Symmetric Minimal Quantum Tomography and Optimal Error Regions

Shang Jiangwei

2013

Symmetric Minimal Quantum Tomography and Optimal Error Regions

SHANG JIANGWEI

B.Sc. (Hons.), National University of Singapore

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Centre for Quantum Technologies

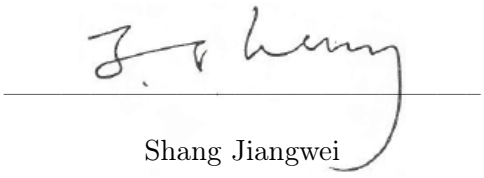
National University of Singapore

2013

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.



Shang Jiangwei

30 Sept 2013

Dedicated to my family, friends

and teachers...

Acknowledgments

First and foremost, I would like to thank my supervisor Prof. Berthold-Georg Englert for his tireless support throughout my undergraduate study as well as Ph.D. candidature in Singapore. I am deeply grateful for your invaluable guidance, as well as your passion, wisdom and insights in Physics that inspires and encourages me always. Thank you for the unconditional support and freedom that is provided through the years and also your integrity and honesty of being an idol that I will follow all along.

I'd like to thank Prof. Feng Yuanping and Prof. Oh Choo Hiap for encouraging and writing me the recommendation letters, and thank Prof. Valerio Scarani and Asst/Prof. Li Wenhui for interviewing and then recommending me into the CQT Ph.D. program. I also want to thank Prof. Vlatko Vedral and Assoc/Prof. Gong Jiangbin for serving in my thesis advisory committee.

Thank you, Asst/Prof. Ng Hui Khoon and Amir Kalev for collaborating with me on much of the work in the past few years, and also for your patience in teaching me and making our discussions effective and enjoyable. Moreover, thank you very much for critical reading of this thesis and giving many valuable comments.

A special thank to Han Rui for being supportive always and to Lee Kean Loon for sharing with me lots of programming skills. I wish to extend my sincere thanks to all my colleagues who gave me the possibility to complete this thesis, especially Assoc/Prof. David Nott, Markus Grassl, Arun Sehrawat, Li Xikun, Tomasz Karpiuk, Zhu Huangjun, Teo Yong Siah and so on.

My gratitude also goes to all my friends in Singapore and China: Zheng Yongming, Cheng Bin, Zheng Lisheng, Li Ang, Li Jinxin, Li Yuxi, Fan Zhitao, Pei Yunbo, Wu Yuanhao, etc. Your friendship is always a source of assurance and support all along.

I would like to acknowledge the financial support from Centre for Quantum Technologies, a Research Centre of Excellence funded by the Ministry of Education and the National Research Foundation of Singapore. I am also grateful to the administrative staff in CQT for providing a comfortable environment and numerous timely help.

Thanks mum, dad, brother, sister-in-law and my two lovely nieces. Nothing would have been possible nor meaningful without your never ending love and support.

Last but not least, thanks again to every one of you who never give up on me!

J. Shang

Singapore, Sept 2013

Contents

Summary

This thesis comprises the study of two basic topics in quantum information science: symmetric minimal quantum tomography and optimal error regions.

We first consider the implementation of the symmetric informationally complete probability-operator measurement (SIC POM) in the Hilbert space of a d-level system in terms of two successive measurements: a diagonal-operator measurement with high-rank outcomes, followed by a rank-1 measurement in a basis chosen in accordance with the result of the first measurement. We show that any Heisenberg-Weyl group-covariant SIC POM can be realized by such a sequence where the second measurement is simply a measurement in the Fourier basis, independent of the result of the first measurement. Furthermore, we study in particular such constructions of SIC POMs in dimensions 2, 3, 4, and 8. Surprisingly, this formulation reveals an operational relation between mutually unbiased bases (MUB) and SIC POMs; the former are used to construct the latter. As a laboratory application of the two-step measurement process, we propose feasible optical experiments that would realize SIC POMs in various dimensions.

The second part of this thesis investigates a simple construction of optimal error regions for quantum state estimation. A point estimator, constructed from the measurement outcomes on a finite number of independently and identically prepared systems, can never be perfectly accurate; it has to be supplemented with an error region that summarizes our uncertainty about the guess. Exploiting the natural correspondence between the size of a region in state space and its prior content, we show that the optimal choices for two types of error regions—the maximum-likelihood region, and the smallest credible region—are both concisely described as the set of all states for which the likelihood (for the given tomographic data) exceeds a threshold value, *i.e.*, a bounded-likelihood region. These error regions are reminiscent of the standard error regions obtained by analyzing the vicinity of the maximum of the likelihood function, a construction valid only when a large number of copies of the state have been observed. Yet, we require no such restriction. This surprisingly simple characterization permits

concise reporting of the error regions even in high-dimensional problems. Besides, our error regions are conceptually different from confidence regions, a subject of recent discussion in the context of quantum state estimation; however, the smallest credible regions can serve as good starting points for constructing confidence regions. We discuss criteria for assigning prior probabilities to regions, and illustrate the concepts and methods with several examples.

List of Tables

List of Figures

List of Symbols[1]

$C(\theta)$	Covariance matrix	22
$C_{\mathcal{R}}$	Credibility of region $\mathcal{R}$	77
$\mathcal{C}_D$	Confidence region for data D	87
d	Dimension of the Hilbert space	*
$D_{\mathrm{B}}(\cdot)$	Bures distance	128
$D_{\mathrm{HS}}(\cdot)$	Hilbert-Schmidt (HS) distance	127
D_{HW}	Heisenberg-Weyl (HW) group	32
$D_{\mathbf{k}}, D_{k_1,k_2}$	Displacement operator of the HW group	33
$D_{\mathrm{tr}}(\cdot)$	Trace distance	126
$\mathrm{E}(\cdot)$	Expectation value	21
f_k	Frequencies	14
$\lvert F\rvert$	Order of finite field F	121
$\hat{H}$	Hamiltonian	11
$\mathcal{H}$	Hilbert space	*
$\hbar$	reduced Planck constant	11
I	identity operator	10
$\mathcal{I}, \mathcal{I}_{jk}$	Fisher information matrix (FIM)	21
$\langle \mathbf{k}, \mathbf{q} \rangle$	$:= k_2 q_1 - k_1 q_2$, the symplectic form	33
$\mathcal{L}(\rho)$	Likelihood functional	15
$L(D\lvert\mathcal{R})$	Region likelihood	77
N	The number of states used in state tomography	*
p_k	Probabilities	11
$\widehat{\mathcal{R}}_{\mathrm{ML}}$	Maximum-likelihood region (MLR)	78
$\mathcal{R}_\lambda$	Bounded-likelihood region (BLR)	81
$S_{\mathcal{R}}$	Size of region $\mathcal{R}$	74

[1] The page number where a symbol is defined is listed at the rightmost column. When the definition is general, the page number is given as *.

$\mathrm{tr}\{\cdot\}$	Trace of an ordinary operator 11
$\mathrm{Var}(\cdot)$	Variance 21
W	Weight matrix 23
X, Z	Cyclic shift operator and phase operator 33
$\mathbb{Z}^+$	Set of positive integers 121
ρ	A generic quantum state ..... *
$\widehat{\rho}$	An estimator of ρ 14
$\widehat{\rho}_{\mathrm{ML}}$	Maximum-likelihood estimator (MLE) 16
ω	$:= \mathrm{e}^{\mathrm{i}2\pi/d}$, fundamental dth root of unity 33
τ	$:= -\mathrm{e}^{\mathrm{i}\pi/d}$ 33
$\sigma_x, \sigma_y, \sigma_z$	Pauli operators ..... *
λ_x	Eigenvalues of ρ 130
Π_k	Measurement outcome 12
Λ_k	Reconstruction operator 15
$\oplus, \ominus$	Field addition and subtraction operations 121
$\odot, \oslash$	Field multiplication and division operations 122

List of Abbreviations

AAPT	Ancilla-assisted process tomography
BLR	Bounded-likelihood region
BM	Bayesian mean
BME	Bayesian mean estimator
BS	Beam splitter
CRLB	Cramér-Rao lower bound
DCQD	Direct characterization of quantum dynamics
EAPT	Entanglement-assisted process tomography
FIM	Fisher information matrix
FT	Fourier transform
HMLE	Hedged maximum-likelihood estimation
HS	Hilbert-Schmidt
HW	Heisenberg-Weyl
HWP	Half-wave plate
IC	Informationally complete
ILS	Iso-likelihood surface
ME	Maximum estimator
ML	Maximum likelihood
MLE	Maximum-likelihood estimator
MLR	Maximum-likelihood region
MSE	Mean square error
MSH	Mean square Hilbert-Schmidt distance
MUB	Mutually unbiased bases
MVU	Minimum variance unbiased
PBS	Polarizing beam splitter
POM	Probability-operator measurement
POVM	Positive operator-valued measure

PPBS	Partially polarizing beam splitter
PVM	Projection-valued measure
PS	Phase shifter
QMT	Quantum measurement tomography
QPT	Quantum process tomography
QST	Quantum state tomography
RLD	Right logarithmic derivative
SCR	Smallest credible region
SIC	Symmetric informationally complete
SLD	Symmetric logarithmic derivative
SQPT	Standard quantum process tomography
TM	Tetrahedron measurement
vNM	von Neumann measurement
WMSE	Weighted mean square error

Chapter 1

Introduction

Quantum mechanics is a mathematical framework for the development of physical theories. On its own, quantum mechanics doesn't tell you what laws a physical system must obey, but it does provide a mathematical and conceptual framework for the development of such laws. As we know, all classical theories, including Newton's mechanics, Maxwell's electromagnetism as well as Einstein's relativity, are deterministic in the sense that the state of the system uniquely determines all phenomena about the system in the future, as well as in the past, at least in principle. However, a fundamental feature of quantum theory is that it is probabilistic, not deterministic [1]. Complete knowledge of the state does not enable us to predict the outcomes of all measurements that could be performed on the system, but only the probabilities of the possible outcomes. In other words, the state does not determine the phenomena about the system.

Generally, there are four postulates in quantum mechanics that provide the connection between the physical world and the mathematical formalism. Here, we only give a global review of these postulates [2, 3], with the more detailed description of them to be given along the course of this thesis. Postulate 1 sets the arena for quantum mechanics, by specifying how the state of an isolated quantum system is to be described. Postulate 2 tells us that the dynamics of closed quantum systems are described by the Schrödinger equation, and thus by unitary evolution. Postulate 3 tells us how to extract information from our quantum systems by giving a prescription for the description of measurement. Postulate 4 tells us how the state spaces of different quantum systems may be combined to give a description of the composite system.

According to Postulate 1, any isolated physical system can be described by a state vector (or a statistical operator) residing in its state space, *i.e.*, the Hilbert space. The

state of a physical system is the mathematical description of our knowledge of it, and provides information on its future and past. Therefore, a state tomographic technique is designed to acquire the complete information of a system, in other words, to achieve the maximum possible knowledge of the state, thus allowing one to make the best probabilistic predictions on the results of any measurement that could be performed on the system [4]. Different from its classical counterpart, the state of a quantum system is confined by the fundamental features of quantum theory, namely the Heisenberg uncertainty relation [5, 6] and the no-cloning theorem [7, 8]. Therefore, it is impossible to infer a generic unknown quantum state from measurements on a single copy of the system; that is, many copies of independently and identically prepared quantum systems are needed for reliable state determination.

Quantum state tomography (also called quantum state estimation; note that we use these two terms interchangeably in this thesis) [4] is a measurement procedure designed to acquire complete information about the state of a given quantum system. It is indispensable to take into account additional constraints, such as the positivity of quantum states, when designing quantum tomographic methods. In addition, the choice of strategies may also depend on the system under consideration and the application in mind. As can be seen, a complete implementation of quantum state tomography involves two basic steps, namely the measurement scheme to get data first, followed by a data processing protocol. One of the main challenges in quantum state tomography is to infer quantum states as efficiently as possible (in terms of, for instance, time consumption) and to optimize the resources necessary to achieve a given accuracy, which can be quantified by various figures of merit, such as the mean trace distance, the mean square Hilbert-Schmidt distance (MSH), the mean fidelity and so on.

Besides its fundamental importance, quantum state tomography is also a crucial component in most, if not all, quantum computation and quantum communication tasks. The characterization of a source of quantum carriers, the verification of the properties of a quantum channel, the monitoring of a transmission line used for quantum key distribution—all three require reliable quantum state tomography, to name just the

most familiar examples. The successful execution of such tasks hinges in part on the ability to assess the state of the system at various stages.

A good quantum state tomographic strategy entails judicial choices on both measurement schemes and data processing protocols for reconstructing the true state. Compared with measurement schemes, there is generally more freedom in choosing the reconstruction methods in practice, and a good choice is the first step towards getting a reliable and efficient estimator. On the other hand, given the measurement results, the optimization of data processing is basically a subject of classical statistical inference, although attention has to be paid to account for any additional quantum constraints, such as the positivity of the density matrices. Therefore, if concentrating solely on the reconstruction methods, quantum state tomography *is* classical state tomography with quantum constraints. Accordingly, quantum mechanicians can benefit much from the methods developed by statisticians.

Since the data have statistical noise, every estimation strategy comes with errors. It is well known in classical statistical inference that the minimal error is determined by the Fisher information matrix [9] through the Cramér-Rao lower bound (CRLB) [10,11]. Therefore, to be statistically meaningful, any point estimator has to be supplemented with error bars of some sort, or error regions beyond dimension one. Many ad-hoc recipes have been proposed for attaching a vicinity of states to an estimator, which usually rely on having a lot of data, involve data resampling, or consider all data that one might have observed. By contrast, in this thesis, we tackle this problem by systematically constructing error regions from the data we actually observed.

In another respect, the main departure of quantum state tomography from its classical counterpart is the choice over measurements, which underlies the difference between quantum information processing and classical information processing. In practice, the set of possible measurements is mainly determined by the experimental apparatus. As technology advances, it is ultimately limited by the basic principles of quantum mechanics. For example, as a consequence of the Heisenberg uncertainty relation [5,6] and the complementarity principle [7, 8], it is impossible to measure two non-commuting

sharp observables simultaneously, which implies that no measurement can extract maximal information about both observables simultaneously. To put it differently, any gain of information about one observable is necessarily accompanied with a loss of information about the other. Therefore, to devise good measurement schemes, it is crucial to balance such information trade-off, which is one of the main challenges in current quantum state tomography theory.

The most natural and useful type of measurements in quantum mechanics is the generalized measurement, which is often referred to as probability-operator measurement (POM) or positive operator-valued measure (POVM). A POM is informationally complete (IC) if any state of the system is determined completely by the probabilities of the possible outcomes. A symmetric IC POM (SIC POM) is an IC POM of a particular kind: In a finite d-dimensional Hilbert space, it is composed of d^2 subnormalized projectors onto pure states with equal pairwise fidelity (the equiangular condition) [12,13]. The high symmetry and high tomographic efficiency of SIC POMs have attracted the attention of many researchers; see, for example, Refs. [12–18]. Besides, SIC POMs are closely related to many other problems in both physics and mathematics, such as quantum cryptography [19,20], MUB [21–24], t-designs and equiangular lines [12,13], and other foundational studies.

All SIC POMs known so far are group covariant in the sense that each of them can be generated from a single state—the fiducial state—under the action of a group composed of unitary operators. Moreover, most known group-covariant SIC POMs are covariant with respect to the Heisenberg-Weyl (HW) group, except for the set of Hoggar lines (in dimension $8 = 2^3$), which is covariant with respect to the three-qubit Pauli group. It seems that there is a deep root for this observation, but the reason is still unclear. Up to now, analytical solutions of HW SIC POMs have been constructed in dimensions 2–16, and 19, 24, 28, 31, 35, 37, 43, 48; numerical solutions with high precision have been found up to dimension 67. All these results strongly support Zauner's conjecture [13] that HW covariant SIC POMs exist in any Hilbert space of finite dimension. In sharp contrast with this wealth of evidence, there is neither

an existence proof nor an efficient way for constructing SIC POMs. What is worse, many basic properties of SIC POMs have remained elusive. The implication of the equiangular condition is largely a mystery, although it looks so simple. In this thesis, we study the construction and implementation of SIC POMs by using what we call the successive-measurement scheme.

Chapter 2 of this thesis presents an overview of quantum state tomography from the theoretical perspective. We start with a brief introduction of the developments in this field and then introduce several basic ingredients in quantum state tomography, such as quantum states and measurements, quantum tomographic methods, Fisher information, and estimation errors. For the tomographic methods, we first present the simplest linear inversion method as well as the well-known maximum-likelihood estimation, followed by several other methods, including the hedged maximum-likelihood estimation, the Bayesian mean estimation, and the minimax mean estimation. We then show the derivation of the Jeffreys prior in Bayesian statistics from the Fisher information.

Chapter 3 deals with the problem of quantum measurements. Based on Postulate 3 of quantum mechanics, we first introduce two general types of quantum measurement, *i.e.*, the projective measurement and the generalized measurement. Then we talk about the basic features of SIC POMs and the construction of group-covariant SIC POMs, followed by the discussion of MUB and the construction of MUB when the dimension d is a prime power. In the last section of this chapter, we present the scheme of successive measurements, using which a few proposals for implementing SIC POMs will be given in the following chapter.

In Chapter 4, we consider the implementation of SIC POMs in the d-dimensional Hilbert space by employing a two-step measurement process: a diagonal-operator measurement with high-rank outcomes, followed by a rank-1 measurement in a basis chosen in accordance with the result of the first measurement [23,24]. By using this scheme, we are able to realize any Heisenberg-Weyl group-covariant SIC POM, where the second measurement is simply a measurement in the Fourier basis, independent of the result of the first measurement. Then, we study the construction of SIC POMs in dimensions

2, 3, 4, and 8 respectively. We find an unexpected operational relation between MUB and SIC POMs; the former are used to construct the latter. In order to implement the two-step measurement process in the laboratory, we also propose feasible optical experiments that would realize SIC POMs in various dimensions.

Chapter 5 considers the construction of optimal error regions for quantum state estimation [25]. Instead of reporting a single point estimator for the actual state of the quantum system for the given data, we intend to assign a region for it. As opposed to standard ad-hoc constructions of error regions, we introduce the maximum-likelihood region—the region of largest likelihood among all regions of the same size—as the natural counterpart of the popular maximum-likelihood point estimator. Here, the size of a region is its prior probability. A related concept is the smallest credible region—the smallest region with pre-chosen posterior probability. For both optimization problems, the optimal region has constant likelihood on its boundary. This surprisingly simple characterization permits concise reporting of the error regions even in high-dimensional problems. We also discuss several criteria for assigning prior probabilities to regions. For illustration, we first apply the method to study the problem of a classical coin. Then in the quantum scenario, we identify optimal error regions for single qubit (confined to the equatorial plane of the Bloch sphere) and two-qubit states from computer-generated data that simulate incomplete tomography with few measured copies.

We close with a short conclusion and outlook in Chapter 6.

Chapter 2

Quantum state tomography

2.1 Introduction

Quantum state tomography (QST) is a procedure for inferring the state of a quantum system from generalized measurements, known as probability-operator measurements (POMs). Owing to the Heisenberg uncertainty relation [5, 6] and the complementarity principle [7, 8], any measurement on a generic quantum system necessarily induces a disturbance, limiting further attempts to extract information from the system. As a result, it is impossible to fully recover the true state of a quantum system if only a finite number of measurements are performed. Quantum state tomography is an important and primitive component in most, if not all, quantum information processing tasks, such as quantum computation, quantum communication, and quantum cryptography, because all these tasks rely heavily on our ability to determine the state of a quantum system at various stages.

The problem of QST can be traced back to Pauli [26] when he asked whether the position distribution and momentum distribution suffice to determine the wave function of a quantum system. However, a systematic study was not initiated until the 1950s when Fano [27] introduced the concept of a quorum. Later, Ivanović [28] explored the state estimation problem from a geometric perspective, with a special emphasis on mutually unbiased measurements. He also constructed a complete set of mutually unbiased measurements when the dimension is a prime, followed by a generalization to prime power dimensions by Wootters and Fields [29].

The advance of experimental techniques and the emergence of quantum information science further stimulated the development of QST. The problem of reconstructing

quantum states from informationally incomplete measurements was addressed by Bužek *et al.* [30, 31], who proposed a method for selecting the most objective estimator by means of Jaynes principle of maximum entropy [32, 33]. Meanwhile, the maximum-likelihood (ML) estimation was advocated by Hradil [4, 34], who developed an efficient algorithm for computing the ML estimator, which avoids the problems of non-positivity and choice ambiguity of the traditional linear estimators. As alternatives to the ML approach, several other methods for state tomography have been developed, including the hedged maximum-likelihood estimation (HMLE) [35–38], the Bayesian mean (BM) estimation [39–45], and the minimax mean estimation [46–50]. These methods are proposed to solve the zero-eigenvalue problem which often occurs in the ML estimation, but may result in additional complications and more computational needs. Meanwhile, several methods have been developed to deal with large quantum systems, such as compressed sensing [51] and direct fidelity estimation [52].

Every statistical inference comes with errors, so how to quantify the efficiency of a state tomographic strategy? This question was first addressed by Helstrom [53, 54], who prompted the introduction of quantum analogs of the Fisher information and the Cramér-Rao lower bound (CRLB) based on the symmetric logarithmic derivative (SLD), and then solved the optimization problem in the one-parameter setting. For the multi-parameter scenario, Yuen and Lax [55] solved the problem of estimating the complex amplitude of coherent signal in Gaussian noise by means of CRLB based on the right logarithmic derivative (RLD), which is often tighter than the SLD bound in the multi-parameter setting. Based on a similar approach, Holevo [56] solved the estimation problem about the mean value of Gaussian states. He also introduced a new quantum Cramér-Rao bound, known as the Holevo bound, which is tighter than both the SLD bound and the RLD bound. However, this bound is generally not easy to calculate since the definition itself involves a tough optimization procedure.

As an extension to QST, quantum process tomography (QPT) focuses on characterizing unknown quantum operations (also called quantum processes or quantum channels) instead of quantum states, which is crucial to ensure the performance of

many quantum information processing protocols. Its development has drawn much inspiration from QST, since mathematically QPT and QST are proved to be equivalent [57]. Introduced by Chuang and Nielsen [58] as well as by Poyatos *et al.* [59] in the late 1990s, the standard QPT (SQPT) involves preparing an ensemble of quantum states and sending them through the process, then using quantum state tomography to identify the resultant states. Several experimental demonstrations of SQPT in NMR [60, 61] and quantum optics systems [62] have been done recently. Other techniques of QPT include the ancilla-assisted process tomography (AAPT) [57, 63] and entanglement-assisted process tomography (EAPT) [64], which make use of an additional ancilla system. All the previous techniques are known as indirect methods for characterization of quantum dynamics, since they require the use of QST to reconstruct the process. In contrast, there are direct methods such as the direct characterization of quantum dynamics (DCQD) [65–68] which provide a full characterization of quantum systems without using state tomography. Reference [69] is a recent survey on all the strategies of QPT and provides a benchmark which is necessary for choosing the scheme that is the most appropriate in a given situation, for given resources.

In a sense complementary to QST and QPT, quantum measurement tomography (QMT) [70, 71] tries to calibrate the measuring apparatus prior to any quantum processing tasks. The strategy is to send in systems of various known states, and use these states to estimate the outcomes of the unknown measurement. Since a measurement can be characterized by a set of POMs, the goal of QMT is to reconstruct these POM outcomes. Inspired by QST, the same strategies, such as the ML estimation [71] and the Bayesian methods, can be used for QMT. Since the observation of several different quantum states by a single measuring apparatus is equivalent to the measurement of several non-commuting observables on many copies of a given quantum state, the ML approach of the QMT can be interpreted as a synthesis of information from mutually incompatible observations [72, 73].

In this chapter, we first review the basic ingredients in QST, such as quantum states and measurements, quantum tomographic methods, Fisher information, and estimation

errors. We then show the derivation of the Jeffreys prior in Bayesian statistics from the Fisher information in Sec. 2.4.1. Some of the topics, like quantum measurements and the Jeffreys prior, will be discussed again in later chapters.

2.2 Quantum states and measurements

2.2.1 Simple systems

Postulate 1 of quantum mechanics says that, associated to any isolated physical system is a complex vector space with inner product known as the *state space*, or the *Hilbert space*, usually denoted by $\mathcal{H}$. All information about the quantum system is encoded in its state vector, which is a unit vector in the system's state space. The knowledge of the state is equivalent to knowing the result of any possible measurement on the system. Mathematically, a pure state is represented by a normalized ket, say $|\psi\rangle$, and any superposition of kets also represents a legitimate state. Since kets that are proportional to each other are physically equivalent, there is a one-to-one correspondence between the pure states and the rays in the Hilbert space.

In general, one can describe the state of a quantum system in the language of a density operator (also called a statistical operator), which is a positive semidefinite matrix of unit trace, usually denoted by ρ. Density operators with rank 1 represent pure states, while those with higher ranks represent mixed states; mathematically, a pure state satisfies $\mathrm{tr}\{\rho^2\} = 1$, but a mixed state has $\mathrm{tr}\{\rho^2\} < 1$. For instance, the density operator of an arbitrary qubit state can be written as

$$\rho = \frac{1}{2}(I + \overrightarrow{r} \cdot \overrightarrow{\sigma}) \,, \tag{2.1}$$

where $\overrightarrow{r}$ is a real three-dimensional vector satisfying $|\overrightarrow{r}| \leq 1$, and $\overrightarrow{\sigma}$ are the Pauli matrices. This state can be visualized in a Bloch ball with Bloch vector $\overrightarrow{r}$, such that all ρs with $|\overrightarrow{r}| = 1$ residing on the surface are pure states and all ρs with $|\overrightarrow{r}| < 1$ inside the sphere are mixed states. When $|\overrightarrow{r}| = 0$, the state becomes $\rho = I/2$, which is called the completely mixed state.

The second Postulate of quantum mechanics gives a prescription for the description of state changes. The evolution of a closed quantum system $|\psi\rangle$ at time t_1 is described by a unitary transformation, such that

$$|\psi'\rangle = U|\psi\rangle\,, \tag{2.2}$$

where $|\psi'\rangle$ is the state at time t_2 and the unitary operator U depends only on the times t_1 and t_2. Or put it differently, Postulate 2 can be described by the time-dependent Schrödinger equation,

$$\mathrm{i}\hbar\frac{\partial|\psi\rangle}{\partial t} = \hat{H}|\psi\rangle\,, \tag{2.3}$$

where $\hbar$ is the reduced Planck constant and $\hat{H}$ is a Hermitian operator known as the Hamiltonian of the closed system. If we know the Hamiltonian of a system, then we understand its dynamics completely, at least in principle. However, in practice, it can be very difficult to figure out the Hamiltonian, and then solve the Schrödinger equation.

A quantum system evolves according to unitary evolution when it is closed. But when the system interacts with the rest of the world, the system is no longer closed, and thus not necessarily subject to unitary evolution. Then Postulate 3 of quantum mechanics provides a way for describing the effects of measurements on quantum systems, according to which, observation in quantum mechanics is an invasive procedure that typically changes the state of the system. A *generalized measurement* in quantum mechanics is described by a set of measurement operators $\{M_k\}$ corresponding to a set of measurement outcomes, which satisfy the completeness condition,

$$\sum_k M_k^\dagger M_k = 1\,. \tag{2.4}$$

Given the initial state of a quantum system ρ, the probability p_k that outcome k occurs is given by the Born rule,

$$p_k = \mathrm{tr}\Big\{M_k \rho M_k^\dagger\Big\}\,. \tag{2.5}$$

As a result of the completeness condition, summation of the probabilities is equal to the

identity, *i.e.*, $\sum_k p_k = 1$, and the post-measurement statistical operator of the system is described as follows

$$\rho' = \frac{M_k \rho M_k^\dagger}{\mathrm{tr}\left\{ M_k \rho M_k^\dagger \right\}} . \tag{2.6}$$

A measurement is a *projective* (or von Neumann) measurement if the measurement operators M_ks are orthogonal projectors. A projective measurement is repeatable in the sense that repeated measurements yield the same outcome as the first one and thus provide no additional information about the original quantum system. If we are interested only in the outcome statistics but not the state after the measurement, the measurement can be effectively described by a set of positive operators $\Pi_k = M_k^\dagger M_k$, with $\sum_k \Pi_k = 1$. In this case, the measurement is referred to as a *probability-operator measurement* (POM), and the set of operators Π_ks may be identified with the outcomes of the measurement. According to Neumark's dilation theorem [74], any POM can be realized as a projective measurement on a larger system.

A measurement is informationally complete (IC) if any state is completely determined by the outcome statistics [75]. In a finite d-dimensional Hilbert space, an IC measurement consists of at least d^2 outcomes. An informationally overcomplete measurement is an IC measurement with more than d^2 outcomes. We will give a more thorough discussion about quantum measurements in Chapter 3, with more emphasis on symmetric IC POMs and mutually unbiased measurements.

2.2.2 Composite systems

Compared with simple systems, a distinctive feature of composite systems is the existence of quantum correlations known as entanglement [76], as emphasized by the famous EPR paradox [77]. Quantum entanglement is not only a characteristic feature of quantum physics but also a crucial resource for many information processing tasks [76], such as quantum teleportation [78], superdense coding [79], quantum key distribution [80], and quantum computation [81]. Its connection with quantum state tomography can be elaborated in two aspects. On one hand, quantum tomographic techniques provide basic means of detecting, quantifying, and characterizing entangle-

ment [76, 82–85]. On the other hand, entanglement is a basic ingredient of collective measurements [86], the most general measurements allowed by quantum mechanics.

Postulate 4 of quantum mechanics describes how the state space of a composite system is built from the state spaces of the component systems. Consider a bipartite composite system as an example. Suppose the Hilbert spaces of two physical systems A and B are $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively, then the Hilbert space $\mathcal{H}$ of the whole system is the tensor product $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$. If we denote the state of the composite system as ρ_{AB}, the reduced density operator for system A is obtained by taking the partial trace over system B, *i.e.*, $\rho_A = \mathrm{tr}_B\{\rho_{AB}\}$. A pure state $\rho \in \mathcal{H}$ is separable if it is a tensor product of the two states in each Hilbert space; otherwise, it is entangled. In other words, a pure state is separable if and only if each reduced state is pure. A mixed state is separable if it can be written as a convex combination of separable pure states [87] and is entangled otherwise. Similar concepts can also be defined for systems composed of more than two parties [76].

A measurement on a composite system is *collective* if it cannot be decomposed into individual measurements on the constituent subsystems. A *separable* measurement is defined if each outcome can be written as a convex combination of tensor products of positive operators, or equivalently, if each outcome corresponds to a separable state, which is not necessarily normalized [88]. A simple example of separable measurements are product measurements, which can be decomposed into independent measurements on the constituent subsystems.

A measurement is *entangled* if it is not separable. A simple example of entangled measurements in the two-qubit setting is the Bell measurement. In practice, it is generally much harder to realize entangled measurements than separable measurements. There is an open question in quantum state tomography theory: By how much can the efficiency be increased with entangled measurements compared with separable measurements? Besides being of practical interest, this question is also of paramount importance in understanding the difference between quantum information processing and classical information processing.

2.3 Quantum tomographic methods

Quantum state tomography is a procedure of inferring the state of a quantum system from measurement outcomes, which originates in classical statistics literature. However, due to the fundamental limitations related to the Heisenberg uncertainty principle [5,6] and the no-cloning theorem [7, 8], it is indispensable to take into account additional constraints, such as the positivity of the quantum state, when designing quantum tomographic methods. In addition, the choice may also depend on the system under consideration and the application in mind. In this section, we review several well-known quantum tomographic methods and briefly comment on each method.

2.3.1 Linear inversion

Linear inversion (sometimes called linear state tomography) is one of the simplest reconstruction methods in state tomography, which was first considered by Fano [27] and followed by many other researchers [17,89–92]. Suppose we are given N identically prepared copies of an unknown quantum system ρ, which are then measured by the POM $\{\Pi_k\}_{k=1}^K$, with $\sum_k \Pi_k = 1$. The probability of getting a particular output k is given by the Born rule: $p_k = \mathrm{tr}\{\rho\Pi_k\}$. Provided that the kth output has been registered n_k times, $\sum_k n_k = N$, then the relative frequency of the output k is $f_k = n_k/N$. In linear inversion, one tries to find an estimator $\hat{\rho}$ that matches the observed frequencies, that is,

$$\mathrm{tr}\{\hat{\rho}\,\Pi_k\} = f_k\,, \quad \forall k\,. \tag{2.7}$$

If the measurement is IC, there exists at most one solution. If, in addition, the measurement is symmetric, there is always (exactly) one solution. Every symmetric POM has, apart from the outcome operators $\{\Pi_k\}_{k=1}^K$, a set of Hermitian, trace-1 operators $\{\Lambda_k\}_{k=1}^K$ with the defining property that $\mathrm{tr}\{\Pi_k\Lambda_l\} = \delta_{kl}$, which is also called the *dual basis*. This allows the expansion of the part of the state measured by the symmetric POM as

$$\hat{\rho} = \sum_{k=1}^{K} f_k\Lambda_k\,. \tag{2.8}$$

Therefore, the Λ_ks are also known as *reconstruction operators*. Once the reconstruction operators are known, the estimator can be computed immediately by applying Eq. (2.8). Since in reality there is generally no estimator that can match the frequencies exactly, the choices for the reconstruction operators may not be unique.

The main advantage of linear inversion is its simplicity. It is a good starting point in theoretical analysis, but not a wise choice in practice due to several major defects. First, the estimator obtained may not be positive semidefinite (may not be physical), which happens quite often if the true state has a very high purity and/or the sample size is small. This problem may be solved by mixing the estimator with some noise (the completely mixed state for example) until it is positive semidefinite. Second, there is generally no systematic strategy to choose the reconstruction operators when the measurement is informationally overcomplete, and the information encoded in the measurement results cannot be extracted optimally if the reconstruction operators are chosen a priori. To solve this problem, we need to change the reconstruction operators adaptively according to the measurement results. Alternatively, we can circumvent the two problems simultaneously by maximizing the likelihood functional (next section).

2.3.2 Maximum-likelihood estimation

First proposed by Fisher [9] in the 1920s, the maximum-likelihood (ML) estimation strategy is an entirely different approach to quantum state tomography compared to the technique of linear inversion. The principle of ML estimation is to seek the quantum state that is most likely to generate the observed data by maximizing the likelihood functional over the state space. The ML estimator (MLE) has become the estimator of choice. During the past decade, it has found extensive applications in quantum state tomography [4, 34, 93, 94] as well as some other areas like entanglement detection [82] and characterization [84].

The ML strategy consists in maximizing the *likelihood functional*, which is defined as follows

$$\mathcal{L}(\rho) = \prod_{k=1}^{K} p_k^{n_k}, \tag{2.9}$$

where $p_k = \mathrm{tr}\{\rho\Pi_k\}$ (Born rule) is the probability of obtaining the outcome k given the true state ρ. In practice, it is more convenient to work with the *log-likelihood functional*,

$$\log \mathcal{L}(\rho) = \sum_{k=1}^{K} n_k \log p_k = N \sum_{k=1}^{K} f_k \log p_k \,. \tag{2.10}$$

The MLE $\widehat{\rho}_{\mathrm{ML}}$ is obtained by maximizing the likelihood functional $\mathcal{L}(\rho)$, or equivalently the log-likelihood functional $\log \mathcal{L}(\rho)$. As a consequence of the Gibbs inequality [95],

$$\sum_k f_k \log p_k \leq \sum_k f_k \log f_k \,, \tag{2.11}$$

the estimator $\hat{\rho}$ obtained by Eq. (2.7) of linear inversion coincides with the MLE $\widehat{\rho}_{\mathrm{ML}}$ if such a state exists.

Generally, it is not an easy task to find a closed formula for the MLE $\widehat{\rho}_{\mathrm{ML}}$. Fortunately, the estimator can be computed efficiently with an algorithm proposed by Hradil [4, 34]. Since the log-likelihood functional $\log \mathcal{L}(\rho)$ is a concave function defined on a convex and closed state space, the search for the MLE turns into a convex optimization problem, which can be solved by using the steepest-ascent method. The starting point for the algorithm can be chosen arbitrarily; usually we take the completely mixed state $\rho_m = 1/d$ for step $m = 0$ in a d-dimensional Hilbert space. Then the MLE $\widehat{\rho}_{\mathrm{ML}}$ can be obtained through the iteration of the following steps:

i. Compute

$$R_m = \sum_k \frac{f_k \Pi_k}{\mathrm{tr}\{\rho_m \Pi_k\}} \,, \tag{2.12}$$

which is a positive semidefinite operator defined by its expansion into the measured POMs.

ii. Choose a small parameter ϵ_m and update the estimator ρ_m according to

$$\rho_{m+1} = \frac{(1+\epsilon_m R_m)\rho_m(1+\epsilon_m R_m)}{\mathrm{tr}\{(1+\epsilon_m R_m)\rho_m(1+\epsilon_m R_m)\}} \,. \tag{2.13}$$

iii. Break out of the iteration if $\mathrm{tr}\{|(R_m - 1)\rho_m|\} \leq \varepsilon$, where $\mathrm{tr}\{|A|\} = \mathrm{tr}\left\{\sqrt{A^\dagger A}\right\}$ is

the *trace norm* for an operator A and ε is a pre-chosen threshold value; otherwise, replace m with $m+1$ and repeat the above steps.

The small parameter ϵ_m may be chosen a priori, for instance, $\epsilon_m = 0.5$ works quite well when d is small. In general, a suitable line optimization procedure for choosing ϵ_m adaptively may help to speed up the algorithm. The MLE $\widehat{\rho}_{\rm ML}$ obtained by the above algorithm is unique if the measurement is IC; otherwise there exists a plateau in the state space on which all states have the same likelihood value, and the estimator is generally not unique. Recently, this problem was solved by Teo *et al.* [96] based on the ML principle and the maximum-entropy principle [32, 33]. An efficient algorithm was developed to compute the most objective estimator—the state with the highest von Neumann entropy among all the states that maximize the likelihood functional.

The ML estimation is by now the most popular state tomography strategy in use and it has many nice features. The MLE is guaranteed to be positive semidefinite and thus represents a legitimate quantum state; it is asymptotically unbiased; it is asymptotically efficient in the sense of attaining the CRLB for a large amount of registered data [9]; it can be computed efficiently with a simple algorithm [34] (for an improved version, see Ref. [97]). However, a major drawback of the ML technique is the zero-eigenvalue problem [45], namely that the MLE is often rank-deficient when the true state has a very high purity or when you get untypical data. These zeros eigenvalues represent unrealistic confidence over certain measurements with only a finite amount of data, which is undesirable for applications such as data compression and cryptography.

2.3.3 Other reconstruction methods

Over the past few years, several alternatives to the ML estimation approach have been proposed, including the hedged maximum-likelihood estimation [35–38], the Bayesian mean (BM) estimation [39–45], and the minimax mean estimation [46–50]. Meanwhile, several methods have been developed to deal with large quantum systems, such as compressed sensing [51] and direct fidelity estimation [52]. In this section, we briefly discuss the first three methods.

2.3.3.1 Hedged maximum-likelihood estimation

Proposed by Blume-Kohout [38], the hedged maximum-likelihood estimation (HMLE) [35–37] can be used as a plug-in substitute for the ML strategy and supplemented to solve the zero-eigenvalue problem which is likely to occur during the application of the ML estimation. This method employs an idea in classical statistical inference known as the "add β" rule, also known as Lidstone's law [98, 99]. In HMLE, instead of maximizing the likelihood functional $\mathcal{L}(\rho)$ itself, the product of $\mathcal{L}(\rho)$ and an additional *hedging functional*

$$h(\rho) = \det(\rho)^{\beta} \tag{2.14}$$

is maximized, where $\det(\cdot)$ represents the determinant and β is called the hedging parameter usually taking values between 0 and 1. By this way, the estimator defined by the maximum of the functional $\mathcal{L}(\rho)h(\rho)$ is guaranteed to have full rank. Since the hedging function $h(\rho)$ and the likelihood functional $\mathcal{L}(\rho)$ are both concave, the estimator can be computed efficiently with a similar algorithm as that used for MLE. These two nice features make HMLE an appealing alternative to the ML estimation. However, the problem with HMLE is that there is no general criterion for choosing the hedging functional, which may depend on both the prior knowledge available and the figure of merit adopted. This contrasts with the classical case, where $\beta \approx 1/2$ is known to be asymptotically optimal in all cases [36].

