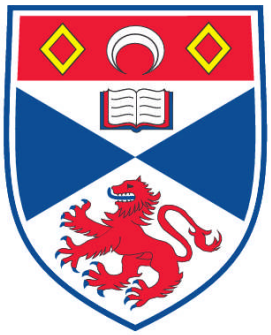
TOPICS IN ESTIMATION OF QUANTUM CHANNELS

Caleb J. O'Loan

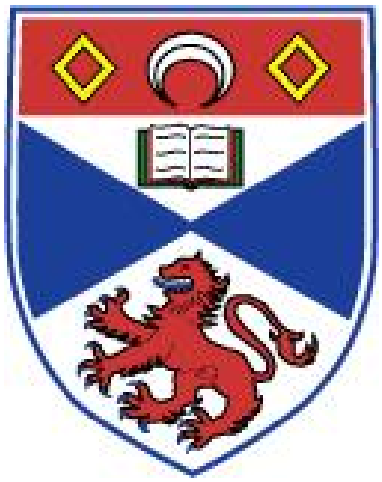
A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews



2009

Topics in Estimation of Quantum Channels

Caleb J. O'Loan



A Thesis submitted
to the University of St Andrews
in application for the degree of
Doctor of Philosophy

8th December 2009

Declaration

I, Caleb J O'Loan, hereby certify that this thesis, which is approximately $23,000$ words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September, 2005 and as a candidate for the degree of Ph.D. in September, 2006; the higher study for which this is a record was carried out in the University of St Andrews between 2005 and 2009.

date ..... signature of candidate .....

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date ..... signature of supervisor .....

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. We have obtained any third-party copyright permissions that may be required in order to allow such access.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to Printed copy and electronic publication of thesis through the University of St Andrews.

date .....

signature of candidate .....

signature of supervisor .....

Acknowledgements

This thesis would not have been completed without the patient and helpful supervision of Prof. Peter E. Jupp. Many thanks go to him. I am also grateful to Dr Terry Rudolph for good advice, for the continued support of my family, and to Richard Gormley for checking the final manuscript.

Contents

Abstract

A quantum channel is a mapping which sends density matrices to density matrices. The estimation of quantum channels is of great importance to the field of quantum information. In this thesis two topics related to estimation of quantum channels are investigated. The first of these is the upper bound of Sarovar and Milburn (2006) on the Fisher information obtainable by measuring the output of a channel. Two questions raised by Sarovar and Milburn about their bound are answered. A Riemannian metric on the space of quantum states is introduced, related to the construction of the Sarovar and Milburn bound. Its properties are characterized.

The second topic investigated is the estimation of unitary channels. The situation is considered in which an experimenter has several non-identical unitary channels that have the same parameter. It is shown that it is possible to improve estimation using the channels together, analogous to the case of identical unitary channels. Also, a new method of phase estimation is given based on a method sketched by Kitaev (1996). Unlike other phase estimation procedures which perform similarly, this procedure requires only very basic experimental resources.

Chapter 1

Mathematical Background

1.1 Overview

This thesis is concerned with estimation of quantum channels. Almost every protocol in quantum information uses quantum channels. They are used in important protocols such as teleportation, Deutsch's algorithm, the Grover search algorithm and the Shor factorization algorithm (Le Bellac, 2006, Chapters 5 and 7). In theory it is assumed that a channel is known precisely, yet in practice this will not generally be the case. Thus the estimation of quantum channels is of fundamental importance to the field of quantum information.

Chapter 1 contains the mathematical and quantum-theoretic background needed to understand the thesis. Chapters 2 and 3 are concerned with the upper bound of Sarovar and Milburn (2006) on the Fisher information obtainable by measuring the output of a channel. Chapters 4 and 5 consider estimation of unitary channels.

In Chapter 1 definitions are given of fundamental objects such as quantum systems, quantum states, quantum measurements and combined systems. Quantum channels are defined and quantum channel estimation introduced. A brief historical background is given of the key developments in channel estimation. Chapter 1 contains also a few small, new results.

Chapter 2 considers work by Sarovar and Milburn (2006), who introduced an upper bound on the Fisher information obtained by measuring the output states of quantum channels. They showed that for certain channels, called quasi-classical channels, their bound is attainable. They asked (i) whether their bound is attainable more generally; (ii) whether or not it is possible to find an explicit expression for measurements attaining this bound. Both of these questions are answered in Chapter 2.

In the process of answering the previous questions, Chapter 2 shows that Sarovar and Milburn's work leads to a new Riemannian metric on the space of quantum states. Chapter 3 considers the questions: What are the properties of this new metric? Is it well defined?

Chapters 4 and 5 are concerned with the cost of estimation of unitary channels. It is known that when there are n identical copies of a unitary channel, there exists (Kahn, 2007) an estimation procedure such that the cost function (a function of the expected fidelity, see (1.90)) is $O(1/n^2)$, instead of the usual $O(1/n)$. Chapter 4 considers the question: If there are n unitary channels which are not identical, but have the same parameter, is an analogous speed-up possible?

Kitaev (1996) sketched an iterative method for phase estimation such that the cost function is $O((\log n/n)^2)$. This method requires only a single copy of a unitary channel and basic measurements. In Chapter 5 it is shown that several attempts to give a detailed method for iterative phase estimation have been unsuccessful. There have been other successful iterative methods, but these require an extra rotation gate capable of performing arbitrary rotations with almost perfect accuracy. Thus Chapter 5 seeks to answer the question: Does a complete iterative phase estimation method exist which requires only a single copy of the unitary and basic measurements?

1.2 System

A *quantum system* is a physical system that obeys the laws of quantum mechanics. The *state* of a quantum system (or *quantum state*, or *quantum state of a system*) is a quantification of the system, which, if known, allows an experimenter to make accurate predictions about the results of any future measurements on that system (Gill, 2001). Since measurement results are probabilistic, knowledge of a quantum state means that, given any measurement, it is possible to work out the long-term relative frequency of the observed outcomes.

A quantum system is represented by a complex Hilbert space $\mathcal{H}$ of dimension d, with a Hermitian inner product. The dimension d is given by the maximum number of distinguishable states in the system. For the spin of an electron, or the polarisation of a photon, $\mathcal{H} = \mathbb{C}^2$. This is because there are only two distinguishable states: spin up and spin down. Any other quantum state can be represented as a complex linear combination of these two states.

In this thesis only finite dimensional complex vector spaces are considered. Any column vector in a complex vector space is denoted by $|\psi\rangle$, the symbol ψ is a label, while $|\cdot\rangle$ denotes that the object is a complex column vector. This

representation of complex vectors is called *Dirac notation*. Given a vector

$$|\psi\rangle = \begin{pmatrix} \psi_1 \\ \psi_2 \\ . \\ . \\ \psi_n \end{pmatrix}, \tag{1.1}$$

its dual $\langle\psi|$ is defined as (with ψ_j^* denoting the complex conjugate of ψ_j)

$$\langle\psi| = (\psi_1^*, \psi_2^*, \ldots, \psi_n^*)\,. \tag{1.2}$$

An *inner product* is a bilinear map that maps a pair of complex vectors to a complex number. Given the vectors $|\psi\rangle$ and $|\phi\rangle$, the inner product between these vectors is denoted by $\langle\psi|\phi\rangle$. There are many different inner products. In this thesis, only the following inner product will be used

$$\langle\psi|\phi\rangle = \sum_{i=1}^{n} \psi_i^* \phi_i \tag{1.3}$$

$$= (\psi_1^*, \psi_2^*, \ldots, \psi_n^*) \begin{pmatrix} \phi_1 \\ \phi_2 \\ . \\ . \\ \phi_n \end{pmatrix}. \tag{1.4}$$

The *norm* of a vector $|\psi\rangle$ can be defined as $\|\psi\| = \sqrt{\langle\psi|\psi\rangle}$. Vectors with norm equal to one are defined as *unit vectors*. Vectors $|\psi\rangle$ and $|\phi\rangle$ are *orthogonal* if their inner product is zero.

Given the vectors $|\psi\rangle$ and $|\phi\rangle$, the *outer product* $|\phi\rangle\langle\psi|$ is given by

$$|\phi\rangle\langle\psi| = \begin{pmatrix} \phi_1 \\ \phi_2 \\ . \\ . \\ \phi_n \end{pmatrix} (\psi_1^*, \psi_2^*, \ldots, \psi_n^*) \tag{1.5}$$

$$= \begin{pmatrix} \phi_1\psi_1^* & \phi_1\psi_2^* & \ldots & \phi_1\psi_n^* \\ \phi_2\psi_1^* & \phi_2\psi_2^* & \ldots & \phi_2\psi_n^* \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \phi_n\psi_1^* & \phi_n\psi_2^* & \ldots & \phi_n\psi_n^* \end{pmatrix}. \tag{1.6}$$

In Dirac notation $\langle\psi|\phi\rangle$ represents a complex number, and $|\phi\rangle\langle\psi|$ a matrix.

A set of vectors $|u_1\rangle, \ldots, |u_n\rangle$ is *orthonormal* if the vectors are normalized and orthogonal, i.e. $\langle u_i|u_j\rangle = \delta_{ij}$. Given a set of orthonormal vectors $|u_1\rangle, \ldots, |u_n\rangle$ in a vector space V, such that $n = \dim V$, this set of vectors forms an orthonormal basis of V. Any vector $|v\rangle$ in V can be written as a scalar multiple of these vectors, i.e.

$$|v\rangle = \sum_{i=1}^{n} v_i|u_i\rangle, \quad v_i = \langle u_i|v\rangle \in \mathbb{C}. \tag{1.7}$$

Lemma 1.1 *Given an orthonormal basis $|u_1\rangle, \ldots, |u_n\rangle$ for a vector space V,*

$$\sum_{i=1}^{n} |u_i\rangle\langle u_i| = \mathbb{I}_n. \tag{1.8}$$

This is called the *completeness relation* (Nielsen and Chuang, 2000, p. 67).

Proof. If $|u_1\rangle, \ldots, |u_n\rangle$ is an orthonormal basis for V, then any $|v\rangle \in V$ can be written as $|v\rangle = \sum_i v_i|u_i\rangle$, where $v_i = \langle u_i|v\rangle$. Now,

$$\left(\sum_{i=1}^{n} |u_i\rangle\langle u_i|\right)|v\rangle = \sum_{i=1}^{n} v_i|u_i\rangle = |v\rangle. \tag{1.9}$$

Since this holds for all $|v\rangle$, the result follows.

The *Hermitian transpose* of a matrix A, denoted by $A^\dagger$, is the matrix found by taking the transpose of A and replacing each entry with its complex conjugate $\left([A^\dagger]_{ij} = [A]^*_{ji}\right)$. A *Hermitian matrix* (also called a *self-adjoint matrix*) is a matrix which is equal to its Hermitian transpose, i.e. B is Hermitian if $B^\dagger = B$. The Pauli matrices, given in (1.15), are examples of Hermitian matrices.

Any matrix B that is Hermitian can be diagonalized, that is written in the form $UDU^\dagger$, where D is a diagonal matrix and U is a unitary matrix, or equivalently in terms of its eigenvalues $\{a_i\}$ and eigenvectors $\{|w_i\rangle\}$ as

$$B = \sum_i a_i|w_i\rangle\langle w_i|.$$

1.3 States

Pure states will now be introduced. These form a subset of the set of all quantum states. A pure state of dimension d can be represented by a d-dimensional complex unit vector $|\psi\rangle$. For real θ, the vectors $|\psi\rangle$ and $e^{i\theta}|\psi\rangle$ represent the same state.

More generally, a d dimensional quantum state is represented by a $d \times d$ matrix ρ, also called a *density matrix*. This is a linear operator which acts on a complex Hilbert space $\mathcal{H}$, is non-negative ($v^T \rho v \geq 0$ for all $v \in \mathbb{R}^d$) and has trace 1. A consequence of being non-negative is that ρ is self-adjoint. The set of states in a complex Hilbert space $\mathcal{H}$ will be denoted by $S(\mathcal{H})$.

A pure state can be referred to either by its state vector $|\psi\rangle$, or by its density matrix $\rho = |\psi\rangle\langle\psi|$. For example,

$$|\psi\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \rho = |\psi\rangle\langle\psi| = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}. \tag{1.10}$$

States which are not pure (have rank greater than one) are called *mixed states*. A simple test for whether a state ρ is pure or mixed is to take the trace of ρ^2. For pure states $\mathrm{tr}\{\rho^2\} = 1$; for mixed states $\mathrm{tr}\{\rho^2\} < 1$.

