, volume 649 of *Lecture Notes in Phys.*, pages 59–112. Springer, Berlin, 2004.

- [11] J. Řeháček, B. G. Englert, and D. Kaszlikowski. Minimal qubit tomography. *Phys. Rev. A*, 70(5):052321, Nov 2004.
- [12] R. A. Johnson and D. W. Wichern. *Applied multivariate statistical analysis*. Pearson Prentice Hall, Upper Saddle River, NJ, sixth edition, 2007.
- [13] E. L. Lehmann and G. Casella. *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1998.
- [14] D. V. Lindley. On a measure of the information provided by an experiment. *Ann. Math. Statist.*, 27:986–1005, 1956.
- [15] A. Magyar, D. Petz, and K. M. Hangos. Bayesian qubit state estimation. *Proceedings of 14th IFAC Symposium on System identification*, pages 949–954, 2006.
- [16] S. Massar and S. Popescu. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.*, 74(8):1259–1263, 1995.
- [17] H. Ohno. Quasi-orthogonal subalgebras of matrix algebras. *Linear Algebra Appl.*, 429(8-9):2146–2158, 2008.
- [18] H. Ohno, D. Petz, and A. Szántó. Quasi-orthogonal subalgebras of 4×4 matrices. *Linear Algebra Appl.*, 425(1):109–118, 2007.
- [19] D. Petz. Complementarity in quantum systems. *Rep. Math. Phys.*, 59(2):209–224, 2007.
- [20] D. Petz. *Quantum information theory and quantum statistics*. Theoretical and Mathematical Physics. Springer-Verlag, Berlin, 2008.
- [21] D. Petz, K. M. Hangos, and A. Magyar. Point estimation of states of finite quantum systems. *J. Phys. A*, 40(28):7955–7969, 2007.
- [22] D. Petz, K. M. Hangos, A. Szántó, and F. Szöllősi. State tomography for two qubits using reduced densities. *J. Phys. A*, 39(34):10901–10907, 2006.
- [23] D. Petz and J. Kahn. Complementary reductions for two qubits. *J. Math. Phys.*, 48(1):012107, 6, 2007.
- [24] W. Pfeifer. *The Lie algebras $\mathfrak{su}(N)$* . Birkhäuser Verlag, Basel, 2003. An introduction. (Chapter 2.1).
- [25] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves. Symmetric informationally complete quantum measurements. *J. Math. Phys.*, 45(6):2171–2180, 2004.
- [26] S. G. Schirmer, T. Zhang, and J. V. Leahy. Orbits of quantum states and geometry of Bloch vectors for n -level systems. *J. Phys. A*, 37(4):1389–1402, 2004.

- [27] T. Tasnádi. Maximal qubit tomography. *arXiv:0803.1946v1*, 2008.
- [28] W. K. Wootters and B. D. Fields. Optimal state-determination by mutually unbiased measurements. *Ann. Physics*, 191(2):363–381, 1989.