2.3.3.2 Bayesian mean estimation

In Bayesian mean (BM) estimation [39–45], one chooses a prior distribution $\pi_0(\rho)\,\mathrm{d}\rho$ over the state space, which represents the estimator's ignorance about the identity of the state and should generally be chosen to be as uninformative as possible. Then the posterior distribution is derived by multiplying the prior with the likelihood functional, that is, $\pi_f(\rho) \propto \mathcal{L}(\rho)\,\pi_0(\rho)\,\mathrm{d}\rho$, which represents the estimator's knowledge. The BM estimator (BME) is the mean state over the posterior, such that,

$$\widehat{\rho}_{\mathrm{BM}} = \int \rho\,\pi_f(\rho)\,\mathrm{d}\rho\,. \tag{2.15}$$

Common choices of the prior include uniform distribution with respect to the Hilbert-Schmidt measure and uniform distribution with respect to the Bures measure [100]. With a suitable choice of the prior, the BM strategy can avoid the zero-eigenvalue problem and is thus more appealing than the ML estimation. In addition, BM estimation often outperforms ML estimation when the sample size is small. The problem with BM estimation is that there is no universal criterion for selecting the prior. While some natural restrictions may be imposed on the prior based on symmetry consideration, say unitary invariance, such restrictions generally cannot specify a unique prior. Another serious problem is the difficulty in computing the estimator even numerically since it involves a high-dimensional integral over the state space. There is still no reliable and efficient algorithm for this purpose (Monte Carlo methods have been proposed).

2.3.3.3 Minimax mean estimation

The minimax mean estimator [46–48] for the trine was proposed by Ng *et al.* [49, 50] very recently. This method generalizes the classical estimator to the quantum problem upon imposing quantum constraints. Firstly, the three-outcome trine measurement (see Sec. 5.5.2.1 in Chapter 5 for more detailed discussions) has outcomes that are sub-normalized projectors onto the eigenstates of σ_x and $(-\sigma_x \pm \sqrt{3}\sigma_y)/2$ with eigenvalues $+1$. It then has the probabilities

$$p_1 = \frac{1}{3}(1+x)\,, \qquad \left.\begin{matrix} p_2 \\ p_3 \end{matrix}\right\} = \frac{1}{6}(2 - x \pm \sqrt{3}\,y)\,, \tag{2.16}$$

with $x = \langle\sigma_x\rangle$, $y = \langle\sigma_y\rangle$ and note the additional constraint, *i.e.*, $\sum_k p_k^2 \leq 1/2$. Now, the mean estimator (ME) $\widehat{\rho}_{\mathrm{ME}}$ is investigated where the mean is taken with a weight function

$$f(p) = \left(\prod_{k=1}^{K} p_k\right)^{\beta-1}. \tag{2.17}$$

Among such mean estimators, an optimal one with the smallest worst-case error (over all physical states)—the minimax mean estimator—is reported. The minimax approach makes use of the mean square error (MSE), defined for state ρ with outcome probabil-

ities p and estimator $\hat{\rho}$ with outcome probabilities $\hat{p}$ as

$$\mathrm{MSE}(\rho, \hat{\rho}) \equiv \sum_{D_N} \mathcal{L}(D_N|\rho) \sum_k [p_k - \hat{p}_k(D_N)]^2, \tag{2.18}$$

where $D_N \sim \{n_1, n_2, ..., n_K\}$ summarizes the detector clicks. Then the optimal value of β is obtained by

$$\min_{\beta} \max_{\rho} \mathrm{MSE}\left(\rho, \widehat{\rho}_{\mathrm{ME}}^{(\beta)}\right). \tag{2.19}$$

However, unlike the classical case, this optimization problem can only be performed numerically because of the quantum constraints. Another problem with this approach is that the resulting minimax mean estimator does not offer much advantage over simpler estimators like the MLE, but the small gain does not warrant the additional complications required to compute it. It is also pertinent to question if the conclusions for the trine hold in higher dimensions. Some other figures of merit (for example, the mean trace distance or relative entropy) rather than the MSE may be explored in future to assess the performance of this estimation strategy.

2.4 Fisher information and estimation errors

The efficiency of an estimation strategy can be quantified by certain measures of information, among which the Fisher information [9] is the most important one. The Fisher information is defined as the expected value of the observed information yielded by a measurement concerning certain parameters of interest. Another concept in statistical inference closely related to the Fisher information is the Cramér-Rao lower bound (CRLB) [10, 11], which quantifies the minimal error with which one can infer these parameters.

Consider the simple example of the estimation of a single parameter θ. Suppose that a family of probability distributions $p(\xi|\theta)$ with measurement outcomes ξ has been registered, based on which the true value of the parameter θ is to be estimated. The partial derivative with respect to θ of the log-likelihood function $\log p(\xi|\theta)$ is called the *score*, which reflects the sensitivity of the log-likelihood function with respect to

the variation of θ. Under certain regularity conditions, the score has a vanishing first moment; its second moment is known as the *Fisher information* [9] and it is given by

$$\mathcal{I}(\theta) = \mathrm{Var}\left(\frac{\partial \log p(\xi|\theta)}{\partial \theta}\right) = \sum_{\xi} p(\xi|\theta)\left(\frac{\partial \log p(\xi|\theta)}{\partial \theta}\right)^2 = \sum_{\xi} \frac{1}{p(\xi|\theta)}\left(\frac{\partial p(\xi|\theta)}{\partial \theta}\right)^2 . \tag{2.20}$$

The Fisher information defined above represents the average sensitivity of the log-likelihood function $\log p(\xi|\theta)$ with respect to the variation of θ. Note that $0 \leq \mathcal{I}(\theta) < \infty$, so intuitively speaking, the larger the Fisher information is, the better one can estimate the value of the parameter θ.

An estimator $\hat{\theta}$ of the parameter θ is *unbiased* if its expectation over ξ is equal to the true value, that is,

$$\mathrm{E}(\hat{\theta} - \theta) \equiv \sum_{\xi} p(\xi|\theta)\left(\hat{\theta}(\xi) - \theta\right) = 0 . \tag{2.21}$$

By taking the derivative of Eq. (2.21) with respect to θ and applying the Cauchy-Schwarz inequality (using the fact that $\sum_{\xi} p(\xi|\theta) = 1$), we obtain the well-known *CRLB* [10, 11], which can be expressed as

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{\mathcal{I}(\theta)} . \tag{2.22}$$

This inequality states that the MSE of any unbiased estimator is bounded from below by the inverse of the Fisher information, no matter in how clever a way an estimator is designed. We can also define the efficiency of an unbiased estimator $\hat{\theta}$,

$$e(\hat{\theta}) = \frac{\mathcal{I}(\theta)^{-1}}{\mathrm{Var}(\hat{\theta})} , \tag{2.23}$$

which measures how close the estimator's variance comes to this lower bound. The CRLB thus gives

$$e(\hat{\theta}) \leq 1 . \tag{2.24}$$

An unbiased estimator which achieves this upper bound is said to be fully efficient.

Such a solution achieves the lowest possible MSE among all the unbiased methods, and is therefore called the minimum variance unbiased (MVU) estimator. The importance of CRLB is that it provides the ultimate resolution of estimation.

In the multi-parameter scenario, the Fisher information takes the form of an $N \times N$ matrix, the Fisher information matrix (FIM), with typical element given as

$$\mathcal{I}_{jk}(\theta) = \mathrm{E}\left[\left(\frac{\partial \log p(\xi|\theta)}{\partial \theta_j}\right)\left(\frac{\partial \log p(\xi|\theta)}{\partial \theta_k}\right)\right]. \tag{2.25}$$

and the CRLB for any unbiased estimator can be written as a matrix inequality,

$$\mathrm{C}(\theta) \geq \mathcal{I}^{-1}(\theta)\,, \tag{2.26}$$

where $\mathrm{C}(\theta)$ is the covariance matrix (or the MSE matrix),

$$\mathrm{C}_{jk}(\theta) = \mathrm{E}\left[\left(\hat{\theta}_j - \theta_j\right)\left(\hat{\theta}_k - \theta_k\right)\right]. \tag{2.27}$$

Since the likelihood function is multiplicative, the FIM is additive; that is, the total FIM for several independent measurements is equal to the sum of the FIMs from each measurement. In particular, the FIM for N identical measurements is N times the FIM for one measurement. Therefore, the covariance matrix of any unbiased estimator based on N measurements satisfies $\mathrm{C}^N(\theta) \geq 1/N\mathcal{I}(\theta)$. According to Fisher's theorem, the CRLB can be saturated asymptotically with the ML estimator for a large amount of registered data (detected particles). Therefore, in the large-sample scenario, the scaled covariance matrix $N\mathrm{C}^N(\theta)$ is generally independent of the sample size. It is also denoted by $C(\theta)$ when there is no confusion.

In practice, it is often more convenient to use a single number rather than a matrix to quantify the error. A common choice is the scaled MSE $\mathrm{tr}\{\mathrm{C}(\theta)\}$; a more general alternative is the weighted MSE (WMSE) $\mathrm{tr}\{\mathrm{W}(\theta)\mathrm{C}(\theta)\}$, where $\mathrm{W}(\theta)$ is a positive semidefinite weight matrix depending on θ. The CRLB implies that $\mathrm{tr}\{\mathrm{W}(\theta)\mathrm{C}(\theta)\} \geq \mathrm{tr}\{\mathrm{W}(\theta)\mathcal{I}^{-1}(\theta)\}$; again this bound can be saturated asymptotically by the ML estimator. However, the problem with the MSE is that it depends on the

parametrization, which is somehow arbitrary. The WMSE, on the other hand, can avoid this problem with a suitable choice of the weight matrix $W(\theta)$.

Let us evaluate the overall performance of the ML estimator with the help of CRLB [4]. The state of any quantum system can be decomposed into an orthonormal basis $\{\Gamma_k\}_{k=1}^{d^2-1}$ of traceless Hermitian operators defined on the Hilbert space of the system,

$$\rho(\theta) = \frac{1}{d} + \sum_{k=1}^{d^2-1} \theta_k \Gamma_k \,, \tag{2.28}$$

where d is the dimension of the Hilbert space. Given a measurement with outcomes Π_ξ, the probability of obtaining outcome ξ is again given by the Born rule $p(\xi|\theta) = \mathrm{tr}\{\rho(\theta)\Pi_\xi\}$; and the likelihood function for a specific measurement is $\mathcal{L} = \prod_\xi \Big(\mathrm{tr}\{\rho(\theta)\Pi_\xi\}\Big)^{n_\xi}$. Then by applying Eq. (2.25), the FIM $\mathcal{I}_{jk}(\theta)$ for the unknown state of Eq. (2.28) is calculated as

$$\mathcal{I}_{jk}(\theta) = N \sum_\xi \mathrm{tr}\left\{\Gamma_j \Pi_\xi\right\} \mathrm{tr}\left\{\Gamma_k \Pi_\xi\right\} / p(\xi|\theta) \,, \tag{2.29}$$

where $N = \sum_\xi n_\xi$ denotes the total number of quantum systems registered. In the asymptotic limit of a large amount of accumulated data, the FIM can be transformed into

$$\mathcal{I}' = \mathrm{U}\,\mathcal{I}\,\mathrm{U}^T, \tag{2.30}$$

where $\mathcal{I}'$ becomes diagonal and the unitary tranformation U is composed of the eigenvectors of the original FIM $\mathcal{I}$. Therefore, once FIM is known, the inverse of it sets a lower bound for the covariance matrix of any unbiased estimator, which is saturated asymptotically by the ML estimator.

2.4.1 Jeffreys prior

In Bayesian statistical inference, a prior is the probability distribution that would express one's belief about an unknown quantity before any data is taken into account. Here, the Fisher information is used to calculate the Jeffreys prior (denoted by $\pi(\theta)$

over the parameter θ) [101,102], which is a standard, non-informative prior distribution on the parameter space that is proportional to the square root of the determinant of the Fisher information, such that,

$$\pi(\theta) \propto \sqrt{\det \mathcal{I}(\theta)} \, . \tag{2.31}$$

The key feature of the Jeffreys prior is that it is form-invariant under reparameterization of the parameter θ, which is a valid reason why this non-informative prior is preferred over others. When using the Jeffreys prior, inferences about the unknown parameter θ depend not just on the probability of the observed data (the likelihood function), but also on the universe of all possible experimental outcomes, as determined by the experimental design, because the Fisher information is computed from an expectation over the chosen universe. One problem of the Jeffreys prior is that sometimes it cannot be normalized, thus one must use an improper prior. But in this thesis, we simply exclude pathological cases of improper priors.

Let's take the classical 3-sided die [50] and the qubits confined to an equatorial plane as examples. The die is described by a probability distribution $\{p_k\}_{k=1}^3$, such that $\sum_{k=1}^3 p_k = 1$ and $p_k \geq 0$ for all k. We can visualize the physical states of the die as points on an equilateral triangle (also known as the regular 2-simplex), with vertices corresponding to the states with outcome probabilities $(p_1, p_2, p_3) = (1,0,0), (0,1,0), \text{and } (0,0,1)$ respectively (see Fig. 2.1). In the figure, we also show the physical qubit states, measured by the three-outcome trine measurement, residing on the disk inscribed within the classical equilateral triangle. Points in the triangle outside of the disk correspond to unphysical states, as the outcome probabilities for the trine measurement have to satisfy an additional constraint, that is, $\sum_k p_k^2 \leq 1/2$.

The Jeffreys prior for the classical 3-sided die is given by

$$\pi(p)\, \mathrm{d}p \propto \frac{\mathrm{d}p_1 \, \mathrm{d}p_2 \, \mathrm{d}p_3}{\sqrt{p_1 p_2 p_3}} \, . \tag{2.32}$$

If we perform a reparameterization of the probabilities $(p) = (q^2)$ for each p_k, then the

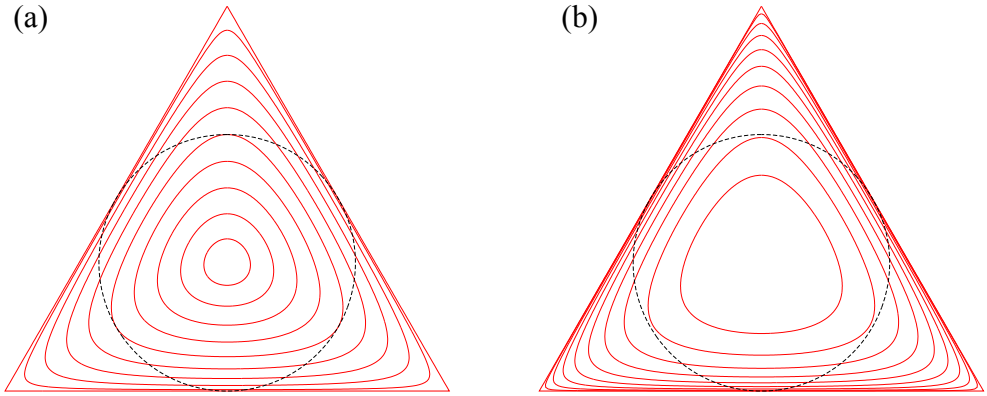


Figure 2.1: Probability distribution for three outcomes by using the Jeffreys prior in the plane. The triangles contain all points (p_1, p_2, p_3) such that $\sum_k p_k = 1$ and $p_k \geq 0$ for all k. The disks contain all points in the triangle that also satisfy the quantum constraint of $\sum_k p_k^2 \leq 1/2$. (a) Contour lines of $S(\alpha)$ with intervals $\alpha/(\pi/6) = 0.1, 0.2, \ldots, 1.0$. (b) Contour lines with $S(\alpha) = 0.1, 0.2, \ldots, 1.0$.

Jeffreys prior in the (p) space transforms into the primitive prior in the (q) space,

$$\pi(q)\,\mathrm{d}q \propto \mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}q_3\,. \tag{2.33}$$

Now, consider the integral of the Jeffreys prior over (p) space, with an additional constraint imposed,

$$S(a) = \frac{1}{2\pi}\int_0^\infty \frac{\mathrm{d}p_1\,\mathrm{d}p_2\,\mathrm{d}p_3}{\sqrt{p_1p_2p_3}}\,\delta(p_1+p_2+p_3-1)\,\eta\big(p_1p_2p_3-a^2\big), \tag{2.34}$$

where the free parameter a is a constant taking values between 0 and $1/\sqrt{27}$, with $S(0) = 1$ and $S(1/\sqrt{27}) = 0$ being the boundary conditions; the symbol $\delta(\cdot)$ denotes Dirac's delta function and $\eta(\cdot)$ is Heaviside's unit step function. The additional constraint $\eta\big(p_1p_2p_3 - a^2\big)$ in Eq. (2.34) has the meaning that the multiplication of the three coordinates for a point (p_1, p_2, p_3) inside the triangle is no less than a^2, which restricts the integration region to a smaller bounded area inside the triangle. Through a reparameterization and transformation of the above integral by using the spherical coordinates, we obtain the following simple form

$$S(\alpha) = \frac{4}{\pi}\int_0^\alpha \mathrm{d}\beta\,\cos(\beta)\,\cos^{-1}\Big(\cos(3\alpha)/\cos(3\beta)\Big), \tag{2.35}$$

where we have $a = \cos(3\alpha)/\sqrt{27}$ and, therefore, $0 \leq \alpha \leq \pi/6$.

Fig. 2.1 shows the probability distribution of the Jeffreys prior for three outcomes in the equilateral plane. In Fig. 2.1(a), ten contour lines of $S(\alpha)$ were plotted with intervals $\alpha/(\pi/6) = 0.1, 0.2, \ldots, 1.0$ respectively; while Fig. 2.1(b) is the plot for ten contour lines with $S(\alpha) = 0.1, 0.2, \ldots, 1.0$ respectively. The difference between these two scenarios gives us a general picture as how the Jeffreys prior affects the probability distributions in the parameter space. We will discuss and use the Jeffreys prior to study the optimal error regions of estimators in Chapter 5.

2.5 Summary

To summarize, this chapter is a general review of the basic ingredients in quantum state tomography, including quantum states and measurements, quantum tomographic methods, Fisher information, and estimation errors. For the tomographic methods, we introduced the simplest linear inversion method, the well-known ML estimation method and several other alternatives to the ML strategy. We also showed the derivation of the Jeffreys prior in Bayesian statistics from the Fisher information. As a popular choice of an unprejudiced prior, we will use the Jeffreys prior in the examples of Chapter 5 to construct error regions.

Chapter 3

Quantum measurements

3.1 Introduction

The framework of quantum mechanics requires a careful definition of measurement (see Ref. [103] for a recent discussion). The issue of measurement lies at the heart of the problem of the interpretation of quantum mechanics, for which there is currently no consensus. Although quantum mechanics has held up to rigorous and thorough experimental testings, many of these experiments are open to different interpretations. There exist a number of contending schools of thought, differing over whether quantum mechanics can be understood to be deterministic, which elements of quantum mechanics can be considered as "real", and many other matters. However, despite the considerable philosophical differences, they almost universally agree on the practical question of what results from a routine quantum-physics laboratory measurement. To describe this, a simple framework to use is what known as the Copenhagen interpretation, the utility of which has been verified countless times, and all the other interpretations (such as the many-worlds interpretation [104]) are necessarily constructed so as to give the same quantitative predictions as this in almost every case.

From a qualitative point of view, the state of a prepared quantum system after measurement is assumed to be an eigenstate of the mathematical operator used to represent that measurement, with the eigenvalue that corresponds to the result of the measurement. Thus, repeated measurements of the same dynamic variable will produce the same result. However, if the preparation of the same system is repeated, subsequent measurements will likely produce different values. By this phenomenon, the measurement process is often said to be random and indeterministic, but there

is considerable dispute over this issue. The expected result of the measurement is in general described by a probability distribution of the measurement outcomes, which is determined by the average or expectation value of the measurement operator over the quantum state of the prepared system.

Postulate 3 of quantum mechanics tells us that all measurements have an associated observable, with the following properties [2]:

i. The observable is a Hermitian (self-adjoint) operator mapping a Hilbert space into itself.

ii. The observable's eigenvectors form an orthonormal basis that span the state space in which that observable exists. Any quantum state can be represented as a superposition of the eigenstates of an observable.

iii. Since Hermitian operators' eigenvalues are real, the possible outcomes of a measurement precisely correspond to the eigenvalues of the observable.

iv. For each eigenvalue there are one or more corresponding eigenvectors. A measurement results in the system being in the eigenstate corresponding to the eigenvalue of the measurement.

Important examples of observables include the Hamiltonian operator $\hat{H}$, the position operator $\hat{x}$, and the momentum operator $\hat{p}$. Two observables commute if and only if there is at least one basis of vectors, each of which is an eigenvector of both operators. Non-commuting observables are said to be incompatible and cannot in general be measured simultaneously, such as the position operator and the momentum operator. In fact, non-commuting observables are related by the Heisenberg uncertainty principle [5, 6], for example, $\Delta(\hat{x})\Delta(\hat{p}) \geq \frac{|\langle[\hat{x},\hat{p}]\rangle|}{2} = \frac{\hbar}{2}$.

In this chapter, however, we are not going to discuss the definition nor interpretation of quantum measurement, but simply give the basic ingredients of it for later use. First, we briefly discuss the two general types of quantum measurement—projective measurement and generalized measurement. Then we introduce the most special case of POM, *i.e.*, SIC POMs as well as the group-covariant SIC POMs. Next we discuss

the MUB and show the derivation of a maximal set of MUB in prime power dimensions. In quantum information theory, both the problems of constructing SIC POMs and a maximal set of MUB are considered to be hard. In the last section of this chapter, we present the scheme of successive measurements, using which several proposals for implementing SIC POMs will be given in the following chapter.

3.2 Projective measurements

The projective measurement (also known as the von Neumann measurement, or simply vNM) is an important special case in the quantum-measurement regime. This measurement scheme, the ancestor of quantum decoherence theory, describes measurements by taking into account the measuring apparatus which is also treated as a quantum object. For many applications of quantum computation and quantum information, we will be concerned primarily with projective measurements [2].

A projective measurement is described by an observable, O, a Hermitian operator on the state space of the system being observed. The observable has a spectral decomposition,

$$O = \sum_k k P_k \,, \tag{3.1}$$

where P_k is the projector onto the eigenspace of O with eigenvalue k. In addition to satisfying the completeness relation, *i.e.*, $\sum_k P_k = I$, the measurement operators also satisfy the condition that P_k are orthogonal projectors, that is, $P_k P_j = \delta_{kj} P_k$. The possible outcomes of the measurement correspond to the eigenvalues, k, of the observable. Upon measuring the state $|\psi\rangle$ of a prepared quantum system, the probability of getting an outcome k is

$$p_k = \langle\psi|P_k|\psi\rangle \,. \tag{3.2}$$

The completeness relation automatically guarantees that the probabilities, *i.e.*, the p_ks, sum to 1. Given the outcome k occurred, the state of the quantum system immediately after the measurement is

$$|\psi_k\rangle = \frac{P_k|\psi\rangle}{\sqrt{p_k}} \,. \tag{3.3}$$

Projective measurements have the unique feature that repeated measurements give the same result. For example, if a measurement gives the result k, then a second identical measurement carried out immediately after gives the same result k with probability 1. Later, we will introduce the scheme of successive measurements and discuss the possible applications of the scheme to construct SIC POMs. The essential idea of successive measurements is to make the intermediate measurements "weak"; in other words, the intermediate measurements should not be vNM, whereas the final measurement is one. By doing so, we make sure that each measurement will yield some new information about the initial quantum system we are interested in. For more discussions on this scheme, see Sec. 3.6 of this chapter as well as the whole Chapter 4.

3.3 Generalized measurements

The quantum measurement postulate, Postulate 3, involves two elements, one being the rule of describing the measurement statistics (the respective probabilities of different possible measurement outcomes) and the other being the rule of describing the post-measurement state of the system. However, for some applications the post-measurement state of the system is of little interest, with the main item of interest being the probabilities of the respective measurement outcomes [2]. In such instances, a more natural and useful type of measurements is the generalized measurement, which is often referred to as probability-operator measurement (POM) or equivalently positive operator-valued measure (POVM). The need for the POM formalism arises from the fact that projective measurements on a larger system, described mathematically by a projection-valued measure (PVM), will act on a subsystem in ways that cannot be described by a PVM on the subsystem alone.

A POM on a quantum system is composed of a set of outcomes. These outcomes are mathematically represented by positive operators Π_k that sum up to the identity, $\Pi_k \geq 0$ with $\sum_k \Pi_k = 1$. The probability of obtaining the outcome k is given by the Born rule: $p_k = \mathrm{tr}\{\rho\Pi_k\}$, where ρ is the pre-measurement statistical operator of the system and $\sum_k p_k = 1$ by the completeness relation. If the kth outcome is found, the

post-measurement statistical operator of the system is given by

$$\rho_k = \frac{1}{p_k} M_k \rho M_k^\dagger \,, \tag{3.4}$$

where M_k is the relevant Kraus operator for the kth outcome, *i.e.*, $\Pi_k = M_k^\dagger M_k$. Note that the decomposition of the Π_ks into the corresponding Kraus operators is not unique; for example, $M_k^\dagger M_k$ is invariant under the unitary transformation $M_k \to U_k M_k$, with different U_ks corresponding in practice to different ways of implementing the measurement. Up to this point, we find that the projective measurement is in fact a special case of POM, with the POM elements being the same as the measurement projectors. However, the repeatability feature of the projective measurement is not generally possessed by a generalized measurement.

3.4 Symmetric informationally complete POMs

A POM is IC if any state of the system is determined completely by the measurement statistics [75, 105, 106]. State tomography strategies infer these probabilities from the data acquired with the aid of the POM. A *symmetric* IC POM (SIC POM) is an IC POM of a particular kind. In a d-dimensional Hilbert space (of kets), it is composed of d^2 outcomes, $\{\Pi_k\}_{k=1}^{d^2}$, which are subnormalized rank-1 projectors onto pure states, $\Pi_k = |\psi_k\rangle\langle\psi_k|/d$, with equal pairwise fidelity [12, 13], such that[1]

$$|\langle\psi_j|\psi_k\rangle|^2 = \frac{d\delta_{jk}+1}{d+1} \,, \quad j,k = 1,2,\ldots,d^2. \tag{3.5}$$

Note that the completeness condition $\sum_{k=1}^{d^2} \Pi_k = 1$ is already implied by the above equation and needs not to be imposed separately. Two SIC POMs are said to be equivalent if there is a unitary operator that maps one SIC POM to the other. As mentioned early, the problem of constructing SIC POMs in any finite dimension is considered to be hard.

[1] One can lift the restriction that the POM outcomes are rank-1 while maintaining the SIC property, but we are not considering this more general situation here.

The high symmetry and high tomographic efficiency of SIC POMs have attracted the attention of many researchers, and a lot of work, both analytical and numerical, has been devoted to the construction of SIC POMs in various dimensions; see, for instance, Refs. [12–18]. Besides, SIC POMs are closely related to many other problems in both physics and mathematics, such as quantum cryptography [19,20], MUB [21–24], t-designs and equiangular lines [12,13], and other foundational studies. Recently, they have also attracted the attention of many experimentalists; for example, qubit SIC POMs [107–109] and qutrit SIC POMs [110] were implemented in the laboratory. In addition, we will propose a novel scheme for realizing HW SIC POMs in any finite dimension by successive measurements [23,24] in the next chapter.

3.4.1 Group-covariant SIC POMs

A group-covariant SIC POM is a measurement which can be generated from a single projector—the *fiducial state*—under the action of a group consisting of unitary operations. Almost all known SIC POMs are covariant with respect to the Heisenberg-Weyl (HW) group (also known as the generalized Pauli group) [12,13,18], except for the set of Hoggar lines (in dimension $8 = 2^3$), which is covariant with respect to the three-qubit Pauli group. Besides the extensive applications they have found in the study of SIC POMs [15,111–113] and MUB [28,114], the HW group and its normalizer—the Clifford group—have also played an important role in quantum information science (see, for example, Ref. [115]). It should be noted that there are different versions of the HW group and, accordingly, different versions of the Clifford group [15,115].

In a d-dimensional Hilbert space, the HW group D_{HW} is composed of d^2 (if one ignores the phase factor) unitary operators, and generated by two operators Z and X

$$Z = \sum_{n=0}^{d-1} |n\rangle\, \omega^n \langle n|\,, \quad X = \sum_{n=0}^{d-1} |n \oplus 1\rangle\langle n|\,, \tag{3.6}$$

where $\omega = \mathrm{e}^{\mathrm{i}2\pi/d}$ is the fundamental dth root of unity and $\oplus$ stands for the sum modulo d. X and Z are the cyclic shift and phase operators respectively, obeying the Weyl

commutation relation

$$ZX = \omega XZ\,, \quad X^d = Z^d = 1\,, \tag{3.7}$$

which determines the HW group up to unitary equivalence and overall phase factors. Their action on the kets $|n\rangle$ of the computational basis is

$$Z|n\rangle = \omega^n|n\rangle\,, \quad X|n\rangle = |n+1\rangle\,. \tag{3.8}$$

All elements of the HW group take on the form

$$D_{k_1,k_2} = \tau^{k_1k_2} X^{k_1} Z^{k_2}\,, \tag{3.9}$$

where $k_1, k_2 = 1, ..., d$ and τ is a primitive dth root of unity when d is odd but a $2d$th root of unity otherwise. These d^2 elements satisfy the following relations [15]:

$$\begin{aligned} D_{\mathbf{k}}^\dagger &= D_{-\mathbf{k}}\,, \\ D_{\mathbf{k}} D_{\mathbf{q}} &= \tau^{\langle \mathbf{k},\mathbf{q}\rangle} D_{\mathbf{k}+\mathbf{q}}\,, \\ D_{\mathbf{k}+d\mathbf{q}} &= \begin{cases} D_{\mathbf{k}} & \text{if } d \text{ is odd}\,, \\ (-1)^{\langle \mathbf{k},\mathbf{q}\rangle} D_{\mathbf{k}} & \text{if } d \text{ is even}\,, \end{cases} \end{aligned} \tag{3.10}$$

where bold face stands for pair of indices, *i.e.*, $\mathbf{k} = (k_1, k_2)$, and $\langle \mathbf{k}, \mathbf{q}\rangle := k_2q_1 - k_1q_2$ is the symplectic form. Note $D_{\mathbf{k}+d\mathbf{q}}$ may differ from $D_{\mathbf{k}}$ by a sign factor if d is even.

According to the definition of a SIC POM, *i.e.*, Eq. (3.5), a fiducial state $|\psi_{\text{fid}}\rangle$ of the HW group in a d-dimensional Hilbert space obeys

$$|\langle\psi_{\text{fid}}|D_{k_1,k_2}|\psi_{\text{fid}}\rangle| = \frac{1}{\sqrt{d+1}} \tag{3.11}$$

for all $(k_1, k_2) \neq (0, 0)$. The d^2 outcomes of the HW SIC POM $\Pi_{k,j}, k, j = 1, ..., d$, take on the following form

$$\Pi_{k,j} = X^k Z^j |\psi_{\text{fid}}\rangle \frac{1}{d} \langle\psi_{\text{fid}}| Z^{j\dagger} X^{k\dagger}\,. \tag{3.12}$$

The fiducial state is chosen such that the Πs satisfy the defining property of a SIC

POM [49] (alternative form of Eq. (3.5), but written using POM outcomes),

$$\mathrm{tr}(\Pi_{k,j}\Pi_{m,n}) = \frac{1}{d^2}\left(\delta_{k,m}\delta_{j,n} + (1-\delta_{k,m}\delta_{j,n})\frac{1}{d+1}\right). \tag{3.13}$$

With the (normalized) fiducial state

$$|\psi_{\mathrm{fid}}\rangle = \sum_{n=0}^{d-1} |n\rangle\alpha_n \tag{3.14}$$

in Eq. (3.12) and applying the transformations Z^j followed by X^k, we obtain

$$\Pi_{k,j} = \frac{1}{d}\sum_{n,m=0}^{d-1} |m\oplus k\rangle\alpha_m\omega^{(m-n)j}\alpha_n^*\langle n\oplus k|\,. \tag{3.15}$$

In Chapter 4, we will show that any HW SIC POM taking the form of Eq. (3.15) can be realized by a two-step measurement scheme: a high-rank diagonal-operator measurement, followed by a projective measurement in the Fourier basis, independent of the result of the first measurement.

Up to now, analytical solutions of HW SIC POMs have been constructed in dimensions 2, 3 [116], 4, 5 [13], 6 [113], 7 [15], 8 [117, 118], 9–15 [18, 118–121], 16 [111], 19 [15], and 24, 28, 31, 35, 37, 43, 48 [18]; numerical solutions with high precision have been found up to dimension 67 [12, 18]. All these results suggest strongly that HW SIC POMs exist in any finite-dimensional Hilbert space, but there is neither a universal recipe for constructing SIC POMs nor a rigorous proof of their existence. What is worse, many basic properties of SIC POMs have remained elusive. Although the equiangular condition looks so simple, its implication is largely a mystery.

When the dimension is a prime power p^k with $k \geq 2$, there is another version of the HW group that is the k-fold tensor product of the usual HW group in prime dimension. This HW group is usually called k-qubit Pauli group when $p = 2$. In dimension 8, the three-qubit Pauli group can generate the set of Hoggar lines (see Sec. 4.6 and Ref. [117]). However, no other multi-qubit Pauli group can generate any SIC POM according to Ref. [122]. The situation is still not clear in the case of odd prime power dimensions.

3.5 Mutually unbiased bases

Mutually unbiased bases (MUB) for quantum degrees of freedom are central to all theoretical investigations and practical explorations of complementary properties. The notion of MUB emerged in the seminal work of Schwinger [123] and it was Ivanović [28] who first explored the idea of applying MUB to the problem of quantum state tomography. The elegant work of Wootters and coworkers [29, 124–126] on MUB has turned it into a cornerstone of quantum information. In addition to playing a vital role in quantum state tomography [28, 29], MUB are also important for many other theoretical studies as well as practical applications, such as the "mean king's problem" [127, 128], quantitative wave-particle duality in multi-path interferometers [129], quantum key distribution [130], quantum teleportation and dense coding [131–133]. See Ref. [114] for a recent review on MUB.

Two orthonormal bases of a Hilbert space are said to be *unbiased* if the transition probability from any state of the first basis to any state of the second basis is independent of the two chosen states. In a finite d-dimensional Hilbert space, the normalized basis states $|a_i\rangle$ and $|b_j\rangle$ of two unbiased bases imply the defining property

$$|\langle a_i|b_j\rangle|^2 = \frac{1}{d} \quad \text{for all } i,j = 1,2,\ldots,d\,. \tag{3.16}$$

Physically speaking, if the system is prepared in a state of the first basis, then all outcomes are equally probable when we conduct a measurement in the second basis. The concept of unbiasedness can be generalized to more than two bases by defining a set of MUB, such that all bases in the set are pairwise unbiased.

Much is known about MUB, but there are also a fair number of important questions that have not been answered in full yet. In a finite d-dimensional Hilbert space, there can be at most $d+1$ MUB, and there exist systematic methods for constructing such a maximal set of MUB if the dimension d is a prime or prime power [28, 29, 114, 134, 135]. In the context of quantum state tomography, a maximal set of MUB is also complete because when we know all the probabilities of transition of a given quantum state

towards the states of the bases of this set—exceptional situations aside, there are $(d+1)(d-1) = d^2 - 1$ independent probabilities—we can reconstruct the statistical operator that characterizes this quantum state; in other words, we can perform full tomography or complete quantum state determination. For other finite dimensions ($N = 6, 10, 12, \dots$), it is still an open problem whether such a maximal set exists or not. Even in the simplest case of dimension six, it remains unknown, although there is quite strong numerical evidence that no more than three MUB exist [113, 136–138]. However, it is always possible to construct a set of at least three MUB in any finite-dimensional space (see Ref. [114] and references therein).

More recently, the problem of the existence of MUB in the infinite-dimensional Hilbert spaces, that is $d \to \infty$, has been addressed. This limit is taken by considering a basic Weyl pair of complementary observables whose eigenbases are conjugated (Fourier transforms of each other) [123]. These conjugated eigenbases are unbiased, and as a manifestation of Bohr's principle of complementarity [139], each Weyl pair is algebraically complete as it suffices for a complete parameterization of the degree of freedom. For infinite-dimensional spaces, different Weyl pairs corresponding to different continuous degrees of freedom can be obtained, and then the maximal set of MUB, since there exist different ways of taking the $d \to \infty$ limit [114, 140].

3.5.1 MUB in prime power dimensions

The construction of maximal sets of MUB in prime power dimensions [114, 135, 141] makes use of the properties of finite fields (see Appendix A). Here we follow Ref. [114] and first introduce the shift operators V_j^i, where the superscript i represents the sets of MUB and the subscript j represents the elements in each set. For dimension $d = p^{\mathrm{M}}$, with p a prime number and $\mathrm{M} \in \mathbb{Z}^+$, we choose the orthonormal set $\{|i\rangle, i = 0, 1, \dots, d-1\}$ as the computational basis and denote the primitive pth root of unity with $\gamma = \mathrm{e}^{\mathrm{i}2\pi/p}$. Then the Fourier transform basis can be defined as

$$|\widetilde{j}\rangle = \frac{1}{\sqrt{d}} \sum_{k=0}^{d-1} |k\rangle \gamma^{\ominus k \odot j} , \tag{3.17}$$

where the field operations $\ominus$ and $\odot$ as well as $\oplus$ below are defined in Appendix A. Clearly, we have $|\langle\widetilde{j}|k\rangle|^2 = 1/d$, meaning that the computational basis and the Fourier transform basis are unbiased (see the scheme in Sec. 4.2 that uses these two bases).

Define the shift operators for the computational basis and the Fourier transform basis respectively as

$$\begin{aligned} V_l^0 &= (V_1^0)^l = \sum_{i=0}^{d-1} |i \oplus l\rangle\langle i| \,, \\ V_0^l &= (V_0^1)^l = \sum_{i=0}^{d-1} |\widetilde{i}\rangle\langle\widetilde{i \oplus l}| \,, \end{aligned} \tag{3.18}$$

where $l = 0, 1, \ldots, d-1$. Immediately, we have the following relations

$$\begin{aligned} &V_l^0|i\rangle = |i \ominus l\rangle \,, \quad V_l^0|\widetilde{i}\rangle = |\widetilde{i}\rangle\gamma^{i\odot l} \,, \\ &V_0^l|\widetilde{i}\rangle = |\widetilde{i \oplus l}\rangle \,, \quad V_0^l|i\rangle = |i\rangle\gamma^{i\odot l} \,. \end{aligned} \tag{3.19}$$

Note that when deriving the relations in Eq. (3.19), we used the equality

$$\sum_{j=0}^{d-1} \gamma^{j\odot i} = d\delta_{i,0} \,, \tag{3.20}$$

which allows us to get a relation between the projector $|i\rangle\langle i|$ with the shift operator V_0^l

$$|i\rangle\langle i| = \frac{1}{d} \sum_{n=0}^{d-1} \left(\gamma^{\ominus i\odot l} V_0^l\right)^n . \tag{3.21}$$

The building blocks of the HW group are obtained through the operator multiplication of the shift operators V_0^j and V_i^0, such that,

$$V_i^j = V_0^j V_i^0 = \gamma^{i\odot j} V_i^0 V_0^j \,, \tag{3.22}$$

with the composition law

$$V_i^j V_k^l = \gamma^{\ominus i\odot l} V_0^j V_0^l V_i^0 V_k^0 = \gamma^{\ominus i\odot l} V_{i\oplus k}^{j\oplus l} \,. \tag{3.23}$$

The orthonormality relation for the HW operators can be derived from the above composition law as

$$\mathrm{tr}\Big\{(V_j^i)^\dagger V_n^m\Big\} = d\delta_{i,m}\delta_{j,n}\,. \tag{3.24}$$

Here, we will not prove that the d^2 orthonormal shift operators give us a maximal set of MUB, but only show the explicit expressions of MUB using V_i^j. Explicitly, the 0th basis is the eigenbasis of V_l^0, namely $|e_i^0\rangle = |\widetilde{i}\rangle$, while the dth MUB is the computational basis $|e_i^d\rangle = |i\rangle$. Generally, the jth state of the ith bases ($i = 0, 1, \ldots, d-1$) can be expressed as

$$|e_j^i\rangle = \frac{1}{\sqrt{d}}\sum_{k=0}^{d-1}|k\rangle\gamma^{\ominus j\odot k}\left(\alpha_{\ominus k}^i\right)^*, \tag{3.25}$$

where α is a complex phase factor, chosen to be symmetric, for instance, $\alpha_l^i = \gamma^{\ominus(i\odot l\odot l)\oslash 2}$ is often used. We can use Eq. (3.25) to verify that these bases are indeed mutually unbiased. It it clear that

$$\langle k|e_j^i\rangle = \frac{1}{\sqrt{d}}\gamma^{\ominus j\odot k}\left(\alpha_{\ominus k}^i\right)^*, \tag{3.26}$$

meaning that the computational basis is unbiased to all the other bases. Generally, we have

$$\langle e_j^i|e_n^m\rangle = \frac{1}{d}\sum_{k=0}^{d-1}\gamma^{\ominus k\odot(n\ominus j)}\alpha_{\ominus k}^i\left(\alpha_{\ominus k}^m\right)^*, \tag{3.27}$$

the square norm of which can be shown to be

$$|\langle e_j^i|e_n^m\rangle|^2 = \frac{1}{d} + \delta_{i,m}\left(\delta_{j,n} - \frac{1}{d}\right). \tag{3.28}$$

Therefore, the set of bases $\{|j\rangle = |e_j^d\rangle, |e_j^i\rangle, i, j = 0, 1, \ldots, d-1\}$ is indeed a maximal set of MUB for dimension $d = p^{\mathrm{M}}$.