Examples of 2-dimensional mixed states are

$$\rho_1 = \begin{pmatrix} 3/4 & 0 \\ 0 & 1/4 \end{pmatrix}, \quad \rho_2 = \begin{pmatrix} 1/2 & -1/6 \\ -1/6 & 1/2 \end{pmatrix}. \tag{1.11}$$

A mixed state can be expressed as a mixture of pure states in many different ways. For instance, the state ρ_1, given in (1.11), can be written as

$$\begin{aligned} \rho_1 &= 3/4 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + 1/4 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\ &= 1/2 \begin{pmatrix} 3/4 & \sqrt{3}/4 \\ \sqrt{3}/4 & 1/4 \end{pmatrix} + 1/2 \begin{pmatrix} 3/4 & -\sqrt{3}/4 \\ -\sqrt{3}/4 & 1/4 \end{pmatrix} \\ &= 1/4 \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} + 1/4 \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix} + 1/2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

The set of 2-dimensional states, which is of great importance to the theory of quantum information, will now be investigated. A 2-dimensional quantum state is called a *qubit*. This is because qubits are the quantum analogue of 'bits' (binary digits). In quantum information qubits are used to store and transmit information. Pure qubits are often expressed in the basis

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{1.12}$$

The state $|\psi\rangle$, given in (1.10), can be expressed as $(|0\rangle + |1\rangle)/\sqrt{2}$.

Any 2-dimensional quantum state can be written, with specific values of x, y and z, as

$$\rho = 1/2 \begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix}, \tag{1.13}$$

where $x^2 + y^2 + z^2 \leq 1$. The set of pure states corresponds to those states for which $x^2 + y^2 + z^2 = 1$. Any 2-dimensional state can be thought of as being a point with Cartesian co-ordinates (x, y, z) contained within a unit ball, known as the *Bloch ball* or *Poincaré ball*. The points on the surface of the ball correspond to pure states; the points within the ball to mixed states.

Alternatively, the state (1.13) can be written in terms of the identity and Pauli matrices

$$\mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{1.14}$$

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{1.15}$$

as

$$\rho = \frac{1}{2}(\mathbb{I} + x\sigma_x + y\sigma_y + z\sigma_z), \quad x, y, z \in \mathbb{R}. \tag{1.16}$$

1.4 Measurements

In this section measurements are introduced. When a state is measured, a single result, m, is observed out of a set Ω of possible outcomes. Associated with each outcome m is a matrix M_m. The set of matrices $\{M_m\}$ constitute a *measurement*. The following conditions are imposed on M_m

$$M_m^\dagger = M_m, \quad M_m \geq 0, \quad \sum_{m\in\Omega} M_m = \mathbb{I}. \tag{1.17}$$

A measurement of this form is called a *Positive Operator Valued Measure*, or *POVM* for short.

Given a state ρ and a measurement $M = \{M_m\}$, the result m is observed with probability given by the Born rule

$$p(m) = \mathrm{tr}\{\rho M_m\}. \tag{1.18}$$

The $p(m)$ defined in (1.18) satisfies

(i) $p(m) \geq 0$, as $\rho \geq 0$ and $M_m \geq 0$,

(ii) $\sum_{m\in\Omega} p(m) = 1$, as

$$\begin{aligned} \sum_{m\in\Omega} p(m) &= \sum_{m\in\Omega} \mathrm{tr}\{\rho M_m\} \\ &= \mathrm{tr}\{\rho \sum_{m\in\Omega} M_m\} \\ &= \mathrm{tr}\{\rho\mathbb{I}\} \\ &= 1. \end{aligned}$$

More generally, (Busch *et al.*, 1995, p. 23) one can consider a non-empty set Ω and a sigma algebra $\mathcal{F}$, which is a collection of subsets of elements of Ω that obeys certain rules. Together Ω and $\mathcal{F}$ give, what is known as, a measured space $(\Omega, \mathcal{F})$. A POVM over a measured space $(\Omega, \mathcal{F})$ is a set $\{M(A_i)\}_{A_i \in \mathcal{F}}$ of operators on $\mathcal{H}$ such that

$$\begin{aligned} M(A_i) &\geq 0, \quad \text{for all } A_i \in \mathcal{F}, && (1.19) \\ M(\cup_i A_i) &= \sum_i M(A_i), \quad \text{for disjoint } A_i, && (1.20) \\ M(\Omega) &= \mathbb{I}. && (1.21) \end{aligned}$$

Applying the measurement M to a state ρ yields outcome i with probability

$$p(i) = \mathrm{tr}\{\rho M(A_i)\}. \tag{1.22}$$

The most commonly used measurements are *Projection Valued Measures* (abbreviated to PVMs). A *PVM* is a POVM with elements, usually written as P_m, which satisfy $P_m P_{m'} = \delta_{mm'} P_m$. (These measurements are also called *projective measurements*.) A PVM $\{P_m\}$ is associated with an *observable*, M, a Hermitian operator on $\mathcal{H}$ (Nielsen and Chuang, 2000, p. 87). The observable has spectral decomposition

$$M = \sum_{m \in \Omega} m P_m.$$

A simple example of a 2-dimensional observable is σ_x

$$\sigma_x = (+1) \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} + (-1) \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}.$$

Measuring this observable corresponds to using the PVM

$$M^x = (M_0, \mathbb{I} - M_0), \quad M_0 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}. \tag{1.23}$$

The term 'measuring in x' refers to using the PVM M^x (for which $M_0 = (\mathbb{I} + \sigma_x)/2$). Similarly, the term 'measuring in y' refers to using the PVM $M^y = (M_0, \mathbb{I} - M_0)$, with $M_0 = (\mathbb{I} + \sigma_y)/2$, and 'measuring in z' to using the PVM $M^z = (M_0, \mathbb{I} - M_0)$, with $M_0 = (\mathbb{I} + \sigma_z)/2$.

A POVM gives only information about data from a measurement. To describe how a state is changed by a measurement it is necessary to use *instruments*. More information on instruments is given in Barndorff-Nielsen *et al.* (2003).

1.5 Combined systems

The *tensor product* is a mathematical operation which can be used to combine vector spaces to form a larger vector space. Given two vector spaces V and W, it is possible to combine them to form the vector space $V \otimes W$, with $\dim(V \otimes W) = \dim V \times \dim W$. Given vectors $|v\rangle \in V$ and $|w\rangle \in W$, the vector $|v\rangle \otimes |w\rangle \in V \otimes W$. The vector $|v\rangle \otimes |w\rangle$ is computed from $|v\rangle$ and $|w\rangle$ in the following way

$$|v\rangle = \begin{pmatrix} v_1 \\ v_2 \\ . \\ . \\ v_m \end{pmatrix}, |w\rangle = \begin{pmatrix} w_1 \\ w_2 \\ . \\ . \\ w_n \end{pmatrix}, \quad |v\rangle \otimes |w\rangle = \begin{pmatrix} v_1 w_1 \\ v_1 w_2 \\ . \\ v_1 w_n \\ v_2 w_1 \\ . \\ v_m w_n \end{pmatrix}.$$

Similarly, given two matrices

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

their tensor product is equal to

$$\begin{aligned} A \otimes B &= \begin{pmatrix} A_{11}B & A_{12}B \\ A_{21}B & A_{22}B \end{pmatrix} \\ &= \begin{pmatrix} A_{11}B_{11} & A_{11}B_{12} & A_{12}B_{11} & A_{12}B_{12} \\ A_{11}B_{21} & A_{11}B_{22} & A_{12}B_{21} & A_{12}B_{22} \\ A_{21}B_{11} & A_{21}B_{12} & A_{22}B_{11} & A_{22}B_{12} \\ A_{21}B_{21} & A_{21}B_{22} & A_{22}B_{21} & A_{22}B_{22} \end{pmatrix}. \end{aligned}$$

More generally, for an $M \times N$ matrix A and a $P \times Q$ matrix B, $A \otimes B$ is an $MP \times NQ$ matrix with entries

$$\begin{aligned} A \otimes B &= \begin{pmatrix} A_{11}B & \cdots & A_{1N}B \\ \cdots & \cdots & \cdots \\ A_{M1}B & \cdots & A_{MN}B \end{pmatrix} \\ &= \begin{pmatrix} A_{11}B_{11} & A_{11}B_{12} & \cdots & A_{1N}B_{1Q} \\ A_{11}B_{21} & A_{11}B_{22} & \cdots & A_{1N}B_{2Q} \\ \cdots & \cdots & \cdots & \cdots \\ A_{M1}B_{P1} & A_{M1}B_{P2} & \cdots & A_{MN}B_{PQ} \end{pmatrix}. \end{aligned}$$

The following rules hold for tensor products

$$\begin{aligned}
(A \otimes B)|v\rangle \otimes |w\rangle &= A|v\rangle \otimes B|w\rangle \\
A \otimes (B + C) &= A \otimes B + A \otimes C \\
(A \otimes B)(C \otimes D) &= (AC \otimes BD) \\
\mathrm{tr}(A \otimes B) &= (\mathrm{tr}A)(\mathrm{tr}B) \\
(A \otimes B)^\dagger &= A^\dagger \otimes B^\dagger.
\end{aligned}$$

Given two different quantum systems represented by Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, the combined system is represented by the tensor product of these Hilbert spaces, i.e. $\mathcal{H}_A \otimes \mathcal{H}_B$, which will be labelled $\mathcal{H}_{A,B}$.

If $\rho^A \in S(\mathcal{H}_A)$ and $\rho^B \in S(\mathcal{H}_B)$, then the *composite state* in $S(\mathcal{H}_{A,B})$ is $\rho^{A,B} = \rho^A \otimes \rho^B$. Given two pure states represented by $|\psi^A\rangle$ and $|\psi^B\rangle$, the composite system is in state represented by $|\psi^{A,B}\rangle = |\psi^A\rangle \otimes |\psi^B\rangle$. For the sake of brevity the sign $\otimes$ will usually be omitted, and the composite state written as $|\psi^A\psi^B\rangle$. Of special interest are states, especially pure ones, which exist in the composite system but cannot be written in the form $|\psi^A\psi^B\rangle$. Examples of such states are the *Bell states*,

$$|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{1.24}$$

$$|\phi^-\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \tag{1.25}$$

$$|\psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \tag{1.26}$$

$$|\psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \tag{1.27}$$

Definition 1.1 *A pure state $|\psi^{A,B}\rangle \in \mathcal{H}_{A,B}$, which cannot be written as $|\psi^A\rangle \otimes |\psi^B\rangle$ is said to be entangled. More generally, we can consider entangled mixed states. A state $\rho \in S(\mathcal{H}_{A,B})$ which cannot be written as a mixture of separable pure states $|\psi_i^A\rangle \otimes |\psi_i^B\rangle \in \mathcal{H}_{A,B}$, i.e. as*

$$\rho = \sum_i p_i |\psi_i^A\rangle\langle\psi_i^A| \otimes |\psi_i^B\rangle\langle\psi_i^B|,$$

is said to be an entangled state.

Definition 1.2 *States that are not entangled are said to be separable.*

1.5.1 Partial trace

The *partial trace* is an important operation when considering combined systems. Given a density matrix $\rho^{AB} \in S(\mathcal{H}_{A,B})$, the state of system $\mathcal{H}_A$ is found by taking the partial trace over $\mathcal{H}_B$,

$$\rho_A = \mathrm{tr}_B\{\rho^{AB}\}, \tag{1.28}$$

where the partial trace tr_B is defined as

$$\begin{aligned} \mathrm{tr}_B\{|\phi_1\rangle\langle\phi_2| \otimes |\psi_1\rangle\langle\psi_2|\} &= |\phi_1\rangle\langle\phi_2|\mathrm{tr}\{|\psi_1\rangle\langle\psi_2|\} \\ &= |\phi_1\rangle\langle\phi_2|\langle\psi_2|\psi_1\rangle. \end{aligned}$$

A state found by taking the partial trace of a larger state on a combined system is known as a *reduced state*.

The density matrix for the Bell state $|\phi^+\rangle$, given in (1.24), is

$$\rho = |\phi^+\rangle\langle\phi^+| = \frac{1}{2}\left(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|\right).$$

Taking the partial trace over $\mathcal{H}_B$ gives

$$\begin{aligned} \rho_A = \mathrm{tr}_B(\rho) &= \frac{1}{2}\Big(|0\rangle\langle 0|\mathrm{tr}(|0\rangle\langle 0|) + |1\rangle\langle 0|\mathrm{tr}(|1\rangle\langle 0|) \\ &+ \quad |0\rangle\langle 1|\mathrm{tr}(|0\rangle\langle 1|) + |1\rangle\langle 1|\mathrm{tr}(|1\rangle\langle 1|)\Big) \\ &= \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|). \end{aligned}$$

Note that, although the composite state is pure, the reduced state is mixed. This is one of the interesting properties of entanglement.

If $\rho^{AB} = \rho \otimes \sigma$ then

$$\rho_A = \mathrm{tr}_B\{\rho \otimes \sigma\} = \rho\,\mathrm{tr}\{\sigma\} = \rho,$$

as would be expected. Similarly, $\rho_B = \mathrm{tr}_A\{\rho^{AB}\} = \sigma$.

1.6 Measurements on several copies of a state

Given a product state of the form $\rho^{\otimes n} = \rho^{(1)} \otimes \cdots \otimes \rho^{(n)}$, where $\rho^{(j)}$ denotes the jth copy of ρ, there are several types of measurements that can be performed. In this section the most common types of measurements will be defined: *collective measurements*, *separable measurements*, *LOCC*, *adaptive measurements* and *separate measurements*. This section is similar to (Ballester, 2005, Section 1.2.6).

1.6.1 Collective measurements

This is the most general type of measurement. If $\dim(\rho) = d$ then $\rho^{\otimes n}$ acts on $\mathbb{C}^{d^n}$. Collective measurements are POVMs whose elements are $d^n \times d^n$ complex matrices satisfying (1.17), (Massar, 2000, Gill, 2008).

1.6.2 Separable measurements

These form a smaller class of measurements. Separable measurements are POVMs whose elements can be expressed as

$$M_m = \sum_{i=1}^{k} M_{m_i}^{(1)} \otimes \cdots \otimes M_{m_i}^{(n)}, \quad M_{m_i}^{(j)} \geq 0. \tag{1.29}$$

The elements M_m must also satisfy (1.17). These measurements do not have a clear physical meaning.

1.6.3 LOCC

Another class of measurements are *Local Operations and Classical Communication* (LOCC) (Nielsen and Chuang, 2000, p. 573). These form a smaller class of measurements than separable measurements as there exist separable measurements that are not LOCC (Bennett *et al.*, 1999). Unlike separable measurements LOCC has a clear physical meaning. Consider the situation in which there are n experimenters, each with a single copy of ρ. Each experimenter can only measure his own copy of ρ but is allowed to communicate with the other experimenters. LOCC measurements are easier to perform than some collective measurements, though the latter may often lead to a far more accurate estimate.

1.6.4 Adaptive measurements

In the field of quantum statistical inference one often comes across the term *adaptive* measurements. Measurements of this type were introduced by Nagaoka (1988, 1989) because the 'optimal' measurements on a single quantum state often depend on the unknown state itself. (Nagaoka (1989) is included in (Hayashi, 2005, pp.125-132). See also Barndorff-Nielsen and Gill (2000) for a discussion of the adaptive measurement strategy.) This dilemma of the optimal estimation strategy depending on the unknown parameter was described by Cochran (1973) as 'You tell me the value of the parameter θ and I promise to design the best experiment for estimating θ.'

In an *adaptable* measurement procedure, n' measurements (with n' small) are performed on the first n' copies of ρ to get a rough estimate $\hat{\rho}$ of the state. Next the POVM which is optimal for $\hat{\rho}^{\otimes n-n'}$ is used on $\rho^{\otimes n-n'}$. An adaptive measurement may be collective but not separable, or separable but not LOCC, or simply LOCC, depending on what the optimal measurment is for $\hat{\rho}^{\otimes n-n'}$.

An example is now given of an adaptive measurement. (The following procedure is LOCC.) Consider the state

$$\rho_\phi = \begin{pmatrix} \cos^2(\theta/2) & \sin(\theta/2)\cos(\theta/2)e^{-i\phi} \\ \sin(\theta/2)\cos(\theta/2)e^{i\phi} & \sin^2(\theta/2) \end{pmatrix},$$

where θ is known. The case $\theta = \pi/2$ is special, as there exists an optimal POVM that does not depend on ϕ. When $\theta \neq \pi/2$, the optimal POVM depends on ϕ, and one such POVM is

$$M = (M_0, \mathbb{I} - M_0), \qquad M_0 = \frac{1}{2}\begin{pmatrix} 1 & -ie^{-i\phi} \\ ie^{i\phi} & 1 \end{pmatrix}.$$

If ρ_ϕ is measured in x (see after (1.23)), outcome 0 is observed with probability $p(0;\phi) = (1 + \sin\theta\cos\phi)/2$, and 1 with probability $p(1;\phi) = (1 - \sin\theta\cos\phi)/2$. Put $N = n'/2$ and let $N_{x=0}$ be the number of times that outcome 0 is observed when ρ_ϕ is measured N times in x. This gives an estimate $N_{x=0}/N$ of $p(0;\phi)$, and since θ is known, an estimate of $\cos\phi$.

If ρ_ϕ is measured in y, outcome 0 is observed with probability $p(0;\phi) = (1+\sin\theta\sin\phi)/2$, and outcome 1 with probability $p(1;\phi) = (1-\sin\theta\sin\phi)/2$. Put $N = n'/2$ and let $N_{y=0}$ be the number of times that outcome 0 is observed when ρ_ϕ is measured N times in y. This gives an estimate $N_{y=0}/N$ of $p(0;\phi)$, and since θ is known, an estimate of $\sin\phi$.

Using estimates of $\cos\phi$ and $\sin\phi$, an estimate $\hat{\phi}$ of ϕ is obtained. The 'optimal' POVM,

$$M = (M_0, \mathbb{I} - M_0), \qquad M_0 = \frac{1}{2}\begin{pmatrix} 1 & -ie^{-i\hat{\phi}} \\ ie^{i\hat{\phi}} & 1 \end{pmatrix}$$

is used on the remaining $n-n'$ copies of ρ_ϕ. Using this measurement, outcome 0 is observed with probability $p(0;\phi) = (1+\sin\theta\sin(\phi-\hat{\phi}))/2$, and outcome 1 with probability $p(1;\phi) = (1-\sin\theta\sin(\phi-\hat{\phi}))/2$. Provided that $\hat{\phi} - \phi \in [-\pi/2, \pi/2]$, the estimate $p(0;\hat{\phi})$ of $p(0;\phi)$ can be used to get a more accurate estimate $\hat{\phi}'$ of ϕ, namely

$$\hat{\phi}' = \hat{\phi} + \arcsin\left(\frac{2p(0;\hat{\phi}) - 1}{\sin\theta}\right).$$

It has been shown by Fujiwara (2006) that, for an adaptive quantum estimation scheme, the sequence of maximum likelihood estimators is strongly consistent and asymptotically efficient.

1.6.5 Separate measurements

These form the smallest class of measurements. A separate measurement is LOCC with no communication. That is, there are n experimenters, each with a copy of ρ, and no communication is allowed between them.

1.7 Quantum Channels

A *quantum channel* is a trace-preserving completely-positive map (TP-CP map) sending density matrices to density matrices. It can be thought of as the quantum analogue of a stochastic mapping. A mapping $\mathcal{F}$ is *positive* if for all $A \geq 0$, $\mathcal{F}(A) \geq 0$. A mapping $\mathcal{F}$ is *completely positive* if for all positive integers k and $B \geq 0$, $(\mathbb{I}_k \otimes \mathcal{F})(B)$ is positive (Nielsen and Chuang, 2000, p. 367).

The mathematical formalism for a quantum channel is originally due to Choi (1975). He showed that a linear map Φ from $\mathcal{M}_n$ to $\mathcal{M}_m$ ($\mathcal{M}_m$ is the set of $m \times m$ complex matrices) is completely positive if and only if it can be written in the form $\Phi(A) = \sum_k E_k A E_k^\dagger$ where E_k are $m \times n$ matrices. For the map Φ to be a quantum channel, it is further required that the mapping is trace-preserving. The map Φ is trace-preserving if and only if $\sum_k E_k^\dagger E_k = \mathbb{I}_n$. Such a set of matrices $E = \{E_k\}$ are known as a set of *Kraus operators*.

Thus, any quantum channel can be represented using Kraus operators E_k as (Kraus, 1983, Nielsen and Chuang, 2000, Bengtsson and Życzkowski, 2006)

$$\rho_0 \mapsto \sum_k E_k \rho_0 E_k^\dagger, \tag{1.30}$$

where

$$\sum_k E_k^\dagger E_k = \mathbb{I}_n. \tag{1.31}$$

The form (1.30) for a general quantum channel can be derived as follows. Consider the composite state formed by the input state $\rho_0 \in S(\mathcal{H})$ and the environment $\rho_{env} \in S(\mathcal{H}_{env})$. Put

$$\rho = \rho_0 \otimes \rho_{env}.$$

Suppose that ρ undergoes unitary evolution, i.e.

$$\begin{aligned} \rho &\mapsto U\rho U^\dagger, \\ &= U(\rho_0 \otimes \rho_{env})U^\dagger, \end{aligned}$$

where $U^\dagger U = UU^\dagger = \mathbb{I}$. It is assumed that ρ_{env} is a pure state, with $\rho_{env} = |0\rangle\langle 0|$, where $|0\rangle, |1\rangle, \ldots, |d-1\rangle$, form a basis of $\mathcal{H}_{env}$. (This is the only place in this thesis where $|0\rangle$ does not refer to $(1,0)^T$.) It is found that ρ_0 has undergone the following transformation

$$\begin{aligned} \rho_0 &\mapsto \mathrm{tr}_{env}\{U(\rho_0 \otimes |0\rangle\langle 0|)U^\dagger\} \\ &= \sum_k \langle k|U|0\rangle \rho_0 \langle 0|U^\dagger|k\rangle \\ &= \sum_k E_k \rho_0 E_k^\dagger, \quad E_k = \langle k|U|0\rangle. \end{aligned}$$

Some examples will now be given of parametric families of quantum channels. A unitary channel is a mapping which transforms a state $\rho \in S(\mathbb{C}^d)$ to the state $U\rho U^\dagger \in S(\mathbb{C}^d)$, where U is a $d \times d$ complex unitary matrix. Chapter 5 considers the problem of estimating the parameter θ in a unitary channel acting on $\mathcal{H} = \mathbb{C}^2$, with unitary matrix

$$U_\theta = \begin{pmatrix} 1 & 0 \\ 0 & e^{i2\pi\theta} \end{pmatrix}. \tag{1.32}$$

If this channel acts on the state ρ, given in (1.10), it produces the output state

$$U_\theta \rho U_\theta^\dagger = 1/2 \begin{pmatrix} 1 & e^{-i2\pi\theta} \\ e^{i2\pi\theta} & 1 \end{pmatrix}.$$

Examples of non-unitary channels (channels with at least two non-zero Kraus operators E_k) are now considered.

The family of depolarizing channels $\mathcal{E} : S(\mathbb{C}^d) \to S(\mathbb{C}^d)$ is the set of mappings (Nielsen and Chuang, 2000, p. 378)

$$\rho_0 \mapsto (1-\epsilon)\rho_0 + \frac{\epsilon}{d}\mathbb{I}_d, \quad 0 < \epsilon < 1. \tag{1.33}$$

A depolarizing channel describes the process in which with probability $1-\epsilon$ the state is left unchanged, and with probability ϵ is replaced by the completely mixed state $\mathbb{I}_d/d$.

The family of 2-dimensional depolarizing channels $\mathcal{E} : S(\mathbb{C}^2) \to S(\mathbb{C}^2)$ have Kraus operators (Nielsen and Chuang, 2000, p. 397)

$$E_0 = \sqrt{1 - \frac{3\epsilon}{4}}\mathbb{I}_2, \quad E_1 = \sqrt{\frac{\epsilon}{4}}\sigma_x, \quad E_2 = \sqrt{\frac{\epsilon}{4}}\sigma_y, \quad E_3 = \sqrt{\frac{\epsilon}{4}}\sigma_z.$$

This family of channels forms a subset of the set of Pauli channels. The family of *Pauli channels* $\mathcal{E} : S(\mathbb{C}^2) \to S(\mathbb{C}^2)$, indexed by the parameters (p_0, p_1, p_2, p_3), with $p_j \geq 0$ and $\sum_{j=0}^{3} p_j = 1$, is the set of channels (Fujiwara and Imai, 2003)

$$\rho_0 \mapsto \sum_{i=0}^{3} p_i \sigma_i \rho_0 \sigma_i, \tag{1.34}$$

where $\sigma_0 = \mathbb{I}_2$ and σ_1, σ_2 and σ_3 are the Pauli matrices (see (1.15)).

The family of *generalized Pauli channels* $\mathcal{E} : S(\mathbb{C}^d) \to S(\mathbb{C}^d)$, indexed by the parameters $(p_0, p_1, p_2, \ldots, p_{d^2-1})$, with $p_j \geq 0$ and $\sum_{j=0}^{d^2-1} p_j = 1$, is the set of channels (Fujiwara and Imai, 2003)

$$\rho_0 \mapsto \sum_{k=0}^{d^2-1} p_k U_k \rho_0 U_k^\dagger, \qquad \mathrm{tr}\{U_k^\dagger U_l\} = d\delta_{kl}. \tag{1.35}$$

The choice of unitary matrices U_j is arbitrary.

The family of *amplitude damping channels* $\mathcal{E} : S(\mathbb{C}^2) \to S(\mathbb{C}^2)$, indexed by the parameter γ, is the set of channels with Kraus operators (Nielsen and Chuang, 2000, p. 380)

$$E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad E_1 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\gamma} \end{pmatrix}. \tag{1.36}$$

This channel describes energy dissipation: every state is brought closer to the pure state $|0\rangle\langle 0|$.