3.6 Successive measurements

As discussed in Sec. 3.2, if a system is subjected to a projective measurement, the statistical operator ρ describing the system would collapse to the state space of the

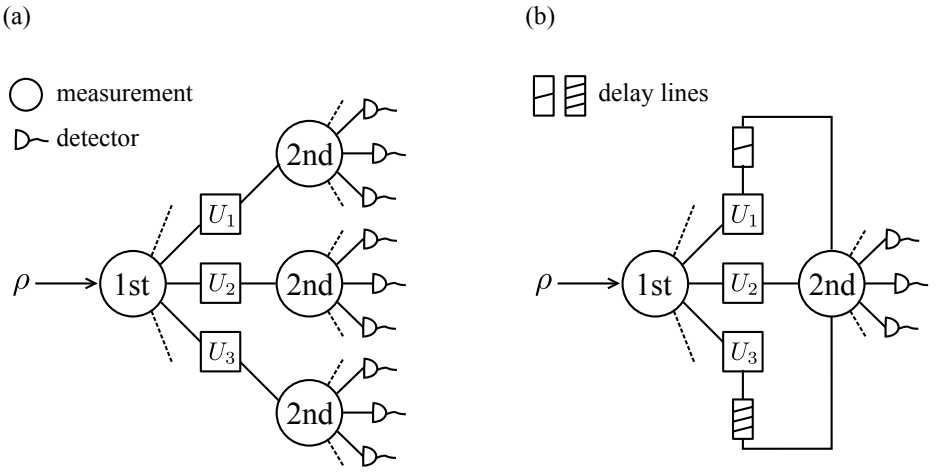


Figure 3.1: A simple sketch for successive measurements. (a) The first measurement is taken to be "weak", and the second measurement is a projective measurement which depends on the actual outcome of the first one. The Us specify the bases for the second measurement. (b) Together with delay lines, the successive nature of the measurement may allow us to use fewer detectors than would have been used otherwise. Here, different sets of outcomes are registered in different temporal domains.

measurement operators. Therefore, a second measurement on the same system would yield no further information, which is not useful for state tomography. Inspired by the idea of "weak measurement" [142], in order to measure the same system many times, the intermediate measurements have to be "weak" followed by a final projective measurement. In this section, we describe the general settings for a sequence of two measurements only, but more complicated settings with more than two successive measurements can be generalized accordingly.

Suppose that a given system is subjected to a sequence of two POMs, the first one has d_1 outcomes $\left\{\mathcal{A}_k = A_k^\dagger A_k\right\}_{k=1}^{d_1}$, followed by a second POM with d_2 outcomes $\left\{\mathcal{B}_j^{(k)} = B_j^{(k)\dagger} B_j^{(k)}\right\}_{j=1}^{d_2}$, where the superscript k indicates that in general the second measurement depends on the actual outcome of the first measurement. Given that the statistical operator for the system prior to the measurements is ρ, then if the nth and mth outcomes for the first and second measurements are found respectively, the post-measurement statistical operator of the system is given by

$$\rho_{n,m} = \frac{B_m^{(n)} A_n \rho A_n^\dagger B_m^{(n)\dagger}}{\mathrm{tr}\left\{\rho A_n^\dagger \mathcal{B}_m^{(n)} A_n\right\}}. \tag{3.29}$$

Following Born's rule, the probability of obtaining the nth and mth outcomes for the first and second measurement is simply given by the denominator of Eq. (3.29) as $p_{n,m} = \mathrm{tr}\Big\{\rho A_n^\dagger \mathcal{B}_m^{(n)} A_n\Big\}$. Accordingly, the two successive measurements are equivalent to a single POM with $d = d_1 d_2$ outcomes that are labeled by a pair of indices $\Pi_{n,m} = A_n^\dagger \mathcal{B}_m^{(n)} A_n$, with $n = 1, \ldots, d_1$ and $m = 1, \ldots, d_2$ respectively. Indeed, summing $\Pi_{n,m}$ over the outcomes labeled by m yields the outcome $\mathcal{A}_n$. Therefore, upon finding the overall outcome $\Pi_{n,m}$, we know that the nth outcome of the first POM and the mth outcome of the second POM are found correspondingly. In Chapter 4, we set $d_1 = d_2 = d$, meaning that the two measurements have the same number of outcomes. As such, we will identify the A_ns and the $\mathcal{B}_m$s such that $\Pi_{n,m}$s make up a SIC POM in the d-dimensional Hilbert space of a qudit by utilizing the structure of Eq. (3.29).

In Fig. (3.1a) we sketch the scheme for successive measurements. We note that the proposed scheme may allow us (depending on the specific experimental realization) to use fewer detectors. Consider the case where the first and the second measurements have the same number of outcomes d. Then, by using delay lines after the first measurement and only d detectors, d^2 outcomes could be registered. Each set of d outcomes is registered in a different time domain as illustrated in Fig. (3.1b). Obviously, such a scheme should use detectors with a rather quick revival time.

3.7 Summary

To conclude, this chapter summarizes the basic ingredients in quantum measurements. We first briefly introduced the two general types of quantum measurement—projective measurement and generalized measurement. Then we had a detailed discussion of SIC POMs and MUB, as well as their constructions in finite dimensions. We emphasize again that the construction of SIC POMs and a complete set of MUB in any finite dimension are both considered to be hard. In the last section, we presented the scheme of successive measurements, using which a few proposals for implementing SIC POMs will be given in the next chapter.

Chapter 4

Symmetric minimal quantum tomography

4.1 Introduction

As discussed in Chapter 3, a lot of work, both analytical and numerical, has been devoted to the construction of SIC POMs in various dimensions due to their high symmetry and high tomographic efficiency; see, for example, Refs. [12–15, 17, 18]. Besides, SIC POMs are closely related to many other problems in both physics and mathematics, such as MUB, equiangular lines, Lie algebras, and so on. Zauner's conjecture [13] states that SIC POMs exist in every finite dimension. While a rigorous proof for this conjecture is still missing, a great deal of numerical evidence suggests strongly that group-covariant SIC POMs indeed exist in all finite-dimensional Hilbert spaces. In this chapter, we consider the implementation of SIC POMs in the Hilbert space of a d-level system by a two-step measurement process: a diagonal-operator measurement with high-rank outcomes, followed by a rank-1 measurement in a basis chosen in accordance with the result of the first measurement. We then proceed to show that any Heisenberg-Weyl (HW) group-covariant SIC POM can be realized by such a sequence where the second measurement is simply a measurement in the Fourier basis, independent of the result of the first measurement.

Nevertheless, in contrast to the major theoretical progress, up to date, all experiments and even proposals for experiments implementing SIC POMs have been limited to the very basic quantum system (qubit) [107, 109], with the exception of the recent experiment by Medendorp *et al.* [110], where a SIC POM for a three-level system was

approximated. This is, in part, due to the fact that there is no systematic procedure for implementing SIC POMs in higher dimensions, in a simple experimental setup.

In Sec. 4.5 of this chapter, we propose an experiment that realizes a SIC POM in the four-dimensional Hilbert space of a qubit pair [23, 24]. The experimental scheme exploits a new approach to SIC POMs that uses a two-step process: a measurement with full-rank outcomes, followed by a projective measurement on a basis that is chosen in accordance with the result of the first measurement. In this work, following the ideas presented in Ref. [23], we explore the possibilities of implementing SIC POMs using a successive-measurement scheme. We start by "breaking" a given SIC POM into two successive measurements, each with d outcomes, with the intention that each measurement will be relatively easy to implement. Unexpectedly, we find that this approach provides a simple, systematic procedure to implement all HW group-covariant SIC POMs. The latter could be realized by first implementing a POM with high-rank outcomes diagonal in a given basis followed by a rank-1 projective measurement, where the basis of the first measurement and the basis of the second measurement are related by the Fourier transform (FT).

Based on this approach, we propose an experimental scheme implementing HW SIC POMs in the Hilbert space of a d-dimensional quantum system (a qudit). In this scheme, the qudit is carried by a single photon as a path qudit, and the implementation is accomplished by means of linear optics (see, for instance, experiments in Refs. [143–145]). In particular, we show that the one-parameter family of nonequivalent HW SIC POMs in dimension 3 could be implemented using the successive-measurement approach in a single experimental setup. Furthermore, we study the construction of the known SIC POMs in dimensions 2 and 8 from two successive measurements. We find that the concept of MUB plays a central role in the construction of SIC POMs in these dimensions—a hint at a possibly profound link between SIC POMs and MUB.

This chapter is organized as follows.[1] Section 4.2 is concerned with the finite-

[1] Note that this chapter is based on Refs. [23, 24], hereby, I sincerely acknowledge the contribution from the other authors of Refs. [23, 24].

dimensional Hilbert spaces. There we discuss the formulation of SIC POMs in general, and the HW SIC POMs in particular, in terms of two successive measurements. Then we study the construction of known SIC POMs in particular dimensions. In Sec. 4.3, we reformulate the SIC POM in dimension 2 (known as the tetrahedron measurement) in terms of successive measurements, and show that the actual implementation of it by Ling *et al.* [107] was indeed carried out using a successive-measurement scheme. We also show how a relation between the SIC POM and MUB in dimension 2 is revealed through this formulation. In Sec. 4.4, we study the decomposition of all known nonequivalent SIC POMs in dimension 3 into two successive measurements. We show that this decomposition allows the implementation of all (known) nonequivalent SIC POMs in dimension 3 with a single experimental setup. In Sec. 4.5, we study the realization of the (known) SIC POMs in dimension 4 by successive measurements. Here we also find an interesting structural and operational relation between MUB and SIC POMs. We briefly describe a proposal for their implementation, using single-photon sources together with passive linear optical elements [23]. In Sec. 4.6, we discuss the construction of the three known, nonequivalent, group-covariant SIC POMs in dimension 8 in terms of successive measurements. We show that the one that is covariant with respect to the three-qubit Pauli group has the same structure as the SIC POMs in the other studied dimensions. Finally, we offer a short summary in Sec. 4.7.

4.2 The general case

In this chapter, we employ the same notations as those used in Chapter 3. The outcomes for a generalized measurement on a quantum system are mathematically represented by a set of positive operators $\{\Pi_k\}$, with $\Pi_k \geq 0$ and $\sum_k \Pi_k = 1$. For the pre-measurement statistical operator ρ of the system, if the kth outcome is found, the post-measurement statistical operator is

$$\rho_k = \frac{1}{p_k} M_k \rho M_k^\dagger \,, \tag{4.1}$$

where $p_k = \mathrm{tr}\{\rho\Pi_k\}$ is the probability of getting the kth outcome and M_k is the

relevant Kraus operator for the kth outcome, such that $\Pi_k = M_k^\dagger M_k$.

We briefly repeat the successive-measurement scheme of Sec. 3.6, but assume now that the two POMs have the same number of outcomes d, such that $\left\{\mathcal{A}_k = A_k^\dagger A_k\right\}_{k=1}^{d}$ followed by $\left\{\mathcal{B}_j^{(k)} = B_j^{(k)\dagger} B_j^{(k)}\right\}_{j=1}^{d}$, where the superscript k indicates that in general the second measurement depends on the actual outcome of the first measurement. Given that the statistical operator for the system prior to the measurements is ρ, then if the nth and mth outcomes for the first and second measurements are found respectively, the post-measurement statistical operator of the system is represented as

$$\rho_{n,m} = \frac{B_m^{(n)} A_n \rho A_n^\dagger B_m^{(n)\dagger}}{\mathrm{tr}\left\{\rho A_n^\dagger \mathcal{B}_m^{(n)} A_n\right\}}\,. \tag{4.2}$$

According to Born's rule, the denominator of the above equation gives the probability of obtaining the nth and mth outcomes for the first and second measurement, that is, $p_{n,m} = \mathrm{tr}\left\{\rho A_n^\dagger \mathcal{B}_m^{(n)} A_n\right\}$. As mentioned before, we may use a single POM (with d^2 outcomes) in the form of $\Pi_{n,m} = A_n^\dagger \mathcal{B}_m^{(n)} A_n$, $n, m = 1, ..., d$, to denote the overall outcome of the two successive measurements. Indeed, summing $\Pi_{n,m}$ over the outcomes labeled by m yields the outcome $\mathcal{A}_n$, and on the other hand, yields the outcome $\mathcal{B}_m$ if summing over n. In what follows, we will identify the A_ns and the $\mathcal{B}_m$s such that $\Pi_{n,m}$s make up a SIC POM in the d-dimensional Hilbert space of a qudit.

4.2.1 HW SIC POMs

Let us begin by showing that *all* SIC POMs which are covariant with respect to the HW group could be realized by a two-step measurement scheme with a rather simple structure—a high-rank, diagonal-operator measurement, followed by a measurement in the Fourier basis.

As discussed in Sec. 3.4.1, a HW SIC POM in a finite d-dimensional Hilbert space has d^2 outcomes $\Pi_{k,j}$, which can be written as

$$\Pi_{k,j} = \frac{1}{d} \sum_{n,m=0}^{d-1} |m \oplus k\rangle \alpha_m \omega^{(m-n)j} \alpha_n^* \langle n \oplus k|\,, \quad k, j = 1, \ldots, d, \tag{4.3}$$

with $\omega = \mathrm{e}^{\mathrm{i}2\pi/d}$ and α_n is related to the fiducial state by

$$|\psi_{\mathrm{fid}}\rangle = \sum_{n=0}^{d-1} |n\rangle \alpha_n \,. \tag{4.4}$$

At this point, we note that the right-hand side of Eq. (4.3) has a two-step measurement structure. The Kraus operators corresponding to the outcomes of the first measurement are given by

$$A_k = \sum_{m=0}^{d-1} |m \oplus k\rangle \alpha_m \langle m \oplus k| \,, \tag{4.5}$$

with $k = 1, ..., d$, while the outcomes of the second measurement are projections onto the eigenstates of the Fourier basis,

$$\mathcal{B}_j = \frac{1}{d} \sum_{m,n=0}^{d-1} |m\rangle \omega^{(m-n)j} \langle n| \,, \tag{4.6}$$

with $j = 1, ..., d$. Indeed, for Eqs. (4.3)–(4.6) we have

$$\Pi_{k,j} = A_k^\dagger \mathcal{B}_j A_k \,, \tag{4.7}$$

so that the HW SIC POM for the fiducial state of Eq. (4.4), when it exists, is realized by a two-step measurement. This demonstrates the case. If we relax the requirement that the sequential measurements in Eq. (4.7) compose a *symmetric* IC POM, one can show [146] that in any finite-dimensional Hilbert space there exist αs such that these measurements are IC.

The mathematical formulation of SIC POMs as a two-step measurement process hints at the possibility for their implementation. Here, we propose an experimental scheme with which any HW SIC POM in a d-dimensional Hilbert space of a qudit could be realized. The qudit is carried by a single photon and is encoded in d spatial alternatives of the photon ("path qudit"). A unitary transformation on the qudit state amounts to sending the photon through a set of beam splitters (BSs) and phase shifters (PSs), similar to the methods presented in Ref. [147].

In this optical setting, the HW SIC POMs are implemented as follows (see Fig. 4.1):

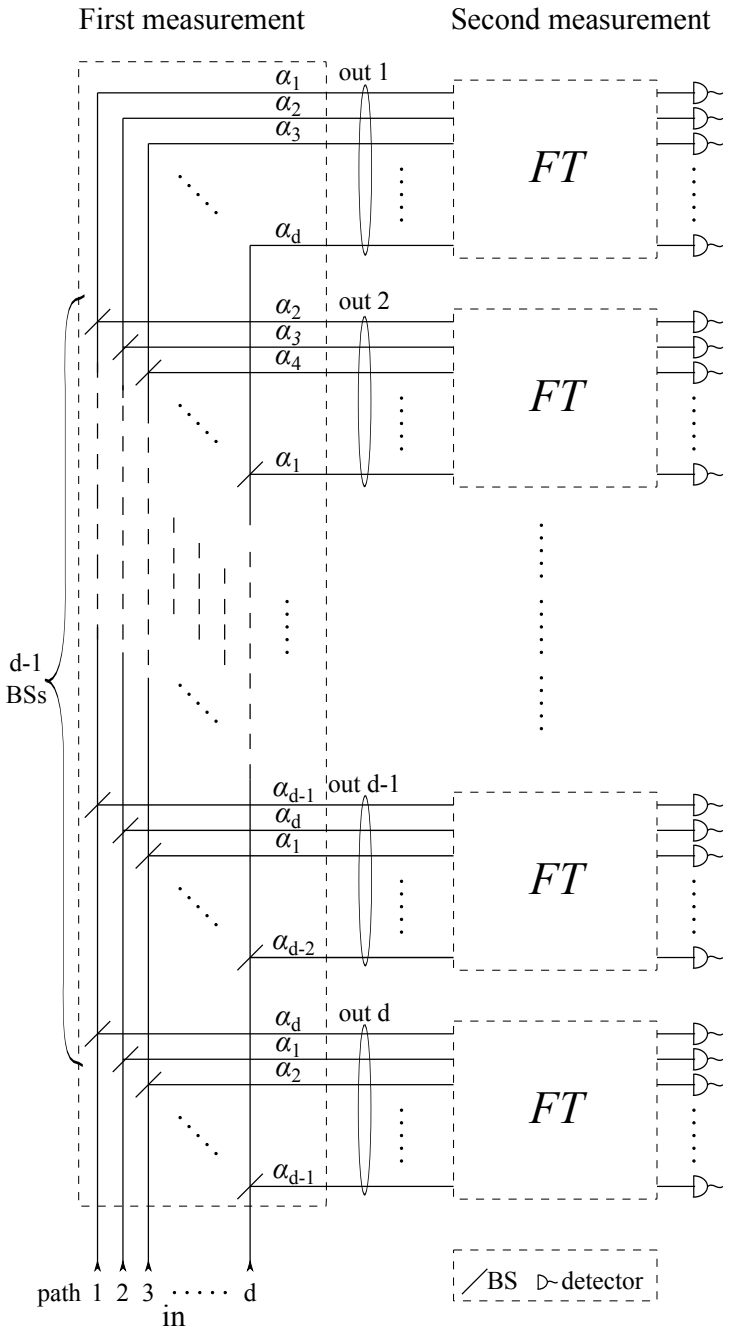


Figure 4.1: An optical implementation of a HW SIC POM using a two-step measurement process.

The first measurement setup is designed to implement the Kraus operators of Eq. (4.5) by appropriately choosing the reflection amplitudes of the $d-1$ BSs at each path. The choice $\prod_{m=1}^{n-1} t_{m,k} \cdot r_{n,k} = \alpha_{k \ominus n}$, where $r_{n,k}$ and $t_{n,k}$ are the reflection and the transmission amplitudes of the nth BS at the kth path (here $n = 1, ..., d-1, k = 1, ..., d$, the BSs

are counted from the entrance port, and the paths are numbered from left to right, as indicated in the figure; we define $t_{0,k} = 1$ for all k), ensures that a photon which enters the apparatus with a path statistical operator ρ exits at port k with the statistical operator $A_k \rho A_k^\dagger \Big/ \mathrm{tr}\Big\{ A_k \rho A_k^\dagger \Big\}$. Upon exiting the first measurement apparatus, the photon is measured in the Fourier basis (indicated in the figure as a black box labeled by FT). This measurement could be realized by a collection of BSs and appropriate PSs [148].

4.2.2 Fuzzy measurements

So far, we considered the decomposition of a given HW SIC POM for a d-level system into a succession of two POMs, each with d outcomes. Now, we would like to follow the reverse path, namely, to start with a given structure for the two POMs, and study under what conditions they compose a SIC POM when measured in succession. In particular, we consider the situation where the first measurement is a "fuzzy measurement", where "fuzzy" means that each of the measurement outcomes corresponds to a projector onto the computational basis $Z_k = |k\rangle\langle k|$, $k = 1, ..., d$, mixed with the identity operator,

$$\mathcal{A}_k = \frac{1}{d}(1 - \lambda) + \lambda Z_k \,, \tag{4.8}$$

whose positivity requires that $-\frac{1}{d-1} \leq \lambda \leq 1$. Up to a unitary transformation, the corresponding Kraus operators for the first measurement are given by

$$A_k = \sqrt{\frac{1-\lambda}{d}} \sum_{j(\neq k)} Z_j + \sqrt{\frac{1+(d-1)\lambda}{d}} Z_k \,. \tag{4.9}$$

For the second measurement, we consider a projective measurement on a basis that is chosen in accordance with the result of the first measurement,

$$\mathcal{B}_j^{(k)} = U_k^\dagger Z_j U_k \equiv |j^{(k)}\rangle\langle j^{(k)}| \,, \tag{4.10}$$

where $|j^{(k)}\rangle$ is the jth state of the kth basis. The unitary operator U_k specifies the basis for the second measurement. It is worth recalling that the outcomes of the first

measurement are invariant under unitary transformations, $A_k \to U_k A_k$. Therefore, we can write down the overall outcomes of the two-step measurement process as

$$\mathcal{M}_{k,j} = A_k^\dagger \mathcal{B}_j^{(k)} A_k \,. \tag{4.11}$$

We now write the necessary and sufficient conditions that the $\mathcal{M}$s in Eq. (4.11) represent a SIC POM, that is, that this ansatz works. Equations (3.13) and (4.9)–(4.11) jointly require that

$$\lambda = \pm\frac{1}{\sqrt{1+d}} \quad \text{and} \quad |\langle m|n^{(m)}\rangle|^2 = \frac{1}{d}\,, \tag{4.12}$$

as well as for $k \neq m$,

$$|\langle n^{(m)}|[\alpha + (\beta-\alpha)(|m\rangle\langle m| + |k\rangle\langle k|)]|j^{(k)}\rangle|^2 = \frac{1}{d}\,, \tag{4.13}$$

with $|n^{(m)}\rangle \equiv U_m^\dagger|n\rangle$, $\alpha = \sqrt{1-\lambda}$, $\beta = \sqrt{1+\lambda(d-1)}$, and all indices take on the values $1, 2, ..., d$. From the condition on λ given right after Eq. (4.8), we get that $\lambda = 1/\sqrt{1+d}$ for $d \geq 4$. Note that the indices k and m label the first measurement while j and n label the second one. Recalling the definition of unbiased bases [114], Eq. (4.12) implies that different bases of the second measurement are unbiased to one of the states from the computational basis.

We are able to solve Eqs. (4.12) and (4.13) for $d = 2$ and 3, and can also show that the known SIC POM in $d = 4$ is a solution for these equations. Unfortunately, we did not manage to solve or prove the existence of a solution for these equations in higher dimensions. In the following sections we discuss the solutions for these equations.

4.3 Dimension 2: A qubit

4.3.1 General construction

We first consider the most general POM in dimension 2 that could be realized by two successive measurements of the following form. We take the first measurement to be a

"fuzzy" projection on the computational basis,

$$A_k^\dagger A_k = \frac{1}{2}(1-\lambda) + \lambda Z_k = \frac{1+\lambda}{2} Z_k + \frac{1-\lambda}{2} Z_{k\oplus 1}\,, \tag{4.14}$$

with $Z_k = |k\rangle\langle k| = \big(1 + (-1)^k \sigma_z\big)/2$, and $k = 0, 1$. The Kraus operators associated (up to a unitary transformation) with these POM elements are given by

$$A_k = \sqrt{\frac{1+\lambda}{2}} Z_k + \sqrt{\frac{1-\lambda}{2}} Z_{k\oplus 1}\,. \tag{4.15}$$

If the second measurement is a projection on the computational basis (modulated by certain unitary transformations), then the overall 4-outcome POM is given by

$$\Pi_{k,j} = \left(\sqrt{\frac{1+\lambda}{2}} Z_k + \sqrt{\frac{1-\lambda}{2}} Z_{k\oplus 1}\right)\Big(U_k^\dagger |j\rangle\langle j| U_k\Big)\left(\sqrt{\frac{1+\lambda}{2}} Z_k + \sqrt{\frac{1-\lambda}{2}} Z_{k\oplus 1}\right), \tag{4.16}$$

with $k, j = 0, 1$. We note that any 2×2 unitary matrix of determinant 1 is of the form

$$U^\dagger = e^{-\mathrm{i}\alpha\sigma_z} e^{-\mathrm{i}\beta\sigma_y} e^{-\mathrm{i}\gamma\sigma_z} \tag{4.17}$$

for some real numbers α, β and γ. Here we can simply take $\gamma = 0$ since the second measurement is diagonal in the z basis. The POM elements are now given by

$$\begin{aligned}\Pi_{k,j} =& \frac{1 + (-1)^{k+j}\lambda\cos(2\beta_k)}{2} \cdot \frac{1}{2} \times \Big(1 + (-1)^k \frac{\lambda + (-1)^{k+j}\cos(2\beta_k)}{1 + (-1)^{k+j}\lambda\cos(2\beta_k)}\,\sigma_z \\ &+ (-1)^j \frac{\sqrt{1-\lambda^2}\sin(2\beta_k)}{1 + (-1)^{k+j}\lambda\cos(2\beta_k)}\big(\cos(2\alpha_k)\,\sigma_x + \sin(2\alpha_k)\,\sigma_y\big)\Big),\end{aligned} \tag{4.18}$$

with $k, j = 0, 1$. One can verify that the Πs correspond to rank-1 projectors. The free parameters λ, α_k and β_k allow us to realize various POMs in two-dimensional Hilbert space with 4 elements.

4.3.2 Tetrahedron measurement

The solution for Eqs. (4.12) and (4.13) in the case of dimension 2 is fairly straightforward since we have $|m\rangle\langle m| + |k\rangle\langle k| = 1$ for $k \neq m$. Accordingly, for $\lambda = \pm 1/\sqrt{3}$,

Eq. (4.13) reads

$$|\langle n^{(m)}|j^{(k)}\rangle|^2 = \frac{1}{2}\,, \tag{4.19}$$

which means that one can realize a SIC POM in dimension 2 by a fuzzy measurement, with the corresponding diagonal Kraus operators

$$\begin{aligned} A_1 &= \mathrm{diag}\left(\sqrt{\frac{1}{2}-\frac{1}{\sqrt{12}}}, \sqrt{\frac{1}{2}+\frac{1}{\sqrt{12}}}\right), \\ A_2 &= \mathrm{diag}\left(\sqrt{\frac{1}{2}+\frac{1}{\sqrt{12}}}, \sqrt{\frac{1}{2}-\frac{1}{\sqrt{12}}}\right), \end{aligned} \tag{4.20}$$

followed by a measurement in a basis which is chosen in accordance with the result of the first measurement. The solution indicates that the two bases of the second measurements must be unbiased to each other and also unbiased to the computational basis. For example, if the As of Eq. (4.20) are diagonal in the σ_3 basis (where the σ_is are the usual Pauli operators), then the two MUB of the second measurements could be the σ_1 basis and the σ_2 basis.

Actually, all SIC POMs for a qubit are unitarily equivalent to the "tetrahedron measurement" (TM), whose outcomes correspond to four vectors that define a tetrahedron (regular 3-simplex) in the Bloch sphere [12,14]. The general form of the SIC POMs for a qubit is

$$\mathcal{T}_{k,j} = \frac{1}{4}\left(1 + (-1)^k\sqrt{\frac{1}{3}}\,\sigma_3 + (-1)^j\sqrt{\frac{2}{3}}\,\sigma_{k+1}\right), \tag{4.21}$$

with $k, j = 0, 1$. The tetrahedron geometry was shown to be the optimal estimation technique when using four-element POMs [14].

As was shown in the previous section, the TM could be realized in a two-step measurement process, for example, by using a setup similar to the one presented in Fig. 4.1. The first measurement is a two-outcome POM given in Eq. (4.8) with $\lambda = 1/\sqrt{3}$ (the negative value of λ yields the "anti-tetrahedron") and in Eq. (4.18) with $\alpha_0 = 0$ and $\beta_0 = \beta_1 = \alpha_1 = \pi/4$, and depending on the actual outcome of the first measurement, the second measurement is a rank-1 projective measurement onto one of two bases which are unbiased to the computational basis and to each other. We could

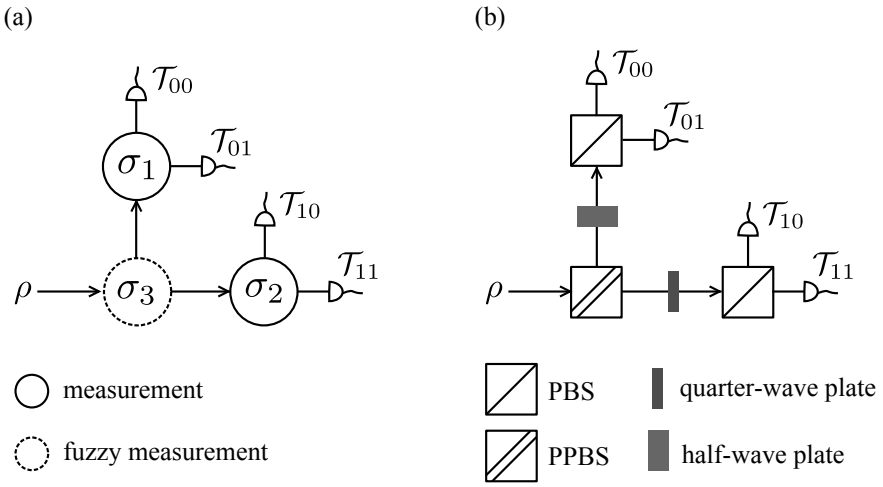


Figure 4.2: An optical implementation of the tetrahedron measurement (polarization qubit) using two successive measurements. (a) The scheme uses a fuzzy measurement of σ_3 followed by either a measurement of σ_1 or of σ_2, depending on the outcome of the first measurement. (b) Optical realization of the tetrahedron measurement that was implemented in Ref. [107]. The fuzzy measurement is realized by a partially polarizing beam splitter and the measurements of σ_1 and σ_2 are realized by the appropriate wave-plates followed by polarizing beam splitters and detectors.

take, for instance, the measurements of σ_1 and σ_2. In Fig. (4.2a), we illustrate the scheme for such a realization.

It is worth noting that the TM was successfully implemented in an optical system [107], where the qubit was encoded in a photon's polarization ("polarization qubit") rather than in a spatial binary alternative (and therefore there was no need to stabilize interferometric loops in the setup). The setup of Ref. [107] also consisted of a sequence of two measurements, quite analogous to what is described above. The fuzzy measurement A_k in the computational basis (horizontal and vertical polarizations) was realized by means of partially polarizing beam splitter (PPBS). The second measurement, *i.e.*, the measurement of σ_1 and σ_2 (depending on whether the photon was transmitted or reflected), was realized by the usual means of wave plates followed by polarizing beam splitters (PBS) and detectors, see Fig. (4.2b). This setup was extended [107], in a straightforward manner, to perform state tomography of many qubits. Each of the qubits passed through the TM. While this POM is IC, it is not symmetric. The specific case of SIC POM in dimension 4 (two-qubit system) will be discussed later.

Actually, with this setup we could use only two detectors for the tomography process, and the four outcomes would be resolved in the time domain by using a delay line between the two arms after the first measurement. This feature could be attractive from experimental point of view.

Here, as a special case of the successive measurement scheme of Fig. 4.1, we present the scheme to realize the TM by using path qubits, see Fig. 4.3. In this setup, the qubit is encoded in a spatial alternative of a single photon ("path qubit"): traveling on the left or on the right. A unitary transformation on the qubit state amounts to sending the photon through a set of beam splitters (BSs) and phase shifters (PSs) [147].

In this optical setting, the TM is implemented as follows: First, two BSs (BS1 and BS2) are used to implement the Kraus operators $A_1 = \mathrm{diag}(t_1, t_2)$ and $A_2 = \mathrm{diag}(r_1, r_2)$, where $\mathrm{diag}(\cdot)$ stands for a diagonal matrix, and t_i and r_i are the transmission and reflection amplitudes of the ith BS. A photon which enters the apparatus with a path statistical operator ρ, exits at port k with the statistical operator $A_k \rho A_k^\dagger \Big/ \mathrm{tr}\Big\{A_k \rho A_k^\dagger\Big\}$. For the values $t_1 = r_2 = \sqrt{\frac{1}{2} - \frac{1}{\sqrt{12}}}$ and $t_2 = r_1 = \sqrt{\frac{1}{2} + \frac{1}{\sqrt{12}}}$, these operators correspond to the measurement outcomes $\mathcal{A}_k = \frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}(-1)^k \sigma_3\right)$ with k=1, 2 (note that the Pauli operator σ_3 is diagonal in the left-right basis). Then a photon that exits the first measurement apparatus at port 1 is measured in the σ_1 basis while a photon that exits at port 2 is a measured in the σ_2 basis. These measurements could be realized by balanced BSs and appropriate PSs, as indicated in the figure.

In the successive measurement construction of POMs for a qubit, there is an operational relation between the TM and the three MUB in dimension 2. The latter are used to construct the former by means of successive measurements. This relation actually stems from the common mathematical structure of the four kets (in the Hilbert space of a qubit) corresponding to the TM and the four kets composing the two bases unbiased to the computational basis and to each other. To see this more clearly, consider the columns of the following matrices:

$$\frac{1}{N}\begin{pmatrix} \chi & \chi \\ 1 & -1 \end{pmatrix}, \quad \frac{1}{N}\begin{pmatrix} \mathrm{i} & -\mathrm{i} \\ \chi & \chi \end{pmatrix}. \tag{4.22}$$

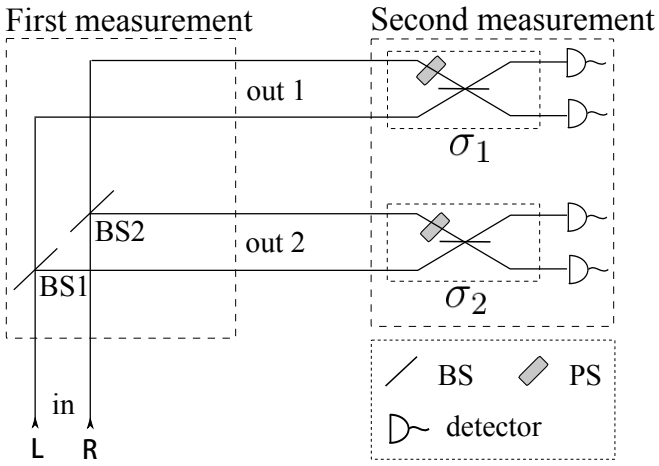


Figure 4.3: An optical implementation of the tetrahedron measurement (path qubit) using two successive measurements.

There are two special cases. For $N = \sqrt{3+\sqrt{3}}$ and $\chi = \sqrt{2+\sqrt{3}}$, the four columns represent the kets corresponding to the TM. While for $N = 1/\sqrt{2}$ and $\chi = 1$, the two columns of each matrix form a basis. The two bases are unbiased to each other and also unbiased to the computational basis. We will see later that similar relations appear in dimensions 3, 4, and 8 as well.

Finally, since the TM is equivalent to a HW SIC POM, it could also be implemented by a two-step process: a measurement with the corresponding Kraus operators of Eq. (4.5) followed by a measurement in the Fourier basis, Eq. (4.6) (call it the σ_1 basis). We note that while the outcomes of the first measurement for this process and the outcomes for the fuzzy measurement in the process discussed in this section are the same, the Kraus operators, and therefore the implementations, are different.

Before moving on to the case of higher dimensions, let us close the present discussion with three remarks: (i) In the above construction, the qubit MUB play a central role: they are used to construct, by means of successive measurements, the SIC POM. We will see later that such a relation appears in other dimensions as well. (ii) A practical implementation of the scheme presented in Fig. 4.3 requires the stabilization of the interferometer loop defined by the four BSs. (iii) SIC POMs for a three-level system could be implemented by using a similar setup, but with allowing the photon to take three different paths, see next section.

4.4 Dimension 3: A qutrit

We focus on constructing all nonequivalent HW SIC POMs for a qutrit using successive measurements. As we shall see, these measurements end up to be a fuzzy measurement in the computational basis, followed by a projective measurement on a basis unbiased to the computational basis.

According to the conditions listed in Eqs. (4.12) and (4.13), the ansatz Eqs. (4.9)–(4.11) yields a SIC POM in dimension 3, if and only if $|\langle m|n^{(m)}\rangle|^2 = 1/3$ and either $\lambda = 1/2$ and

$$|\langle n^{(m)}|(1 + |m\rangle\langle m| + |k\rangle\langle k|)|j^{(k)}\rangle|^2 = 1\,, \tag{4.23}$$

with $k \neq m$, or $\lambda = -1/2$ and

$$|\langle n^{(m)}|l\rangle\langle l|j^{(k)}\rangle|^2 = \frac{1}{9}\,, \tag{4.24}$$

with $k \neq m \neq l$. Whereas Eq. (4.23) does not have a solution, Eq. (4.24) can be solved. One possible solution is $|\langle n^{(m)}|l\rangle|^2 = |\langle l|j^{(k)}\rangle|^2 = 1/3$. This implies that a SIC POM in dimension 3 could be broken into a sequence of a fuzzy measurement with $\lambda = -1/2$, such that

$$A_k^\dagger = \frac{1}{\sqrt{2}}(|k \oplus 1\rangle\langle k \oplus 1| + |k \oplus 2\rangle\langle k \oplus 2|)\,, \tag{4.25}$$

with $k = 1, 2, 3$, followed by a projective measurement onto a basis unbiased to the computational basis (in which the As are diagonal).

Actually, in dimension 3, there exists a one-parameter family of nonequivalent HW SIC POMs [12, 13, 15],

$$\begin{aligned} \Pi_{k,j} &= \frac{1}{9}|\phi_{k,j}(t)\rangle\langle\phi_{k,j}(t)|\,, \\ |\phi_{k,j}(t)\rangle &= \frac{1}{\sqrt{2}}\Big(|k\rangle - e^{\mathrm{i}2t}\omega^j|k \oplus 1\rangle\Big), \end{aligned} \tag{4.26}$$

where $k, j = 1, 2, 3$, $0 \leq t \leq \pi/6$, $\omega = \mathrm{e}^{\mathrm{i}2\pi/3}$, and the symbol $\oplus$ stands for addition modulo 3. This continuum of SIC POMs could be realized using our ansatz in the following way: First, a fuzzy measurement, with the corresponding Kraus operators of

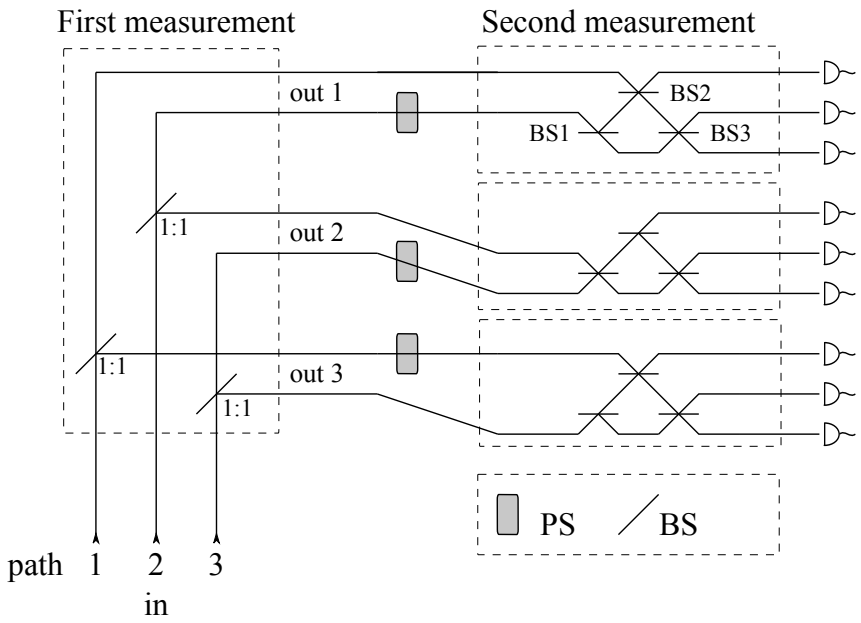


Figure 4.4: An optical implementation of the one-parameter family of nonequivalent SIC POMs for a path qutrit.

Eq. (4.25), is carried out. Then, if the kth outcome is found, the system goes through the following diagonal unitary transformation

$$U_k^\dagger = |k\rangle\langle k| - |k\oplus 1\rangle e^{\mathrm{i}2t}\langle k\oplus 1| + |k\oplus 2\rangle\langle k\oplus 2|\,. \tag{4.27}$$

And last, the system is subjected to a projective measurement onto a basis unbiased to the computational basis, say the Fourier basis; cf. Eq. (4.6) with $d = 3$.

This procedure implements the SIC POMs in Eq. (4.26) for all t. From an operational point of view, this result shows that the entire family of nonequivalent SIC POMs could, in principle, be realized in a single setup. In Fig. 4.4, we present such an implementation in an optical setting for a path qutrit. The balanced $(1:1)$ BSs in the first part are used to implement the fuzzy measurement, then the unitary transformations of Eq. (4.27) are implemented by path dependent PSs placed in the appropriate paths, and finally the Fourier transformation is applied to the state of the qutrit after which the path of the photon is detected. The Fourier transformation is implemented using three BSs, BS1, BS2, and BS3, which implement the transformations by the unitary operators $(\sigma_3+\sigma_1)/\sqrt{2}$, $(\sigma_3+\sqrt{2}\sigma_1)/\sqrt{3}$, and $(\sigma_3+\sigma_2)/\sqrt{2}$, respectively.

In the above construction, we chose the Fourier basis for the second measurement independent of the outcome of the first measurement. However, the same family of SIC POMs (or its unitary equivalent) could be realized with a different choice for the basis for the second measurement, as long as Eq. (4.24) is obeyed. For example, one may use three MUB which are also unbiased to the computational basis,

$$\begin{aligned}
\mathfrak{B}_1 &= \left\{ \begin{array}{l} \frac{1}{\sqrt{3}}(|0\rangle + |1\rangle + |2\rangle) \\ \frac{1}{\sqrt{3}}(|0\rangle + \omega|1\rangle + \omega^2|2\rangle) \\ \frac{1}{\sqrt{3}}(|0\rangle + \omega^2|1\rangle + \omega|2\rangle) \end{array} \right\}, \\
\mathfrak{B}_2 &= \left\{ \begin{array}{l} \frac{1}{\sqrt{3}}(|0\rangle + \omega^2|1\rangle + \omega^2|2\rangle) \\ \frac{1}{\sqrt{3}}(\omega^2|0\rangle + |1\rangle + \omega^2|2\rangle) \\ \frac{1}{\sqrt{3}}(\omega^2|0\rangle + \omega^2|1\rangle + |2\rangle) \end{array} \right\}, \\
\mathfrak{B}_3 &= \left\{ \begin{array}{l} \frac{1}{\sqrt{3}}(|0\rangle + \omega|1\rangle + \omega|2\rangle) \\ \frac{1}{\sqrt{3}}(\omega|0\rangle + |1\rangle + \omega|2\rangle) \\ \frac{1}{\sqrt{3}}(\omega|0\rangle + \omega|1\rangle + |2\rangle) \end{array} \right\}.
\end{aligned} \tag{4.28}$$

These bases can be used for all values of the parameter t.

As mentioned in Sec. 4.1, an experiment was recently proposed by Medendorp *et al.* [110], where a SIC POM for a three-level system was approximated. But for our proposal of Fig. 4.4, the one-parameter family of nonequivalent SIC POMs for a path qutrit can be realized exactly without approximation. As a further note, the necessity of an adequate measurement of qutrits is caused not only by fundamental interest but also by some potential applications. For example, it has been shown that the key distribution in quantum cryptography is associated with the dimensionality of the Hilbert space for the states in use (see, for instance, Refs. [149, 150]). From this point of view only, qutrits are expected to play a more important role than qubits.

4.5 Dimension 4: Two qubits

Higher dimensional systems, like dimensions 4 and 8, offer advantages such as increased security in a range of quantum information protocols, greater channel capacity for quan-

tum communication, novel fundamental tests of quantum mechanics, and more efficient quantum gates [2]. Optically, such systems have been realized using polarization and transverse spatial modes. However, in each case, state transformation techniques have proved difficult to realize. In fact, performing such transformations is a significant problem in a range of physical architectures.

In this section, we propose an experiment that realizes a SIC POM in the four-dimensional Hilbert space of a qubit pair. The qubit pair is carried by a single photon as a polarization qubit and a path qubit. The implementation of the SIC POM is accomplished with the means of linear optics. The experimental scheme exploits our approach to SIC POMs that uses a two-step process: a measurement with full-rank outcomes, followed by a projective measurement on a basis that is chosen in accordance with the result of the first measurement. The basis of the first measurement and the four bases of the second measurements are pairwise unbiased—a hint at a possibly profound link between SIC POMs and mutually unbiased bases.

In dimension 4, there is only one known HW SIC POM, and all the other known SIC POMs are unitarily equivalent to it [15]. Written in a compact form, the HW SIC POM is given as

$$\begin{aligned} \Pi_{k,j} &= \frac{1}{16}\tilde{X}^k\tilde{Z}^j|\phi\rangle\langle\phi|\tilde{Z}^{j\dagger}\tilde{X}^{k\dagger}, \\ |\phi\rangle &= \frac{1}{N}\left(\chi|0\rangle + \sum_{m=1}^{3}|m\rangle\right), \end{aligned} \tag{4.29}$$

with $N = \sqrt{5+\sqrt{5}}$, $\chi = \sqrt{2+\sqrt{5}}$, $k,j = 1,2,3,4$, and the generators of the HW group appear in the form,

$$\tilde{Z} = \mathrm{e}^{\mathrm{i}\frac{\pi}{4}}\begin{pmatrix} 0 & 1 & 0 & 0 \\ -\mathrm{i} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\mathrm{i} \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad \tilde{X} = \mathrm{e}^{\mathrm{i}\frac{\pi}{4}}\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\mathrm{i} & 0 & 0 & 0 \\ 0 & \mathrm{i} & 0 & 0 \end{pmatrix}. \tag{4.30}$$

Thus, this SIC POM is composed of 16 subnormalized projectors onto 16 (fiducial)

kets. The latter are represented in the following matrices as columns [151],

$$\frac{1}{N}\begin{pmatrix} \chi & \chi & \chi & \chi \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}, \quad \frac{1}{N}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ \mathrm{i}\chi & \mathrm{i}\chi & -\mathrm{i}\chi & -\mathrm{i}\chi \\ -\mathrm{i} & \mathrm{i} & \mathrm{i} & -\mathrm{i} \end{pmatrix},$$
$$\frac{1}{N}\begin{pmatrix} 1 & 1 & 1 & 1 \\ \mathrm{i}\chi & -\mathrm{i}\chi & \mathrm{i}\chi & -\mathrm{i}\chi \\ \mathrm{i} & \mathrm{i} & -\mathrm{i} & -\mathrm{i} \\ -1 & 1 & 1 & -1 \end{pmatrix}, \frac{1}{N}\begin{pmatrix} 1 & 1 & 1 & 1 \\ \mathrm{i} & -\mathrm{i} & \mathrm{i} & -\mathrm{i} \\ 1 & 1 & -1 & -1 \\ -\mathrm{i}\chi & \mathrm{i}\chi & \mathrm{i}\chi & -\mathrm{i}\chi \end{pmatrix}. \tag{4.31}$$

These matrices have a unique structure. Each of them could be written as a diagonal matrix times a unitary matrix. The set of bases, corresponding to each unitary matrix together with the computational basis form the complete set of MUB in dimension 4. To be more specific, the diagonal matrices are

$$\begin{aligned} A_1 &= \frac{1}{N}\,\mathrm{diag}(\chi,1,1,1)\,, \quad A_3 = \frac{1}{N}\,\mathrm{diag}(1,1,\chi,1)\,, \\ A_2 &= \frac{1}{N}\,\mathrm{diag}(1,\chi,1,1)\,, \quad A_4 = \frac{1}{N}\,\mathrm{diag}(1,1,1,\chi)\,, \end{aligned} \tag{4.32}$$

and the unitary matrices are

$$\mathcal{U}_1 = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}, \quad \mathcal{U}_3 = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ \mathrm{i} & \mathrm{i} & -\mathrm{i} & -\mathrm{i} \\ -\mathrm{i} & \mathrm{i} & \mathrm{i} & -\mathrm{i} \end{pmatrix},$$
$$\mathcal{U}_2 = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ \mathrm{i} & -\mathrm{i} & \mathrm{i} & -\mathrm{i} \\ \mathrm{i} & \mathrm{i} & -\mathrm{i} & -\mathrm{i} \\ -1 & 1 & 1 & -1 \end{pmatrix}, \quad \mathcal{U}_4 = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ \mathrm{i} & -\mathrm{i} & \mathrm{i} & -\mathrm{i} \\ 1 & 1 & -1 & -1 \\ -\mathrm{i} & \mathrm{i} & \mathrm{i} & -\mathrm{i} \end{pmatrix}. \tag{4.33}$$

Noting that $\sum_j A_j^\dagger A_j = 1$, we identify the As with the Kraus operators of a measurement. Actually, the four operations of Eq. (4.33) transform the computational basis

into the MUB, such that

$$
\begin{aligned}
\mathfrak{B}_1 &= \left\{ \begin{array}{c} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \end{array} \right\} \otimes \left\{ \begin{array}{c} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \end{array} \right\}, \\
\mathfrak{B}_2 &= \left\{ \begin{array}{c} \frac{1}{\sqrt{2}}(|0\rangle + \mathrm{i}|1\rangle) \\ \frac{1}{\sqrt{2}}(|0\rangle - \mathrm{i}|1\rangle) \end{array} \right\} \otimes \left\{ \begin{array}{c} \frac{1}{\sqrt{2}}(|0\rangle + \mathrm{i}|1\rangle) \\ \frac{1}{\sqrt{2}}(|0\rangle - \mathrm{i}|1\rangle) \end{array} \right\}, \\
\mathfrak{B}_3 &= \textsc{cz} \left\{ \begin{array}{c} \frac{1}{\sqrt{2}}(|0\rangle + \mathrm{i}|1\rangle) \\ \frac{1}{\sqrt{2}}(|0\rangle - \mathrm{i}|1\rangle) \end{array} \right\} \otimes \left\{ \begin{array}{c} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \end{array} \right\}, \\
\mathfrak{B}_4 &= \textsc{cz} \left\{ \begin{array}{c} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \end{array} \right\} \otimes \left\{ \begin{array}{c} \frac{1}{\sqrt{2}}(|0\rangle + \mathrm{i}|1\rangle) \\ \frac{1}{\sqrt{2}}(|0\rangle - \mathrm{i}|1\rangle) \end{array} \right\},
\end{aligned}
\tag{4.34}
$$

where CZ stands for the controlled-Z (phase flip) operation, *i.e.*, $\textsc{cz} = \mathrm{diag}(1, 1, 1, -1)$. The bases $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are composed of product states, while the bases $\mathfrak{B}_3$ and $\mathfrak{B}_4$ consist of maximally entangled states.

The structural relation between the fiducial kets, Eq. (4.31), and the kets that compose the four MUB in dimension 4, Eq. (4.33), is now clear. For $N = \sqrt{5 + \sqrt{5}}$ and $\chi = \sqrt{2 + \sqrt{5}}$, the columns of each matrix in Eq. (4.31) form the 16 fiducial kets, while for $N = 1/2$ and $\chi = 1$, the columns of each matrix in Eq. (4.33) form a basis. These bases are mutually unbiased to each other and also unbiased to the computational basis; cf. Eq. (4.34).

The structure of the fiducial vectors in Eq. (4.31) allows us to implement the SIC POM by two successive measurements: a measurement whose Kraus operators are given in Eq. (4.32), and depending on the measurement outcome, a measurement in one of the MUB of Eq. (4.34). We should not fail to mention that the Kraus operators of Eq. (4.32) correspond to a fuzzy measurement with $\lambda = 1/\sqrt{5}$ for $d = 4$; cf. Eq. (4.9). Next, we propose an optical implementation for this scheme.

4.5.1 Experiment proposal

Up to date, the SIC POM in dimension 4 has not been realized in laboratories (partly) due to its complexity. A state tomography of two qubits was realized by using the

SIC POMs (the TM) for a single qubit, *e.g.*, in Ref. [107], or by measurements of the complete set of MUB in dimension 4 as in Ref. [152]. The former is not symmetric and the latter is nor symmetric neither minimal.

Our proposal [23, 24] for implementation is based on the methods presented in Ref. [147] where the two qubits, a polarization qubit and a path qubit, are encoded in a single photon. (We choose here to use a polarization qubit instead of another path qubit in order to avoid as many interferometric loops as possible in the optical setup.) We consider the vertical (v) and horizontal (h) polarizations as the basic alternative of the polarization qubit, and traveling on the left (L) or on the right (R) as the basic alternative of the path qubit. A unitary transformation on the two-qubit state amounts to sending the photon through a set of passive linear optical elements (optical plates) that unitarily change the state of the path and polarization qubits [147]. In particular, in order to realize the fuzzy measurement, two more path-qubits were used as ancillae. With this scheme at hand, the tetrahedron measurement for a (polarization) qubit has already been realized [107] where the path qubit played the role of an ancillary qubit system (a meter).

Let us describe the scheme to realize the SIC POM for a two-qubit system—the polarization and path qubits encoded in a single photon state. For this purpose, we would need the following optical elements:

- Half-wave plate (HWP): A HWP with its major axis at an angle θ to the optical axis, transforms the polarization according to the unitary

$$U_{\mathrm{HWP}}(\theta) = -\mathrm{i}\begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix}. \tag{4.35}$$

  And in our scheme, we only need the HWP with an angle $\theta = \pi/8$, which effects the transition

$$U_{\mathrm{HWP}} = -\frac{\mathrm{i}}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{4.36}$$

- Polarization dependent phase-shifter (PS): It transforms the polarization accord-

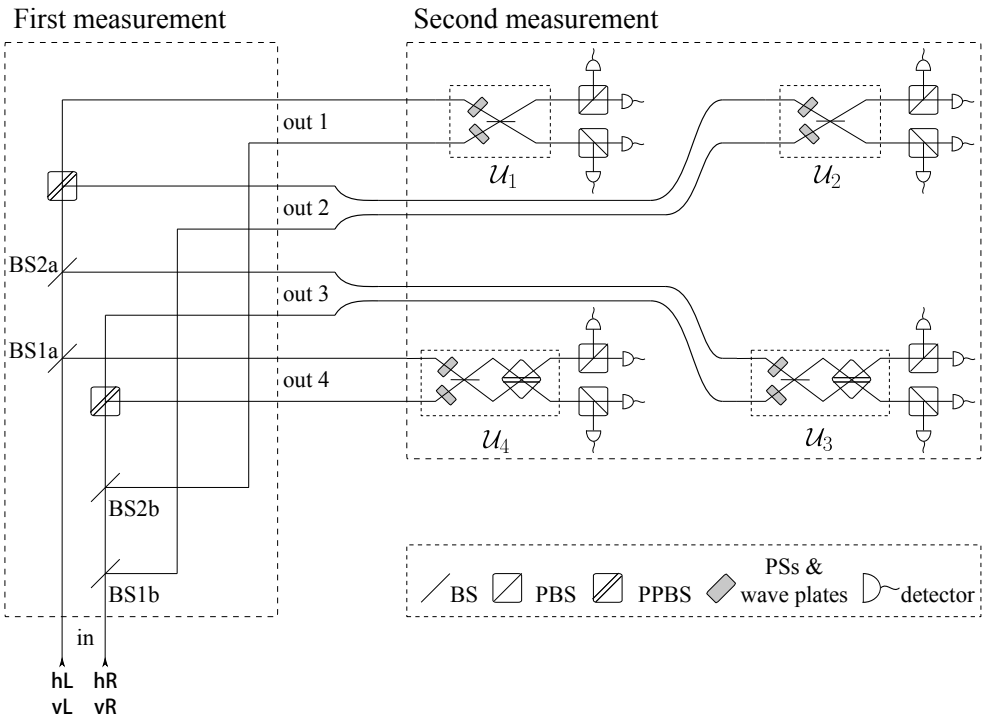


Figure 4.5: A successive-measurement scheme for realizing the SIC POM of a qubit pair. Here the two-qubit state is encoded in the spatial-polarization state of a single photon.

ing to the unitary

$$U_{\rm PS}(\phi) = \begin{pmatrix} 1 & 0 \\ 0 & {\rm e}^{{\rm i}\phi} \end{pmatrix}. \tag{4.37}$$

- Beam splitter (BS): A BS splits the electromagnetic field into two spatial modes with a given reflection and transmission coefficients. Here we need the use of a balanced BS whose action on the state of the path qubit is given by the unitary transformation

$$U_{\rm BS} = \frac{\rm i}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{4.38}$$

  Notice that in the definition of the BS above, we included a global phase factor.

- Path dependent phase-shifter (PS): An interferometric phase shift in the right path amounts to the unitary $U_{\rm PS}$ given above.

- Partially Polarizing BS (PPBS): This is a BS whose reflection and transmission

coefficients depend on the polarization. Its action corresponds to a joint unitary transformation on the polarization-path qubits. In the present context, it suffices to consider a PPBS with real amplitude division coefficients r and t that obey the unitarity condition $r^2 + t^2 = 1$ for the vertical and horizontal polarizations,

$$U_{\text{PPBS}} = \begin{pmatrix} r_v & t_v & 0 & 0 \\ -t_v & r_v & 0 & 0 \\ 0 & 0 & r_h & t_h \\ 0 & 0 & t_h & -r_h \end{pmatrix}. \tag{4.39}$$

This is a block-diagonal matrix, with the blocks transforming the vertical or horizontal polarization, respectively. Two cases of interest are (i) $r_v = r_h = 1$: the controlled-z gate, and (ii) $r_v = t_h = 1$: the polarizing beam splitter (PBS) which totally reflects (transmits) vertically (horizontally) polarized light, written explicitly as

$$U_{\text{PBS}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{4.40}$$

The Kraus operators for the first measurement are listed in Eq. (4.32). Their realization is schematically drawn in Fig. 4.5 at the 'first measurement' part. For each port, we set the parameters of the different optical elements such that a photon which enters the apparatus with a polarization-path statistical operators ρ, exits at port k with the two-qubit statistical operator $A_k \rho A_k^\dagger \Big/ \text{tr}\Big\{ A_k \rho A_k^\dagger \Big\}$. Thus, a projective measurement (with 4 possible outcomes) on the ancillary qubits effectively produces the desired POM on the two-qubit system $\{A_k\}$. To be more specific, the apparatus is configured such that the beam splitters BS1a and BS1b have the same properties and so have beam splitters BS2a and BS2b. The PPBSs on the left and right arms also have the same properties. The reflection coefficient of BS1a and BS1b is $r_1 = 1/N$. The reflection coefficient r_2 of BS2a and BS2b satisfies $t_1 r_2 = 1/N$, that is, $r_2 = 1/\sqrt{N^2 - 1}$, where t_1 is the transmission coefficient of BS1a(b). Setting $r_v = t_h = y$ in Eq. (4.39),

the two PPBSs transform vertically polarized incident light $|\mathsf{v}\rangle$ to the polarizations $y|\mathsf{v}\rangle$ and $\sqrt{1-y^2}|\mathsf{v}\rangle$ in the reflected and transmitted arms, respectively, and horizontally polarized light $|\mathsf{h}\rangle$ to the polarizations $\sqrt{1-y^2}|\mathsf{h}\rangle$ in reflection and $y|\mathsf{h}\rangle$ in transmission. The amplitude division coefficient y is chosen such that $t_1 + t_2 y = 1/N$, and therefore, $y = 1/\sqrt{N^2-2}$, where t_2 is the transmission coefficient of BS2a(b). These settings ensure that the measurement of Eq. (4.32) is realized.

To complete the measurement scheme, a second measurement is taking place. This measurement depends on the actual outcome of the first measurement, namely, on the output port where the photon exits. For photons emerging from the kth port, basis $\mathfrak{B}_k$ of Eq. (4.34) is measured. In order to measure in a given basis, $\mathfrak{B}_k$, we first apply a unitary operation $\mathcal{U}_k$ of Eq. (4.33) that transforms the basis $\mathfrak{B}_k$ into the computational basis and then measure in the computational basis by using PBSs and photodetectors, as illustrated in Fig. 4.5 at the 'second measurement' part.

To implement the unitary transformations of Eq. (4.33), one could use either a single, specially designed, birefringent material, or a sequence of wave plates and PPBSs. Considering the latter option, these unitary transformations can be represented as

$$\begin{aligned}
\mathcal{U}_1 &= U_{\mathrm{HWP}} \otimes U_{\mathrm{BS}}, \\
\mathcal{U}_2 &= (U_{\mathrm{PS}} U_{\mathrm{HWP}}) \otimes (U_{\mathrm{PS}} U_{\mathrm{BS}}), \\
\mathcal{U}_3 &= \mathrm{CZ}\Big((U_{\mathrm{PS}} U_{\mathrm{HWP}}) \otimes U_{\mathrm{BS}} \Big), \\
\mathcal{U}_4 &= \mathrm{CZ}\Big(U_{\mathrm{HWP}} \otimes (U_{\mathrm{PS}} U_{\mathrm{BS}}) \Big),
\end{aligned} \tag{4.41}$$

where CZ is the controlled-Z gate, $U_{\mathrm{PS}} = \mathrm{diag}(1, \mathrm{i})$ shifts the phase of the path and polarization qubits by $\pi/2$, and U_{HWP} and U_{BS} together implement the Hadamard gate,

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \tag{4.42}$$

for the polarization qubit and the path qubit, respectively. For this aim, we use a HWP with its major axis at an angle $\pi/8$ to the optical axis, and a balanced BS.

We see that the unitary transformations $\mathcal{U}_1$ and $\mathcal{U}_2$ can be decomposed into a

tensor product of two unitary transformations, one for the path qubit and one for the polarization qubit. However, the unitary transformations $\mathcal{U}_3$ and $\mathcal{U}_4$ are not of that kind and could be realized, for example, by using PPBSs together with a Mach-Zehnder interferometer. This closes our proposal.

4.6 Dimension 8: Three qubits

In dimension 8, there are three known nonequivalent SIC POMs. One of them is covariant with respect to the three-qubit Pauli group, an alternative version of the HW group [13, 117]; while the other two are covariant with respect to the HW group [18]. According to the result presented in Sec. 4.2, the latter two could be realized by a diagonal-operator measurement followed by a measurement in the Fourier basis of the three qubits. Interestingly, the former SIC POM (also known as Hoggar's SIC POM or Hoggar lines [117]) is the only exception known so far that is not covariant with respect to the HW group. But as we will see in what follows, this SIC POM could be broken into a diagonal-operator measurement followed by projective measurements in eight MUB, similar to what happens in dimensions 2, 3, and 4.

Hoggar's SIC POM is composed of (subnormalized) projectors onto 64 kets. The latter are constructed from the action of the three-qubit Pauli group elements on a fiducial vector $|\phi\rangle$,

$$\left|\begin{smallmatrix}(k,l,m)\\(n,r,s)\end{smallmatrix}\right\rangle = Z_1^n X_1^k \otimes Z_2^r X_2^l \otimes Z_3^s X_3^m |\phi\rangle\,, \tag{4.43}$$

with all indices take on the values 1, 2. Here $Z = \sigma_3$ and $X = \sigma_1$ are the generators of the Pauli group in dimension 2, and their subscripts in Eq. (4.43) label the degree of freedom on which they act. In what follows, we omit this subscript when no ambiguity arises. We refer to the basis in which σ_3 is diagonal as the computational basis. In this basis, the fiducial ket $|\phi\rangle$ is represented by

$$|\phi\rangle = \frac{1}{\sqrt{6}}(r, 0, -\omega_8^*, \omega_8^*, -\omega_8, \omega_8^*, 0, 0)^\intercal\,, \tag{4.44}$$

where $\omega_8 = \mathrm{e}^{\mathrm{i}2\pi/8} = \sqrt{\mathrm{i}}$ is the fundamental eighth root of unity and $r = \omega_8 + \omega_8^* = \sqrt{2}$.

Table 4.1: Hoggar's SIC POM for dimension 8, which is covariant with respect to the three-qubit Pauli group. Matrix of complex 2-vectors (a, b) (denoted by the letters "O, D, S, R") gives the 64 lines, where $\omega_8 = e^{i2\pi/8} = \sqrt{i}$ and $r = \omega_8 + \omega_8^* = \sqrt{2}$.

Row	O	D	S	R	
1	$(0,\ 0)$	$(\omega_8,\ \omega_8^*)$	$(\omega_8,\ -\omega_8)$	$(0,\ r)$	Type 1: ODSR
2	$(0,\ 0)$	$(\omega_8^*,\ -\omega_8)$	$(\omega_8,\ \omega_8)$	$(r,\ 0)$	Type 2: DORS
3	$(0,\ 0)$	$(\omega_8^*,\ \omega_8)$	$(\omega_8^*,\ \omega_8^*)$	$(0,\ ir)$	Type 3: SROD
4	$(0,\ 0)$	$(\omega_8,\ -\omega_8^*)$	$(\omega_8^*,\ -\omega_8^*)$	$(ir,\ 0)$	Type 4: RSDO

This set of 64 SIC POM vectors in dimension 8 can also be read off from Hoggar's paper [117] about the quaternionic polytope. We modified the table by using our familiar notations, which is given in Table 4.1. A given row in the table exhibits a complex 8-vector v, with $|v| = 6$. Then by coordinate sign changes or by inserting $\pm$ signs in each row, we are able to obtain a total number of $4 \times 4 = 16$ vectors for each row and altogether $16 \times 4 = 64$ vectors.

This SIC POM can be broken into two successive measurements. The Kraus operators corresponding to the first measurement are

$$
\begin{aligned}
A_1 &= \frac{2}{\sqrt{3}}\,\mathrm{diag}(0, -ir, \omega_8^*, i\omega_8^*, -\omega_8^*, -i\omega_8, 0, 0)\,,\\
A_2 &= \frac{2}{\sqrt{3}}\,\mathrm{diag}(-\omega_8^*, \omega_8^*, -ir, 0, 0, 0, i\omega_8, i\omega_8^*)\,,\\
A_3 &= \frac{2}{\sqrt{3}}\,\mathrm{diag}(\omega_8^*, i\omega_8^*, 0, r, 0, 0, -i\omega_8^*, \omega_8)\,,\\
A_4 &= \frac{2}{\sqrt{3}}\,\mathrm{diag}(-\omega_8, -\omega_8^*, 0, 0, -ir, 0, -i\omega_8^*, -i\omega_8^*)\,,\\
A_5 &= \frac{2}{\sqrt{3}}\,\mathrm{diag}(-\omega_8^*, i\omega_8, 0, 0, 0, r, -i\omega_8^*, \omega_8^*)\,,\\
A_6 &= \frac{2}{\sqrt{3}}\,\mathrm{diag}(0, 0, i\omega_8, i\omega_8^*, i\omega_8^*, i\omega_8^*, r, 0)\,,\\
A_7 &= \frac{2}{\sqrt{3}}\,\mathrm{diag}(0, 0, i\omega_8^*, \omega_8, -i\omega_8^*, \omega_8^*, 0, ir)\,,\\
A_8 &= \frac{2}{\sqrt{3}}\,\mathrm{diag}(r, 0, -\omega_8^*, \omega_8^*, -\omega_8, -\omega_8^*, 0, 0)\,.
\end{aligned}
\tag{4.45}
$$

The basis for the second measurement is chosen in accordance with the result of the first measurement. The eight different bases for the second measurement, together

with the computational basis, form a complete set of MUB in dimension 8. This set is constructed as follows [153]. Consider the two unbiased bases in dimension 2—the ones that correspond to the Pauli matrices σ_1 and σ_2 denoted as column vectors

$$\mathbf{O} \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{I} \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{pmatrix}. \tag{4.46}$$

Then the full set of MUB vectors (recognized as columns of the matrices) in dimension 8 can be simply constructed as

$$\mathrm{MUB8} = \begin{Bmatrix} \mathbf{OOO} & \mathbf{U{\cdot}OOI} & \mathbf{V{\cdot}OIO} & \mathbf{W{\cdot}OII} \\ \mathbf{III} & \mathbf{U{\cdot}IIO} & \mathbf{V{\cdot}IOI} & \mathbf{W{\cdot}IOO} \end{Bmatrix}, \tag{4.47}$$

where the unitary transformations $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ are the controlled-z (phase flip) operations for three-qubit gate, given by

$$\textsc{cz for 3-qubit} \begin{cases} \mathbf{U} = \mathrm{diag}(1,1,1,1,1,-1,-1,1)\,, \\ \mathbf{V} = \mathrm{diag}(1,1,1,-1,1,-1,1,1)\,, \\ \mathbf{W} = \mathrm{diag}(1,1,1,-1,1,1,-1,1)\,. \end{cases} \tag{4.48}$$

The eight sets of MUB vectors comprise the bases for the second measurement. Then, the full set of 64 SIC POMs in dimension 8 represented as columns are

$$\mathrm{SIC8} = \begin{Bmatrix} A_1.\mathbf{OOO} & A_3.\mathbf{U{\cdot}IIO} & A_5.\mathbf{W{\cdot}IOO} & A_7.\mathbf{V{\cdot}OIO} \\ A_2.\mathbf{U{\cdot}OOI} & A_4.\mathbf{III} & A_6.\mathbf{V{\cdot}IOI} & A_8.\mathbf{V{\cdot}OII} \end{Bmatrix}. \tag{4.49}$$

Or, in a more compact form, we may label the unbiased bases in dimension 2 by $\left\{|e_v^b\rangle\right\}$ where $v = 1, 2$ labels the vector in basis $b = 1, 2$. Furthermore, consider the operator

$$\mathcal{G}(k,l,m) = \frac{1}{2}(1 + \sigma_3^k \otimes \sigma_3^l \otimes \sigma_3^m + \sigma_3^{1-k} \otimes \sigma_3^{1-l} \otimes \sigma_3^{1-m} - \sigma_3 \otimes \sigma_3 \otimes \sigma_3)\,, \tag{4.50}$$

with $k, l, m = 1, 2$. The states defined by a fixed triplet (k, l, m),

$$\left|e^{(k,l,m)}_{(n,r,s)}\right\rangle = \mathcal{G}(k,l,m)|e_n^m\rangle \otimes |e_r^k\rangle \otimes |e_s^l\rangle\,, \tag{4.51}$$

form a basis, while bases with different (k, l, m) triplets are mutually unbiased to each other. These bases together with the computational basis form a complete set of MUB for three qubits (dimension 8). One can verify that the 64 fiducial vectors in Eq. (4.43) could be written as follows

$$\left|{}^{(k,l,m)}_{(n,r,s)}\right\rangle = A_{(k,l,m)} \left|e^{(k,l,m)}_{(n,r,s)}\right\rangle, \tag{4.52}$$

where the index of the Kraus operators is written in binary representation, where m is the least significant bit. Indeed, the last equation implies that Hoggar's SIC POM can be realized by a measurement with the corresponding Kraus operators $A_{(k,l,m)}$ and depending on the result, followed by a measurement in one of the MUB.

Finally, we note that the above construction of the SIC POM in dimension 8 is different in two points from the constructions given for the SIC POMs in dimensions 2, 3, and 4. First, the SIC POMs in dimensions 2, 3, and 4 are covariant with respect to the HW group while the Hoggar's SIC POM is covariant with respect to the three-qubit Pauli group. And second, in dimensions 2, 3, and 4 the first measurement is a fuzzy measurement while in dimension 8 this is not the case.

4.7 Summary

SIC POMs are considered to be hard to implement. Here, we are proposing to implement them by breaking the measurement process into two steps, having in mind that each step should be rather easy to implement. Based on this idea, we presented a systematic procedure that implements HW SIC POMs in finite-dimensional systems. The implementation is accomplished by a diagonal-operator measurement with high-rank outcomes followed by a rank-1 measurement in the Fourier basis. As an example, we have considered the realization of HW SIC POMs for a path qudit encoded in a single photon. Moreover, we found that if we take the first measurement to be a fuzzy measurement and we let the bases for the second measurement to be chosen in accordance with the result of the first measurement, then in the particular studies cases (dimensions 2, 3, and 4) an operational link between SIC POMs and MUB appears, that is,

the MUB are used to implement the SIC POMs in the successive measurement scheme. A similar link was found in dimension 8 as well, but here the first measurement was not of the fuzzy kind.

Moreover, we proposed a feasible experimental scheme that implements the SIC POM for a two-qubit system. Our scheme uses linear optical elements and photodetectors, and is, therefore, well within the reach of existing technology. The proposal is based on a successive-measurement approach to SIC POMs. We found that the SIC POM for the qubit pair corresponds to a POM diagonal in the computational basis, followed by projections onto bases which are mutually unbiased. We observed that this unique construction is owed to a structural relation between the fiducial vectors and the MUB in dimension 4.

On a more general note, we believe that it would be interesting to learn, if and how this scheme can be generalized to higher dimensions. Such a study could be of a theoretical and a practical use; it might teach us about the SIC POMs' structure in high dimensions and provide new ideas for implementing them.

There is still an open question as to the generality of such a relation and its origin. Currently it is unclear whether the successive measurement approach will provide a reasonable scheme for implementing SIC POMs in arbitrary dimensions and thus reveal their structure in high-dimensional Hilbert spaces.

Chapter 5

Optimal error regions of estimators

5.1 Introduction

Quantum state estimation (see, for example, Chapter 2 of this thesis and Ref. [4]) is central to many, if not all, tasks that process quantum information. In the typical situation that we are considering, a source emits several independently and identically prepared quantum-information carriers, which are measured one-by-one by an apparatus that realizes a probability-operator measurement (POM), suitably designed to extract the wanted information. The POM has a number of outcomes, with detectors that register individual information carriers (photons in the majority of current experiments), and the data consist of the observed sequence of detection events ("clicks").[1]

The quantum state to be estimated is described by a statistical operator, the *state*, and the data can be used to determine an *estimator* for the state—another state that, so one hopes, approximates the actual state well. There are various strategies for finding such an estimator. Thanks to the efficient methods that Hradil, Řeháček, and their collaborators developed for calculating maximum-likelihood estimators (MLEs, reviewed in Chapter 3 of Ref. [4]; see also Ref. [97] and Chapter 2 of current thesis), MLEs have become the estimators of choice. For the given data, the MLE is the state for which the data are more likely than for any other state.

Whether one prefers the MLE or a point estimator found by another method, the data have statistical noise and, therefore, one needs to supplement the point estimator with error bars of some sort—*error regions*, more generally, for higher-dimensional

[1] It is advisable to verify that the observed sequence does not have systematic correlations that speak against the assumption of independently and identically prepared quantum-information carriers.

problems. Many recipes, often ad-hoc in nature, have been proposed for attaching a vicinity of states to an estimator. These usually rely on having a lot of data (see Refs. [154] and [155] for examples in quantum state estimation), involve data resampling (see, for instance, Ref. [156]), or consider all data that one might have observed (see Refs. [157, 158], and Sec. 5.3.4 on confidence regions of current chapter). By contrast, we systematically construct error regions from the data *actually* observed [25].

For this purpose, we propose *maximum-likelihood regions* (MLRs) and *smallest credible regions* (SCRs). These are regions in the space of quantum states (more precisely: in the reconstruction space; see Sec. 5.2.1). The MLR is that region of pre-chosen size, for which the given data are more likely than for any other region of the same size. The SCR is the smallest region with pre-chosen credibility—the credibility of a region being its posterior probability, that is: the probability of finding the actual state in the region, conditioned on the data (see, for example, Ref. [159]). Whether one chooses the MLR or the SCR as the optimal error region depends on the situation at hand.

Central to both concepts is the notion of the *size* of a region. In fact, some notion of size must underlie *any* useful definition of error regions, since one usually aims at reporting an error region that is not unnecessarily large—a judgement that can only be made with a suitable concept of size. We agree with Evans, Guttman, and Swartz [160] that, in the context of state estimation, it is most natural to measure the size of a region by its prior—before any data are at hand—probability of finding the actual state in the region: Regions with the same prior probability are considered as having the same size. The size of a region hence expresses the relative importance of that region of states.

The identification "size ≡ prior probability" is also technically possible because both quantities simply add when disjoint regions are combined into a single region. While for some tasks one prefers not to assign a prior,[2] since state estimation expresses our best attempt at guessing the state, any prior information we possess should be taken into account in the estimation process, alongside the data. Much guidance on choosing

[2]For tasks like quantum key distribution, one may want to adopt a different attitude, and assume the worst possible scenario, rather than relying on one's information to assign a prior. Then, the confidence regions of Refs. [157] and [158] are appropriate as error regions.

priors can be found in standard statistics literature; in Sec. 5.4, we provide a summary that focuses on points relevant in quantum contexts. Ultimately, the choice of prior is up to the user, but it should be *consistent*: The estimation results should be dominated by the data, not the prior, if many copies of the state are measured.

As we show below, the problems of finding the MLR and the SCR are duals of each other. In both cases, the optimal regions contain all states for which the likelihood of the data exceeds a threshold value. This provides a simple and concise way of communicating one's uncertainty of the estimate. That the optimal error regions possess such a simple description is surprising, since our construction imposes no restriction on the shape of the regions to be considered. The shape of the optimal regions are uniquely determined by the likelihood function, in sharp contrast to the arbitrariness in the shape of a confidence region (see Sec. 5.3.4), a concept that is the subject of recent discussion [157, 158]. Yet the two are not unrelated: Our SCRs provide natural starting points for the construction of the confidence regions considered in Ref. [157].

While the chosen MLR or SCR depends on the prior, the set of candidate regions is prior-independent: It depends only on the likelihood function for the given data. Also reassuring is the fact that every MLR or SCR is a small vicinity of the MLE, in the respective limits of small size or small credibility. This is reminiscent of standard ellipsoidal error regions constructed around the MLE, but which are applicable only in the limit of a large amount of data when the central limit theorem can be invoked and the uncertainty can be characterized by the Fisher information [154].

Here is a brief outline of this chapter.[3] We set the stage in Sec. 5.2 where we introduce the reconstruction space, discuss the size of a region, and define the various joint and conditional probabilities. Equipped with these tools, we then formulate in Secs. 5.3.1 and 5.3.2 the optimization problems that identify the MLRs and SCRs and find their solutions in Sec. 5.3.3; this is followed by remarks on confidence regions in Sec. 5.3.4. Criteria for choosing unprejudiced priors are the subject of Sec. 5.4. We

[3] Note that this chapter is based on Ref. [25], hereby, I sincerely acknowledge the contribution from the other authors of Ref. [25].

illustrate the matter by several examples in Sec. 5.5, and close with an outlook.