The family of *generalized damping channels* $\mathcal{E} : S(\mathbb{C}^2) \to S(\mathbb{C}^2)$ (Nielsen and Chuang, 2000, p. 382), indexed by the parameters γ, p is the set of channels with Kraus operators

$$\begin{aligned} E_0 &= \sqrt{p} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad E_1 = \sqrt{p} \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\gamma} \end{pmatrix} \\ E_2 &= \sqrt{1-p} \begin{pmatrix} \sqrt{\gamma} & 0 \\ 0 & 1 \end{pmatrix}, \quad E_3 = \sqrt{1-p} \begin{pmatrix} 0 & 0 \\ \sqrt{1-\gamma} & 0 \end{pmatrix}. \end{aligned} \tag{1.37}$$

The parameter $p \in [0, 1]$ represents the temperature of the environment.

1.8 Fisher information

1.8.1 One-parameter case

Given a univariate family of probability distributions with probability density functions $p(x; \theta)$, the *Fisher information*, introduced by Fisher (1922), is

defined as

$$F_\theta \equiv \int p(x;\theta)\left(\frac{\partial \ln p(x;\theta)}{\partial\theta}\right)^2 dx \tag{1.38}$$

$$= \int \frac{1}{p(x;\theta)}\left(\frac{\partial p(x;\theta)}{\partial\theta}\right)^2 dx. \tag{1.39}$$

Intuitively, Fisher information gives a measure of the amount of 'information' about θ contained in an observation. If the random variable X is discrete with probabilities $p(1;\theta),\ldots,p(n;\theta)$, then the Fisher information can be expressed as

$$F_\theta = \sum_{m=1}^{n} \frac{1}{p(m;\theta)}\left(\frac{dp(m;\theta)}{d\theta}\right)^2.$$

Proposition 1.1 *The Fisher information from n i.i.d. observations $X_1, X_2, \ldots X_n$ is equal to nF_θ where F_θ is the Fisher information from a single observation X_j.*

Proof. The Fisher information for a single observation X_j can be written as

$$F_\theta = -E\left[\frac{d^2 l(\theta;x)}{d\theta^2}\right], \tag{1.40}$$

where $l(\theta;x) = \log L(\theta;x)$ is the log-likelihood (the natural logarithm of the likelihood function). For the case of n observations,

$$L(\theta;x_1,\ldots,x_n) \equiv \prod_{i=1}^{n} p(x_i;\theta). \tag{1.41}$$

Thus

$$l(\theta;x_1,\ldots,x_n) = \sum_{i=1}^{n} \log p(x_i;\theta), \tag{1.42}$$

and the Fisher information from n observations is equal to

$$F_\theta^{(n)} = -E\left[\sum_{i=1}^{n} \frac{d^2 l(\theta;x_i)}{d\theta^2}\right] \tag{1.43}$$

$$= nF_\theta. \tag{1.44}$$

The importance of Fisher information is seen in the *Cramér–Rao inequality*. This states that the mean square error of an unbiased estimator t is greater than or equal to the reciprocal of the Fisher information, i.e.

$$E[(\hat{\theta}-\theta)^2] \geq \frac{1}{F_\theta}. \tag{1.45}$$

The right hand side of (1.45) is known as the *Cramér–Rao bound*. Under mild regularity conditions for $p(x;\theta)$, using a maximum likelihood estimator, as the number of observations $n \to \infty$ (Van der Vaart, 1998, p. 63)

$$\sqrt{n}(\hat{\theta} - \theta) \rightsquigarrow \mathcal{N}(0, F_\theta^{-1}), \tag{1.46}$$

and so, assuming the estimator is unbiased,

$$nE[(\hat{\theta} - \theta)^2] \to \frac{1}{F_\theta}. \tag{1.47}$$

(The symbol $\rightsquigarrow$ denotes convergence in distribution.) The larger the Fisher information, the more accurately the unknown parameter can be estimated. A standard approach to estimation of a parameter, in a known family of distributions, is to use the maximum likelihood estimator. Consequently the result (1.47) is of great importance: it enables the asymptotic behaviour of an estimate to be quantified.

1.8.2 Multi-parameter case

Given a p-parameter family of probability distribution with probability density functions $p(x;\theta^1,\ldots,\theta^p)$, the *Fisher information*, is the $p \times p$ matrix F_θ with entries

$$\begin{aligned}(F_\theta)_{jk} &\equiv \int p(x;\theta)\left(\frac{\partial \ln p(x;\theta)}{\partial \theta^j}\right)\left(\frac{\partial \ln p(x;\theta)}{\partial \theta^k}\right)dx \\ &= \int \frac{1}{p(x;\theta)}\left(\frac{\partial p(x;\theta)}{\partial \theta^j}\right)\left(\frac{\partial p(x;\theta)}{\partial \theta^k}\right)dx.\end{aligned}$$

The Cramér–Rao inequality becomes a matrix inequality. This states that the mean square error of an unbiased estimator t is greater than or equal to the inverse of the Fisher information, i.e.

$$E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] \geq F_\theta^{-1}. \tag{1.48}$$

This means that the matrix $E[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^T] - F_\theta^{-1}$ is positive semi-definite, i.e. for all $v \in \mathbb{R}^p$,

$$v^T(E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] - F_\theta^{-1})v \geq 0.$$

1.9 Quantum information

Definition 1.3 *A Riemannian metric on a manifold $\mathcal{M}$ is a mathematical object that assigns smoothly to each point x of $\mathcal{M}$, and each coordinate system*

$\theta = (\theta^1, \ldots, \theta^p)$ round x, a positive semi-definite $p \times p$ matrix $g_\theta(x)$ such that, for another coordinate system $\phi = (\phi^1, \ldots, \phi^p)$,

$$g_\phi(x) = \left(\frac{d\theta}{d\phi}\right) g_\theta(x) \left(\frac{d\theta}{d\phi}\right)^T, \tag{1.49}$$

or, in terms of elements of $g_\phi(x)$,

$$g_\phi(x)_{ij} = \sum_{k,l} g_\theta(x)_{kl} \frac{d\theta^i}{d\phi^k} \frac{d\theta^j}{d\phi^l}. \tag{1.50}$$

It has been shown by Morozova and Čencov (1990) that, up to a constant factor, the Fisher information is the unique monotone Riemannian metric on Θ. Several types of *quantum information* have been suggested as quantum versions of Fisher information (Petz and Sudár, 1996), defined from a parametric family of states ρ_θ. As the Fisher information is a Riemannian metric, any quantum analogue of Fisher information should also be a Riemannian metric. The following properties are important when considering Riemannian metrics.

Invariance.

Two parametric families of states ρ_θ and σ_θ are said to be *equivalent* ($\rho_\theta \sim \sigma_\theta$) (Petz and Sudár, 1996) if there exist two fixed TP-CP maps $\mathcal{E}, \mathcal{F}$ such that

$$\rho_\theta = \mathcal{E}(\sigma_\theta), \qquad \sigma_\theta = \mathcal{F}(\rho_\theta).$$

The Riemannian metric J is said to be *invariant* (Petz and Sudár, 1996) if

$$\rho_\theta \sim \sigma_\theta \quad \text{implies} \quad J(\rho_\theta) = J(\sigma_\theta).$$

Monotonicity.

The Riemannian metric J is said to be *monotone* (Petz and Sudár, 1996) if

$$J(\rho_\theta) \geq J(\mathcal{E}(\rho_\theta))$$

for all TP-CP maps $\mathcal{E}$.

A well-defined Riemannian metric must be invariant and it is desirable that it is monotone. (If a metric is monotone then it is also invariant.) It has been shown by Petz and Sudár (1996) that there is no unique monotone quantum information quantity. The most frequently encountered monotone metrics in recent literature are the *Symmetric Logarithmic Derivative (SLD)*, *Right Logarithmic Derivative (RLD)* and *Kubo-Mori-Bogoliubov (KMB)* metrics, which are defined in (1.51) – (1.54) and (1.59) – (1.62).

Given a one-parameter family of states ρ_θ, these quantum information quantities can be expressed as

$$H^x = \mathrm{tr}\{\lambda_x^\dagger \rho \lambda_x\}, \tag{1.51}$$

where the parameter θ has been suppressed, and the quantum scores λ_x (the quantum analogues of the logarithmic derivative $dl/d\theta$) are defined as the solutions to the following matrix equations

$$\frac{d\rho}{d\theta} = \frac{1}{2}(\rho\lambda_{SLD} + \lambda_{SLD}\rho) \tag{1.52}$$

$$\frac{d\rho}{d\theta} = \rho\lambda_{RLD} \tag{1.53}$$

$$\lambda_{KMB} = \frac{d\log\rho}{d\theta}. \tag{1.54}$$

These all satisfy

$$\mathrm{tr}\{\rho\lambda_x\} = 0. \tag{1.55}$$

The quantum scores λ_{RLD} and λ_{KMB} are defined only when ρ_θ has full rank, i.e. when ρ_θ is invertible. When ρ_θ does not have full rank, λ_{SLD} is not defined uniquely, though H^{SLD} does not depend on the choice of λ_{SLD}.

The SLD quantum information is the most commonly used quantum information quantity. It is the minimum among the set of monotone quantum information quantities (Petz and Sudár, 1996). In this thesis the SLD quantum information will be denoted simply by H or H_θ, and the SLD quantum score by λ or λ_θ.

The SLD quantum information for n copies of the state ρ_θ (i.e. $\rho_\theta^{(n)} = \rho_\theta \otimes \rho_\theta \otimes \cdots \otimes \rho_\theta$) is n times the SLD quantum information for the state ρ_θ, that is

$$H(\rho_\theta^{(n)}) = nH_\theta(\rho_\theta). \tag{1.56}$$

To see this, note that in this case

$$\frac{d\rho^{(n)}}{d\theta} = \frac{d\rho}{d\theta}\otimes\rho\otimes\cdots\otimes\rho + \rho\otimes\frac{d\rho}{d\theta}\otimes\rho\otimes\cdots\otimes\rho + \rho\otimes\cdots\otimes\rho\otimes\frac{d\rho}{d\theta}. \tag{1.57}$$

Let λ be a possible solution for the SLD score for ρ. Then a possible solution for the SLD score for $\rho^{(n)}$ is

$$\lambda^{(n)} = \lambda\otimes\mathbb{I}\otimes\cdots\otimes\mathbb{I} + \mathbb{I}\otimes\lambda\otimes\mathbb{I}\otimes\cdots\otimes\mathbb{I} + \cdots + \mathbb{I}\otimes\cdots\otimes\mathbb{I}\otimes\lambda. \tag{1.58}$$

The SLD quantum information can be written as

$$\begin{aligned} H(\rho_\theta^{(n)}) &= \mathrm{tr}\left\{\frac{d\rho^{(n)}}{d\theta}\lambda^{(n)}\right\} \\ &= \mathrm{tr}\{\frac{d\rho}{d\theta}\lambda\otimes\rho\otimes\cdots\otimes\rho\} \\ &+ \mathrm{tr}\{\frac{d\rho}{d\theta}\otimes\rho\lambda\otimes\cdots\otimes\rho\} \\ &\vdots \\ &+ \mathrm{tr}\{\rho\otimes\cdots\otimes\rho\otimes\frac{d\rho}{d\theta}\lambda\}. \end{aligned}$$

Since $\mathrm{tr}\{\rho\lambda\}=0$, the only non-zero terms are those of the form $\mathrm{tr}\{\rho\otimes\cdots\otimes\rho\otimes(d\rho/d\theta)\lambda\otimes\rho\cdots\otimes\rho\}$. As there are n terms of this form, (1.56) holds.

1.9.1 Multi-parameter models

Given a p-parameter family of quantum states $\rho(\theta^1,\ldots,\theta^p)$, the quantum analogues of the logarithmic derivative $(\partial l/\partial\theta^j)$ are defined as the solutions to the following matrix equations

$$\frac{\partial\rho}{\partial\theta^j} = \frac{1}{2}(\rho\lambda^j_{SLD}+\lambda^j_{SLD}\rho) \tag{1.59}$$

$$\frac{\partial\rho}{\partial\theta^j} = \rho\lambda^j_{RLD} \tag{1.60}$$

$$\lambda_{KMB} = \frac{\partial\log\rho}{\partial\theta^j}. \tag{1.61}$$

These all satisfy

$$\mathrm{tr}\{\rho\lambda^j_x\}=0.$$

The quantum informations are $p\times p$ matrices with entries

$$(H^x_\theta)_{jk} = \Re\mathrm{tr}\{\lambda^{\dagger j}_x\rho\lambda^k_x\}. \tag{1.62}$$

1.10 Braunstein-Caves inequality

Consider the parametric statistical model resulting from a measurement $M=\{M_m\}$ of a parametric family of states ρ_θ. This has the probability function

$$p(m;\theta)=\mathrm{tr}\{\rho_\theta M_m\}, \tag{1.63}$$

which gives Fisher information F_θ^M. Braunstein and Caves (1994) proved the inequality

$$F_\theta^M \leq H_\theta \tag{1.64}$$

for the one-parameter case. They showed that, in the one-parameter case, the SLD quantum information gives the maximum Fisher information that can be obtained from measuring a model ρ_θ.