5.2 Setting the stage

5.2.1 Reconstruction space

The K outcomes Π_1, Π_2, ..., Π_K of the POM, with which the data are acquired, are positive Hilbert-space operators that decompose the identity,

$$\sum_{k=1}^{K} \Pi_k = 1 \quad \text{with } \Pi_k \geq 0 \text{ for } k = 1, 2, \ldots, K. \tag{5.1}$$

If the state ρ (a statistical operator) describes the system, then the probability p_k that the kth detector will click for the next copy to be measured is

$$p_k = \mathrm{tr}\{\rho\Pi_k\} = \langle \Pi_k \rangle \,, \tag{5.2}$$

which is the Born rule, of course. Here, the state ρ can be any positive operator with unit trace,

$$\rho \geq 0 \,, \quad \mathrm{tr}\{\rho\} = 1 \,. \tag{5.3}$$

The positivity of ρ and its normalization to unity ensure the positivity of the p_ks and their normalization

$$p_k \geq 0 \,, \quad \sum_{k=1}^{K} p_k = 1 \,. \tag{5.4}$$

Probabilities $p = (p_1, p_2, \ldots, p_K)$ for which there is a state ρ such that Eq. (5.2) holds, are *permissible* probabilities. They make up the *probability space*.

The probability space for a K-outcome POM is usually smaller than that of a tossed K-sided die because not all positive p_ks with unit sum are permissible. The quantum nature of the state estimation problem enters *only* in these additional restrictions on p: Quantum state estimation is standard statistical state estimation with constraints of quantum-mechanical origin. The rich methods of statistical inference immediately apply, modified where necessary to account for the restricted probability space.

Whereas p is uniquely determined by ρ in accordance with Eq. (5.2), the converse is true only if the POM is informationally complete (IC). In any case, there is always a *reconstruction space* $\mathcal{R}_0$, a set of ρs that contains exactly one ρ for each permissible p, consistent with the Born rule. If there is more than one reconstruction space, it does not matter which one we choose. As an example, consider a harmonic oscillator with its infinite-dimensional state space. If the POM has two outcomes with p_1 equal to the probability of finding the oscillator in its ground state, and $p_2 = 1 - p_1$, the reconstruction space is the set of convex combinations of the projector to the ground state and another state with no ground-state component. In this situation, there are very many reconstruction spaces to choose from, because *any* other state serves the purpose, and all one can infer from the data is an estimate of the ground-state probability.

Since the probability space is unique, while there can be many different reconstruction spaces, it is often more convenient to work in the probability space. In particular, the probability space has the desirable property that it is always convex; it is, however, not always possible to find a convex reconstruction space. The primary objective of state estimation is then to find an estimator, or a region of estimators, for the probabilities p. The conversion of p into a state ρ can be performed later, if at all. At this stage, if the POM is not IC, one must invoke additional criteria—beyond what the data tell us—for a unique mapping $p \to \rho$. For example, one could follow Jaynes's guidance [32,33] and maximize the entropy [96] (see also Chapter 6 of Ref. [4]). Following the tradition in this topic, however, we will formally work in a reconstruction space $\mathcal{R}_0$ although all actual calculations are performed in the probability space. Estimators are states in $\mathcal{R}_0$, and regions are sets of states there.

5.2.2 Size and prior content of a region

Prior to acquiring any data, we assign equal probabilities to equivalent alternatives. For instance, if we split the reconstruction space in two, it is equally likely that the actual state is in either half and, therefore, each half should carry a prior probability of $1/2$, provided that the splitting-in-two is fair, that is: the two pieces are of equal size.

A preconceived notion of size is taken for granted here. Further fair splitting, into more disjoint regions of equal size, then suggests rather strongly that the prior probability of a region should be proportional to its size. We take this suggestion seriously: Scale all region sizes such that the whole reconstruction space has unit size because the actual state is surely somewhere in the state space, and then the size of a region *is* its prior probability—its "prior content" if we borrow terminology from Bayesian statistics.

As mentioned already in Sec. 5.1, it is technically possible to identify the size of a region with its prior probability, because both quantities simply add if disjoint regions are combined into a single region. There is no room for mathematical inconsistencies here, unless we begin with a region-to-size mapping for which the reconstruction space cannot be normalized to unit size, so that we would obtain improper prior probabilities. We are not interested in pathological cases of this or other kinds and just exclude them. Should an improper prior be useful in a particular context, it should come about as the limit of a well-defined sequence of proper priors.

The above line of reasoning can be reversed. Should we have established each region's prior probability with other means (perhaps invoking symmetry arguments or taking into account that the source under investigation is designed to emit the information carriers in a certain target state; see Sec. 5.4), then we accept this as the natural measure of the region's size [160]. After all, the reconstruction space is an abstract construct that is often not endowed with a self-suggesting unique metric. Instead, a region's prior probability—the quantity that matters most in the present context of statistical inference—offers a natural notion of size. This relieves us of the need to invoke additional, possibly artificial, criteria for the assignment of size, for instance, one that has more to do with a simple parameterization of the state space than the relative importance of different regions in terms of our prior expectations.

We denote by $(\mathrm{d}\rho)$ the size of the infinitesimal vicinity of state ρ in $\mathcal{R}_0$. The size $S_\mathcal{R}$ of a region $\mathcal{R} \subseteq \mathcal{R}_0$ is then obtained by integrating over the region,

$$S_\mathcal{R} = \int_\mathcal{R} (\mathrm{d}\rho) \quad \text{with} \int_{\mathcal{R}_0} (\mathrm{d}\rho) = 1\,, \tag{5.5}$$

where the latter integration covers all of the reconstruction space, *i.e.*, $\mathcal{R}_0$.

By construction, $S_{\mathcal{R}}$ does not depend on the parameterization that we use for the numerical representation of $(\mathrm{d}\rho)$. The primary parameterization is in terms of the probabilities,

$$(\mathrm{d}\rho) = (\mathrm{d}p)\, w(p) \quad \text{with } (\mathrm{d}p) = \mathrm{d}p_1\, \mathrm{d}p_2 \cdots \mathrm{d}p_K, \tag{5.6}$$

where the prior density $w(p)$ is nonzero for all permissible probabilities and vanishes for all non-permissible ones. In particular, $w(p)$ always contains

$$w_0(p) = \eta(p_1)\eta(p_2)\cdots\eta(p_K)\,\delta\left(\sum_k p_k - 1\right) \tag{5.7}$$

as a factor and so enforces the constraints that the probabilities are positive and have unit sum, where the symbol $\eta(\cdot)$ denotes Heaviside's unit step function and $\delta(\cdot)$ is Dirac's delta function. If there are no other constraints, we have the probability space of a K-sided die. For genuine quantum measurements, however, there are additional constraints, some accounted for by more delta-function factors, others by step functions. The delta-function constraints reduce the dimension of the reconstruction space from $K-1$ to the number of independent probabilities. Accordingly, there is a factor of constraint $w_{\mathrm{cstr}}(p)$ [containing $\omega_0(p)$] that specifies the probability space and appears in all possible priors. In particular, there are two specific priors we will employ as examples below: the *primitive prior*

$$(\mathrm{d}\rho) \propto (\mathrm{d}p)\, w_{\mathrm{cstr}}(p)\,, \tag{5.8}$$

and the *Jeffreys prior* (see, for instance, Ref. [161] and Sec. 2.4.1 of Chapter 2)

$$(\mathrm{d}\rho) \propto (\mathrm{d}p)\, w_{\mathrm{cstr}}(p)\, \frac{1}{\sqrt{p_1 p_2 \cdots p_K}}\,, \tag{5.9}$$

which is a popular choice of an unprejudiced prior [162].

For the harmonic-oscillator example in Sec. 5.2.1, which has the same probability space as a tossed coin, the factor $w_0(p)$ selects the line segment with $0 \le p_1 = 1 - p_2 \le 1$

in the p_1p_2 plane. If we choose the primitive prior $(\mathrm{d}\rho) = (\mathrm{d}p)\, w_0(p)$, the subsegment with $a \leq p_1 \leq b$ has size $b - a$. For the Jeffreys prior

$$(\mathrm{d}\rho) = (\mathrm{d}p)\, w_0(p) \frac{1}{\pi\sqrt{p_1 p_2}}\,, \tag{5.10}$$

the same subsegment has size $\frac{2}{\pi}[\sin^{-1}(\sqrt{b}) - \sin^{-1}(\sqrt{a})]$.

In this example, and also in those we use for illustration in Secs. 5.5.1 and 5.5.2 below, it is easy to state quite explicitly the restrictions on the set of permissible probabilities that follow from the Born rule; in other situations, including the examples of Sec. 5.5.3, this is more difficult; in yet more complicated situations it could be impossible. It is, however, possible to check numerically if a certain $\tilde{p} = (\tilde{p}_1, \ldots, \tilde{p}_K)$ is permissible. For example, one calculates a MLE (which can be done efficiently) for relative frequencies $n_k/N = \tilde{p}_k$ (see below), and if the resulting probabilities p are such that $p = \tilde{p}$, then $\tilde{p}$ is permissible; otherwise it is not. This is also why state estimation is often done by searching for a statistical operator in a suitable state space. For practical reasons, it may be necessary to truncate the full state space—which can be, and often is, infinite-dimensional—to a test space of manageable size. With such a truncation, one accepts that not all permissible probabilities are investigated. Therefore, a criterion for judging if the test space is large enough is to verify that the estimated probabilities do not change significantly when the space is enlarged. Examples for the artifacts that result from test spaces that are too small can be found in Ref. [163].

5.2.3 Point likelihood, region likelihood, credibility

The data D consist of a sequence of detector clicks, with n_k clicks in total of the kth detector after measuring $N = n_1 + n_2 + \cdots + n_K$ copies of the state.[4] The probability of obtaining D, if ρ is the state, is the familiar *point likelihood*

$$L(D|\rho) = p_1^{n_1} p_2^{n_2} \cdots p_K^{n_K}\,. \tag{5.11}$$

[4] One can account for detector inefficiencies and dark counts, but such technical details, important for practical applications, are immaterial to the current discussion.

It attains its maximal value when ρ is the MLE $\widehat{\rho}_{\mathrm{ML}}$,

$$\max_{\rho} L(D|\rho) = L(D|\widehat{\rho}_{\mathrm{ML}})\,, \tag{5.12}$$

where $\widehat{\rho}_{\mathrm{ML}}$ is in the reconstruction space $\mathcal{R}_0$, but the maximum could be taken over all states. As we can see, the MLE is fully determined by the relative frequencies n_k/N.

The joint probability of finding the state ρ in the region $\mathcal{R}$ and obtaining the data D is then

$$\mathrm{prob}(D \wedge \mathcal{R}) = \int_{\mathcal{R}} (\mathrm{d}\rho)\, L(D|\rho)\,. \tag{5.13}$$

If $\mathcal{R} = \mathcal{R}_0$, we have the *prior likelihood* $L(D)$,

$$\mathrm{prob}(D \wedge \mathcal{R}_0) = L(D) = \int_{\mathcal{R}_0} (\mathrm{d}\rho)\, L(D|\rho)\,. \tag{5.14}$$

Since one of the click sequences is surely observed, the likelihoods of Eqs. (5.11) and (5.14) have unit sum,

$$\begin{aligned} \sum_D L(D|\rho) &= \sum_{n_1,\ldots,n_K} \frac{N!\,\delta_{N,n_1+n_2+\cdots+n_K}}{n_1!\,n_2!\,\cdots\,n_K!} p_1^{n_1} p_2^{n_2} \cdots p_K^{n_K} \\ &= (p_1 + p_2 + \cdots + p_K)^N = 1\,, \\ \sum_D L(D) &= \int_{\mathcal{R}_0} (\mathrm{d}\rho) = 1\,, \end{aligned} \tag{5.15}$$

where the summation is taken over all possible data for N clicks and the multinomial factor is the number of sequences with the same counts of detector clicks.

We factor the joint probability $\mathrm{prob}(D \wedge \mathcal{R})$ in two different ways,

$$\mathrm{prob}(D \wedge \mathcal{R}) = L(D|\mathcal{R}) S_{\mathcal{R}} = C_{\mathcal{R}}(D) L(D)\,, \tag{5.16}$$

and so identify the *region likelihood* $L(D|\mathcal{R})$ and the *credibility* $C_{\mathcal{R}}(D)$. Both quantities are conditional probabilities: $L(D|\mathcal{R})$ is the probability of obtaining the data D if the actual state is in the region $\mathcal{R}$; $C_{\mathcal{R}}(D)$ is the probability that the actual state is in the region $\mathcal{R}$ if the data D were obtained—the posterior probability of $\mathcal{R}$.

5.3 Optimal error regions

5.3.1 Maximum-likelihood regions

Instead of looking for the MLE, the single point in the reconstruction space that has the largest likelihood for the given data D, we desire a region with the largest likelihood—the MLR. For this purpose, we maximize the region likelihood $L(D|\mathcal{R})$ under the constraint that only regions with a pre-chosen size s participate in the competition, with $0 < s < 1$; an unconstrained maximization of $L(D|\mathcal{R})$ is not meaningful because it gives the limiting region that consists of nothing but the point $\widehat{\rho}_{\text{ML}}$. The resulting MLR $\widehat{\mathcal{R}}_{\text{ML}}$ is a function of the data D and the size s, but we wish to not overload the notation and will keep these dependences implicit, just like the notation does not explicitly indicate the D dependence of the MLE $\widehat{\rho}_{\text{ML}}$.

The MLR analog of the MLE definition in Eq. (5.12) is then

$$\max_{\mathcal{R}\subseteq\mathcal{R}_0} L(D|\mathcal{R}) = L(D|\widehat{\mathcal{R}}_{\text{ML}}) \quad \text{with } S_{\mathcal{R}} = s\,. \tag{5.17}$$

Since all competing regions have the same size, we can equivalently maximize the joint probability under the size constraint,

$$\max_{\mathcal{R}\subseteq\mathcal{R}_0} \text{prob}(D \wedge \mathcal{R}) = \text{prob}(D \wedge \widehat{\mathcal{R}}_{\text{ML}}) \quad \text{with } S_{\mathcal{R}} = s\,. \tag{5.18}$$

The answer to this maximization problem is given in Corollary 4 of Ref. [160], which we translate into our present context as follows:

> The MLRs of various sizes s consist of all states ρ for which the point likelihood exceeds a threshold value, with higher thresholds for smaller sizes. (5.19)

This corollary is justified by a detailed proof of considerable mathematical sophistication in Ref. [160]. Here we proceed to offer an alternative argument that is perhaps more accessible to the working physicists.

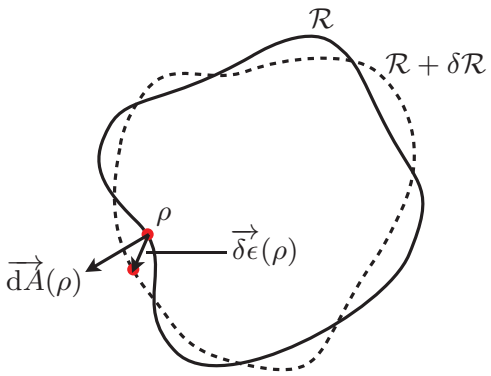


Figure 5.1: Infinitesimal variation of region $\mathcal{R}$. The boundary $\partial\mathcal{R}$ of region $\mathcal{R}$ (solid line) is deformed to become the boundary of region $\mathcal{R}+\delta\mathcal{R}$ (dashed line). $\overrightarrow{\mathrm{d}A}(\rho)$ is the vectorial surface element of $\partial\mathcal{R}$ at ρ, and $\overrightarrow{\delta\epsilon}(\rho)$ is the infinitesimal displacement of ρ.

Owing to the maximum property of the MLR and its fixed size, both $S_{\mathcal{R}}$ and $\mathrm{prob}(D\wedge\mathcal{R})$ must be stationary under infinitesimal variations $\delta\mathcal{R}$ of the region $\mathcal{R}$. Such an infinitesimal variation is achieved by deforming the boundary $\partial\mathcal{R}$ of the region, as illustrated in Fig. 5.1. The resulting change in the size $S_{\mathcal{R}}$ vanishes for all permissible deformations,

$$\delta S_{\mathcal{R}} = \int_{\partial\mathcal{R}} \overrightarrow{\mathrm{d}A}(\rho)\cdot\overrightarrow{\delta\epsilon}(\rho) = 0\,. \tag{5.20}$$

Here, $\overrightarrow{\mathrm{d}A}(\rho)$ is the vectorial surface element of the boundary $\partial\mathcal{R}$ at point ρ in the reconstruction space, and $\overrightarrow{\delta\epsilon}(\rho)$ is the infinitesimal displacement of the point ρ that deforms $\mathcal{R}$ into $\mathcal{R}+\delta\mathcal{R}$.

The corresponding change in $\mathrm{prob}(D\wedge\mathcal{R})$ is

$$\delta\mathrm{prob}(D\wedge\mathcal{R}) = \int_{\partial\mathcal{R}} \overrightarrow{\mathrm{d}A}(\rho)\cdot\overrightarrow{\delta\epsilon}(\rho)\,L(D|\rho) = 0\,, \tag{5.21}$$

which attains the indicated value of 0 at the extremum $\mathcal{R}=\widehat{\mathcal{R}}_{\mathrm{ML}}$. If we have the situation sketched in the top-left plot of Fig. 5.2, where $\widehat{\mathcal{R}}_{\mathrm{ML}}$ is completely in the interior of the reconstruction space, both Eqs. (5.20) and (5.21) must hold simultaneously for arbitrary infinitesimal deformation $\delta\mathcal{R}$. This is possible only if the point likelihood $L(D|\rho)$ is constant on the boundary $\partial\widehat{\mathcal{R}}_{\mathrm{ML}}$ of $\widehat{\mathcal{R}}_{\mathrm{ML}}$, for an $\widehat{\mathcal{R}}_{\mathrm{ML}}$ entirely contained inside $\mathcal{R}_0$ (so that $\overrightarrow{\delta\epsilon}(\rho)$ can have any direction), that is: $\partial\widehat{\mathcal{R}}_{\mathrm{ML}}$ is an *iso-likelihood surface*

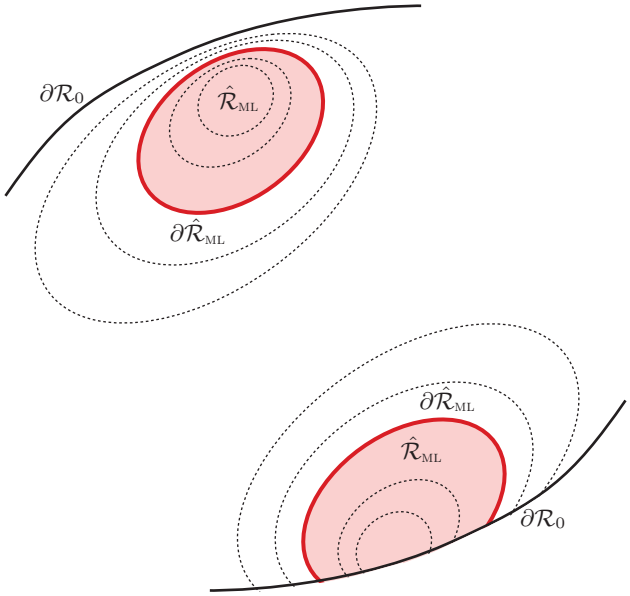


Figure 5.2: MLRs of two different kinds. In the top-left sketch, $\widehat{\mathcal{R}}_{\mathrm{ML}}$ is completely contained inside the reconstruction space; while in the bottom-right sketch, the boundary $\partial\widehat{\mathcal{R}}_{\mathrm{ML}}$ of $\widehat{\mathcal{R}}_{\mathrm{ML}}$ contains a part of the surface $\partial\mathcal{R}_0$ of the reconstruction space. Dotted lines indicate iso-likelihood surfaces, that is: surfaces on which the point likelihood is constant.

(ILS). Furthermore, $\widehat{\mathcal{R}}_{\mathrm{ML}}$ must correspond to the interior of this ILS (as opposed to its complement in the reconstruction space), since the concavity of the logarithm of the point likelihood implies that the interior necessarily has larger likelihood values than its complement.[5]

If the boundary $\partial\widehat{\mathcal{R}}_{\mathrm{ML}}$ of $\widehat{\mathcal{R}}_{\mathrm{ML}}$ contains a part of the surface $\partial\mathcal{R}_0$ of the reconstruction space, which is the situation on the bottom-right in Fig. 5.2, all interior points on $\partial\widehat{\mathcal{R}}_{\mathrm{ML}}$ must still lie on an ILS, or else we can always deform $\partial\widehat{\mathcal{R}}_{\mathrm{ML}}$ to attain a larger value of the region likelihood with a permissible choice of $\overrightarrow{\delta\epsilon}(\rho)$. On the $\partial\mathcal{R}_0$ part of $\partial\widehat{\mathcal{R}}_{\mathrm{ML}}$, the point likelihood $L(D|\rho)$ has larger values than the constant value on the interior

[5]The negative logarithm of the point likelihood is N times the sum of the relative entropy between the probabilities p and the frequencies ν, and the Shannon entropy of the frequencies (see Table 5.1),

$$-\frac{1}{N}\log L(D|\rho) = \sum_k \nu_k \log\frac{\nu_k}{p_k} - \sum_k \nu_k \log\nu_k\,.$$

Since the relative entropy is a convex function of the probabilities, the logarithm of the point likelihood is a concave function of p.

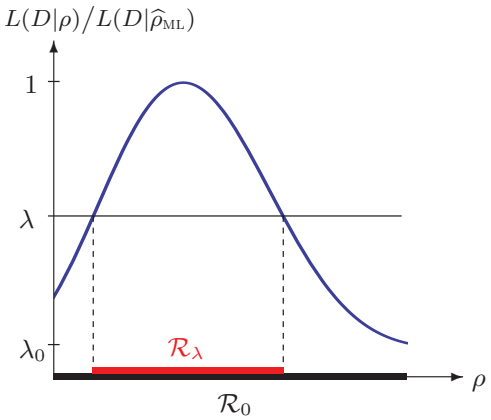


Figure 5.3: Illustration of a BLR: $\mathcal{R}_0$ is the reconstruction space; the region $\mathcal{R}_\lambda$ is a BLR, delineated by the threshold value $\lambda L(D|\widehat{\rho}_{\rm ML})$; λ_0 marks the minimum ratio $L(D|\rho)/L(D|\widehat{\rho}_{\rm ML})$ over $\mathcal{R}_0$.

part of the boundary, because ILSs that are inside $\widehat{\mathcal{R}}_{\rm ML}$ (dashed lines in Fig. 5.2) and have endpoints in $\partial\mathcal{R}_0$ assign their larger likelihood values to these points. Therefore, deforming the $\partial\mathcal{R}_0$ part of $\partial\widehat{\mathcal{R}}_{\rm ML}$ inwards, with the change in size compensated for by an outwards deformation of the interior part of $\partial\widehat{\mathcal{R}}_{\rm ML}$, decreases the value of the region likelihood. And since outwards deformations of $\partial\mathcal{R}_0$ are not possible, a region with an ILS as interior part of the boundary, supplemented by a part of $\partial\mathcal{R}_0$, is a possible MLR, indeed.

In summary, the MLRs of various sizes s consist of all states ρ for which the point likelihood $L(D|\rho)$ exceeds a certain threshold value, with higher thresholds for smaller sizes. Now, it is expedient to specify the threshold value as a fraction of the maximum value $L(D|\widehat{\rho}_{\rm ML})$ of the point likelihood; see Fig. 5.3. Denoting this fraction by λ, the characteristic function of the corresponding *bounded-likelihood region* (BLR) $\mathcal{R}_\lambda$ is the step function

$$\chi_\lambda(\rho) = \eta\Big(L(D|\rho) - \lambda L(D|\widehat{\rho}_{\rm ML})\Big), \tag{5.22}$$

where

$$\chi_\lambda(\rho) = \begin{cases} 1 & \text{if } \rho \text{ is in } \mathcal{R} \\ 0 & \text{else} \end{cases} \tag{5.23}$$

is the characteristic function of region $\mathcal{R}$. BLRs have appeared previously in standard statistical analysis; see Ref. [164] and references therein.

The BLR $\mathcal{R}_\lambda$ has the size

$$s_\lambda = \int_{\mathcal{R}_0} (\mathrm{d}\rho)\, \chi_\lambda(\rho)\,, \tag{5.24}$$

and we have $\mathcal{R}_\lambda = \mathcal{R}_0$ and $s_\lambda = s_0 = 1$ for $\lambda \leq \lambda_0$ with $\lambda_0 \geq 0$ given by

$$\min_\rho L(D|\rho) = \lambda_0 L(D|\widehat{\rho}_{\mathrm{ML}})\,. \tag{5.25}$$

As λ increases from λ_0 to 1, s_λ decreases monotonically from 1 to 0. Note that λ_0 may not be 0. Since λ_0 marks the minimum ratio $L(D|\rho)/L(D|\widehat{\rho}_{\mathrm{ML}})$ over $\mathcal{R}_0$, it will be finite if the minimum value of the point likelihood in a reconstruction space is not 0. Remember that the whole reconstruction space has unit size (corresponding to λ_0). This is the situation sketched in Fig. 5.3 where we have $\lambda_0 > 0$, which is the more general scenario. The size s specified in Eq. (5.17) is obtained for an intermediate λ value, and the corresponding BLR is the looked-for MLR.

The MLE is contained in all MLRs. In the $s \to 0$ limit, the MLR becomes an infinitesimal vicinity of the MLE and the region likelihood of the limit region is equal to the point likelihood of the MLE, $L(D|\widehat{\mathcal{R}}_{\mathrm{ML}}) \to L(D|\widehat{\rho}_{\mathrm{ML}})$.

5.3.2 Smallest credible regions

The MLR is the region for which the observed data are particularly likely. With a reversal of emphasis, we now look for a region that contains the actual state with high probability. Ultimately, this is the SCR $\widehat{\mathcal{R}}_{\mathrm{sc}}$—the smallest region for which the credibility has the pre-chosen value c. For the given D, the optimization problem

$$\min_{\mathcal{R}\subseteq\mathcal{R}_0} S_{\mathcal{R}} = S_{\widehat{\mathcal{R}}_{\mathrm{sc}}} \quad \text{with } C_{\mathcal{R}}(D) = c \tag{5.26}$$

is dual to that of Eqs. (5.17) and (5.18). Here we minimize the size for given joint probability; there we maximize the joint probability for given size. It follows that the BLRs of Eq. (5.22) are not only the MLRs, they are also the SCRs: Each MLR is a SCR, each SCR is a MLR.

The BLR $\mathcal{R}_\lambda$ has the credibility

$$c_\lambda = \frac{1}{L(D)} \int_{\mathcal{R}_0} (\mathrm{d}\rho)\, \chi_\lambda(\rho) L(D|\rho)\,, \tag{5.27}$$

which, just like s_λ, decreases monotonically from 1 to 0 as λ increases from λ_0 to 1. The credibility c specified in Eq. (5.26) is obtained for an intermediate value, and the corresponding BLR is the looked-for SCR.

That the general definitions of the MLR and the SCR, which allow for regions of arbitrary shapes, permit such a simple characterization in terms of BLRs is remarkable. BLRs are reminiscent of standard ellipsoidal error regions constructed by analyzing the neighborhood of the peak of the likelihood function—a procedure justified only for large enough N for the central limit theorem to apply (see, for instance, Ref. [154]); yet, our result employs no such assumption. Also surprising is that, while λ depends on the choice of prior, the set of regions that enter the competition is independent of that choice; the prior enters only in the size, region likelihood, and credibility of the MLR/SCR.

Once the data are obtained, there is *the* MLR and *the* SCR for these data, and other MLRs or SCRs associated with unobserved data play no role. This is in sharp contrast to confidence regions, whose construction requires consideration of all data that could have been obtained, since the confidence level is a property of the entire set of confidence regions, one for each possible data (see Sec. 5.3.4). Nevertheless, they are not unrelated: Christandl and Renner [157] showed that high-credibility regions offer starting points for constructing confidence regions—a set of SCRs with high credibility immediately suggests itself—and Blume-Kohout [158] argued that BLRs can be good confidence regions.

5.3.3 Reporting error regions

The responses of the size s_λ and the credibility c_λ of a BLR to an infinitesimal change of λ are linked by

$$L(D) \frac{\partial}{\partial \lambda} c_\lambda = L(D|\widehat{\rho}_{\mathrm{ML}}) \lambda \frac{\partial}{\partial \lambda} s_\lambda\,. \tag{5.28}$$

Therefore, once s_λ is known as a function of λ, we obtain c_λ by integrating Eq. (5.28), such that

$$c_\lambda = \frac{\lambda s_\lambda + \int_\lambda^1 \mathrm{d}\lambda'\, s_{\lambda'}}{\int_0^1 \mathrm{d}\lambda'\, s_{\lambda'}}\,, \tag{5.29}$$

in deriving which, we also used a relation between $L(D)$ and s_λ, *i.e.*,

$$L(D) = L(D|\widehat{\rho}_{\mathrm{ML}}) \int_0^1 \mathrm{d}\lambda'\, s_{\lambda'}\,. \tag{5.30}$$

This is, of course, consistent with the limiting values for $\lambda \leq \lambda_0$ and $\lambda = 1$, and also establishes that, for all intermediate values, the credibility of a BLR is larger than its size (see Fig. 5.7, the insets of Fig. 5.8, and Fig. 5.9 in Sec. 5.5),

$$c_\lambda > s_\lambda \quad \text{for } 0 < \lambda < 1\,. \tag{5.31}$$

Further, Eqs. (5.28) and (5.29) tell us that in the $\lambda \to 1$ limit (L'Hôpital's rule may be applied here), when both s_λ and c_λ vanish, their ratio is finite and exceeds unity,

$$\frac{c_\lambda}{s_\lambda} \to \frac{1}{\int_0^1 \mathrm{d}\lambda'\, s_{\lambda'}} = \frac{L(D|\widehat{\rho}_{\mathrm{ML}})}{L(D)} > 1 \quad \text{for } \lambda \to 1\,. \tag{5.32}$$

We note that this relation provides the value of $L(D)$, since the maximal value $L(D|\widehat{\rho}_{\mathrm{ML}})$ of the point likelihood is computed earlier as it is needed for identifying the BLRs.

Relation (5.31) is also an immediate consequence of the following two inequalities

$$\begin{aligned} \mathrm{prob}(D \wedge \mathcal{R}_\lambda) &> s_\lambda \lambda L(D|\widehat{\rho}_{\mathrm{ML}})\,, \\ \mathrm{prob}(D \wedge \mathcal{R}_\lambda) &> L(D) - (1 - s_\lambda)\lambda L(D|\widehat{\rho}_{\mathrm{ML}})\,, \end{aligned} \tag{5.33}$$

which in turn follow from

$$\begin{aligned} \chi_\lambda(\rho) L(D|\rho) &\geq \chi_\lambda(\rho) \lambda L(D|\widehat{\rho}_{\mathrm{ML}})\,, \\ [1 - \chi_\lambda(\rho)] L(D|\rho) &\leq [1 - \chi_\lambda(\rho)] \lambda L(D|\widehat{\rho}_{\mathrm{ML}})\,, \end{aligned} \tag{5.34}$$

with the equal sign holding only on the (interior part of the) boundary of $\mathcal{R}_\lambda$. The

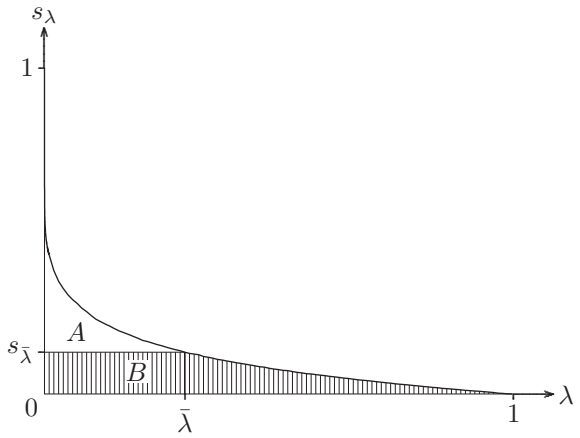


Figure 5.4: Geometrical meaning of the relation (5.29) between the size s_λ and the credibility c_λ. For the chosen value of λ, say $\bar\lambda$, the horizontal line from $(0, s_{\bar\lambda})$ to $(\bar\lambda, s_{\bar\lambda})$ divides the area under the graph of s_λ into the two pieces A and B indicated in the plot. The credibility is the fractional size of area B, that is $c_{\bar\lambda} = B/(A+B)$.

inequalities (5.34) state the defining property of the BLR: Inside the region, the point likelihood is larger than its value on the interior boundary; outside it is less than that. Now, by using Eq.(5.29), we write down the difference between c_λ and s_λ,

$$c_\lambda - s_\lambda = \frac{1}{\displaystyle\int_0^1 \mathrm{d}\lambda'\, s_{\lambda'}} \left[s_\lambda \left(\lambda - \int_0^\lambda \mathrm{d}\lambda'\, s_{\lambda'} \right) + (1 - s_\lambda) \int_\lambda^1 \mathrm{d}\lambda'\, s_{\lambda'} \right]. \tag{5.35}$$

Since each term on the right hand side of the equality is always positive except when $\lambda = 0$, we have the inequality Eq. (5.31) satisfied for all the values $0 < \lambda < 1$.

Inasmuch as the value of s_λ quantifies our prior belief that the actual state is in the region $\mathcal{R}_\lambda$, we are surprised when the data tell us that the probability for finding the state in that region is larger. Accordingly, the SCR is the region for which we are most surprised for the given prior belief. Moreover, if we wish to be quantitative about these beliefs, we can use the number $10 \log_{10}(C_\mathcal{R}/S_\mathcal{R})$ to measure the evidence for the hypothesis that the actual state is in region $\mathcal{R}$ (in units of dB). Then there is more evidence in favor of the BLR $\mathcal{R}_\lambda$ than for any other region of the same credibility. This matter and other aspects of Bayesian inference based on the concept of relative surprise are discussed in Ref. [160].

Relation (5.29) has a simple geometrical meaning in terms of areas under the graph

of s_λ, as explained in Fig. 5.4. This relation is also of considerable practical importance because we only need to evaluate the integrals of Eq. (5.24), but not those of Eqs. (5.27) and (5.14). Since the latter integrals require well-tailored Monte Carlo methods to handle the typically sharply peaked point likelihood, the numerical effort is substantially reduced if we only need to evaluate the integral of Eq. (5.24). Indeed, error regions for the observed data are then concisely communicated by reporting s_λ and c_λ as functions of λ. With these, the end user interested in the MLR with the size s of his liking or the SCR of her wanted credibility c can determine the required value of λ. It is then easy to check whether a state is inside the specified error region. The example of Sec. 5.5.3 illustrates the matter for an 8-dimensional reconstruction space, for which the error regions would be impossible to visualize, but can still be easily specified through reporting the s_λ and c_λ values.

Once more, we use the harmonic-oscillator example of Sec. 5.2.1 for an illustration. Suppose, $N = 2$ copies have been measured, and we obtained one click each for the two outcomes, so that the point likelihood is $p_1 p_2$. In this situation, we have $\lambda_0 = 0$ and $\chi_\lambda(\rho) = \eta(4p_1p_2 - \lambda)$, so that $\big|p_1 - p_2\big| \leq \sqrt{1-\lambda}$ for the BLR $\mathcal{R}_\lambda$. This gives

$$\begin{aligned} s_\lambda &= \sqrt{1-\lambda}\,, \\ c_\lambda &= \frac{1}{2}(2+\lambda)\sqrt{1-\lambda} \end{aligned} \tag{5.36}$$

for the primitive prior, and

$$\begin{aligned} s_\lambda &= 1 - \frac{2}{\pi}\sin^{-1}(\sqrt{\lambda})\,, \\ c_\lambda &= 1 - \frac{2}{\pi}\sin^{-1}(\sqrt{\lambda}) + \frac{2}{\pi}\sqrt{\lambda(1-\lambda)} \end{aligned} \tag{5.37}$$

for the Jeffreys prior.

5.3.4 Confidence regions

The confidence regions that were recently studied by Christandl and Renner [157], and independently by Blume-Kohout [158], are markedly different from the MLRs and the

SCRs. The MLR and the SCR represent inferences drawn about the unknown state ρ from the data D that have actually been observed. By contrast, confidence regions are a set of regions, one region for each possible data, whether observed or not, from the measurement of N copies. The confidence regions would contain *any* state in, at least, a certain fraction of many N-copy measurements, if the many measurements were performed. This fraction is defined as the confidence level.

When denoting by $\mathcal{C}_D$ the confidence region for data D, the confidence level γ of the set $\mathbf{C}$ of $\mathcal{C}_D$s for all conceivable data (for fixed N) is

$$\gamma(\mathbf{C}) = \min_{\rho} \sum_{D} L(D|\rho)\, \eta_{\mathcal{C}_D}(\rho)\,, \tag{5.38}$$

where $\eta_{\mathcal{C}_D}(\rho) = 1$ if ρ is in $\mathcal{C}_D$ and 0 otherwise; the minimum is reached in the "worst case". For example, in the security analysis of a protocol for quantum key distribution, one wishes a large value of γ to protect against an adversary who controls the source and prepares the quantum-information carriers in the state that is best for her.

Any set $\mathbf{C}$, for which γ has the desired value, serves the purpose. A smaller set $\mathbf{C}'$, in the sense that $\mathcal{C}'_D$ is contained in $\mathcal{C}_D$ for all D, is preferable, but usually there is no smallest set of confidence regions. Here, "smaller" is solely in this inclusion sense, with no reference to a quantification of the size of a region and, therefore, there is no necessity of specifying the prior probability of any region. Since the transition from set $\mathbf{C}$ to the smaller set $\mathbf{C}'$ requires the shrinking of some of the $\mathcal{C}_D$s without enlarging even a single one, it is easily possible to have two sets of confidence regions with the same confidence level and neither set smaller than the other.

For illustration, we consider the harmonic-oscillator example of Sec. 5.2.1 yet another time. Figure 5.5 shows two sets of confidence regions ($\gamma = 0.8$) and the corresponding three SCRs ($c = 0.8$) for the primitive prior and the Jeffreys prior. Both sets of confidence regions are optimal in the sense that one cannot shrink even one of the regions without decreasing the confidence level, but neither set is smaller than the other. In the absence of additional criteria that specify a preference, both work equally well as sets of confidence regions. This generic non-uniqueness of confidence regions,

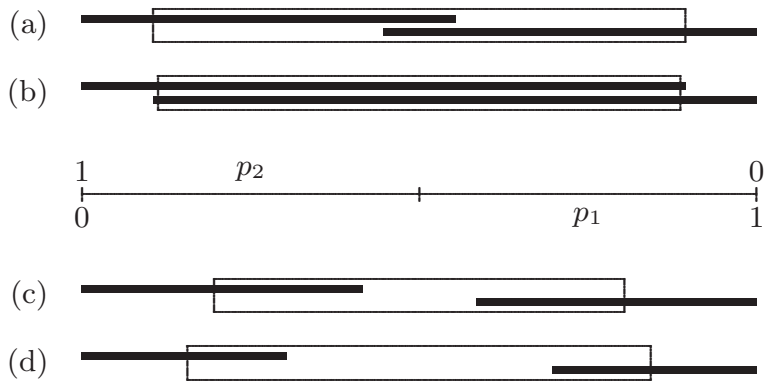


Figure 5.5: Confidence regions and smallest credible regions. The bars indicate intervals of $p_1 = 1 - p_2$ for the harmonic-oscillator example of Sec. 5.2.1, which has the reconstruction space of a tossed coin. Two copies are measured. The left solid bars indicate the regions for $(n_1, n_2) = (0, 2)$ counts; the right solid bars are for $(n_1, n_2) = (2, 0)$; and the central open bars are for $(n_1, n_2) = (1, 1)$. Cases (a) and (b) show two sets of confidence regions for confidence level $\gamma = 0.8$. Regions (c) and (d) are the SCRs for the primitive prior and the Jeffreys prior, respectively, both with a credibility $c = 0.8$.

and the arbitrariness associated with it, are in marked contrast to the SCRs, which are always unique.

We also observe in this example that confidence regions tend to overlap a lot, which is indeed unavoidable if a large confidence level is desired. By contrast, the SCRs for different data usually do not overlap unless the data are quite similar. In Fig. 5.5, there is no overlap of the SCRs for the data $(n_1, n_2) = (0, 2)$ and $(2, 0)$.

Another important difference of considerable concern in all practical applications is the following. Once the data are obtained, there is *the* MLR and *the* SCR for these data, and it plays no role what other MLRs or SCRs are associated with different data that have not been observed. To find a confidence region for the actual data, however, one must first specify the whole set $\mathbf{C}$ of confidence regions because the confidence level of Eq. (5.38) is a property of the whole set.

5.4 Choosing the prior

The assignment of prior probabilities to regions in the reconstruction space should be done in an unprejudiced manner while taking into account all prior information that might be available. We cannot do justice to the rich literature on this subject

and are content with noting that Ref. [162] reviews various approaches to constructing unprejudiced priors. Here, we discuss some criteria that are useful when choosing a prior, illustrating with examples familiar in quantum contexts.

A general remark is this: The chosen prior should give some weight to (almost) all states, and it should not give extremely high weight to states in some part of the state space and extremely low weight to other states. This is to say that the prior should be *consistent* in the sense that the credibility of a region—its posterior content—is dominated by the data, rather than by the prior, if a reasonably large number N of copies is measured. For the examples of Fig. 5.8 in Sec. 5.5.2, $N = 24$ is close to being "reasonably large", while $N = 2$ in Fig. 5.5 is clearly not. Also, $N = 60$ in Sec. 5.5.3 is not large enough to ensure data dominance, because the λ values in Table 5.3 for the primitive prior are much smaller than those for the Jeffreys prior.

Below, we describe a few criteria for choosing priors. We begin in Sec. 5.4.1 with the common choice of a uniform prior; Sec. 5.4.2 discusses priors motivated by the utility of the estimated state; Sec. 5.4.3 invokes symmetry arguments to restrict considerations to priors that possess some symmetry properties; Sec. 5.4.4 presents form-invariant prior constructions; Sec. 5.4.5 deals with the situation where one has a target state in mind; and Sec. 5.4.6 is about priors induced by marginalization of full-state-space priors according to what the data can tell us.

5.4.1 Uniformity

The time-honored strategy of choosing a uniform prior on $\mathcal{R}_0$ in which all states are treated equally gets us into a circular argument. Our identification of the size of a region with its prior content amounts to assigning equal probabilities to regions of equal sizes, prior to acquiring any data. But that just means that we now have to declare how we measure the size of a region without prejudice, and we are again faced with the original question about a uniform prior.

In fact, there is no unique meaning of the uniformity of a prior. In the sense that each prior tells us how to quantify the size of a region, each prior is uniform with respect

to its induced size measure. To illustrate, reconsider the harmonic-oscillator example of Sec. 5.2.1. For the primitive prior of Eq. (5.8), the parameterization

$$\begin{aligned} p_1 &= \frac{1}{2}(v+u)\,, \quad p_2 = \frac{1}{2}(v-u)\,, \\ \mathrm{d}p_1\,\mathrm{d}p_2 &= \mathrm{d}u\,\mathrm{d}v\,\frac{1}{2} \end{aligned} \tag{5.39}$$

gives

$$\begin{aligned} (\mathrm{d}\rho) \;&=\; \mathrm{d}u\,\mathrm{d}v\,\frac{1}{2}\eta(v+u)\eta(v-u)\delta(v-1) \\ &\to\; \mathrm{d}u\,\frac{1}{2} \quad \text{with } -1 \leq u \leq 1\,, \end{aligned} \tag{5.40}$$

where we integrate over v in the last step and so observe that the primitive prior is uniform in u, that is: the size of the region $u_1 < u < u_2$ is proportional to $u_2 - u_1$. Likewise, the parameterization

$$\begin{aligned} p_1 &= v(\sin\alpha)^2\,, \quad p_2 = v(\cos\alpha)^2\,, \\ \mathrm{d}p_1\,\mathrm{d}p_2 &= \mathrm{d}\alpha\,\mathrm{d}v\,v\sin(2\alpha) \end{aligned} \tag{5.41}$$

gives

$$(\mathrm{d}\rho) \to \mathrm{d}\alpha\,\frac{2}{\pi} \quad \text{with } 0 \leq \alpha \leq \frac{\pi}{2} \tag{5.42}$$

for the Jeffreys prior of Eq. (5.9), which is uniform in the parameter α, instead. Other priors can be treated analogously, each of them yielding a uniform prior in an appropriate single parameter.

Visualization of the uniformity for qubit priors can be found in Fig. 5.6. Plot (b) shows uniform tiling of the unit disk by tiles of equal size. Here size is measured by the primitive prior of Eq. (5.72), which is uniform in x and y, and also in r^2 and φ (the latter is used for the plot). Plots (c1) and (c2) show uniform tilings of the unit disk for the Jeffreys prior for the four-outcome POM of Eq. (5.73), while plots (d1) and (d2) show those for the three-outcome POM of Eq. (5.74). The crosshair symmetry of the four-outcome POM and the trine symmetry of the three-outcome POM are manifest

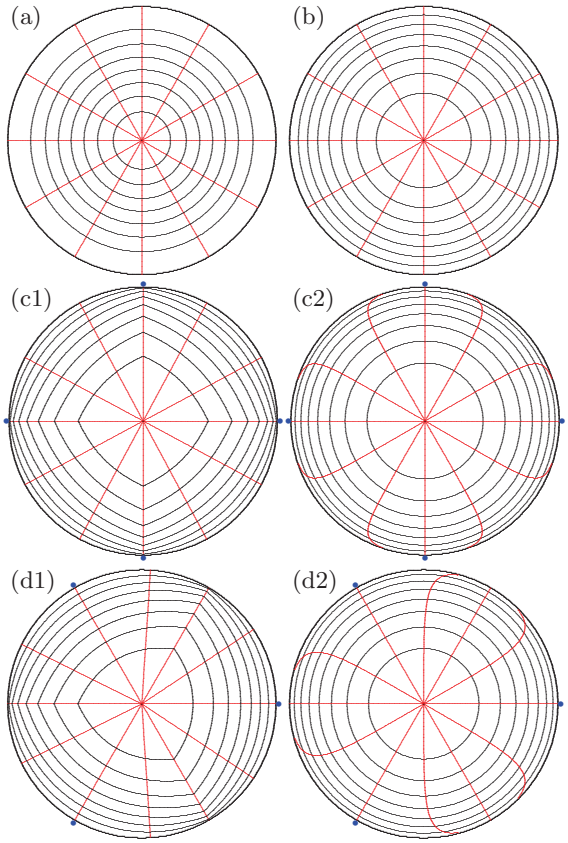


Figure 5.6: Uniform tilings of the unit disk for four different priors. The disk is in the xy plane, with the x axis horizontal, the y axis vertical, and the disk center at $x = y = 0$. Tiling (a) is for the marginal prior of Eq. (5.52); tiling (b) depicts the primitive prior of Eq. (5.72); tilings (c1) and (c2) illustrate the Jeffreys prior of Eq. (5.73) with the blue dots (•) just outside the unit circle indicating the four directions onto which the POM outcomes project; and tilings (d1)and (d2) are for the Jeffreys prior of Eq. (5.74), the blue dots marking the three directions of the trine projectors. In each tiling, we identify 96 regions of equal size by dividing the disk into eight "tree rings" of equal size and twelve "pie slices" of equal size. In the tilings (a), (b), (c1), and (d1), the boundaries of the pie slices are (red) rays and an arc of the unit circle; in the tilings (a), (b), (c2), and (d2), the tree rings have concentric circles as their boundaries.

in their respective uniform tilings.

The parameterizations in Eqs. (5.39) and (5.41), and the tilings of Fig. 5.6 exhibit in which explicit sense the primitive prior and the Jeffreys prior are uniform. But the priors are what they are, irrespective of how they are parameterized. They are explicitly uniform in a particular parameterization and implicitly uniform in all others. Uniformity, it follows, cannot serve as a principle that distinguishes one prior from another.

This ubiquity of uniform priors for a continuous set of infinitesimal probabilities is in marked contrast to situations in which prior probabilities are assigned to a finite number of discrete possibilities, such as the 38 pockets of a double-zero roulette wheel. Uniform probabilities of 1/38 suggest themselves, are meaningful, and clearly distinguished from other priors, all of which have a bias. Uniformity in a particularly natural parameterization of the probability space might also be meaningful. This, however, invokes a notion of "natural" that others may not share.

5.4.2 Utility

In many applications, estimating the state is not a purpose in itself, but only an intermediate step on the way to determining some particular properties of the physical system. The objective is then to find the value of a parameter that quantifies the *utility* of the state.

For example, one could be interested in the fidelity of the actual state with a target state, or in an entanglement measure of a two-partite state, or in another quantity that tells us how useful are the quantum-information carriers for their intended task. In a situation of this kind, one should, if possible, use a prior that is uniform in the utility parameter of interest. Contrary to the situation of the previous section, where requiring uniformity in $\mathcal{R}_0$ may be ill-advised because uniformity is a parameterization-dependent notion, here we specify uniformity for the parameter we are interested in.

To illustrate, consider a single qubit. Suppose the utility parameter is the purity $\xi(\rho) = \mathrm{tr}\{\rho^2\}$ of the state ρ. With the Bloch-ball representation of a qubit state, $\rho = \frac{1}{2}(1 + \boldsymbol{\varrho} \cdot \boldsymbol{\sigma})$, where $\boldsymbol{\varrho} = \mathrm{tr}\{\boldsymbol{\sigma}\rho\} = \langle \boldsymbol{\sigma} \rangle$ is the Bloch vector and $\boldsymbol{\sigma}$ is the vector of Pauli matrices, the purity is given by

$$\xi(\rho) = \frac{1}{2}(1 + \varrho^2) \quad \text{with } \varrho = |\boldsymbol{\varrho}| \,. \tag{5.43}$$

A prior uniform in purity induces a prior on the state space according to

$$(\mathrm{d}\rho) \propto \mathrm{d}\xi \, \mathrm{d}\Omega \propto \varrho \, \mathrm{d}\varrho \, \mathrm{d}\Omega \,, \tag{5.44}$$

where we parameterize the Bloch ball by spherical coordinates (ϱ, θ, ϕ). Here, $\mathrm{d}\Omega$ is the prior for the angular coordinates; the prior for the radial coordinate ϱ is fixed by our choice of uniformity in ξ. Irrespective of what we choose for $\mathrm{d}\Omega$, the marginal prior for ϱ is uniform in ξ.

If one can quantify the utility of an estimator by a cost function, an optimal prior can be selected by a minimax strategy: For each prior in the competition one determines the maximum of the cost function over the states in the reconstruction space, and then chooses the prior for which the maximum cost is minimal. In classical statistics, such minimax strategies are common (see, for instance, Chapter 5 in Ref. [165]); for an example in the context of quantum state estimation, see Ref. [50].

5.4.3 Symmetry

Symmetry considerations are often helpful in narrowing the search for the appropriate prior. For a particularly instructive example, see section 12.4.4 in Jaynes's posthumous book [166].

Returning to the uniform-in-purity prior of Eq. (5.44), one can invoke rotational symmetry in favor of the usual solid-angle element, $\mathrm{d}\Omega = \sin\theta \mathrm{d}\theta\, \mathrm{d}\phi$, as the choice of angular prior. The reasoning is as follows: The purity of a qubit state does not change under unitary transformations; unitarily equivalent states have the same purity. Now, regions that are turned into each other by a unitary transformation have identical radial content whereas the angular dependences are related by a rotation. Invariance under rotations, in turn, requires that the prior is proportional to the solid angle, hence the identification of $\mathrm{d}\Omega$ with the differential of the solid angle. Note that the resulting prior element $(\mathrm{d}\rho)$ is different from the usual Euclidean volume element, $\varrho^2 \mathrm{d}\varrho\, \sin\theta \mathrm{d}\theta\, \mathrm{d}\phi$, which would be natural if the Bloch ball were an object in the physical three-dimensional space. But it ain't.

Symmetry arguments can be very helpful if used carefully and not blindly. For a fairly tossed coin, the prior should not be affected if the probabilities for heads and tails are interchanged, $w(p_1, p_2) = w(p_2, p_1)$. However, for the harmonic-oscillator

example of Sec. 5.2.1, which has the same reconstruction space as the coin, there is poor justification for requiring this symmetry because the two probabilities—of finding the oscillator in its ground state, or not—are not on equal footing.

5.4.4 Invariance

When one speaks of an *invariant prior*, one does not mean the invariance under a change of parameterization—all priors are invariant in this respect (see Sec. 5.4.1)—but rather a *form-invariant* construction in terms of a quantity that, preferably, has an invariant significance. We consider two particular constructions that make use of the metric induced by the response of the selected function to infinitesimal changes of its variables.

The first construction begins with a quantity $F(p)$ that is a function of all probabilities $p = (p_1, \ldots, p_K)$. We include the square root of the determinant of the dyadic second derivative in the prior density as a factor,

$$(\mathrm{d}\rho) = (\mathrm{d}p) \left| \det \left\{ \left(\frac{\partial^2 F}{\partial p_j \, \partial p_k} \right)_{jk} \right\} \right|^{1/2} w_{\mathrm{cstr}}(p) \,, \tag{5.45}$$

where $w_{\mathrm{cstr}}(p)$ contains all the delta-function and step-function factors of constraint as well as the normalization factor that ensures the unit size of the reconstruction space. The prior defined by Eq. (5.45) is invariant in the sense that a change of parameterization, from p to α, say, does not affect its structure,

$$(\mathrm{d}\rho) = (\mathrm{d}\alpha) \left| \det \left\{ \left(\frac{\partial^2 F}{\partial \alpha_j \, \partial \alpha_k} \right)_{jk} \right\} \right|^{1/2} w_{\mathrm{cstr}}\big(p(\alpha)\big) \,, \tag{5.46}$$

because the various Jacobian determinants for the reparameterization take care of each other. Since $w_{\mathrm{cstr}}(p)$ enforces all constraints, the p_ks are independent variables when $F(p)$ and $G(p, \nu)$ are differentiated in Eq. (5.45) and Eq. (5.47), respectively.

For the second construction, we use a data-dependent function $G(p, \nu)$ of the probabilities p and the frequencies $\nu = (\nu_1, \nu_2, \ldots, \nu_K)$ with $\nu_j = n_j/N$. Here, the square root of the determinant of the expected value of the dyadic square of the p-gradient of

Table 5.1: Form-invariant priors constructed by one of the two methods described in the text. The "$\sqrt{\det}$" column gives the p-dependent factors only and omits all p-independent constants. The first method of Eq. (5.45) proceeds from functions of the probabilities that have extremal values when all probabilities are equal or all vanish save one. The second method of Eq. (5.47) uses functions that quantify how similar are the probabilities and the frequencies. The "hedged prior" is named in analogy to the "hedged likelihood" [38].

method	primary function	$\sqrt{\det}$
1st	$-\sum_k p_k \log p_k$ (Shannon entropy)	$\frac{1}{\sqrt{p_1 p_2 \cdots p_K}}$ (Jeffreys prior)
1st	$\sum_k p_k^2$ (purity)	1 (primitive prior)
2nd	$\sum_k \nu_k p_k$ (inner product)	$\sqrt{p_1 p_2 \cdots p_K}$ (hedged prior)
2nd	$\sum_k \nu_k \log(\nu_k/p_k)$ (relative entropy)	$\frac{1}{\sqrt{p_1 p_2 \cdots p_K}}$ (Jeffreys prior)

G is a factor in the prior density,

$$(\mathrm{d}\rho) = (\mathrm{d}p) \left| \det \left\{ \overline{\left(\frac{\partial G}{\partial p_j} \frac{\partial G}{\partial p_k} \right)_{jk}} \right\} \right|^{1/2} w_{\mathrm{cstr}}(p) \,, \tag{5.47}$$

where $\overline{f(\nu)}$ denotes the expected value of $f(\nu)$,

$$\overline{f(\nu)} = \sum_D L(D|\rho) f(\nu) \,. \tag{5.48}$$

We have, in particular, the generating function

$$\overline{\exp\left(\sum_{k=1}^{K} a_k \nu_k \right)} = \left(\sum_{k=1}^{K} \mathrm{e}^{a_k/N} p_k \right)^N \tag{5.49}$$

for the expected values of products of the ν_ks. The prior defined by Eq. (5.47) is form-invariant in the same sense, and for the same reason, as the prior of Eq. (5.45).

Table 5.1 reports a few examples of "$\sqrt{\det}$" factors constructed by one of these two methods. It is worth noting that the Jeffreys prior can be obtained from the entropy of the probabilities by the first method as well as from the relative entropy between the probabilities and the frequencies by the second method. The latter is a variant of Jeffreys's original derivation [161] in terms of the Fisher information.

5.4.5 Conjugation

Sometimes there are reasons to expect that the actual state is close to a certain target state with probabilities $t = (t_1, t_2, \ldots, t_K)$. This is the situation, for example, when a source is designed to emit the quantum-information carriers in a particular state. A *conjugate prior*

$$(\mathrm{d}\rho) = (\mathrm{d}p) \left(p_1^{t_1} p_2^{t_2} \cdots p_K^{t_K}\right)^{\beta} w_{\mathrm{cstr}}(p) \quad \text{with } \beta > 0 \tag{5.50}$$

could then be a natural choice. Such priors are called "conjugate" in standard statistics literature because the $(\cdots)^{\beta}$ factor has the same structure as the point likelihood: a product of powers of the detection probabilities. The $(\cdots)^{\beta}$ factor is maximal for $p = t$, and the peak is narrower when β is larger.

The conjugate prior can be understood as the "mock posterior" for the primitive prior that results from pretending that β copies have been measured in the past and data obtained that are most typical for the target state. Therefore, a conjugate prior is quite natural to express the expectation that the apparatus is functioning well. The posterior content of a region will be data-dominated only if N is much larger than β.

In this context, it may be worth noting that the Bayesian mean state (see Sec. 2.3.3.2 in Chapter 2),

$$\widehat{\rho}_{\mathrm{BM}} = \int_{\mathcal{R}_0} (\mathrm{d}\rho)\, \rho\,, \tag{5.51}$$

computed with the conjugate prior above, is usually not the target state unless β is large. One could construct priors for which $\widehat{\rho}_{\mathrm{BM}}$ is the target state, but the presence of the $w_{\mathrm{cstr}}(p)$ factor requires a case-by-case construction.

5.4.6 Marginalization

All priors used as examples—the ones in Eqs. (5.40), (5.42) and (5.50), and Table 5.1—have in common that they are defined in terms of the probabilities and, therefore, they refer to the particular POM with which the data are collected. While this takes duly into account the significance of the data, it does not seem to square with the point of view that prior probabilities are solely a property of the physical processes that put the quantum-information carriers into the state that is then diagnosed by the POM.

When adopting this viewpoint, one begins with a prior density defined on the entire state space. In addition to the parameters that specify the reconstruction space (essentially the probabilities p), this full-space prior will depend on parameters whose values are not determined by the data. There could be very many nuisance parameters of this kind, as illustrated by the somewhat extreme harmonic-oscillator example of Sec. 5.2.1. Upon integrating the full-space prior over the nuisance parameters, one obtains a *marginal prior* on the reconstruction space. As a function on the reconstruction space, the marginal prior is naturally parameterized in terms of the probabilities and so fits into the formalism we are using throughout.

We note that the invoking of "additional criteria" for a unique mapping from p to ρ, as mentioned at the end of Sec. 5.2.1, is exactly what would be required if one wishes to report estimated values of the nuisance parameters. That, however, goes beyond making statements that are solidly supported by the data and is, therefore, outside the scope of our present discussion.

The symmetric uniform-in-purity prior of Secs. 5.4.2 and 5.4.3 provides an example for marginalization if the POM only gives information about $x = \langle\sigma_x\rangle$ and $y = \langle\sigma_y\rangle$, but not about $z = \langle\sigma_z\rangle$. We express the full-space prior in cartesian coordinates, integrate over z, and arrive at

$$\begin{aligned}(\mathrm{d}\rho) &= \mathrm{d}x\,\mathrm{d}y\,\frac{1}{2\pi}\int\limits_{-\infty}^{\infty}\mathrm{d}z\,\frac{\eta(1-x^2-y^2-z^2)}{\sqrt{x^2+y^2+z^2}} \\ &= \mathrm{d}x\,\mathrm{d}y\,\frac{1}{\pi}\eta(1-x^2-y^2)\cosh^{-1}\frac{1}{\sqrt{x^2+y^2}}\,. \end{aligned} \tag{5.52}$$

This marginal prior is a function on the unit disk in the xy plane, which is the natural choice of reconstruction space here. When one expresses $(\mathrm{d}\rho)$ in polar coordinates, $x + \mathrm{i}y = r\mathrm{e}^{\mathrm{i}\varphi}$, one sees that $(\mathrm{d}\rho)$ is uniform in φ and in $r^2 \cosh^{-1}(1/r) - \sqrt{1-r^2}$, which increases monotonically from -1 to 0 on the way from the center of the disk at $r = 0$ to the unit circle where $r = 1$. Plot (a) in Fig. 5.6 illustrates the matter.

5.5 Examples

In this section, we first apply the method of constructing MLRs and SCRs to study the problem of a classical coin. Then for illustrations in the quantum scenario, we identify optimal error regions for single qubit (confined to the equatorial plane of the Bloch sphere) and two-qubit states from computer-generated data that simulate incomplete tomography with few measured copies.

5.5.1 The classical coin

As the simplest example, we consider the classical coin and try to find out the probability p that heads (or tails) turn up by tossing a biased coin. We use a general normalized prior function, parameterized by a factor β, *i.e.*,

$$\pi(p, \beta) = \frac{(2\beta+1)!}{(\beta!)^2}[p(1-p)]^\beta, \quad \text{with } \beta > -1, \tag{5.53}$$

which corresponds to the primitive prior if $\beta = 0$, the Jeffreys prior for two outcomes if $\beta = -1/2$, and the hedged prior if $\beta = 1/2$. Then for a BLR $\mathcal{R}_\lambda$ with $0 < \lambda < 1$, the size of it using the above prior is calculated through Eq. (5.24), such that

$$\begin{aligned} s_\lambda(\beta) &= \int_{\mathcal{R}_\lambda} (\mathrm{d}p)\, \pi(p, \beta) \\ &= \frac{(2\beta+1)!}{(\beta!)^2}\Big(B_{p_2} - B_{p_1}\Big)(\beta+1, \beta+1)\,, \end{aligned} \tag{5.54}$$

where $B_x(a, b) = \displaystyle\int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t$ is the incomplete beta function and the BLR $\mathcal{R}_\lambda$ is simply the interval $[p_1, p_2]$ in the real unit line and has the boundary condition being

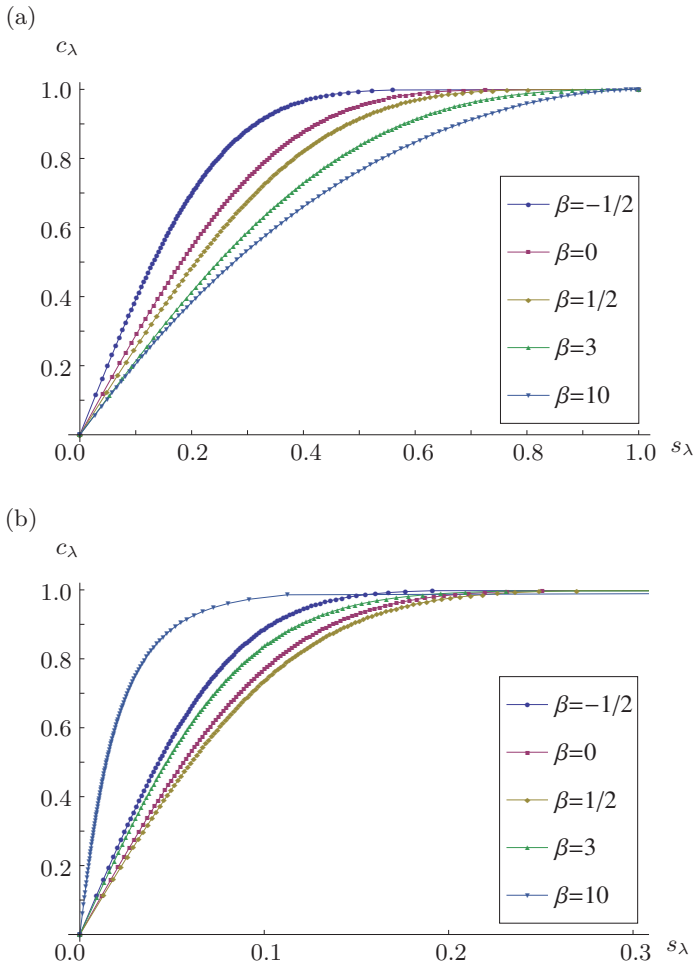


Figure 5.7: Plots of the credibility c_λ versus the size s_λ for the BLRs of two simulated experiments of coin tossing by using various β values of the prior, *i.e.*, Eq. (5.53). Plot (a) is for $N = 100$ total tosses; plot (b) is for $N = 10$ total tosses with the s_λ values shown up to 0.3 only, since $c_\lambda = 1.0$ for all values of $s_\lambda \geq 0.3$.

$L(D|p_1) = L(D|p_2)$. By applying Eq. (5.27), the corresponding credibility is

$$\begin{aligned} c_\lambda(\beta) &= \frac{\displaystyle\int_{\mathcal{R}_\lambda} (\mathrm{d}p)\, \pi(p,\beta)\, L(D|p)}{\displaystyle\int_{\mathcal{R}_0} (\mathrm{d}p)\, \pi(p,\beta)\, L(D|p)} \\ &= \frac{\Big(B_{p_2} - B_{p_1}\Big)(n+\beta+1, N-n+\beta+1)}{B(n+\beta+1, N-n+\beta+1)}\,, \end{aligned} \tag{5.55}$$

assuming that n heads (or tails) occur out of N total tosses and $B(a,b) = \dfrac{(a-1)!(b-1)!}{(a+b-1)!}$ is the (complete) beta function. Another way to calculate the credibil-

ity is to use Eq. (5.29), in which the pre-obtained values of the size s_λ can be used to calculate c_λ without invoking the normally highly-peaked likelihood function.

In Figs. 5.7(a) and 5.7(b), we show the plots of the credibility c_λ versus the size s_λ for the BLRs of two simulated experiments by using various β values of the prior, *i.e.*, Eq. (5.53). Figure 5.7(a) was generated by simulating a total number of $N = 100$ tosses; while Fig. 5.7(b) was done by using $N = 10$ tosses only. As can be seen from the figures, the ratio c_λ/s_λ (gradients of the plots) in Fig. 5.7(b) is much larger than that in Fig. 5.7(a). The reason is that with a large N, the likelihood function $L(D|p)$ is highly peaked, which plays a dominant role when calculating the credibility c_λ. However, the size s_λ of a BLR is solely determined by the prior.

5.5.2 Incomplete single-qubit tomography

For a first illustration in the quantum scenario, we consider the simplest situation that exhibits the typical features: The quantum-information carriers have a qubit degree of freedom (confined to the equatorial plane of the Bloch sphere), which is measured by one of two standard POMs that are not informationally complete.

5.5.2.1 POMs and priors

For both POMs, the unit disk in the xy plane suggests itself for the reconstruction space $\mathcal{R}_0$. The first POM is the crosshair measurement ($K = 4$) that is built from four pure states symmetrically arranged in the xy plane of the Bloch sphere, subtending angles of $\pi/2$ between pairs of states,

$$|\psi_k\rangle\langle\psi_k| = \frac{1}{2}(1 + \sigma_x \cos\phi_k + \sigma_y \sin\phi_k), \quad \text{with } \phi_k \equiv \phi_0 + (k-1)\frac{\pi}{2}\,, \tag{5.56}$$

for $k = 1, 2, 3, 4$, where σs are the usual Pauli operators. It it easy to check that these four states are complete, such that

$$\frac{1}{2}\sum_{k=1}^{4} |\psi_k\rangle\langle\psi_k| = 1, \tag{5.57}$$

but they are not pairwise linearly independent, meaning that the POM outcomes constructed from them are not symmetric (not a symmetric POM),

$$\Pi_k \equiv |\psi_k\rangle \frac{1}{2} \langle \psi_k|, \quad \text{with } k = 1, 2, 3, 4. \tag{5.58}$$

As is required for is a physical POM, Eq. (5.57) ensures that $\sum_{k=1}^{4} \Pi_k = 1$. Every physical state lying in the xy plane of the Bloch sphere can be described in polar coordinates as

$$\rho = \frac{1}{2} \left[1 + r(\sigma_z \cos\phi + \sigma_x \sin\phi) \right], \quad \text{with } 0 \leq r \leq 1 \text{ and } 0 \leq \phi < 2\pi. \tag{5.59}$$

Then the outcome probabilities for state ρ are given by

$$p_k = \mathrm{tr}\{\rho \Pi_k\} = \frac{1}{4}[1 + r\cos(\phi - \phi_k)], \quad k = 1, 2, 3, 4. \tag{5.60}$$

For the simulation, we make the phase $\phi_0 = 0$ being constant and combine the projective measurements of σ_x and σ_y into the four-outcome POM ($K = 4$) with probabilities

$$\left.\begin{matrix} p_1 \\ p_2 \end{matrix}\right\} = \frac{1}{4}(1 \pm x), \qquad \left.\begin{matrix} p_3 \\ p_4 \end{matrix}\right\} = \frac{1}{4}(1 \pm y), \tag{5.61}$$

with $x = \langle \sigma_x \rangle$ and $y = \langle \sigma_y \rangle$. Notice that we have $\sum_{k=1}^{4} p_k = 1$, $p_1 + p_2 = p_3 + p_4 = 1/2$, and also an additional constraint for this POM, such that

$$\frac{1}{4} \leq \sum_{k=1}^{4} p_k^2 = \frac{1}{4}\left(1 + \frac{1}{2} r^2\right) \leq \frac{3}{8}. \tag{5.62}$$

Therefore, the permissible probabilities are identified by

$$w_{\mathrm{cstr}}(p) \doteq \eta(p)\, \delta(p_1 + p_2 - \tfrac{1}{2})\, \delta(p_3 + p_4 - \tfrac{1}{2})\, \eta(3 - 8p^2), \tag{5.63}$$

where

$$\eta(p) = \prod_{k=1}^{K} \eta(p_k) \quad \text{and} \quad p^2 = \sum_{k=1}^{K} p_k^2. \tag{5.64}$$

The dotted equal sign in Eq. (5.63) stands for "equal up to a multiplicative constant", namely the factor that ensures the unit size of the reconstruction space.

The second POM (introduced briefly in Sec. 2.4.1 of Chapter 2) that we use is the three-outcome trine measurement ($K = 3$), which is indeed a symmetric POM with three POM outcomes built from three pure states symmetrically arranged in the xy plane of the Bloch sphere, subtending angles of $2\pi/3$ between pairs of states, such that

$$|\psi_k\rangle\langle\psi_k| = \frac{1}{2}(1 + \sigma_z \cos\phi_k + \sigma_x \sin\phi_k), \quad \text{with } \phi_k \equiv \phi_0 + (k-1)\frac{2\pi}{3}\,, \tag{5.65}$$

for $k = 1, 2, 3$. These trine states are pairwise linearly independent and complete since

$$|\langle\psi_k|\psi_l\rangle|^2 = \frac{3}{4}\delta_{kl} + \frac{1}{4} \quad \text{and} \quad \frac{2}{3}\sum_{k=1}^{3}|\psi_k\rangle\langle\psi_k| = 1. \tag{5.66}$$

The outcomes of the trine POM are

$$\Pi_k \equiv |\psi_k\rangle\frac{2}{3}\langle\psi_k|, \quad \text{with } k = 1, 2, 3. \tag{5.67}$$

Equation (5.66) ensures that $\sum_{k=1}^{3}\Pi_k = 1$ and the outcome probabilities are

$$p_k = \mathrm{tr}\{\rho\Pi_k\} = \frac{1}{3}[1 + r\cos(\phi - \phi_k)], \quad k = 1, 2, 3. \tag{5.68}$$

Notice that we have $\sum_{k=1}^{3} p_k = 1$, and the trine outcome probabilities also satisfy the following constraint,

$$\frac{1}{3} \le \sum_{k=1}^{3} p_k^2 = \frac{1}{3}\left(1 + \frac{1}{2}r^2\right) \le \frac{1}{2}\,, \tag{5.69}$$

for all the physical qubit states. For the simulation, we again make the phase $\phi_0 = 0$ being constant, in which case the POM outcomes can be represented by subnormalized projectors on the eigenstates of σ_x and $(-\sigma_x \pm \sqrt{3}\,\sigma_y)/2$ with eigenvalue $+1$. It then has the probabilities

$$p_1 = \frac{1}{3}(1 + x)\,, \qquad \left.\begin{matrix} p_2 \\ p_3 \end{matrix}\right\} = \frac{1}{6}(2 - x \pm \sqrt{3}\,y)\,, \tag{5.70}$$

for which

$$w_{\mathrm{cstr}}(p) \doteq \eta(p)\,\delta(p_1+p_2+p_3-1)\,\eta(1-2p^2) \tag{5.71}$$

summarizes the constraints for the trine measurement that the permissible values of p_1, p_2, p_3 should obey.

Both POMs have the same expression for the primitive prior,

$$(\mathrm{d}\rho) = \mathrm{d}x\,\mathrm{d}y\,\frac{1}{\pi}\eta(1-x^2-y^2) = \mathrm{d}(r^2)\,\frac{\mathrm{d}\varphi}{2\pi}\,, \tag{5.72}$$

where $x+\mathrm{i}y = r\mathrm{e}^{\mathrm{i}\varphi}$ with $0 \le r \le 1$ and φ covers any convenient range of 2π. This prior is uniform in x and y, and also uniform in r^2 and φ. The polar-coordinate version is the more natural parameterization of the unit disk, which is used for the plot (b) in Fig. 5.6.

The Jeffreys prior for the four-outcome POM is

$$(\mathrm{d}\rho) = \frac{2}{\pi^2}\frac{\mathrm{d}r\,r\,\mathrm{d}\varphi}{\sqrt{1-r^2+\frac{1}{4}r^4\sin(2\varphi)^2}}\,. \tag{5.73}$$

Plots (c1) and (c2) in Fig. 5.6 show uniform tilings of the unit disk for this prior. For the three-outcome POM, we have the Jeffreys prior

$$(\mathrm{d}\rho) = \frac{1}{4\pi-24\sin^{-1}(1/3)}\frac{\mathrm{d}r\,r\,\mathrm{d}\varphi}{\sqrt{1-\frac{3}{4}r^2+\frac{1}{4}r^3\cos(3\varphi)}}\,, \tag{5.74}$$

and the tilings of plots (d1) and (d2) in Fig. 5.6. The crosshair symmetry of the four-outcome POM and the trine symmetry of the three-outcome POM are manifest in their respective uniform tilings.

5.5.2.2 Computer-generated data

Before jumping to the simulated experiments, we need some mathematical tools to simplify the calculation. We follow Ref. [50] to define the generalized *moments* for a

K-outcome measurement as

$$\begin{aligned} M_\beta^\lambda(n_1, n_2, \ldots, n_K) &= 2\int_0^\infty (\mathrm{d}p)\, \omega(p)\, \pi(p,\beta)\, \chi_\lambda(p) \prod_{k=1}^K p_k^{n_k} \\ &= 2\int_{\mathcal{R}_\lambda} (\mathrm{d}p)\, \omega(p)\, \pi(p,\beta) \prod_{k=1}^K p_k^{n_k}, \quad 0 \le \lambda \le 1, \end{aligned} \tag{5.75}$$

where $\omega(p)$ is the characteristic function accounting for all the physicality constraints, and $\pi(p,\beta)$ is the prior function with an additional parameter β, for instance, form like the following (unnormalized),

$$\pi(p,\beta) \propto \left(\prod_{k=1}^K p_k\right)^\beta. \tag{5.76}$$

The values of β can take any real number as long as all the integrals for the moments exist, *i.e.*, $\beta = -1/2$ gives the Jeffreys prior and $\beta = 0$ is the primitive prior. The BLR $\mathcal{R}_\lambda$ is determined by the step function $\chi_\lambda(p) = \eta(L(D|p) - \lambda L(D|\widehat{p}_{\mathrm{ML}}))$, which is the same as the defined $\mathcal{R}_\lambda$ using $\chi_\lambda(\rho)$ in Eq. (5.22).

The moments have permutation symmetry [50], such that

$$M_\beta^\lambda(n_1, n_2, \ldots, n_K) = M_\beta^\lambda(n_2, n_3, \ldots, n_K, n_1) = \cdots = M_\beta^\lambda(n_K, n_1, \ldots, n_{K-1}), \tag{5.77}$$

and obey a sum rule [50],

$$\begin{aligned} &M_\beta^\lambda(n_1+1, n_2, \ldots, n_K) + M_\beta^\lambda(n_1, n_2+1, \ldots, n_K) + \cdots \\ &\quad \cdots + M_\beta^\lambda(n_1, n_2, \ldots, n_K+1) = M_\beta^\lambda(n_1, n_2, \ldots, n_K), \end{aligned} \tag{5.78}$$

since we have $\sum_{k=1}^K p_k = 1$. With the moments defined in Eq. (5.75), the size s_λ and the credibility c_λ of a BLR $\mathcal{R}_\lambda$ for a K-outcome measurement can be simply expressed as the following,

$$s_\lambda(\beta) = \frac{M_\beta^\lambda(0,0,\ldots,0)}{M_\beta^0(0,0,\ldots,0)}, \quad c_\lambda(\beta) = \frac{M_\beta^\lambda(n_1, n_2, \ldots, n_K)}{M_\beta^0(n_1, n_2, \ldots, n_K)}. \tag{5.79}$$

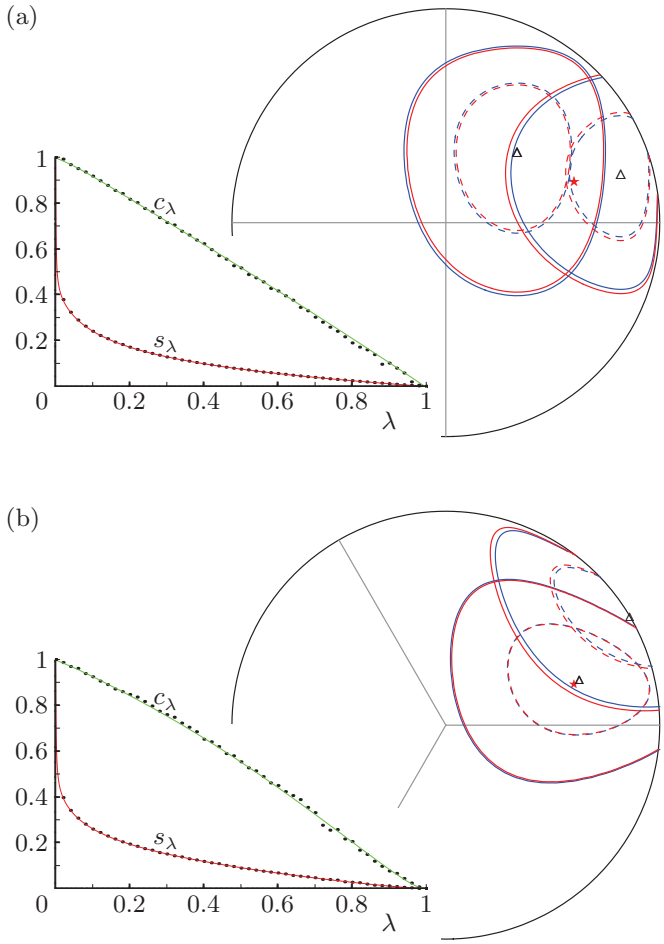


Figure 5.8: Smallest credible regions for simulated experiments. Twenty-four copies are measured by the POMs of Sec. 5.5.2.1, which have the unit disk of Fig. 5.6 as the reconstruction space. Plot (a) is for the four-outcome POM with the crosshair indicating the orientations of the two projective measurements. Plot (b) is for the three-outcome measurement with the orientation of the trine indicated. The red star ($\star$) at $(x, y) = (0.6, 0.2)$ marks the actual state that was used for the simulation. For each POM, there are SCRs for the data of two simulated experiments, with black triangles ($\triangle$) indicating the respective MLEs. The boundaries of the SCRs with credibility $c = 0.9$ are traced by the continuous lines; all of these SCRs contain the actual state. The dashed lines are the boundaries of the SCRs with credibility $c = 0.5$; the actual state is inside half of these SCRs. Red lines are for the primitive prior of Eq. (5.72), the blue lines are for the Jeffreys priors of Eqs. (5.73) and (5.74), respectively. — The insets in the lower left corners show the size s_λ and the credibility c_λ for the BLRs of two simulated experiments. Inset (a) is for $(6, 3, 10, 5)$ counts for the four-outcome POM and the Jeffreys prior; inset (b) is for $(13, 7, 4)$ counts for the three-outcome POM and the primitive prior. The dots show the values computed with a Monte Carlo algorithm. There is much more scatter in the c_λ values than the s_λ values. The red lines are fits to the s_λ values, with the fits using twice as many values than there are dots in the insets. The green lines that approximate the c_λ values are obtained from the red lines with the aid of Eq. (5.29).