1.10.1 Multi-parameter Braunstein-Caves inequality

Theorem 1.1 *Let F_θ^M be the Fisher information given by a measurement M on a parameterised quantum model $\{\rho_\theta : \theta = \theta^1, \ldots, \theta^p \in \mathbb{R}^p\}$, with SLD quantum information H_θ. Then*

$$F_\theta^M \leq H_\theta. \tag{1.65}$$

This means that the matrix $H_\theta - F_\theta^M$ is non-negative, which is equivalent to

$$\sum_{j,k} x_j x_k (F_\theta^M)_{j,k} \leq \sum_{jk} x_j x_k (H_\theta)_{jk}, \tag{1.66}$$

for all vectors $x = (x_1, \ldots, x_p) \in \mathbb{R}^p$.

Not only does (1.65) give an upper bound on the Fisher information, but the proof gives necessary and sufficient conditions for equality. The following proof is similar to that in (Ballester, 2005, p. 26).

Proof.

Denote by Ω^+ the set of outcomes which occur with non-zero probability.

Then

$$\begin{aligned}
\sum_{j,k} x_j x_k (F_\theta^M)_{jk} &= \sum_{j,k} x_j x_k \sum_{m\in\Omega^+} \frac{1}{p(m;\theta)} \left(\frac{\partial p(m;\theta)}{\partial \theta^j}\right)\left(\frac{\partial p(m;\theta)}{\partial \theta^k}\right) \\
&= \sum_{j,k} x_j x_k \sum_{m\in\Omega^+} \frac{1}{\mathrm{tr}\{\rho M_m\}} \left(\mathrm{tr}\left\{\frac{\partial \rho}{\partial \theta^j} M_m\right\}\right)\left(\mathrm{tr}\left\{\frac{\partial \rho}{\partial \theta^k} M_m\right\}\right) \\
&= \sum_{j,k} x_j x_k \sum_{m\in\Omega^+} \frac{1}{\mathrm{tr}\{\rho M_m\}} \left(\Re\mathrm{tr}\{\lambda^j \rho M_m\}\right)\left(\Re\mathrm{tr}\{\lambda^k \rho M_m\}\right) \\
&\leq \sum_{j,k} x_j x_k \sum_{m\in\Omega^+} \frac{1}{\mathrm{tr}\{\rho M_m\}} |\mathrm{tr}\{\lambda^j \rho M_m\}||\mathrm{tr}\{\lambda^k \rho M_m\}| \\
&= \sum_{m\in\Omega^+} \frac{1}{\mathrm{tr}\{\rho M_m\}} |\mathrm{tr}\{\lambda \rho M_m\}|^2, \quad \text{where } \lambda = \sum_j x_j \lambda^j, \\
&= \sum_{m\in\Omega^+} \frac{1}{\mathrm{tr}\{\rho M_m\}} |\mathrm{tr}\{(M_m^{1/2}\lambda\rho^{1/2})(M_m^{1/2}\rho^{1/2})^\dagger\}|^2 \\
&\leq \sum_{m\in\Omega^+} \frac{1}{\mathrm{tr}\{\rho M_m\}} \mathrm{tr}\{\rho M_m\}\mathrm{tr}\{M_m \lambda\rho\lambda\} \\
&= \sum_{m\in\Omega^+} \mathrm{tr}\{M_m \lambda\rho\lambda\} \\
&\leq \mathrm{tr}\{\lambda\rho\lambda\} \qquad (1.67) \\
&= \sum_{j,k} x_j x_k \mathrm{tr}\{\lambda^j \rho \lambda^k\} \qquad (1.68) \\
&= \sum_{j,k} x_j x_k (H_\theta)_{jk}.
\end{aligned}$$

Inequality (1.67) follows from the fact that $\sum_{m\in\Omega^+} M_m \leq \mathbb{I}$, since $\sum_{m\in\Omega} M_m = \mathbb{I}$ and $M_m \geq 0$.

Theorem 1.2 *Equality holds in (1.65) if and only if*

$$M_m^{1/2}\lambda^j\rho^{1/2} = \xi_m^j M_m^{1/2}\rho^{1/2}, \quad \xi_m^j \in \mathbb{R} \quad \forall j, m. \qquad (1.69)$$

The following proof is similar to that in (Ballester, 2005, p. 27).
Proof.

Equality holds in (1.65) if and only if the following three conditions are met

$$\begin{aligned} \Im\mathrm{tr}\{\lambda^j \rho M_m\} &= 0 \quad \forall j, m, && (1.70)\\ M_m^{1/2}\lambda\rho^{1/2} &= z_m M_m^{1/2}\rho^{1/2}, \quad \text{some } z_m \in \mathbb{C}, \quad \forall m, \lambda, && (1.71)\\ \sum_{m\in\Omega^+} \mathrm{tr}\{M_m\lambda\rho\lambda\} &= \mathrm{tr}\{\lambda\rho\lambda\}, \quad \forall m, \lambda. && (1.72) \end{aligned}$$

Since

$$\mathrm{tr}\{\lambda\rho\lambda\} = \sum_{m\in\Omega^+} \mathrm{tr}\{M_m\lambda\rho\lambda\} + \sum_{m\in\Omega\setminus\Omega^+} \mathrm{tr}\{M_m\lambda\rho\lambda\},$$

equality holds in (1.65) if and only if the following three conditions are met

$$\begin{aligned} \Im\mathrm{tr}\{\lambda^j \rho M_m\} &= 0 \quad \forall j, m, && (1.73)\\ M_m^{1/2}\lambda^j\rho^{1/2} &= z_m^j M_m^{1/2}\rho^{1/2}, \quad \text{some } z_m^j \in \mathbb{C}, \quad \forall j, m, && (1.74)\\ \sum_{m\in\Omega\setminus\Omega^+} \mathrm{tr}\{M_m\lambda^j\rho\lambda^k\} &= 0, \quad \forall j, k, m. && (1.75) \end{aligned}$$

Theorem 1.2 will be proved by showing that

(i) if (1.69) holds, then (1.73), (1.74) and (1.75) hold, and thus equality holds in (1.65) (consequently (1.69) is a sufficient condition);

(ii) the conditions (1.73) and (1.74) both hold only if (1.69) holds (thus (1.69) is a necessary condition).

(i) If (1.69) holds, then (1.74) obviously holds. Pre-multiplying (1.69) by $M_m^{1/2}$, post-multiplying by $\rho^{1/2}$ and taking the trace shows that (1.73) also holds. Note that

$$\mathrm{tr}\{A^\dagger A\} = 0 \quad \text{if and only if} \quad A = 0. \tag{1.76}$$

For $m \in \Omega \setminus \Omega^+$, $p_m = 0$, and so, since $p_m = \mathrm{tr}\{(M_m^{1/2}\rho^{1/2})^\dagger (M_m^{1/2}\rho^{1/2})\}$, by (1.76) it follows that $M_m^{1/2}\rho^{1/2} = 0$. If (1.69) holds, then for $m \in \Omega \setminus \Omega^+$, $M_m^{1/2}\lambda^j\rho^{1/2} = 0$ and so (1.75) holds.

(ii) First, it will be assumed that (1.74) holds. Pre-multiplying (1.74) by $M_m^{1/2}$, post-multiplying by $\rho^{1/2}$ and taking the trace gives

$$\mathrm{tr}\{M_m\lambda^j\rho\} = z_m^j p_m, \quad \forall j, m. \tag{1.77}$$

For condition (1.73) to hold, z_m^j must be real. Thus (1.73) and (1.74) both hold only if (1.69) holds.

1.10.2 Equality

For one-parameter models, equality holds in (1.64) if and only if

$$M_m^{1/2}\lambda\rho^{1/2} = \xi_m M_m^{1/2}\rho^{1/2}, \quad \forall m, \quad \xi_m \in \mathbb{R}. \tag{1.78}$$

As λ is self-adjoint, it can be written as

$$\lambda = \sum_i \mu_i |e_i\rangle\langle e_i|.$$

The POVM $M = \{M_i = |e_i\rangle\langle e_i|\}$ satisfies (1.78) and so, using this POVM, equality holds in (1.64). It has been shown by Barndorff-Nielsen and Gill (2000) that, in general, the optimal POVM will depend on the unknown parameter. To get around this, an adaptive measurement scheme can be used, as described in Section 1.6.4. There are a few families of states for which the optimal POVM does not depend on the parameter, such as the set of states corresponding to the 'equator' of the Bloch ball, given by the set of density matrices

$$\rho_\theta = 1/2 \begin{pmatrix} 1 & e^{-i2\pi\theta} \\ e^{i2\pi\theta} & 1 \end{pmatrix}, \quad \theta \in [0, 1),$$

and sets of quasi-classical states. *Quasi-classical states* are defined as sets of states for which the eigenvectors $\{|w_i\rangle\}$ are known, i.e. families of states of the form

$$\rho_\theta = \sum_{i=1}^{d} p_i(\theta)|w_i\rangle\langle w_i|.$$

In the multi-dimensional case there exist sets of states for which the bound (1.65) is not attainable even using an adaptive scheme (Barndorff-Nielsen and Gill, 2000). In Section 4.2, it is shown that for any non-degenerate set of pure states (these can be parameterised by a maximum of $2(d-1)$ parameters), equality holds in (1.65) only if the number of parameters $p \leq d-1$.

The fact that the SLD quantum information is not in general attainable means that it cannot in general be used to find the optimal estimation method for quantum states. The problem of optimally estimating n identical quantum states has recently been solved by Guţă *et al.* (2007), Kahn and Guţă (2009). The solutions presented in these papers are based on *quantum local asymptotic normality*: given n copies of a state, as $n \to \infty$ the joint state converges to a statistical model consisting of a classical Gaussian distribution and a quantum Gaussian distribution. The optimal estimation procedure for these models is known, having been solved by Yuen and Lax

(1973), Holevo (1982). Quantum local asymptotic normality was first studied in Hayashi (2003a,b) and used for estimation in Hayashi and Matsumoto (2004). It was later made more rigorous by (Guţă and Kahn, 2006, Guţă and Jencova, 2007).

The optimal estimation of qubits has been solved explicitly in the Bayesian set-up, in the particular case of an invariant prior in (Bagan *et al.*, 2006).

1.10.3 Equality in the case of pure states

Putting together (1.48) and (1.65) gives the *quantum Cramér–Rao inequality*

$$E[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^T] \geq H_\theta^{-1}. \tag{1.79}$$

A result of Matsumoto (1997) will now be considered. In the case of pure states, it gives a concise necessary and sufficient condition for the existence of a POVM such that equality holds in (1.79).

Theorem 1.3 *Let $\{\rho_\theta : \theta \in \Theta\}$ be a parameterised family of pure states with $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$. Then there exists a POVM and estimator such that equality holds in (1.79) at $\theta = \theta_0$, if and only if*

$$\Im\langle l_j(\theta_0)|l_k(\theta_0)\rangle = 0, \quad \forall j,k, \tag{1.80}$$

where $|l_j(\theta)\rangle = \lambda_\theta^j|\psi_\theta\rangle$ (Matsumoto, 1997, 2002, Fujiwara, 2002).

In Section 4.2, it is shown that condition (1.80) is equivalent to the simpler condition

$$\Im\langle\psi^{(j)}(\theta_0)|\psi^{(k)}(\theta_0)\rangle = 0, \quad \forall j,k, \quad |\psi^{(j)}\rangle = \frac{\partial|\psi\rangle}{\partial\theta^j}. \tag{1.81}$$

When (1.80) is satisfied, a POVM giving equality in (1.79) is given explicitly by (Ballester, 2004a)

$$\begin{aligned} M_m &= |b_m\rangle\langle b_m|, \qquad m = 1,\ldots,p+1, \\ M_{p+2} &= \mathbb{I} - \sum_{m=1}^{p+1} M_m, \\ |b_m\rangle &= \sum_{n=1}^{p+1} O_{mn}|v_n\rangle, \\ |v_m\rangle &= \sum_n (H^{-1/2})_{mn}|l_n\rangle, \quad |v_{p+1}\rangle = |\psi\rangle, \end{aligned} \tag{1.82}$$

with O a $(p+1)\times(p+1)$ real orthogonal matrix satisfying $O_{m,p+1} \neq 0$. That this POVM does indeed give equality in (1.79) can be seen from Lemma 9 of Fujiwara (2002).

An original proof of the necessity part of Theorem 1.3 will now be given.