The denominators in the two expressions with superscripts $\lambda = 0$ serve as normalization purposes, even if the prior distributions may not be normalized previously. When the parameter $\lambda = 0$, we are taking the whole state space into account, which gives the size and the credibility all equal to 1; if $\lambda = 1$, the region converges to a single point, which is exactly the MLE.

With the previous tools at hand, we write down the moments for the four-outcome POM with the Jeffreys prior as well as the primitive prior, in polar coordinates,

$$\begin{aligned} M^{\lambda}_{1/2}(n_1, n_2, n_3, n_4) &= \frac{1}{\pi} \int_{\mathcal{R}_\lambda} r \, \mathrm{d}r \, \mathrm{d}\phi \prod_{k=1}^{4} p_k^{n_k - 1/2}, \\ M^{\lambda}_{1}(n_1, n_2, n_3, n_4) &= \frac{1}{\pi} \int_{\mathcal{R}_\lambda} r \, \mathrm{d}r \, \mathrm{d}\phi \prod_{k=1}^{4} p_k^{n_k}, \end{aligned} \tag{5.80}$$

and for the trine POM with these two priors, respectively,

$$\begin{aligned} M^{\lambda}_{1/2}(n_1, n_2, n_3) &= \frac{1}{\pi} \int_{\mathcal{R}_\lambda} r \, \mathrm{d}r \, \mathrm{d}\phi \prod_{k=1}^{3} p_k^{n_k - 1/2}, \\ M^{\lambda}_{1}(n_1, n_2, n_3) &= \frac{1}{\pi} \int_{\mathcal{R}_\lambda} r \, \mathrm{d}r \, \mathrm{d}\phi \prod_{k=1}^{3} p_k^{n_k}. \end{aligned} \tag{5.81}$$

Then the size s_λ and the credibility c_λ are calculated by applying Eq. (5.79).

Figures 5.8(a) and 5.8(b) show SCRs obtained for simulated experiments in which $N = 24$ copies of a qubit state are measured. The actual state used for the simulation has $x = 0.6$ and $y = 0.2$. Its position in the reconstruction space is indicated by the red star ($\star$). In Fig. 5.8(a), we see the SCRs for the four-outcome POM. Two measurements were simulated, with $(n_1, n_2, n_3, n_4) = (8, 5, 10, 1)$ and $(6, 3, 10, 5)$ clicks of the detectors, respectively, and the triangles ($\triangle$) show the positions of the corresponding MLEs. For each data, the plot reports the SCRs with credibility $c = 0.5$ and $c = 0.9$, both for the primitive prior of Eq. (5.72) and for the Jeffreys prior of Eq. (5.73). The actual state is inside two of the four SCRs with credibility $c = 0.5$ and is contained in all four SCRs with credibility $c = 0.9$.

Not unexpectedly, we get quite different regions for the two rather different sets of

detector click counts. Yet, we observe that the choice of prior has little effect on the SCRs, although the total number of measured copies is too small for relying on the consistency of the priors. The same remarks apply to the SCRs for the three-outcome POM in Fig. 5.8(b); here we counted $(n_1, n_2, n_3) = (15, 8, 1)$ and $(13, 7, 4)$ detector clicks in the simulated experiments.

In Sec. 5.3.3, we remarked that the estimator regions are properly communicated by reporting s_λ and c_λ as functions of λ. This is accomplished by the insets in Fig. 5.8 for two of the four simulated experiments. The dots give the values obtained by numerical integration that uses an (adapted) Monte Carlo algorithm. The scatter of these numerical values confirms the expected: The computation of s_λ only requires sampling the probability space in accordance with the prior and determining the fraction of the sample that is in $\mathcal{R}_\lambda$; for the computation of c_λ we need to add the values of $L(D|\rho)$ for the sample points inside $\mathcal{R}_\lambda$; and since $L(D|\rho)$ is a sharply peaked function of the probabilities, the s_λ values are more trustworthy than the c_λ values for the same computational effort. The line fitted to the s_λ values is a Padé approximant (see, for example, section 5.12 in Ref. [167]) that have taken the analytic forms near $\lambda = \lambda_0 = 0$ and $\lambda = 1$ into account (see below). The line approximating the c_λ values is then computed in accordance with Eq. (5.29).

5.5.2.3 Analytic forms of s_λ near $\lambda = 0$ and $\lambda = 1$

The behaviors of s_λ near $\lambda = 0$ and $\lambda = 1$ can be analyzed in the general situation of having a probability space with dimension d_p. There are actually two different, yet quite similar scenarios. Here we only consider the case when (a small vicinity of) the MLE is contained inside the probability space. When $\lambda \lesssim 1$, it's easy to show that the size s_λ takes on the following form,

$$s_\lambda \propto \left(\log \frac{1}{\lambda}\right)^{d_\mathrm{p}/2} \ll 1, \tag{5.82}$$

which can be rewritten as

$$\log\left(1 - s_\lambda^{2/d_\mathrm{p}}\right) \propto \log \lambda\,. \tag{5.83}$$

Whereas when $\lambda \gtrsim 0$, the size s_λ is approximated by

$$s_\lambda \propto 1 - a\lambda^\alpha, \tag{5.84}$$

where a and α are two undetermined free factors, and again we rewrite it as

$$\log\left(1 - s_\lambda^{2/d_\mathrm{p}}\right) \propto \log\frac{2a}{d_\mathrm{p}} + \alpha\log\lambda\,. \tag{5.85}$$

By now, we see that both of the relations (5.83) and (5.85) are linear between $\log\left(1 - s_\lambda^{2/d_\mathrm{p}}\right)$ and $\log\lambda$, knowing which greatly helps interpret the data (see Fig. 5.9). Depending on the likelihood function, s_λ changes abruptly near $\lambda = 0$, while near $\lambda = 1$ the data are not trustworthy because of noise. The scenario when the MLE lies on the boundary of the probability space has quite similar results, with all the entries of d_p in (5.82)–(5.85) being replaced by $(d_\mathrm{p} + 1)$.

5.5.3 Incomplete two-qubit tomography

For a second illustration, we consider the situations that arise in the quantum-key-distribution schemes by Bennett and Brassard (BB84 [168]) and the trine-antitrine (TAT) scheme of Ref. [169]. Both schemes can be implemented by having a source of entangled qubit pairs distribute one qubit each to the two communicating parties. Prior to any key generation, the two-qubit state emitted by the source needs to be characterized. It is desirable to achieve quantum state estimation with reliable error regions without sacrificing many data that are then not available for the key generation.

5.5.3.1 POMs and computer-generated data

In the BB84 scheme, each qubit is measured by the crosshair POM of Eq. (5.61); the resulting two-qubit POM has sixteen outcomes that obey eight constraints that give delta-function factors in $w_\mathrm{cstr}(p)$. In the TAT scheme, one qubit is measured by the trine POM of Eq. (5.70) and the other qubit by the antitrine POM that has the signs of x and y reversed in Eq. (5.70); the resulting two-qubit POM has nine outcomes subject

to the single delta-function constraint of unit sum. Accordingly, the probability space is eight-dimensional for both schemes,[6] and we cannot report the SCRs by showing the optimal error regions in the reconstruction space with plots, as was possible for the two-dimensional probability space in Fig. 5.8. Therefore, we employ the strategy of Sec. 5.3.3 and report the size s_λ and the credibility c_λ of the respective BLRs as functions of λ.

For the generation of the simulated data, we first add noise to the singlet state by putting it through a random Pauli channel, which is used as a simple model for noise in a communication protocol. The channel acts on an input state as

$$\rho \to \sum_{jk} r_{jk} \left(\sigma_j \otimes \sigma_k\right) \rho \left(\sigma_j \otimes \sigma_k\right), \quad \text{with } j,k = 0,x,y,z, \tag{5.86}$$

where σ_0 denotes the single-qubit identity operator, and the r_{jk}s are sixteen randomly chosen probabilities. The 60 copies of the true state come from passing 60 copies of the singlet state through one instance of the random Pauli channel, *i.e.*, the r_{jk}s are randomly picked once, with r_{00} given a higher weight of 0.7 to simulate weak noise. The resulting true state has the probabilities for the two-qubit POMs given in the top row of Table 5.2. For example, the "12" entry in the 4×4 table for the double-crosshair POM is the probability for outcome $\Pi_1 \otimes \Pi_2 = \frac{1}{4}(1+\sigma_x) \otimes \frac{1}{4}(1-\sigma_x)$. The "11" entry of the 3×3 table for the trine-antitrine POM is 16/9 times that number. Note that all marginal probabilities (sums of rows and sums of columns) are equal; this is so because the reduced single-qubit states of the true state are completely mixed. For the same reason, both tables are symmetric and the lower-left and upper-right 2×2 subtables of the 4×4 table have entries of $1/16 = 0.0625$. More generally, there is a one-to-one correspondence between the 16 permissible probabilities in the 4×4 table and the 9 permissible probabilities in the 3×3 table, because all table entries are determined by

[6] In actual experiments, the probability space is nine-dimensional because one must account for the no-click probability of the qubit pairs that do not give rise to coincidence clicks. Further, the state estimation could also exploit the data collected for single-qubit detection without the coincidental detection of the partner qubit. Consistent with the footnote in Sec. 5.2.3, we are here content with the idealized situation of perfect detection devices, because our objective is to give an example for a higher-dimensional space, rather than evaluating real experimental data.

Table 5.2: Computer-generated data for the estimation of a two-qubit state from measuring 60 identically prepared copies. The first row gives the joint probabilities of the true state. The broken second row shows the number of detector-click pairs obtained in the simulated experiment (and their expected values) together with the single-qubit marginals. The third row reports the joint probabilities of the MLEs for the data in the second row. In each row, we have a 4×4 table on the left for the double-crosshair POM of the BB84 scenario and a 3×3 table on the right for the 9-outcome POM of the TAT scheme. The rows of a 4×4 table for the double-crosshair POM refer to the four Π_js of the first qubit in the pair and the columns refer to the Π_ks of the second qubit; entry "jk" is the probability for outcome $\Pi_j \otimes \Pi_k$. Analogously, entry "jk" in a 3×3 table for the TAT scheme is the expectation value of $\Pi_j \otimes \Pi_k$ with trine outcome Π_j and antitrine outcome Π_k.

	double-crosshair POM				trine-antitrine POM		
true-state probabilities	1	2	3	4	1	2	3
1	0.0206	0.1044	0.0625	0.0625	0.1856	0.0739	0.0739
2	0.1044	0.0206	0.0625	0.0625	0.0739	0.1848	0.0747
3	0.0625	0.0625	0.0212	0.1038	0.0739	0.0747	0.1848
4	0.0625	0.0625	0.1038	0.0212			

computer-generated data (expected number of clicks)	1	2	3	4	
1	0 (1.24)	4 (6.26)	6 (3.75)	4 (3.75)	14 (15)
2	6 (6.26)	3 (1.24)	8 (3.75)	4 (3.75)	23 (15)
3	3 (3.75)	1 (3.75)	0 (1.27)	8 (6.23)	12 (15)
4	1 (3.75)	7 (3.75)	4 (6.23)	1 (1.27)	13 (15)
	10 (15)	15 (15)	18 (15)	17 (15)	

	1	2	3	
1	11 (11.14)	4 (4.43)	5 (4.43)	20 (20)
2	2 (4.43)	10 (11.09)	5 (4.48)	17 (20)
3	4 (4.43)	6 (4.48)	13 (11.09)	23 (20)
	17 (20)	20 (20)	23 (20)	

MLE probabilities	1	2	3	4	1	2	3
1	0.0056	0.1012	0.0497	0.0571	0.1833	0.0667	0.0833
2	0.0939	0.0493	0.0821	0.0611	0.0333	0.1667	0.0833
3	0.0630	0.0344	0.0025	0.0949	0.0667	0.1000	0.2167
4	0.0365	0.1160	0.1293	0.0232			

the expectation values of $A \otimes B$ with $A, B = 1, \sigma_x, \sigma_y$.

Simulated measurements of 60 qubit pairs in the true state for each POM produced the counts of detector-click pairs in the second row of Table 5.2; expected values are

given in parentheses. Owing to the statistical fluctuations, the tables of counts are not symmetric[7] and the marginal counts are not equal.

The third row of Table 5.2 shows the corresponding MLE probabilities. These probabilities are equal to the relative frequencies of the counts for the 9-outcome POM, but are different from the relative frequencies for the 16-outcome POM. This tells us that the computer-generated data are not typical for the double-crosshair POM, whereas we have typical data for the trine-antitrine POM.

5.5.3.2 Size and credibility of the BLRs

As noted in Sec. 5.3.3, the primary task of the data evaluation is the computation of the multi-dimensional integrals that give the size s_λ of the BLRs for the whole range of $0 < \lambda < 1$. For the data in Table 5.2, these are integrals over eight-dimensional regions. We used a random-sampling technique for this purpose.

As a preparation, we generated a random sample of 648 785 permissible sets of probabilities, uniformly distributed in accordance with the primitive prior (see Sec. 5.5.3.3 below). In view of the one-to-one correspondence between the permissible probabilities of the 16-outcome POM and the 9-outcome POM, the same random sample can be, and was, used for both POMs.

The actual data processing consists of two steps. In the first step, we determine the size s_λ for the 161 values of λ with $-\log_{10}\lambda = 0.0\,(0.1)\,16.0$. This requires a simple counting of how many samples are inside the BLR $\mathcal{R}_\lambda$ if the primitive prior is used. In the case of the Jeffreys prior, one adds the weights $(p_1 p_2 \cdots)^{-1/2}$ of the samples inside the BLR. The correct normalization follows from $s_{\lambda=0} = 1$.

In the second step, the integrals needed in Eq. (5.29) are evaluated, for which a simple linear interpolation between adjacent (λ, s_λ) pairs is sufficiently accurate. Then, c_λ is known as a function of λ and the λ values for which we have 99% or 95% credibility are determined.

[7] One usually restores the symmetry by the so-called "twirling" before the key generation protocol is executed. The characterization of the source, however, should be done without the twirling.

Table 5.3: Threshold λ values for 99% and 95% credibility for the data of Table 5.2 and Fig. 5.9, and the sizes of the respective BLRs. The true state is inside the $\mathcal{R}_\lambda$s with $\lambda < 3.368 \times 10^{-3}$ for the 16-outcome POM (with its untypical data), and inside the BLRs with $\lambda < 0.2486$ for the 9-outcome POM.

	16-outcome POM			9-outcome POM		
primitive prior	λ	s_λ	c_λ	λ	s_λ	c_λ
	6.70×10^{-5}	0.0279	0.99	1.92×10^{-4}	0.0601	0.99
	6.03×10^{-4}	0.0106	0.95	1.44×10^{-3}	0.0268	0.95
Jeffreys prior	λ	s_λ	c_λ	λ	s_λ	c_λ
	1.73×10^{-4}	0.0374	0.99	6.74×10^{-5}	0.0186	0.99
	1.35×10^{-3}	0.0161	0.95	6.20×10^{-4}	0.0070	0.95

We show s_λ and c_λ as functions of λ in Fig. 5.9. Table 5.3 reports the λ values of the 99% and 95% credibility thresholds. We observe that for the 16-outcome POM, the true state is inside the SCRs with 99% credibility for both the primitive prior and the Jeffreys prior, whereas it is inside the 95% SCR only for the primitive prior but not for the Jeffreys prior. This is more evidence that these data are untypical. By contrast, for the 9-outcome POM, the true state is inside all SCRs for both priors and both values of the credibility.

Typicality, or lack thereof, can also be noticed in Fig. 5.9. Since the Jeffreys prior gives more weight to the regions near the boundary of the probability space than the primitive prior, and less weight to regions deep inside, one expects that the values of s_λ for the primitive prior are larger than those for the Jeffreys prior if the data are typical and, accordingly, the MLE is not close to the boundary. This is indeed the case for the trine-antitrine data, but not for the double-crosshair data.

5.5.3.3 Numerical effort

The two steps of data evaluation, the computation of the size s_λ and then the credibility c_λ, take a few seconds of CPU time. The preparation of the random sample of permissible probabilities, which could be done ahead of the data taking, lasts much longer. For each potential sample of probabilities, we first generate nine random numbers $x_1, x_2, \ldots, x_9$ uniformly and independently between 0 and 1. Then, the nine

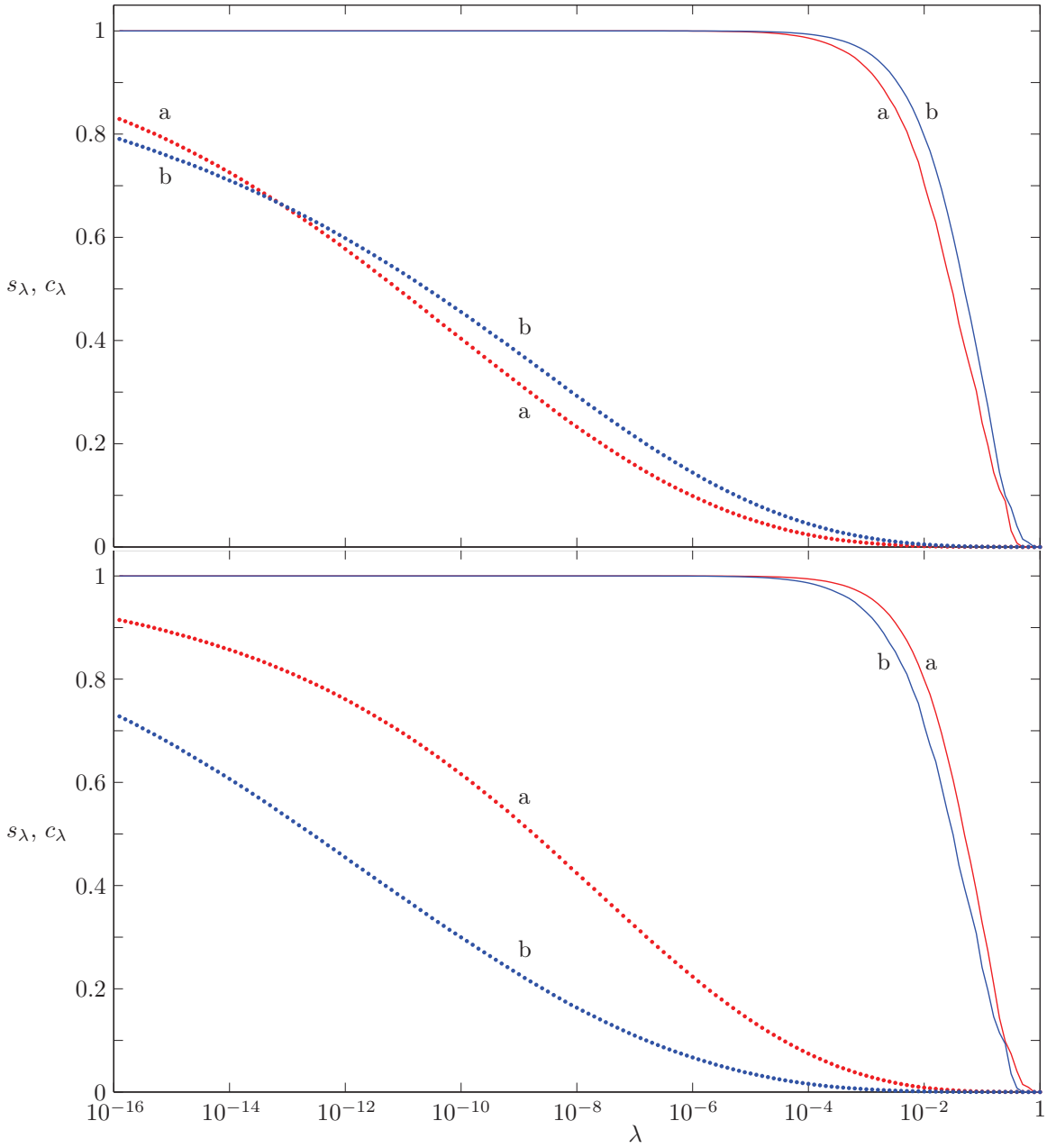


Figure 5.9: The size s_λ (dotted lines) and the credibility c_λ (solid lines) as functions of λ for the data of Table 5.2. The top plot is for the double-crosshair POM, the bottom plot is for the trine-antitrine POM; curves 'a' are for the primitive prior, curves 'b' are for the Jeffreys prior. The abscissa is linear in $\log\lambda$. For $\lambda \lesssim 1$, the BLRs are so small that only very few sample points are inside and the sizes s_λ have comparatively large fluctuation errors. This statistical noise is visible in the bottom-right corners of the plots. It has no bearing, however, on the accuracy of the credibility c_λ in the important range of smaller λ values, as one notes upon recalling Fig. 5.4.

probabilities are chosen according to

$$p_k = \frac{\log x_k}{\sum_{k=1}^{9} \log x_k}\,, \quad \text{with } k = 1, 2, ..., 9, \tag{5.87}$$

which constitute a sample in the eight-dimensional simplex of the classical nine-sided die, and the samples are distributed in accordance with the primitive prior. This can be checked easily by showing that

$$\begin{aligned} &\mathrm{d}p_1 \cdots \mathrm{d}p_9 \int_0^1 \mathrm{d}x_1 \cdots \int_0^1 \mathrm{d}x_9\, \delta\left(p_1 - \frac{\log x_1}{\sum_{k=1}^{9} \log x_k}\right) \cdots \delta\left(p_9 - \frac{\log x_9}{\sum_{k=1}^{9} \log x_k}\right) \\ &= \mathrm{d}p_1 \cdots \mathrm{d}p_9\, \delta\left(\sum_{k=1}^{9} p_k - 1\right). \end{aligned} \tag{5.88}$$

The sample $p = (p_1, p_2, \ldots, p_9)$ is accepted if it is a permissible set of probabilities for the trine-antitrine POM with its nine outcomes. Whereas the generation of another sample p is fast, the test of permissibility is the part that consumes most of the CPU time. After identifying the candidate p with the relative frequencies of a measurement with the 9-outcome POM, we calculate a MLE for these frequencies. If the probabilities of the MLE are equal to p, this sample probability is accepted, otherwise it is rejected. In the sampling of this example, only 9.27% of the 7×10^6 candidate probabilities generated were accepted.

Since random-sampling techniques are the methods of choice, our sample of 648 785 probabilities took almost 100 hours of CPU time[8] on a standard desktop (Intel i7-870 CPU, using one of the four cores and 8 GB RAM). The procedure of random sampling that we employed was simple and reliable but not optimized for speed. There is clearly much room for improvement. For instance, one may try to parallelize the sampling over many different computers and later combine into a single dataset. The chance that a candidate probability is permissible can be much increased by cleverer Monte Carlo methods where one makes use of information at the current physical point to

[8]Since the completion of this thesis, we have improved the sampling algorithm and thereby reduced the CPU time consumed to less than 10% of what it was before. This speed-up results from a faster test of permissibility that is still easy to implement. Further improvements are likely.

stay within the physical state space. It is also worth noting that this computational time is an overhead that is incurred only once and the sampling can be done ahead of any actual data-taking in the laboratory.

5.6 Summary

For the given data and chosen size or credibility, the MLR or the SCR is a neighborhood of the MLE. In this sense, one can regard them as systematically constructed error regions for the MLE. While there are efficient methods for computing the MLE [4, 97], we are currently lacking equally efficient algorithms for finding the MLR and the SCR. Progress on this front is needed before one can apply the concepts of MLRs and SCRs to situations in which the reconstruction space is of high dimension. Upon recalling that IC POMs for two-qubit systems already have a 15-dimensional reconstruction space, the need for powerful numerical schemes is utterly plain.

In many applications, one is interested in a few parameters only, perhaps a single one, such as the concurrence of a two-qubit state or its fidelity with a target state (see, for example, Refs. [52, 170–174]). It may then be possible to reduce the dimensionality of the problem by marginalizing the nuisance parameters, preferably proceeding from a utility-based prior. A variant of the methodology described here can be used to determine small regions of high credibility in the few-parameter space of interest, without first determining SCRs in the reconstruction space.

Even after such a reduction, there remains the challenge of evaluating the multi-dimensional integrals that tell us the size of the BLRs, and then their credibility, so that we can identify the looked-for MLR and SCR. For this purpose, one needs good sampling strategies [160]. Markov-chain Monte Carlo (MCMC) methods, such as the Metropolis-Hastings algorithm, suggest themselves (see, for example, section 15.8 in Ref. [167]). After surveying many kinds of MCMC strategies, we found that the Langevin MC method might suit our problem better than others. Efforts in this direction will continue. It is also suggestive to rely on the data themselves for guidance. The full sequence of detector clicks identifies the MLE of the data, and subsequences—

chosen randomly or systematically—have their own MLEs. These bootstrapped MLEs are expected to accumulate in the vicinity of the full-data MLE and may so provide a useful sampling method.

We close this chapter with a general observation. MLEs, MLRs, SCRs, and confidence regions are concepts of statistics, even if the terminology is not universal. As we have seen, the quantum aspect of the state estimation problem enters only through the Born rule which restricts the probabilities to those obtainable from a POM and a bona fide statistical operator. Except for these restrictions, there is no difference between state estimation in quantum mechanics and standard statistics. Accordingly, quantum mechanicians can benefit much from the methods developed by statisticians.

Chapter 6

Conclusion and Outlook

In quantum information theory, the problem of constructing SIC POMs in various dimensions is considered to be hard, especially when the dimension d grows larger. Zauner's conjecture states that SIC POMs exist in every finite dimension. Although a great deal of numerical evidence strongly supports it (at least for the group-covariant SIC POMs), a rigorous proof for this conjecture is still missing. In the last two decades, a lot of work, both analytical and numerical, has been devoted to the construction of SIC POMs in various dimensions, not only because SIC POMs have the nice property of high symmetry and high tomographic efficiency, but also because they are closely related to many other problems in both physics and mathematics, such as MUB, equiangular lines, Lie algebras, and so on. Therefore, a deeper and more thorough understanding of SIC POMs may also help solve these problems.

Nevertheless, in contrast to the major theoretical progress, another aspect concerning SIC POMs is their implementation. Up to date, all experiments and even proposals for experiments implementing SIC POMs have been limited to the very basic quantum system of a qubit, with the exception of the recent experiment by Medendorp *et al.* [110], where a SIC POM for a three-level system was approximated. This is, in part, due to the fact that there is no systematic procedure for implementing SIC POMs in higher dimensions in a simple experimental setup.

As a contribution to solve the above problems, in this thesis, we introduced the successive-measurement scheme. We propose to implement the SIC POMs by breaking the measurement process into two steps, having in mind that each step should be relatively easy to implement. Based on this idea, we present a systematic procedure that implements all HW SIC POMs in finite-dimensional systems. The implementation

is accomplished by a diagonal-operator measurement with high-rank outcomes followed by a rank-1 measurement in the Fourier basis. As an example, we have considered the realization of HW SIC POMs for a path qudit encoded in a single photon. Moreover, we found that if we take the first measurement to be fuzzy and let the bases for the second measurement be chosen in accordance with the result of the first measurement, then in the particular cases studied (dimensions 2, 3, and 4) an operational link between SIC POMs and MUB appears: The MUB are used to construct the SIC POMs. A similar link was found in dimension 8 as well, but in this case the first measurement was not of the fuzzy kind.

Moreover, we propose a feasible experimental scheme that implements the SIC POM for a two-qubit system. Our scheme uses linear optical elements and photodetectors, and is, therefore, well within the reach of current technology. The proposal is also based on the successive-measurement approach to SIC POMs. We found that the SIC POM for the qubit pair corresponds to a POM diagonal in the computational basis, followed by projections onto bases which are mutually unbiased. We observe that this unique construction is due to a structural relation between the fiducial vectors and the MUB in dimension 4.

On a more general note, we believe that it would be interesting to learn if and how this scheme can be generalized to higher dimensions. Such a study could be of theoretical as well as practical use; it might teach us about the SIC POMs' structure in high dimensions and provide new ideas for implementing them. Besides, the relation we found between SIC POMs and MUB may provide a hint to prove the existence of SIC POMs (at least, we hope) in prime power dimensions, in which circumstance a complete set of MUB does exist and can be easily constructed. In addition, there is still an open question as to the generality of such a relation and its origin. Currently, it is unclear whether the successive-measurement approach will provide a reasonable scheme for implementing SIC POMs in arbitrary dimensions and thus reveal their structure in high-dimensional Hilbert spaces. We have tried to construct the SIC POMs in dimension 16 using the successive-measurement scheme, but with no success.

The second main topic of this thesis investigates optimal error regions of estimators for quantum state tomography. A point estimator is a state that represents one's best guess of the actual state of the unknown quantum system for the given data. To be statistically meaningful, estimators have to be endowed with error regions, the generalization of "error bars" beyond one dimension. For this purpose, we propose maximum-likelihood regions (MLRs) and smallest credible regions (SCRs). These are regions in the space of quantum states. The MLR is that region of pre-chosen size, for which the given data are more likely than for any other region of the same size. The SCR is the smallest region with pre-chosen credibility—the credibility of a region being its posterior probability, that is: the probability of finding the actual state in the region, conditioned on the data. Whether one chooses the MLR or the SCR as the optimal error region depends on the situation at hand.

Central to both concepts is the notion of the size of a region. In the context of state estimation, it is most natural to measure the size of a region by its prior—before any data are at hand—probability of finding the actual state in the region: Regions with the same prior probability are considered as having the same size. The size of a region hence expresses the relative importance of that region of states. Ultimately, the choice of prior is up to the user, but it should be consistent: The estimation results should be dominated by the data, not the prior, if many copies of the state are measured.

As we show, the problems of finding the MLR and the SCR are duals of each other. In both cases, the optimal regions contain all states for which the likelihood of the data exceeds a threshold value. This provides a simple and concise way of communicating one's uncertainty of the estimate. That the optimal error regions possess such a simple description is surprising, since our construction imposes no restriction on the shape of the regions to be considered. The shape of the optimal regions are uniquely determined by the likelihood function, in sharp contrast to the arbitrariness in the shape of a confidence region. Yet the two are not unrelated: Our SCRs provide natural starting points for the construction of the confidence regions.

While the chosen MLR or SCR depends on the prior, the set of candidate regions

is prior-independent: It depends only on the likelihood function for the given data. Also reassuring is the fact that every MLR or SCR is a small vicinity of the MLE, in the respective limits of small size or small credibility. This is reminiscent of standard ellipsoidal error regions constructed around the MLE, but which are applicable only in the limit of a large amount of data when the central limit theorem can be invoked and the uncertainty can be characterized by the Fisher information.

While there are efficient methods for computing the MLE, we are currently lacking equally efficient algorithms for finding the MLR and the SCR—there remains the challenge of evaluating the multi-dimensional integrals that give s_λ. For this, one needs good sampling strategies. Markov-chain Monte Carlo methods, such as the Metropolis-Hastings algorithm, suggest themselves. Progress in this aspect has been made, and in the mean time we are seeking possible applications of the algorithm.

Often, only a few parameters computed from the state are of interest. It is then possible to reduce the dimensionality of the problem by discarding nuisance parameters. A variant of the methodology described here can be used to determine small regions of high credibility in the few-parameter space of interest, without first determining SCRs in the reconstruction space. This direction is currently under investigation.

Appendix A

Finite Fields

In abstract algebra, a *finite field* or *Galois field* is a field that contains a finite number of elements. Finite fields are important in many subjects, such as number theory, coding theory, cryptography, and quantum error correction [175]. The construction of maximal sets of MUB in prime power dimensions [114, 135] also makes use of the properties of finite fields. For the purpose of this thesis (Sec. 3.5 to be specific), we give a very brief description of finite fields. More details on this topic can be found in, for instance, Refs. [114, 176].

The number of elements of a finite field is a prime power, and for $d = p^{\mathrm{M}}$, with p a prime number and $\mathrm{M} \in \mathbb{Z}^+$, there exists one and only one field F (up to isomorphism) with order $|F| = d$. In particular, a field P of prime order p can be identified with the field $\mathbb{Z}/p\mathbb{Z}$ of residues modulo p, and a field F with $d = p^{\mathrm{M}}$ can be regarded as the splitting field over P of the polynomial $x^d - x$. More explicitly, every element i of F can be represented by M-tuples $(i_0, i_1, \ldots, i_{\mathrm{M}-1})$ of integers, with each integer running from 0 to $p-1$, that we get from the p-ary expansion of i:

$$i = (i_0, i_1, \ldots, i_{\mathrm{M}-1}) \quad \text{if} \quad i = \sum_{n=0}^{\mathrm{M}-1} i_n p^n. \tag{A.1}$$

Each field is characterized by two operations, an addition and a multiplication, that we shall denote by $\oplus$ and $\odot$ respectively. The field addition operation $\oplus$ is equivalent to the component-wise addition modulo p, that is

$$i = j \oplus k \Leftrightarrow i_n = j_n + k_n \,(\mathrm{mod}\, p)\,, \tag{A.2}$$

for $n = 0, 1, \ldots, \mathrm{M} - 1$. As a consequence, the summation in Eq. (A.1) is also a field summation, such that,

$$i = (i_0 p^0) \oplus (i_1 p^1) \oplus \cdots \oplus (i_{\mathrm{M}-1} p^{\mathrm{M}-1}) = \bigoplus_{n=0}^{\mathrm{M}-1} i_n p^n. \tag{A.3}$$

The inverse of element i relative to the field addition operation is denoted as $\ominus i$, and one may also consider the symbol $\ominus$ as the field subtraction operation.

Unfortunately, there is no similarly simple convention for the field multiplication operation $\odot$ as that for the addition operation $\oplus$, except for $d = p$ and $d = 4$. However, in view of the associative and distributive nature of $\odot$, that is: $(a \odot b) \odot c = a \odot (b \odot c)$ and $(a \oplus c) \odot c = (a \odot c) \oplus (b \odot c)$, respectively, we only need to state the values of $p^j \odot p^k$. For $\mathrm{M} = 1$ and $d = p$, the field multiplication is just multiplication modulo p. For $\mathrm{M} > 1$, we have the Galois construction

$$p^j \odot p^k = \begin{cases} p^{j+k} & \text{if } j + k < \mathrm{M}\,, \\ \displaystyle\sum_{l=0}^{\mathrm{M}-1} \mu_l p^l & \text{if } j + k = \mathrm{M}\,, \\ p \odot (p^{j-1} \odot p^k) & \text{recursively, if } j + k > \mathrm{M}\,, \end{cases} \tag{A.4}$$

where the coefficients $\mu_l \in \mathbb{Z}/p\mathbb{Z}$ that define the $j + k = \mathrm{M}$ products are restricted by the following requirement that

$$x \mapsto x^{\mathrm{M}} - \sum_{l=0}^{\mathrm{M}-1} \mu_l x^l, \tag{A.5}$$

which is an irreducible polynomial over the Galois field with p elements. Similarly as the addition operation, one may define the inverse of a nonzero element i relative to the multiplication operation $\odot$ to be $\oslash i$, and treat the symbol $\oslash$ as the field division operation.

Appendix B

Quantum gates

Analogous to the way a classical computer is built from an electrical circuit containing wires and logic gates, a quantum computer is built from a quantum circuit containing wires and elementary quantum gates to carry around and manipulate the quantum information. Basically, quantum gates amount to unitary transformations of the quantum states. In this appendix, we describe some simple single qubit gates as well as controlled gates for several qubits. More details see, for example, Refs. [2, 153].

B.1 Single qubit gates

As the simplest quantum system of all, a single qubit is represented by a vector $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ parameterized by two complex numbers satisfying $\left|\alpha\right|^2 + \left|\beta\right|^2 = 1$. Thus operations on a single qubit are described by 2×2 unitary matrices, of which Pauli matrices are the most familiar ones:

$$X \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \quad Y \equiv \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}; \quad Z \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{B.1}$$

Note that Pauli-X gate is also called the quantum NOT gate as the role of $|0\rangle$ and $|1\rangle$ in state $|\psi\rangle$ will be interchanged after the operation; Pauli-Z gate flips the sign of $|1\rangle$ to give $-|1\rangle$, while leaves $|0\rangle$ unchanged. Three other quantum gates also play an important role, the Hadamard gate (denoted by H), phase gate (denoted by S), and $\pi/8$ gate (denoted by T):

$$H \equiv \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}; \quad S \equiv \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{i} \end{bmatrix}; \quad T \equiv \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{e}^{\mathrm{i}\pi/4} \end{bmatrix}. \tag{B.2}$$

The Pauli-Z gate, phase gate, and $\pi/8$ gate are three special cases of the family of the phase shift gates (denoted by R):

$$R \equiv \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{e}^{\mathrm{i}\theta} \end{bmatrix}, \tag{B.3}$$

with $\theta = \pi$ being the Pauli-Z gate, $\theta = \pi/2$ being the phase gate S, and $\theta = \pi/4$ being the $\pi/8$ gate T.

A couple of useful algebraic relations are that $S = T^2$, $H = (X + Z)/\sqrt{2}$ and $H^2 = S^\dagger S = T^\dagger T = I_2$, where I_2 is the identity matrix in dimension 2. Here, we use the following chart to illustrate the operations of these single qubit gates on the qubit vector $|\psi\rangle$:

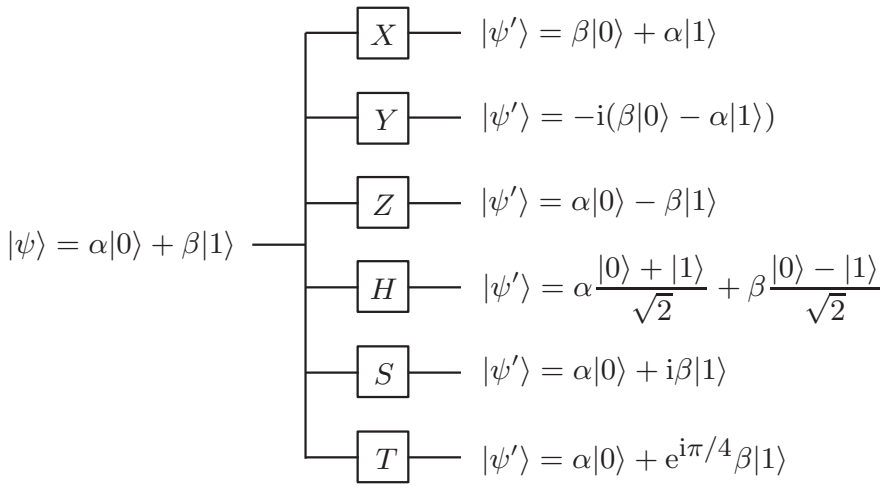


Figure B.1: Symbols of the most common single qubit gates as well as their actions on the qubit vector $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$.