Lemma 1.2 *Condition (1.80) is a necessary condition for equality in (1.79).*

Proof. For equality in (1.79) it is necessary that equality holds in (1.65). Equality holds in (1.65) if and only if

$$M_m^{1/2}\lambda^k\rho^{1/2} = \xi_m^k M_m^{1/2}\rho^{1/2}, \quad \xi_m^k \in \mathbb{R} \quad \forall k, m. \tag{1.83}$$

For pure states, (1.83) becomes

$$M_m^{1/2}|l_k\rangle\langle\psi| = \xi_m^k M_m^{1/2}|\psi\rangle\langle\psi|, \quad \forall k, m.$$

Thus equality holds in (1.65) if and only if

$$M_m^{1/2}|l_k\rangle = \xi_m^k M_m^{1/2}|\psi\rangle, \quad \forall k, m. \tag{1.84}$$

Taking the transpose of (1.84) gives

$$\langle l_j|M_m^{1/2} = \xi_m^j\langle\psi|M_m^{1/2}, \quad \forall j, m. \tag{1.85}$$

Pre-multiplying the left hand side of (1.84) by the left hand side of (1.85), and the right hand side of (1.84) by the right hand side of (1.85) gives the necessary condition

$$\langle l_j|M_m|l_k\rangle = \xi_m^j\xi_m^k p_m, \quad \forall j, k, m.$$

Summing over m, and using the result $\sum_m M_m = \mathbb{I}$, gives

$$\langle l_j|l_k\rangle = \sum_m \xi_m^j\xi_m^k p_m.$$

As $\xi_m^j\xi_m^k$ and p_m are all real, it follows that $\langle l_j|l_k\rangle$ is real and (1.80) is a necessary condition for equality in (1.65), and thus a necessary condition for equality in (1.79).

That (1.80) is a sufficient condition for equality in (1.79), follows from Ballester's result that if (1.80) holds, then the POVM given in (1.82) gives equality in (1.65).

1.10.4 Attainable measurements - the 2-dimensional case

For quantum statistical models with $\mathcal{H} = \mathbb{C}^2$, equality holds in the Braunstein-Caves inequality (1.64) only if every element of the POVM $M = \{M_k\}$ has rank 1. This was shown for pure states by Barndorff-Nielsen and Gill (2000), and for mixed states by Luati (2004). An original proof of this result, which includes mixed and pure state models, will now be given. A necessary condition for equality in (1.64) is

$$M_m^{1/2}\lambda\rho^{1/2} = \xi_m M_m^{1/2}\rho^{1/2}, \quad \xi_m \in \mathbb{R}, \quad \forall m.$$

Pre-multiplying by $M_m^{1/2}$ and post-multiplying by $\rho^{1/2}$ gives

$$M_m\lambda\rho = \xi_m M_m\rho, \quad \forall m.$$

Then

$$M_m A_m = 0, \qquad \forall m, \tag{1.86}$$

where

$$A_m = \lambda\rho - \xi_m\rho.$$

Now, from (1.86), it is seen that M_m is singular unless $A_m = 0$ for all θ. It will be assumed that $A_m = 0$ for all θ. If this is so then

$$\lambda\rho = \xi_m\rho. \tag{1.87}$$

Taking the trace of (1.87) gives

$$\begin{aligned} \mathrm{tr}\{\lambda\rho\} &= \xi_m \mathrm{tr}\{\rho\} \\ 0 &= \xi_m. \end{aligned} \tag{1.88}$$

Thus from (1.87) and (1.88),

$$\lambda\rho = 0,$$

and so

$$\lambda\rho + (\lambda\rho)^\dagger = 2\frac{d\rho}{d\theta} = 0.$$

Thus, if $A_m = 0$, the model does not depend on θ. Assuming that the model does depend on θ, it follows that $A_m \neq 0$ and so M_m is singular. A consequence of this is that in the 2-dimensional case, the elements of attainable measurements have rank 1.

1.11 Estimation

Many quantum information processes can be represented as quantum channels. In practice, quantum channels are not known *a priori* and estimating them is of great importance.

There are several ways to estimate a quantum channel. One approach is quantum process tomography, which is discussed in chapter 8 of Nielsen and Chuang (2000). For this approach it is necessary to estimate how the channel acts on different bases of the Hilbert space plus linear combinations thereof. A problem with this method is that in many practical situations it is not possible to prepare these input states in the laboratory (de Martini *et al.*, 2003).

Another approach is to assume that the channel comes from a given parametric family of channels (Fujiwara, 2001, 2002, 2004, Fujiwara and Imai, 2003, Ballester, 2004a,b, Sarovar and Milburn, 2006). (The latter approach will be followed in this thesis.) A family of channels parametrized by a real parameter θ can be represented by Kraus operators depending on θ as

$$\rho_0 \mapsto \sum_k E_k(\theta)\rho_0 E_k^\dagger(\theta). \tag{1.89}$$

When estimating a quantum channel, there are many different factors to consider: how should the channels be arranged, and what combination of input state, POVM and estimator is best. The idea of finding the optimal input state was considered by Acín *et al.* (2001).

In general, for a parametric family of channels, different input states lead to different families of output states. The input state is chosen such that the family of output states has the maximum attainable SLD quantum information. The measurement which gives equality in (1.65) is chosen (an adaptive measurement may be needed), and the maximum likelihood estimator used.

In this thesis the performance of an estimation procedure is usually measured either by the mean square error $E[(\hat{\theta} - \theta)^2]$ or for unitary channels, where $U_{\hat{\theta}}$ is the estimate of the unitary matrix U_θ, by

$$1 - \langle F(U_{\hat{\theta}}, U_\theta)\rangle = 1 - \frac{\left\langle |\mathrm{tr}\{U_\theta^{-1}U_{\hat{\theta}}|^2 \right\rangle}{d^2}, \tag{1.90}$$

where $\langle\cdot\rangle$ denotes expectation. Often this cost function will be denoted simply by $1 - \langle F\rangle$. Given a family of channels $\mathcal{E}(\theta)$, an estimate $\hat{\theta}$ of a parameter θ will depend on n, the number of times the channel $\mathcal{E}(\theta)$ is used. Similarly, an estimate $U_{\hat{\theta}}$ of a unitary matrix U_θ will also depend on n. It is of interest to see how rapidly $\hat{\theta}$ approaches θ, and $U_{\hat{\theta}}$ approaches U_θ, as $n \to \infty$. The 'big

O' notation will be used for this purpose. It is said that '$f(n)$ is $O(g(n))$' if there exist constants c and n_0 such that for all $n > n_0$, $f(n) \leq cg(n)$ (Nielsen and Chuang, 2000, p. 136). That is, for large n, up to an unimportant factor, the function $g(n)$ is an upper bound on $f(n)$.

1.11.1 Important developments in channel estimation

Here a brief review is given of the major advances in the estimation of quantum channels.

A channel $\mathcal{E} : S(\mathbb{C}^d) \mapsto S(\mathbb{C}^d)$, can be extended to a channel $\mathbb{I} \otimes \mathcal{E} : S(\mathbb{C}^{d^2}) \to S(\mathbb{C}^{d^2})$ by

$$\rho_1 \mapsto (\mathbb{I}_d \otimes \mathcal{E})(\rho_1), \quad \rho_1 \in S(\mathbb{C}^{d^2}). \tag{1.91}$$

For many channels $\mathcal{E}$, when using (1.91), a maximally entangled input state is optimal, in terms of Fisher information. Often the Fisher information is significantly greater than can be obtained from the unextended channel $\mathcal{E} : S(\mathbb{C}^d) \mapsto S(\mathbb{C}^d)$. This was shown for a completely unknown unitary matrix in $SU(2)$ by Fujiwara (2002), and $SU(d)$ (close to the identity) by Ballester (2004b). This has also been shown for several non-unitary channels, in particular the 2-dimensional depolarizing channel (Fujiwara, 2001) and, more generally, the generalized Pauli channels (Fujiwara and Imai, 2003).

Another advantage of the extended channel $\mathbb{I} \otimes \mathcal{E}$ is that, using a maximally entangled input state, the output states are in one-to-one correspondence with the channel. This means that, in contrast to quantum tomography, the experimenter does not require many different input states: it is enough to have many copies of a maximally entangled state.

Using the extension (1.91) with a maximally entangled input state the mean square error and $1 - \langle F \rangle$ are $O(1/n)$ (Hayashi, 2006a). This rate at which $1 - \langle F \rangle$ approaches zero is known as the *standard quantum limit* (de Burgh and Bartlett, 2005), but can be surpassed (Hayashi, 2006a, Kahn, 2007, Imai and Fujiwara, 2007).

Another major step in estimation, when n copies of a channel are available, was the idea of using the following extension with an entangled input state, so that

$$\rho_2 \mapsto \mathcal{E}^{\otimes n}(\rho_2), \quad \rho_2 \in S(\mathbb{C}^{d^n}). \tag{1.92}$$

One of the first clear uses of this method for estimation was by Huelga *et al.* (1997).

Using the experimental setup (1.92), it has been shown that it is possible to estimate a unitary matrix with $1 - \langle F \rangle = O(1/n^2)$. This has been shown for estimation of an unknown unitary matrix in $SU(2)$ by Hayashi (2006a)

and $SU(d)$ by Kahn (2007). This rate at which $1 - \langle F \rangle$ approaches zero is known as the *Heisenberg limit* (Giovannetti *et al.*, 2004) and cannot be surpassed (Kahn, 2007).

A *reference frame* is a specific coordinate system. Estimation of a unitary matrix in $SU(2)$ is equivalent to the problem of transmiting a 3-dimensional reference frame from Alice to Bob. Alice encodes information about her reference frame in quantum particles, and then sends these to Bob. Bob measures the quantum particles, and from his results estimates Alice's reference frame. It has been shown that it is possible to do this with $1-\langle F \rangle = O(1/n^2)$ (Bagan *et al.*, 2004a,b, Chiribella *et al.*, 2004).

For most channels it is not possible to surpass the standard quantum limit. This has been shown for generalized Pauli channels by Fujiwara and Imai (2003). Recently it has been shown (Fujiwara and Imai, 2008) that for most channels, given n copies and using the setup (1.92), the SLD quantum information is $O(n)$. A consequence of this is that, from the quantum Cramér-Rao inequality (1.79), for these channels, the mean square error is $O(1/n)$.

It is also possible to use a channel repeatedly on the same input state, i.e.

$$\rho_0 \mapsto \mathcal{E}^n(\rho_0). \tag{1.93}$$

Kitaev (1996) suggested an l-stage iterative estimation scheme for the unitary matrix (1.32). At the kth stage U_θ acts 2^{k-1} times on the same input state. At each stage, several measurements are made. Using this information, an estimate $\hat{\theta}$ of θ is obtained satisfying $\Pr(|\hat{\theta} - \theta|_1 \leq 1/2^{l+2}) \geq 1 - \epsilon$. The value of ϵ can be made arbitrarily small by doing more measurements at each stage.

For a similar estimation scheme, Rudolph and Grover (2003) showed that, by choosing $\epsilon = 1/2^{2l}$, $1 - \langle F \rangle = O((\log n/n)^2)$. The advantage of these estimation schemes is their simplicity: they require no entanglement and only a single copy of $\mathcal{E}$. In spite of this, $1 - \langle F \rangle$ is still close to the Heisenberg limit.

This thesis contains, as far as the author is aware, the first complete method for iterative estimation similar to that of Kitaev (1996). It is also shown that an extension similar to (1.92) can be used to estimate n non-identical channels, with an entangled input state. This results in an increase in the rate at which the mean square error decreases, relative to using a separable state.

Chapter 2

Attainability of the information bound of Sarovar and Milburn

2.1 Introduction

The problem of estimating non-unitary quantum channels is more difficult than that of estimating unitary channels. The output states of non-unitary channels are mixed, and the SLD quantum information is generally more cumbersome to compute. Also, for multi-parameter families of mixed states, there is no known analogue of Matsumoto's condition (1.80) for equality in the Quantum Cramér–Rao inequality (1.79); neither is there a known method for computing the optimal POVM.

Sarovar and Milburn (2006) introduced an upper bound on the Fisher information obtained from measuring the output states of a parameterised family of channels. They also gave necessary and sufficient conditions for equality. Their bound depends on the Kraus operators of the channel and not on the set of output states. In this chapter it is shown that this bound is not generally attainable, and consequently does not generally give the optimal POVM. Thus the attempt of Sarovar and Milburn to find the optimal estimation strategy for non-unitary quantum channels is not succesful. (The work in this chapter has been published in O'Loan (2007).)

The problem of how to express the SLD quantum information of a noisy channel in terms of its Kraus operators has recently been solved by Fujiwara and Imai (2008) for the extended channel $\mathbb{I}_d \otimes \mathcal{E} : S(\mathbb{C}^{d^2}) \mapsto S(\mathbb{C}^{d^2})$. This puts an upper bound on the SLD quantum information for the unextended channel $\mathcal{E} : S(\mathbb{C}^d) \mapsto S(\mathbb{C}^d)$, but this bound will not, in general, be attainable.