B.2 Controlled gates

Controlled operations may be applied to two or more qubits, but here we only consider the quantum gates with two input qubits. Generally, a two-qubit controlled-U gate has the following form

$$C_{\boldsymbol{U}} = \begin{bmatrix} \boldsymbol{I}_2 & 0 \\ 0 & \boldsymbol{U} \end{bmatrix} = |0\rangle\langle 0| \otimes \boldsymbol{I}_2 + |1\rangle\langle 1| \otimes \boldsymbol{U}. \tag{B.4}$$

In terms of the computational basis, the action of the controlled-U gate is represented as $|c\rangle|t\rangle \rightarrow |c\rangle U|t\rangle$; when the control qubit $|c\rangle = |0\rangle$, the target qubit $|t\rangle$ passes through the gate unchanged, whereas when the control qubit $|c\rangle = |1\rangle$, the operator U is applied to the target qubit.

*Controlled-*NOT (CNOT or XOR) gate: This is the prototypical controlled gate for two qubits, which in matrix form is

$$\text{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{B.5}$$

Writing it compactly, the action of the CNOT gate on two qubits is given by $\text{CNOT}|c\rangle|t\rangle = |c\rangle|c \oplus t\rangle$ with $c, t \in \{0, 1\}$.

Controlled-phase (CP) gate: In Secs. 4.5 and 4.6 of Chapter 4, we meet the controlled-Z (CZ) gate, which in matrix representation is

$$\text{CZ} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{B.6}$$

In a compact form, the action of the CZ gate on two qubits is $\text{CZ}|c\rangle|t\rangle = |c\rangle(-1)^c|t\rangle$ with $c, t \in \{0, 1\}$.

A set of gates is said to be universal for quantum computation if any unitary operation may be approximated to arbitrary accuracy by a quantum circuit involving only those gates [2]. It has been shown that the following three sets of gates are universal: (1) two-level unitary gates, (2) single qubit and CNOT gates, and (3) Hadamard, phase, CNOT and $\pi/8$ gates.

Appendix C

Distance and distinguishability measures

In quantum state tomography, how good the estimator is has to be answered by using certain distance or distinguishability measures. In this appendix, we briefly review some most often used candidates, such as the trace distance, the Hilbert-Schmidt distance, the fidelity, the Bures distance and the relative entropy. See Refs. [2,100,177] for more detailed discussions.

C.1 Trace distance and Hilbert-Schmidt distance

Classically, for two probability distributions $\{p(x)\}$ and $\{q(x)\}$ over the same index set x, we define the *trace distance* as follows

$$D_{\mathrm{cl}}(p(x), q(x)) \equiv \frac{1}{2} \sum_{x} |p(x) - q(x)| \,. \tag{C.1}$$

This quantity is also known as the L_1 *distance* or *Kolmogorov distance*. Analogously, the *trace distance* between two quantum states ρ and σ is defined as

$$D_{\mathrm{tr}}(\rho, \sigma) \equiv \frac{1}{2} \mathrm{tr}\{|\rho - \sigma|\} \,. \tag{C.2}$$

It is one of the most common figures of merit used in quantum state tomography, especially in experiments, because it has a nice operational interpretation, which is best manifested in a state discrimination problem. The trace distance between two given states determines how well they can be distinguished from each other by the optimal

strategy. Notice also that if ρ and σ commute then the quantum trace distance between them is equal to the classical trace distance between the eigenvalues of ρ and σ.

The *Hilbert-Schmidt (HS) distance* between ρ and σ is induced by the HS inner product among operators

$$D_{\mathrm{HS}}(\rho,\sigma) \equiv \sqrt{\mathrm{tr}\{(\rho-\sigma)^2\}}\,. \tag{C.3}$$

It is the Euclidean distance between ρ and σ viewed as vectors in the space of Hermitian operators. When both ρ and σ are diagonal, it reduces to the L_2 *distance* between the diagonals of the two states, respectively.

C.2 Fidelity and Bures distance

The notion of *fidelity* also originates in classical probability theory, which, for two probability distributions $\{p(x)\}$ and $\{q(x)\}$, is defined by

$$F_{\mathrm{cl}}(p(x),q(x)) \equiv \sum_x \sqrt{p(x)q(x)}\,. \tag{C.4}$$

Similarly, we define the *fidelity* between two quantum states ρ and σ as [2][1]

$$F_{\mathrm{qu}}(\rho,\sigma) \equiv \mathrm{tr}\left\{\sqrt{\rho^{1/2}\sigma\rho^{1/2}}\right\}, \tag{C.5}$$

or equivalently,

$$F_{\mathrm{qu}}(\rho,\sigma) \equiv \mathrm{tr}\left\{\left|\rho^{1/2}\sigma^{1/2}\right|\right\}. \tag{C.6}$$

Clearly, the second definition shows that the fidelity is symmetric with respect to the two states. Notice that both the classical and quantum versions of fidelity are not metrics, although they do give rise to other useful metrics. There are three special cases where it is possible to give more explicit formulae for the fidelity. The first is

[1] It should be noted that some authors define the fidelity with a square [100, 177], which is called the squared fidelity according to our definition.

when ρ and σ commute, the quantum fidelity $F_{\mathrm{qu}}(\rho,\sigma)$ reduces to the classical fidelity between the eigenvalue distributions of ρ and σ. The second one is when σ is a pure state $|\psi\rangle$, the formula can be simplified to $F_{\mathrm{qu}}(\rho,|\psi\rangle\langle\psi|)=\sqrt{\langle\psi|\rho|\psi\rangle}$. The third special case concerns the fidelity between two single-qubit states [178],

$$F_{\mathrm{qu}}(\rho(\boldsymbol{r}_1),\sigma(\boldsymbol{r}_2)) = \frac{1}{2}\Big(1+\boldsymbol{r}_1\cdot\boldsymbol{r}_2+\sqrt{(1-\boldsymbol{r}_1\cdot\boldsymbol{r}_1)(1-\boldsymbol{r}_2\cdot\boldsymbol{r}_2)}\Big), \tag{C.7}$$

where we have used the Bloch-ball representation for qubits of Eq. (2.1), and $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the Bloch vectors for the two qubits respectively. For a nice geometric observation of the fidelity between these two single-qubit states, see Ref. [179].

According to Uhlmann's theorem [177], $F_{\mathrm{qu}}(\rho,\sigma)$ is equal to the maximal transition probability between *purifications* of ρ and σ,

$$F_{\mathrm{qu}}(\rho,\sigma) = \max_{|\psi_\rho\rangle,|\psi_\sigma\rangle} |\langle\psi_\rho|\psi_\sigma\rangle|\,, \tag{C.8}$$

where $|\psi_\rho\rangle$ is a purification of ρ. Therefore, in general, the fidelity is the maximum overlap between purifications. This formula makes it clear that the fidelity is symmetric in its inputs, $F_{\mathrm{qu}}(\rho,\sigma)=F_{\mathrm{qu}}(\sigma,\rho)$, and that the fidelity is bounded between 0 and 1. The minimum is attained if and only if ρ and σ have support on orthogonal subspaces, meaning that they are perfectly distinguishable; while the maximum is saturated if and only if $\rho=\sigma$, which can be seen from Uhlmann's formula [2]. Other important properties enjoyed by quantum fidelity include unitary invariance, concavity, multiplicativity and joint concavity; see, for instance, Ref. [180].

The fidelity can be used to define the *Bures distance* $D_{\mathrm{B}}(\rho,\sigma)$,

$$D_{\mathrm{B}}^2(\rho,\sigma) \equiv 2-2F_{\mathrm{qu}}(\rho,\sigma)\,, \tag{C.9}$$

which is a metric on density operators. When both ρ and σ are diagonal, the Bures distance reduces to the *Hellinger distance* between the diagonals of ρ and σ [2]. When $\rho=\mathrm{diag}(\lambda_0,\lambda_1,\ldots,\lambda_{d-1})$ has full rank and σ is infinitesimally apart, the Bures

distance is explicitly given by [181]

$$D_{\mathrm{B}}^2(\rho, \rho + d\rho) = \frac{1}{2} \sum_{j,k} \frac{|\langle j|\mathrm{d}\rho|k\rangle|^2}{\lambda_j + \lambda_k} \,. \tag{C.10}$$

Like its classical counterpart, the infinitesimal Bures distance has a clear operational meaning as it determines how well two nearby quantum states can be distinguished [182]. In addition, the Riemannian metric defined by this distance is equivalent to the metric defined by the SLD quantum Fisher information matrix [182].

The fidelity $F_{\mathrm{qu}}(\rho, \sigma)$ and the trace distance $D_{\mathrm{tr}}(\rho, \sigma)$ are closely related, despite their very different forms. Here, we simply report the relationship between them, which is given by

$$1 - F_{\mathrm{qu}}(\rho, \sigma) \leq D_{\mathrm{tr}}(\rho, \sigma) \leq \sqrt{1 - F_{\mathrm{qu}}(\rho, \sigma)^2} \,. \tag{C.11}$$

See Ref. [2] for a rigorous proof of this inequality. The implication is that the trace distance and the fidelity are qualitatively equivalent measures of closeness for quantum states. In many circumstances, it does not matter whether the trace distance or the fidelity is used to quantify distance, since results about one may be used to deduce equivalent results about the other.

C.3 Relative entropy

Entropy is a key concept of quantum information theory [2]. It measures how much uncertainty there is in the state of a physical system. Here, we only review its applications in distinguishing quantum states, *i.e.*, the relative entropy.

The *relative entropy* or *Kullback-Leibler divergence* is a very useful entropy-like measure of the closeness of two probability distributions, $\{p(x)\}$ and $\{q(x)\}$, over the same index set x. Being a non-symmetric measure, the relative entropy of $\{p(x)\}$ to $\{q(x)\}$ is defined by

$$H(p(x)||q(x)) \equiv \sum_x p(x) \log \frac{p(x)}{q(x)} \equiv -H(X) - \sum_x p(x) \log q(x) \,, \tag{C.12}$$

where $H(X) \equiv -\sum_x p(x) \log p(x)$ is known as the *Shannon entropy* associated with the probability distribution $\{p(x)\}$. Note that we define $-0 \log 0 \equiv 0$ and $-p(x) \log 0 \equiv +\infty$ if $p(x) > 0$. The relative entropy is non-negative, $H(p(x)||q(x)) \geq 0$, with equality holds if and only if $p(x) = q(x)$ for all x.

In quantum mechanics, we have the *von Neumann entropy* of a quantum state ρ defined analogously as the Shannon entropy, such that

$$S(\rho) \equiv -\mathrm{tr}\{\rho \log \rho\}\,, \tag{C.13}$$

with the logarithms based on 2. If λ_x are the eigenvalues of ρ, then the von Neumann entropy can be re-expressed as

$$S(\rho) = -\sum_x \lambda_x \log \lambda_x\,, \tag{C.14}$$

where again we define $0 \log 0 \equiv 0$, as that for the Shannon entropy. Then the *quantum relative entropy* of ρ to σ is define by

$$S(\rho||\sigma) \equiv \mathrm{tr}\{\rho \log \rho\} - \mathrm{tr}\{\rho \log \sigma\}. \tag{C.15}$$

Similarly, the quantum version of the relative entropy is non-negative, $S(\rho||\sigma) \geq 0$ (known as the *Klein's inequality*), with equality holds if and only if $\rho = \sigma$.

Bibliography

[1] B.-G. Englert. *Lectures on Quantum Mechanics.* World Scientific Publishing Co. Pte. Ltd., Singapore, 2006.

[2] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, Cambridge, 2010.

[3] D. J. Griffiths. *Introduction to Quantum Mechanics.* 2nd ed., Pearson Prentice Hall, NJ, 2004.

[4] M. G. A. Paris and J. Řeháček (Eds.). *Quantum State Estimation*, volume 649 of *Lecture Notes in Physics.* Springer-Verlag, Berlin Heidelberg, 2004.

[5] W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeit. für Physik*, 43:172–198, 1927.

[6] W. Heisenberg. *The Physical Principles of Quantum Theory.* The University of Chicago Press, Chicago, IL, 1930.

[7] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.

[8] H. P. Yuen. Amplification of quantum states and noiseless photon amplifiers. *Phys. Lett. A*, 113(8):405–407, 1986.

[9] R. A. Fisher. Theory of statistical estimation. *Math. Proc. Cambr. Phil. Soc.*, 22:700–725, 1925.

[10] C. R. Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Cal. Math. Soc.*, 37:81–91, 1945.

[11] H. Cramér. *Mathematical Methods of Statistics.* Princeton University Press, Princeton, NJ, 1946.

[12] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves. Symmetric informationally complete quantum measurements. *J. Math. Phys.*, 45:2171, 2004.

[13] G. Zauner. Quantum designs: Foundations of a noncommutative design theory. *Int. J. Quantum Inf.*, 9:445–507, 2011.

[14] J. Řeháček, B.-G. Englert, and D. Kaszlikowski. Minimal qubit tomography. *Phys. Rev. A*, 70:052321, 2004.

[15] D. M. Appleby. Symmetric informationally complete-positive operator valued measures and the extended Clifford group. *J. Math. Phys.*, 46:052107, 2005.

[16] H. Zhu and B.-G. Englert. Quantum state tomography with fully symmetric measurements and product measurements. *Phys. Rev. A*, 84:022327, 2011.

[17] A. J. Scott. Tight informationally complete quantum measurements. *J. Phys. A: Math. Gen.*, 39:13507, 2006.

[18] A. J. Scott and M. Grassl. Symmetric informationally complete positive-operator-valued measures: A new computer study. *J. Math. Phys.*, 51:042203, 2010.

[19] T. Durt, C. Kurtsiefer, A. Lamas-Linares, and A. Ling. Wigner tomography of two-qubit states and quantum cryptography. *Phys. Rev. A*, 78:042338, 2008.

[20] B.-G. Englert, D. Kaszlikowski, H. K. Ng, W. K. Chua, J. Řeháček, and J. Anders. Efficient and robust quantum key distribution with minimal state tomography. *eprint arXiv:quant-ph/0412075v4*, 2008.

[21] O. Albouy and M. R. Kibler. A unified approach to SIC-POVMs and MUBs. *J. Russ. Laser Res.*, 28:429–438, 2007.

[22] D. M. Appleby. SIC-POVMs and MUBs: Geometrical relationships in prime dimension. *AIP Conf. Proc.*, 1101:223, 2009.

[23] A. Kalev, J. Shang, and B.-G. Englert. Experimental proposal for symmetric minimal two-qubit state tomography. *Phys. Rev. A*, 85:052115, 2012.

[24] A. Kalev, J. Shang, and B.-G. Englert. Symmetric minimal quantum tomography by successive measurements. *Phys. Rev. A*, 85:052116, 2012.

[25] J. Shang, H. K. Ng, A. Sehrawat, X. Li, and B.-G. Englert. Optimal error regions for quantum state estimation. *New J. Phys.*, 15:123026, 2013.

[26] W. Pauli. *General Principles of Quantum Mechanics*. Springer-Verlag, Berlin, 1990.

[27] U. Fano. Description of states in quantum mechanics by density matrix and operator techniques. *Rev. Mod. Phys.*, 29:74–93, 1957.

[28] I. D. Ivanović. Geometrical description of quantal state determination. *J. Phys. A: Math. Gen.*, 14:3241, 1981.

[29] W. K. Wootters and B. D. Fields. Optimal state-determination by mutually unbiased measurements. *Ann. Phys.*, 191:363–381, 1989.

[30] V. Bužek, G. Drobný, G. Adam, R. Derka, and P. L. Knight. Reconstruction of quantum states of spin systems via the Jaynes principle of maximum entropy. *J. Mod. Opt.*, 44(11-12):2607–2627, 1997.

[31] V. Bužek, G. Drobný, R. Derka, G. Adam, and H. Wiedemann. Quantum state reconstruction from incomplete data. *Chaos, Solitons Fractals*, 10(6):981–1074, 1999.

[32] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, 1957.

[33] E. T. Jaynes. Information theory and statistical mechanics. II. *Phys. Rev.*, 108:171–190, 1957.

[34] Z. Hradil. Quantum-state estimation. *Phys. Rev. A*, 55:R1561–R1564, 1997.

[35] B. Clarke and A. Barron. Jeffreys' prior is asymptotically least favorable under entropy risk. *J. Statist. Plann. Inference*, 41:37–60, 1994.

[36] R. E. Krichevskiy. Laplace's law of succession and universal encoding. *IEEE Trans. Inf. Theory*, 44:296–303, 1998.

[37] D. Braess, J. Forster, T. Sauer, and H. Simon. *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 2002.

[38] R. Blume-Kohout. Hedged maximum likelihood quantum state estimation. *Phys. Rev. Lett.*, 105:200504, 2010.

[39] B. P. Carlin and T. A. Louis. *Bayesian Methods for Data Analysis (Texts in Statistical Science)*. 3rd ed., Chapman & Hall/CRC, New York, 2008.

[40] C. W. Helstrom. *Quantum Detection and Estimation Theory*. Academic Press, New York, 1976.

[41] K. R. W. Jones. Principles of quantum inference. *Ann. Phys.*, 207:140–170, 1991.

[42] R. Schack, T. A. Brun, and C. M. Caves. Quantum Bayes rule. *Phys. Rev. A*, 64:014305, 2001.

[43] F. Neri. Quantum Bayesian methods and subsequent measurements. *Phys. Rev. A*, 72:062306, 2005.

[44] F. Tanaka and F. Komaki. Bayesian predictive density operators for exchangeable quantum-statistical models. *Phys. Rev. A*, 71:052323, 2005.

[45] R. Blume-Kohout. Optimal, reliable estimation of quantum states. *New J. Phys.*, 12:043034, 2010.

[46] G. M. D'Ariano, M. F. Sacchi, and J. Kahn. Minimax quantum-state discrimination. *Phys. Rev. A*, 72:032310, 2005.

[47] G. M. D'Ariano, M. F. Sacchi, and J. Kahn. Minimax discrimination of two Pauli channels. *Phys. Rev. A*, 72:052302, 2005.

[48] M. Guţă and L. Artiles. Minimax estimation of the Wigner function in quantum homodyne tomography with ideal detectors. *Math. Meth. Stat.*, 16:1–15, 2007.

[49] H. K. Ng and B.-G. Englert. A simple minimax estimator for quantum states. *Int. J. Quantum Inf.*, 10:1250038, 2012.

[50] H. K. Ng, K. T. B. Phuah, and B.-G. Englert. Minimax mean estimator for the trine. *New J. Phys.*, 14:085007, 2012.

[51] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, 2010.

[52] S. T. Flammia and Y.-K. Liu. Direct fidelity estimation from few Pauli measurements. *Phys. Rev. Lett.*, 106:230501, 2011.

[53] C. W. Helstrom. Minimum mean-squared error of estimates in quantum statistics. *Phys. Lett. A*, 25:101–102, 1967.

[54] C. W. Helstrom. The minimum variance of estimates in quantum signal detection. *IEEE Trans. Inf. Theory*, 14(2):234–242, 1968.

[55] H. P. Yuen and M. Lax. Multiple-parameter quantum estimation and measurement of nonselfadjoint observables. *IEEE Trans. Inf. Theory*, 19(6):740–750, 1973.

[56] A. S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory.* North-Holland, Amsterdam, 1982.

[57] J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. O'Brien, M. A. Nielsen, and A. G. White. Ancilla-assisted quantum process tomography. *Phys. Rev. Lett.*, 90:193601, 2003.

[58] I. L. Chuang and M. A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt.*, 44(11–12):2455–2467, 1997.

[59] J. F. Poyatos, J. I. Cirac, and P. Zoller. Complete characterization of a quantum process: The two-bit quantum gate. *Phys. Rev. Lett.*, 78:390–393, 1997.

[60] A. M. Childs, I. L. Chuang, and D. W. Leung. Realization of quantum process tomography in NMR. *Phys. Rev. A*, 64:012314, 2001.

[61] N. Boulant, T. F. Havel, M. A. Pravia, and D. G. Cory. Robust method for estimating the Lindblad operators of a dissipative quantum process from measurements of the density operator at multiple time points. *Phys. Rev. A*, 67:042322, 2003.

[62] M. W. Mitchell, C. W. Ellenor, S. Schneider, and A. M. Steinberg. Diagnosis, prescription, and prognosis of a Bell-state filter by quantum process tomography. *Phys. Rev. Lett.*, 91:120402, 2003.

[63] D. W. Leung. Choi's proof as a recipe for quantum process tomography. *J. Math. Phys.*, 44:528–533, 2003.

[64] F. De Martini, A. Mazzei, M. Ricci, and G. M. D'Ariano. Exploiting quantum parallelism of entanglement for a complete experimental quantum characterization of a single-qubit device. *Phys. Rev. A*, 67:062307, 2003.

[65] M. Mohseni and D. A. Lidar. Direct characterization of quantum dynamics. *Phys. Rev. Lett.*, 97:170501, 2006.

[66] M. Mohseni and D. A. Lidar. Direct characterization of quantum dynamics: General theory. *Phys. Rev. A*, 75:062331, 2007.

[67] Z.-W. Wang, Y.-S. Zhang, Y.-F. Huang, X.-F. Ren, and G.-C. Guo. Experimental realization of direct characterization of quantum dynamics. *Phys. Rev. A*, 75:044304, 2007.

[68] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme. Symmetrized characterization of noisy quantum processes. *Science*, 317:1893–1896, 2007.

[69] M. Mohseni, A. T. Rezakhani, and D. A. Lidar. Quantum-process tomography: Resource analysis of different strategies. *Phys. Rev. A*, 77:032322, 2008.

[70] A. Luis and L. L. Sánchez-Soto. Complete characterization of arbitrary quantum measurement processes. *Phys. Rev. Lett.*, 83:3573–3576, 1999.

[71] J. Fiurášek. Maximum-likelihood estimation of quantum measurement. *Phys. Rev. A*, 64:024102, 2001.

[72] Z. Hradil, J. Summhammer, and H. Rauch. Quantum tomography as normalization of incompatible observations. *Phys. Lett. A*, 261(1):20–24, 1999.

[73] Z. Hradil and J. Summhammer. Quantum theory of incompatible observations. *J. Phys. A: Math. Gen*, 33:7607, 2000.

[74] A. Peres. Neumark's theorem and quantum inseparability. *Found. Phys.*, 20:1441–1453, 1990.

[75] P. Busch. Informationally complete sets of physical quantities. *Int. J. Theor. Phys.*, 30:1217, 1991.

[76] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, 2009.

[77] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, 1935.

[78] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, 1993.

[79] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.*, 69:2881–2884, 1992.

[80] A. Ekert. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67:661–663, 1991.

[81] R. Raussendorf and H. J. Briegel. A one-way quantum computer. *Phys. Rev. Lett.*, 86:5188–5191, 2001.

[82] R. Blume-Kohout, J. O. S. Yin, and S. J. van Enk. Entanglement verification with finite data. *Phys. Rev. Lett.*, 105:170501, 2010.

[83] M. Bourennane, M. Eibl, C. Kurtsiefer, S. Gaertner, H. Weinfurter, O. Gühne, P. Hyllus, D. Bruß, M. Lewenstein, and A. Sanpera. Experimental detection of multipartite entanglement using witness operators. *Phys. Rev. Lett.*, 92:087902, 2004.

[84] L. Chen, H. Zhu, and T.-C. Wei. Connections of geometric measure of entanglement of pure symmetric states to quantum state estimation. *Phys. Rev. A*, 83:012305, 2011.

[85] H. Zhu, Y. S. Teo, and B.-G. Englert. Minimal tomography with entanglement witnesses. *Phys. Rev. A*, 81:052339, 2010.

[86] S. Massar and S. Popescu. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.*, 74:1259–1263, 1995.

[87] R. F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40:4277–4281, 1989.

[88] H. Zhu. *Quantum state estimation and symmetric informationally complete POMs.* Ph.D. thesis (Singapore 2012).

[89] W. Band and J. Park. The empirical determination of quantum states. *Found. Phys.*, 1:133–144, 1970.

[90] J. Park and W. Band. A general method of empirical state determination in quantum physics: Part I. *Found. Phys.*, 1:211–216, 1971.

[91] W. Band and J. Park. A general method of empirical state determination in quantum physics: Part II. *Found. Phys.*, 1:339–357, 1971.

[92] W. Gale, E. Guth, and G. T. Trammell. Determination of the quantum state by measurements. *Phys. Rev.*, 165:1434–1436, 1968.

[93] A. I. Lvovsky and M. G. Raymer. Continuous-variable optical quantum-state tomography. *Rev. Mod. Phys.*, 81:299, 2009.

[94] J. Řeháček, Z. Hradil, and M. Ježek. Iterative algorithm for reconstruction of entangled states. *Phys. Rev. A*, 63:040303, 2001.

[95] C. M. Caves and P. D. Drummond. Quantum limits on bosonic communication rates. *Rev. Mod. Phys.*, 66:481–537, 1994.

[96] Y. S. Teo, H. Zhu, B.-G. Englert, J. Řeháček, and Z. Hradil. Quantum-state reconstruction by maximizing likelihood and entropy. *Phys. Rev. Lett.*, 107:020404, 2011.

[97] Y. S. Teo. *Numerical estimation schemes for quantum tomography*. Ph.D. thesis (Singapore 2012); eprint arXiv:1302:3399 [quant-ph] (2013).

[98] G. J. Lidstone. Note on the general case of the Bayes–Laplace formula for inductive or a posteriori probabilities. *Trans. Fac. Actuaries*, 8:182–192, 1920.

[99] E. S. Ristad. A natural law of succession. *eprint arXiv:cmp-lg/9508012*, 1995.

[100] I. Bengtsson and K. Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, Cambridge, 2006.

[101] H. Jeffreys. *Theory of Probability*. Oxford University Press, Oxford, 1939.

[102] H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186:453–461, 1946.

[103] B.-G. Englert. On quantum theory. *eprint arXiv:1308.5290 [quant-ph]*, 2013.

[104] L. Vaidman. Many-worlds interpretation of quantum mechanics. *Stanford Encyclopedia of Philosophy*, 2002. http://plato.stanford.edu/entries/qm-manyworlds/#Teg98.

[105] E. Prugovečki. Information-theoretical aspects of quantum measurement. *Int. J. Theor. Phys.*, 16:321, 1977.

[106] G. M. D'Ariano, P. Perinotti, and M. F. Sacchi. Informationally complete measurements and group representation. *J. Opt. B: Quantum Semiclass. Opt.*, 6:S487, 2004.

[107] A. Ling, K. P. Soh, A. Lamas-Linares, and C. Kurtsiefer. Experimental polarization state tomography using optimal polarizers. *Phys. Rev. A*, 74:022309, 2006.

[108] J. Du, M. Sun, X. Peng, and T. Durt. Realization of entanglement-assisted qubit-covariant symmetric-informationally-complete positive-operator-valued measurements. *Phys. Rev. A*, 74:042341, 2006.

[109] W. M. Pimenta, B. Marques, M. A. Carvalho, M. R. Barros, J. G. Fonseca, J. Ferraz, M. Terra Cunha, and S. Pádua. Minimal state tomography of spatial qubits using a spatial light modulator. *Opt. Express*, 18:24423, 2010.

[110] Z. E. D. Medendorp, F. A. Torres-Ruiz, L. K. Shalm, G. N. M. Tabia, C. A. Fuchs, and A. M. Steinberg. Experimental characterization of qutrits using symmetric informationally complete positive operator-valued measurements. *Phys. Rev. A*, 83:051801(R), 2011.

[111] D. M. Appleby, I. Bengtsson, S. Brierley, M. Grassl, D. Gross, and J.-Å. Larsson. The monomial representations of the Clifford group. *Quant. Inf. Comput.*, 12:0404–0431, 2012.

[112] S. T. Flammia. On SIC-POVMs in prime dimensions. *J. Phys. A: Math. Gen.*, 39:13483, 2006.

[113] M. Grassl. On SIC-POVMs and MUBs in dimension 6. In *Proceedings of the 2004 ERATO Conference on Quantum Information Science*, pages 60–61. Tokyo, 2004. Available online: `http://arxiv.org/abs/quant-ph/0406175`.

[114] T. Durt, B.-G. Englert, I. Bengtsson, and K. Życzkowski. On mutually unbiased bases. *Int. J. Quantum Inf.*, 8:535, 2010.

[115] D. Gottesman. Theory of fault-tolerant quantum computation. *Phys. Rev. A*, 57:127–137, 1998.

[116] P. Delsarte, J. M. Goethals, and J. J. Seidel. Bounds for systems of lines, and Jacobi polynomials. *Philips Res. Rep.*, 30:91, 1975.

[117] S. G. Hoggar. 64 lines from a quaternionic polytope. *Geom. Det.*, 69:287, 1998.

[118] M. Grassl. Tomography of quantum states in small dimensions. *Electron. Notes Discrete Math.*, 20:151, 2005.

[119] M. Grassl. Finding equiangular lines in complex space. In *MAGMA 2006 Conference*. Technische Universität Berlin, 2006. Available online: http://magma.maths.usyd.edu.au/Magma2006/.

[120] M. Grassl. Computing equiangular lines in complex space. *Lect. Notes Comput. Sci.*, 5393:89–104, 2008.

[121] M. Grassl. Seeking symmetries of SIC-POVMs. In *Seeking SICs: A Workshop on Quantum Frames and Designs*. Perimeter Institute, Waterloo, 2008. Available online: http://pirsa.org/08100069/.

[122] C. Godsil and A. Roy. Equiangular lines, mutually unbiased bases, and spin models. *Eur. J. Combinator.*, 30:246–262, 2009.

[123] J. Schwinger. Unitary operator bases. *Proc. Natl. Acad. Sci. USA*, 46(4):570–579, 1960.

[124] W. K. Wootters. A wigner-function formulation of finite-state quantum mechanics. *Ann. Phys.*, 176:1–21, 1987.

[125] K. S. Gibbons, M. J. Hoffman, and W. K. Wootters. Discrete phase space based on finite fields. *Phys. Rev. A*, 70:062101, 2004.

[126] K. S. Gibbons, M. J. Hoffman, and W. K. Wootters. Quantum measurements and finite geometry. *Found. Phys.*, 36:112–126, 2006.

[127] L. Vaidman, Y. Aharonov, and D. Z. Albert. How to ascertain the values of σ_x, σ_y, and σ_z of a spin-1/2 particle. *Phys. Rev. Lett.*, 58:1385–1387, 1987.

[128] B.-G. Englert and Y. Aharonov. The mean king's problem: prime degrees of freedom. *Phys. Lett. A*, 284:1–5, 2001.

[129] B.-G. Englert, D. Kaszlikowski, L. C. Kwek, and W. H. Chee. Wave-particle duality in multi-path interferometers: General concepts and three-path interferometers. *Int. J. Quantum Inf.*, 6:129, 2008.

[130] N. J. Cerf, M. Bourennane, A. Karlsson, and N. Gisin. Security of quantum key distribution using d-level systems. *Phys. Rev. Lett.*, 88:127902, 2002.

[131] T. Durt. If $1 = 2 \oplus 3$, then $1 = 2 \odot 3$: Bell states, finite groups, and mutually unbiased bases, a unifying approach. *eprint arXiv:quant-ph/0401046v2*, 2012.

[132] A. B. Klimov, D. Sych, L. L. Sánchez-Soto, and G. Leuchs. Mutually unbiased bases and generalized Bell states. *Phys. Rev. A*, 79:052101, 2009.

[133] M. Revzen. Maximally entangled states via mutual unbiased collective bases. *Phys. Rev. A*, 81:012113, 2010.

[134] S. Bandyopadhyay, P. O. Boykin, V. Roychowdhury, and F. Vatan. A new proof for the existence of mutually unbiased bases. *Algorithmica*, 34:512, 2002.

[135] T. Durt. About mutually unbiased bases in even and odd prime power dimensions. *J. Phys. A: Math. Gen.*, 38(23):5267, 2001.

[136] P. Butterley and W. Hall. Numerical evidence for the maximum number of mutually unbiased bases in dimension six. *Phys. Lett. A*, 369:5, 2007.

[137] S. Brierley and S. Weigert. Maximal sets of mutually unbiased quantum states in dimension 6. *Phys. Rev. A*, 78:042312, 2008.

[138] P. Raynal, X. Lü, and B.-G. Englert. Mutually unbiased bases in six dimensions: The four most distant bases. *Phys. Rev. A*, 83:062303, 2011.

[139] N. Bohr. The quantum postulate and the recent development of atomic theory. *Nature*, 121:580–590, 1928.

[140] X. Lü, P. Raynal, and B.-G. Englert. Mutually unbiased bases for the rotor degree of freedom. *Phys. Rev. A*, 85:052316, 2012.

[141] X. Lü. *A study on mutually unbiased bases*. Ph.D. thesis (Singapore 2012).

[142] Y. Aharonov, D. Z. Albert, and L. Vaidman. How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100. *Phys. Rev. Lett.*, 60:1351–1354, 1988.

[143] G. Weihs, M. Reck, H. Weinfurter, and A. Zeilinger. Two-photon interference in optical fiber multiports. *Phys. Rev. A*, 54:893–897, 1996.

[144] A. Peruzzo, A. Laing, A. Politi, T. Rudolph, and J. L. O'Brien. Multimode quantum interference of photons in multiport integrated devices. *Nat. Commun.*, 2:244, 2011.

[145] J. C. F. Matthews, A. Politi, D. Bonneau, and J. L. O'Brien. Heralding two-photon and four-photon path entanglement on a chip. *Phys. Rev. Lett.*, 107:163602, 2011.

[146] C. Carmeli, T. Heinosaari, and A. Toigo. Informationally complete joint measurements on finite quantum systems. *Phys. Rev. A*, 85:012109, 2012.

[147] B.-G. Englert, C. Kurtsiefer, and H. Weinfurter. Universal unitary gate for single-photon 2-qubit states. *Phys. Rev. A*, 63:032303, 2001.

[148] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.*, 73:58, 1994.

[149] H. Bechmann-Pasquinucci and W. Tittel. Quantum cryptography using larger alphabets. *Phys. Rev. A*, 61:062308, 2000.

[150] L. Sheridan and V. Scarani. Security proof for quantum key distribution using qudit systems. *Phys. Rev. A*, 82:030301(R), 2010.

[151] I. Bengtsson. From SICs and MUBs to Eddington. *J. Phys.: Conf. Ser.*, 254:012007, 2010.

[152] R. B. A. Adamson and A. M. Setinberg. Quantum state estimation with mutually unbiased bases. *Phys. Rev. Lett.*, 105:030406, 2010.

[153] I. Tselniker, M. Nazarathy, and M. Orenstein. Mutually unbiased bases in 4, 8, and 16 dimensions generated by means of controlled-phase gates with application to entangled-photon QKD protocols. *IEEE J. Sel. Top. Quant.*, 15:1713–1723, 2009.

[154] J. Řeháček, D. Mogilevtsev, and Z. Hradil. Tomography for quantum diagnostics. *New J. Phys.*, 10:043022, 2008.

[155] K. M. R. Audenaert and S. Scheel. Quantum tomographic reconstruction with error bars: a Kalman filter approach. *New J. Phys.*, 11:023028, 2009.

[156] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap.* Chapman & Hall/CRC Monographs on Statistics & Applied Probability, New York, 1993.

[157] M. Christandl and R. Renner. Reliable quantum state tomography. *Phys. Rev. Lett.*, 109:120403, 2012.

[158] R. Blume-Kohout. Robust error bars for quantum tomography. *eprint arXiv:1202.5270 [quant-ph]*, 2012.

[159] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis.* 2nd ed., Springer-Verlag, New York, 1985.

[160] M. J. Evans, I. Guttman, and T. Swartz. Optimality and computations for relative surprise inferences. *Can. J. Stat.*, 34:113, 2006.

[161] H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. London Series A*, 186:453, 1946.

[162] R. E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *Proc. Roy. Soc. London Series A*, 91:1343, 1996.

[163] Y. S. Teo, B. Stoklasa, B.-G. Englert, J. Řeháček, and Z. Hradil. Incomplete quantum state estimation: A comprehensive study. *Phys. Rev. A*, 85:042317, 2012.

[164] L. A. Wasserman. A robust Bayesian interpretation of likelihood regions. *Ann. Stat.*, 17:1387, 1989.

[165] E. L. Lehmann and G. Gasella. *Theory of Point Estimation.* 2nd ed., Springer-Verlag, Berlin, 1998.

[166] E. T. Jaynes. *Probability Theory—The Logic of Science.* Cambridge University Press, Cambridge, 2003.

[167] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes: The Art of Scientific Computing.* 3rd ed., Cambridge University Press, Cambridge, 2007.

[168] C. H. Bennett and G. Brassard. Quantum cryptography: public key distribution and coin tossing. In *IEEE Conf. Computers, Systems, and Signal Processing*, volume 175. Bangalore, India (New York: IEEE), 1984.

[169] G. N. M. Tabia and B.-G. Englert. Efficient quantum key distribution with trines of reference-frame-free qubits. *Phys. Lett. A*, 375:817–822, 2011.

[170] P. Horodecki and A. Ekert. Method for direct detection of quantum entanglement. *Phys. Rev. Lett.*, 89:127902, 2002.

[171] S. P. Walborn, P. H. Souto Ribeiro, L. Davidovich, F. Mintert, and A. Buchleitner. Experimental determination of entanglement with a single measurement. *Nature*, 440:1022–1024, 2006.

[172] C. Schmid, N. Kiesel, W. Wieczorek, H. Weinfurter, F. Mintert, and A. Buchleitner. Experimental direct observation of mixed state entanglement. *Phys. Rev. Lett.*, 101:260505, 2008.

[173] T. Brun. Measuring polynomial functions of states. *Quant. Inf. Comp.*, 4:401, 2004.

[174] H. Carteret. Exact interferometers for the concurrence and reidual 3-tangle. *eprint arXiv:quant-ph/0309212*, 2003.

[175] A. R. Calderbank, E. M. Reins, P. W. Shor, and N. J. A. Sloane. Quantum error correction via codes over GF(4). *IEEE Trans. Inform. Theory*, 44:1369–1387, 1998.

[176] G. Karpilovski. *Field Theory.* Marcel Dekker Inc., New York and Basel, 1988.

[177] A. Uhlmann. The "transition probability" in the state space of a *-algebra. *Rep. Math. Phys.*, 9(2):273–279, 1976.

[178] R. Jozsa. Fidelity for mixed quantum states. *J. Mod. Opt.*, 41(12):2315–2323, 1994.

[179] J.-L. Chen, L. Fu, A. A. Ungar, and X.-G. Zhao. Geometric observation for Bures fidelity between two states of a qubit. *Phys. Rev. A*, 65:024303, 2002.

[180] J. A. Miszczak, Z. Puchala, P. Horodecki, A. Uhlmann, and K. Życzkowski. Sub- and super-fidelity as bounds for quantum fidelity. *Quantum Info. Comput.*, 9(1):103–130, 2009.

[181] M. Hübner. Explicit computation of the Bures distance for density matrices. *Phys. Lett. A*, 163:239–242, 1992.

[182] S. L. Braunstein and C. M. Caves. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.*, 72:3439–3443, 1994.

List of Publications

1. J. Shang, H. K. Ng, A. Sehrawat, X. Li, and B.-G. Englert. Optimal error regions for quantum state estimation. *New J. Phys.*, 15:123026, 2013; eprint arXiv:1302.4081v2 [quant-ph] (2013).

2. A. Kalev, J. Shang, and B.-G. Englert. Symmetric minimal quantum tomography by successive measurements. *Phys. Rev. A*, 85:052116, 2012; eprint arXiv:1203.1677v1 [quant-ph] (2012).

3. A. Kalev, J. Shang, and B.-G. Englert. Experimental proposal for symmetric minimal two-qubit state tomography. *Phys. Rev. A*, 85:052115, 2012; eprint arXiv:1203.1675v1 [quant-ph] (2012).

4. J. Shang, K. L. Lee, and B.-G. Englert. **SeCQC**: An open-source program code for the numerical **Se**arch for the classical **C**apacity of **Q**uantum **C**hannels. eprint arXiv:1108.0226v1 [quant-ph] (2011).
   URL: `http://www.quantumlah.org/publications/software/SeCQC/`.

5. K. L. Lee, J. Shang, W. K. Chua, S. Y. Looi, and B.-G. Englert. **SOMIM**: An open-source program code for the numerical **S**earch for **O**ptimal **M**easurements by an **I**terative **M**ethod. eprint arXiv:0805.2847v2 [quant-ph] (2011).
   URL: `http://www.quantumlah.org/publications/software/SOMIM/`.

Index