2.1.1 The approach of Sarovar and Milburn

Sarovar and Milburn looked at estimating one-parameter quantum channels of the form

$$\rho_0 \mapsto \sum_k E_k(\theta)\rho_0 E_k^\dagger(\theta), \tag{2.1}$$

(see (1.30)). The input state ρ_0 is a known pure state, and is chosen such that the output state is in one-to-one correspondence with the channel. Since a specific value of θ corresponds to a specific channel, estimation of the channel reduces to a parameter estimation problem. Sarovar and Milburn were interested in finding the maximal Fisher information that can be obtained by measuring the output states of the set of channels (2.1). They were also interested in finding POVMs that attain this bound. First, Sarovar and Milburn derived the inequality

$$F_\theta^M \leq C_E(\theta). \tag{2.2}$$

where E denotes a set of Kraus operators $\{E_k\}$ and

$$C_E(\theta) = 4\sum_k \mathrm{tr}\{E_k'(\theta)\rho_0 E_k'^\dagger(\theta)\}, \qquad E_k'(\theta) = \frac{d}{d\theta}E_k(\theta). \tag{2.3}$$

However, it was noted that $C_E(\theta)$ depends on the Kraus representation E (Sarovar and Milburn, 2006). For any channel $\mathcal{E}$, the Kraus representation is not unique. Given a unitary matrix $U = [u_{jk}]$ then the set of operators $\{F_j\}$ given by

$$F_j = \sum_k u_{jk}E_k,$$

lead to the same quantum channel (Nielsen and Chuang, 2000, p. 372). That is, for all ρ_0,

$$\sum_k E_k\rho_0 E_k^\dagger = \sum_j F_j\rho_0 F_j^\dagger.$$

To obtain a bound which depends only on the channel and not on the Kraus representation, Sarovar and Milburn chose the bound given by the canonical Kraus operators. *Canonical Kraus operators* $\{\Upsilon_k(\theta)\}$ are defined as Kraus operators satisfying

$$\mathrm{tr}\{\Upsilon_k(\theta)\rho_0\Upsilon_j^\dagger(\theta)\} = \delta_{jk}p_k(\theta), \qquad \forall j,k. \tag{2.4}$$

From (2.2) it follows that

$$F_\theta^M \leq C_\Upsilon(\theta). \tag{2.5}$$

Remark 2.1 *The canonical Kraus operators are unique only up to a choice of phase (see p. 267 of Bengtsson and Życzkowski (2006)). In Chapter 3 it is shown that this leads to ambiguity in the bound $C_\Upsilon(\theta)$. However, this does not affect the results in this chapter.*

Throughout the rest of this chapter the right hand side of (2.5) will be referred to as the *SM bound*. The bound (2.5) is said to be *uniformly attainable* if, for all θ in Θ, there exists a POVM M, possibly depending on θ, such that $F_\theta^M = C_\Upsilon(\theta)$. If this bound is not uniformly attainable, then no bound of the form (2.3) is uniformly attainable (Sarovar and Milburn, 2006). To achieve equality in (2.5) the POVM $\{M_m\}$ must satisfy

$$M_m^{1/2}\Upsilon_k'(\theta)\rho_0^{1/2} = \xi_m(\theta)M_m^{1/2}\Upsilon_k(\theta)\rho_0^{1/2}, \qquad \forall m,k, \tag{2.6}$$

for some real $\xi_m(\theta)$. (This condition is analogous to (1.69).) For channels with quasi-classical output states (see Section 1.10.2), it was shown in Sarovar and Milburn (2006) that this bound is attainable. Channels of this type will be called *quasi-classical channels*. Sarovar and Milburn asked

(i) whether their bound (2.5) is attainable more generally,

(ii) whether explicit expressions for optimal POVMs can be derived from the attainability conditions (2.6).

It is very important for an upper bound on Fisher information to be attainable, otherwise it gives an unrealistic view of how well a parameter can be estimated.

2.2 One-parameter channels

In this Chapter the extended channel will be considered, i.e.

$$\rho_0 \mapsto \mathbb{I}_d \otimes \mathcal{E}(\rho_0), \qquad \rho_0 \in \mathcal{S}(\mathbb{C}^{d^2}). \tag{2.7}$$

The canonical Kruas operators $\{\Upsilon_k(\theta)\}$ are $d^2 \times d^2$ Kraus operators satisfying (2.4).

When the input state is pure, with $\rho_0 = |\psi_0\rangle\langle\psi_0|$, condition (2.4) for the canonical Kraus decomposition is equivalent to the condition

$$\langle v_j(\theta)|v_k(\theta)\rangle = \delta_{jk}p_k(\theta), \qquad \text{where} \quad |v_k(\theta)\rangle = \Upsilon_k(\theta)|\psi_0\rangle. \tag{2.8}$$

The output state is

$$\rho_\theta = \sum_k |v_k(\theta)\rangle\langle v_k(\theta)|.$$

This can be rewritten as

$$\rho_\theta = \sum_k p_k(\theta)|w_k(\theta)\rangle\langle w_k(\theta)|, \qquad |w_k(\theta)\rangle = \frac{1}{\sqrt{p_k(\theta)}}|v_k(\theta)\rangle. \tag{2.9}$$

Thus the canonical decomposition leads to the spectral decomposition of the output state (Sarovar and Milburn, 2006).

Proposition 2.1 *The SM bound, $C_\Upsilon(\theta)$, can be expressed as (omitting θ)*

$$\begin{aligned} C_\Upsilon &= \sum_{k,p_k\neq 0} \frac{p_k'^2}{p_k} + \sum_{j<k,p_j+p_k>0} 4(p_j+p_k)|\langle w_j'|w_k\rangle|^2 \\ &+ 4\sum_{k,p_k\neq 0} p_k|\langle w_k'|w_k\rangle|^2. \end{aligned} \tag{2.10}$$

Proof. For simplicity, it is assumed that for all $p_j(\theta)$ either

(i) $p_j(\theta) > 0$ for all θ,

(ii) $p_j(\theta) = 0$ for all θ.

When $p_j(\theta) = 0$ for all θ, it follows from (2.8) and (2.9) that

$$\begin{aligned} \Upsilon_j|\psi_0\rangle &= \sqrt{p_j}|w_j\rangle = 0, \\ \Upsilon_j'|\psi_0\rangle &= 0, \\ \mathrm{tr}\{\Upsilon_j^{'}\rho_0\Upsilon_j^{\dagger'}\} &= \langle\psi_0|\Upsilon_j^{\dagger'}\Upsilon_j'|\psi_0\rangle = 0. \end{aligned}$$

When $p_j(\theta) > 0$, for all θ,

$$\begin{aligned} \Upsilon_j|\psi_0\rangle &= \sqrt{p_j}|w_j\rangle, \\ \Upsilon_j'|\psi_0\rangle &= \frac{p_j'}{2\sqrt{p_j}}|w_j\rangle + \sqrt{p_j}|w_j'\rangle. \end{aligned}$$

Then

$$\begin{aligned} \langle\psi_0|\Upsilon_j^{\dagger'}\Upsilon_j'|\psi_0\rangle &= \left(\frac{p_j'}{2\sqrt{p_j}}\langle w_j| + \sqrt{p_j}\langle w_j'|\right)\left(\frac{p_j'}{2\sqrt{p_j}}|w_j\rangle + \sqrt{p_j}|w_j'\rangle\right), \\ &= \frac{p_j'^2}{4p_j} + \frac{p_j'}{2}\left(\langle w_j'|w_j\rangle + \langle w_j|w_j'\rangle\right) + p_j\langle w_j'|w_j'\rangle. \end{aligned} \tag{2.11}$$

The right hand side of (2.11) can be simplified, because

$$\langle w_j'|w_j\rangle + \langle w_j|w_j'\rangle = \frac{\partial}{\partial\theta}\mathrm{tr}\{\rho_j\} = 0, \qquad \rho_j = |w_j\rangle\langle w_j|. \tag{2.12}$$

(It follows from (2.12) that $\langle w_j'|w_j\rangle$ is purely imaginary.) Thus

$$C_\Upsilon = 4\sum_{j,p_j\neq 0}\left(\frac{p_j'^2}{4p_j} + p_j\langle w_j'|w_j'\rangle\right).$$

Inserting the identity $\mathbb{I}_d = \sum_{k=1}^d |w_k\rangle\langle w_k|$ into $\langle w_j'|w_j'\rangle$ gives

$$\begin{aligned} C_\Upsilon &= \sum_{j,p_j\neq 0}\frac{p_j'^2}{p_j} + \sum_{j,k,p_j\neq 0} 4p_j\langle w_j'|w_k\rangle\langle w_k|w_j'\rangle, \\ &= \sum_{j,p_j\neq 0}\frac{p_j'^2}{p_j} + \sum_{j,k,p_j\neq 0} 4p_j|\langle w_j'|w_k\rangle|^2. \end{aligned} \tag{2.13}$$

The right hand side of (2.13) will be re-written in (2.18). Since

$$\langle w_j|w_k\rangle = \delta_{jk},$$

it follows that

$$\begin{aligned} \frac{\partial}{\partial\theta}\langle w_j|w_k\rangle &= \langle w_j'|w_k\rangle + \langle w_j|w_k'\rangle = 0, \\ \langle w_j'|w_k\rangle &= -\langle w_j|w_k'\rangle, \qquad (2.14) \\ |\langle w_j'|w_k\rangle|^2 &= \langle w_j'|w_k\rangle\langle w_k|w_j'\rangle \\ &= (-\langle w_j|w_k'\rangle)(-\langle w_k'|w_j\rangle) = |\langle w_k'|w_j\rangle|^2. \qquad (2.15) \end{aligned}$$

Now,

$$\begin{aligned} \sum_{j,k,p_j\neq 0} 4p_j|\langle w_j'|w_k\rangle|^2 &= \sum_{j<k,p_j\neq 0} 4p_j|\langle w_j'|w_k\rangle|^2 + \sum_{k<j,p_j\neq 0} 4p_j|\langle w_j'|w_k\rangle|^2 \\ &+ \sum_{j=k,p_j\neq 0} 4p_j|\langle w_j'|w_k\rangle|^2. \end{aligned} \tag{2.16}$$

Swapping the indices j and k in the second term and using (2.15) simplifies (2.16) further to

$$\sum_{j,k,p_j\neq 0} 4p_j|\langle w_j'|w_k\rangle|^2 = \sum_{j<k,p_j+p_k\neq 0} 4(p_j+p_k)|\langle w_j'|w_k\rangle|^2 + \sum_{j,p_j\neq 0} 4p_j|\langle w_j'|w_j\rangle|^2. \tag{2.17}$$

Thus, from (2.13) and (2.17), the SM bound $C_\Upsilon(\theta)$ can be rewritten as

$$\begin{aligned} C_\Upsilon &= \sum_{j,p_j\neq 0} \frac{p_j'^2}{p_j} + \sum_{j<k,p_j+p_k>0} 4(p_j+p_k)|\langle w_j'|w_k\rangle|^2 \\ &+ 4\sum_{p_k\neq 0} p_k|\langle w_k'|w_k\rangle|^2. \end{aligned} \tag{2.18}$$

Remark 2.2 *It can be seen that $C_\Upsilon(\theta)$ can be described solely in terms of the family of output states. The SM bound was originally derived as an upper bound on the Fisher information for a one-parameter family of quantum channels. Since any parametric family of quantum states can be written in the form*

$$\rho_\theta = \sum_k p_k(\theta)|w_k(\theta)\rangle\langle w_k(\theta)|,$$

$C_\Upsilon(\theta)$ can be extended to an upper bound on the Fisher information for one-parameter families of states.

It can be seen from the form of (2.18) that $C_\Upsilon(\theta)$ is a Riemannian metric on a 1-dimensional manifold (see Section 1.9).

Proposition 2.2 *The SLD quantum information can be written as (omitting θ)*

$$H = \sum_{k,p_k\neq 0} \frac{p_k'^2}{p_k} + \sum_{j<k,p_j+p_k>0} 4\frac{(p_j-p_k)^2}{p_j+p_k}|\langle w_j'|w_k\rangle|^2. \tag{2.19}$$

Proof. The SLD is defined as any self-adjoint solution λ of the matrix equation

$$\frac{d\rho}{d\theta} = \frac{1}{2}\left(\rho\lambda + \lambda\rho\right). \tag{2.20}$$

The SLD quantum information is defined as

$$H = \mathrm{tr}\{\rho\lambda^2\}.$$

Substituting (2.9) into (2.20) gives

$$\begin{aligned} &\sum_{i=1} \{p_i'|w_i\rangle\langle w_i| + p_i(|w_i'\rangle\langle w_i| + |w_i\rangle\langle w_i'|)\} \\ &= \frac{1}{2}\left(\sum_l p_l|w_l\rangle\langle w_l|\lambda + \lambda\sum_m p_m|w_m\rangle\langle w_m|\right). \end{aligned} \tag{2.21}$$

From (2.21) the components of the SLD are calculated. First, the diagonal elements λ_{jj} are considered. Pre-multiplying (2.21) by $\langle w_j|$ and post-multiplying $|w_j\rangle$ gives, on the left hand side,

$$p'_j + p_j(\langle w_j|w'_j\rangle + \langle w'_j|w_j\rangle) = p'_j$$

by (2.12), and on the right hand side

$$p_j\langle w_j|\lambda|w_j\rangle.$$

Hence, provided that $p_j > 0$,

$$\lambda_{jj} = \frac{p'_j}{p_j}.$$

The diagonal elements λ_{jj} are not defined when $p_j = 0$. In this case, a particular solution of λ is chosen for which $\lambda_{jj} = 0$. Next, the off-diagonal components λ_{jk} are considered. Pre-multiplying (2.21) by $\langle w_j|$ and post-multiplying by $|w_k\rangle$ gives, on the left hand side

$$0 + p_k\langle w_j|w'_k\rangle + p_j\langle w'_j|w_k\rangle = (p_j - p_k)\langle w'_j|w_k\rangle,$$

by (2.14), and on the right hand side

$$\frac{1}{2}(p_j + p_k)\langle w_j|\lambda|w_k\rangle.$$

Thus, provided that $p_j + p_k > 0$,

$$\lambda_{jk} = \frac{2(p_j - p_k)\langle w'_j|w_k\rangle}{p_j + p_k}.$$

The entries λ_{jk} are not defined when $p_j + p_k = 0$. Again a particular solution of λ is chosen for which $\lambda_{jk} = 0$, when $p_j + p_k = 0$. This gives the following particular solution of the SLD

$$\tilde{\lambda} = \sum_{k,p_k\neq 0} \frac{p'_k}{p_k}|w_k\rangle\langle w_k| + \sum_{j\neq k, p_j+p_k>0} 2\frac{p_j - p_k}{p_j + p_k}\langle w'_j|w_k\rangle|w_j\rangle\langle w_k|. \tag{2.22}$$

Denote by $\tilde{\lambda}^{2*}$ the part of $\tilde{\lambda}^2$ which makes a non-zero contribution to $\mathrm{tr}\{\rho\tilde{\lambda}^2\}$. Only terms of the form $z_k|w_k\rangle\langle w_k|$, with $z_k \in \mathbb{C}$, in $\tilde{\lambda}^2$ will contribute to $\mathrm{tr}\{\rho\tilde{\lambda}^2\}$. Thus,

$$\begin{aligned}
\tilde{\lambda}^{2*} &= \sum_{k,p_k\neq 0}\left(\frac{p'_k}{p_k}\right)^2|w_k\rangle\langle w_k| + \sum_{j\neq k,p_j+p_k>0} 4\frac{p_j - p_k}{p_j + p_k}\frac{p_k - p_j}{p_k + p_j}\langle w'_j|w_k\rangle\langle w'_k|w_j\rangle|w_j\rangle\langle w_j| \\
&= \sum_{k,p_k\neq 0}\left(\frac{p'_k}{p_k}\right)^2|w_k\rangle\langle w_k| + \sum_{j\neq k,p_j+p_k>0} 4\left(\frac{p_j - p_k}{p_j + p_k}\right)^2|\langle w'_j|w_k\rangle|^2|w_j\rangle\langle w_j|,
\end{aligned}$$

using (2.14). This gives

$$H = \sum_{k,p_k\neq 0} \frac{p_k'^2}{p_k} + \sum_{j\neq k, p_j+p_k>0} 4p_j \left(\frac{p_j-p_k}{p_j+p_k}\right)^2 |\langle w_j'|w_k\rangle|^2. \tag{2.23}$$

The second term on the right hand side of (2.23) can be rewritten as

$$\begin{aligned}\sum_{j\neq k, p_j+p_k>0} 4p_j \left(\frac{p_j-p_k}{p_j+p_k}\right)^2 |\langle w_j'|w_k\rangle|^2 &= \sum_{j<k, p_j+p_k>0} 4p_j \left(\frac{p_j-p_k}{p_j+p_k}\right)^2 |\langle w_j'|w_k\rangle|^2 \\ &+ \sum_{k<j, p_j+p_k>0} 4p_j \left(\frac{p_j-p_k}{p_j+p_k}\right)^2 |\langle w_j'|w_k\rangle|^2.\end{aligned}$$

Swapping the indices, j and k, in the second term on the right hand side of the above equation and using (2.15) gives

$$\sum_{j\neq k, p_j+p_k>0} 4p_j \left(\frac{p_j-p_k}{p_j+p_k}\right)^2 |\langle w_j'|w_k\rangle|^2 = \sum_{j<k, p_j+p_k>0} 4\frac{(p_j-p_k)^2}{p_j+p_k} |\langle w_j'|w_k\rangle|^2. \tag{2.24}$$

The required result (2.19) follows from (2.23) and (2.24).

Theorem 2.1

$$H_\theta \leq C_\Upsilon(\theta), \tag{2.25}$$

with equality if and only if

$$\langle w_j'|w_k\rangle = 0, \quad \forall j,k \quad \text{with} \quad p_j, p_k > 0. \tag{2.26}$$

Proof. The first terms in H_θ, given in (2.19), and $C_\Upsilon(\theta)$, given in (2.10), are identical. Thus

$$C_\Upsilon(\theta) - H_\theta = A_C(\theta) - A_H(\theta) + B_C(\theta),$$

where (omitting θ)

$$\begin{aligned} A_H &= \sum_{j<k, p_j+p_k>0} 4\frac{(p_j-p_k)^2}{p_j+p_k} |\langle w_j'|w_k\rangle|^2, \\ A_C &= \sum_{j<k, p_j+p_k>0} 4(p_j+p_k) |\langle w_j'|w_k\rangle|^2, \\ B_C &= 4\sum_{k,p_k\neq 0} p_k |\langle w_k'|w_k\rangle|^2. \end{aligned}$$

The terms A_C and A_H are symmetric in j and k due to (2.15). Now

$$\begin{aligned} A_C - A_H &= 2 \sum_{j\neq k, p_j+p_k>0} \frac{(p_j+p_k)^2-(p_j-p_k)^2}{p_j+p_k} |\langle w_j'|w_k\rangle|^2, \\ &= 8 \sum_{j\neq k, p_j+p_k>0} \frac{p_j p_k}{p_j+p_k} |\langle w_j'|w_k\rangle|^2. \end{aligned}$$

Changing the range of the summation to $j \neq k$ where $p_j, p_k > 0$, and adding B_C gives

$$C_\Upsilon - H = 8 \sum_{j,k,p_j,p_k>0} \frac{p_j p_k}{p_j+p_k} |\langle w_j'|w_k\rangle|^2. \tag{2.27}$$

Since the right hand side of (2.27) is non-negative, (2.25) follows.

Equality holds in (2.25) if and only if the right hand side of (2.27) is zero, which holds if and only if (2.26) holds.

Lemma 2.1 *For channels, with output states, for which $p_j(\theta) > 0$ for all j and θ, the bound (2.25) is achievable if and only if the channel is quasi-classical.*

Proof. Equality holds in (2.25) if and only if (2.26) is satisfied. When $p_j(\theta) > 0$ for all j and θ, condition (2.26) is satisfied if and only if $|w_j'\rangle$ has zero components along every vector $|w_k\rangle$. This is possible only if $|w_j'\rangle = 0$ and hence the channel is quasi-classical.

Lemma 2.2 *For unitary channels, the bound (2.25) is achievable if and only if*

$$\mathrm{tr}\{U_\theta \rho_0 U_\theta'^\dagger\} = 0. \tag{2.28}$$

Proof. Equality holds in (2.25) if and only if (2.26) is satisfied. For unitary channels there is only one non-zero p_j and $|w_j\rangle = U_\theta|\psi_0\rangle$, where $\rho_0 = |\psi_0\rangle\langle\psi_0|$. Condition (2.26) is satisfied if and only if $\langle w_j'|w_j\rangle = 0$. This is equivalent to (2.28).

Remark 2.3 *Note that, for the most common unitary channels – those of the form $\exp(i\theta H)$, with H a self-adjoint matrix – condition (2.28) is satisfied.*

Example 2.1 *There exist channels which are neither quasi-classical or unitary for which equality holds in (2.25). The channel with an arbitray pure input state and output states*

$$\rho_\theta = \theta^2 |w_1(\theta)\rangle\langle w_1(\theta)| + (1-\theta^2)|w_2(\theta)\rangle\langle w_2(\theta)|, \quad 0 < \theta < 1,$$

where

$$|w_1(\theta)\rangle = (\theta, \sqrt{1-\theta^2}, 0)^T, \quad |w_2(\theta)\rangle = (0,0,1)^T,$$

satisfies (2.26), and so equality holds in (2.25).

Theorem 2.2

$$F_\theta^M \leq C_\Upsilon(\theta), \tag{2.29}$$

with equality if and only if

$$\langle w_j'|w_k\rangle = 0, \quad \forall j,k \quad \text{with} \quad p_j, p_k > 0. \tag{2.30}$$

Proof. Inequality (2.29) follows from (1.64) and (2.25). Equality holds in (2.29) if and only if there is equality in both (1.64) and (2.25). For one-parameter families of states it is always possible to find a POVM M_θ, depending on θ, which achieves equality in (1.64) (Braunstein and Caves, 1994). However, equality holds in (2.25) if and only if (2.30) is satisfied.

Theorem 2.3

$$H_\theta \leq C_E(\theta), \tag{2.31}$$

with equality if and only if the set of output states satisfies (2.26), and a fixed unitary matrix $U = [u_{jk}]$ exists such that the Kraus operators E_j are related to the canonical Kraus operators Υ_k by

$$E_j(\theta) = \sum_k u_{jk} \Upsilon_k(\theta).$$

Proof. Inequality (2.31) will be proved by considering two cases:

(i) When equality is attainable in (2.2), it is attainable also in (2.5) (Sarovar and Milburn, 2006). In this case, $C_\Upsilon(\theta) \leq C_E(\theta)$ for all other sets of Kraus operators $E = \{E_j\}$ (Sarovar and Milburn, 2006). Inequality (2.25) gives $H_\theta \leq C_E(\theta)$.

(ii) When (2.2) is not attainable, $F_\theta^M < C_E(\theta)$ for all M. For one-parameter families of states there always exists a measurement M_θ such that $F_\theta^{M_\theta} = H_\theta$. Thus $H_\theta = F_\theta^{M_\theta} < C_E(\theta)$.

Equality holds in (2.31) only if the bound given by the canonical Kraus operators C_Υ is attainable. The bound C_Υ is attainable if and only if the set of ouput states satisfies (2.26). It has been shown (Nielsen and Chuang,

2000, p. 372) that if two sets of Kraus operators $\{E_j\}$ and $\{F_k\}$ lead to the same quantum channel then they must be related by

$$E_j = \sum_k u_{jk} F_k, \tag{2.32}$$

where $U = [u_{jk}]$ is a unitary matrix. When C_Υ is attainable (Sarovar and Milburn, 2006),

$$C_E = C_\Upsilon + 4 \sum_{jk} p_j |u'_{jk}|^2.$$

Thus for equality in (2.31) it is further required that $\sum_{jk} p_j |u'_{jk}|^2 = 0$. This is satisfied if and only if a unitary matrix $U = [u_{jk}]$ exists satisfying (2.32) that does not depend on θ.

Remark 2.4 *Condition (2.6) cannot be used generally to test for optimality of POVMs. Condition (2.6) is a necessary and sufficient condition for equality between the Fisher information and the SM bound. Since it is not generally possible to achieve equality between the Fisher information and the SM bound, condition (2.6) cannot be achieved for general models. Thus it cannot be used generally to test for POVMs giving maximal Fisher information.*

2.3 Multi-parameter channels

2.3.1 The multi-parameter SM bound

The SM bound for a multi-parameter family of channels will be defined as the matrix $C_\Upsilon(\theta)$ with entries

$$C_\Upsilon(\theta)_{jk} = 4 \sum_l \Re\mathrm{tr}\left\{ \Upsilon_l(\theta)^{(j)} \rho_0 \Upsilon_l(\theta)^{(k)\dagger} \right\}, \qquad \Upsilon_l(\theta)^{(k)} = \frac{\partial}{\partial \theta^k} \Upsilon_l(\theta). \tag{2.33}$$

Proposition 2.3 *For θ and v in $\mathbb{R}^p$, and $t \to 0$,*

$$\frac{d}{dt} \Upsilon_k(\theta + tv) = \sum_l \Upsilon_k(\theta)^{(l)} v^l + O(t), \tag{2.34}$$

$$\tilde{\lambda}_t = \sum_l \tilde{\lambda}_\theta^{(l)} v^l + O(t), \tag{2.35}$$

where $\tilde{\lambda}_t$ is defined by (2.22) with respect to the parameter t, $\tilde{\lambda}_\theta^{(l)}$ is defined by (2.22) with respect to the parameter θ^l and v^l is the lth component of the vector v.

Proof.
Put $\phi(t) = \theta + tv$, with components $\phi^l(t) = \theta^l + tv^l$. Using the chain rule to differentiate $\Upsilon_k(\phi(t))$ gives

$$\frac{d}{dt}\Upsilon_k(\phi(t)) = \sum_l \frac{\partial \Upsilon_k(\phi)}{\partial \phi^l}\frac{\partial \phi^l}{\partial t}. \tag{2.36}$$

Now,

$$\begin{aligned}\frac{\partial \Upsilon_k(\phi)}{\partial \phi^l} &= \left.\frac{\partial \Upsilon_k(\phi)}{\partial \phi^l}\right|_{t=0} + O(t) = \frac{\partial \Upsilon_k(\theta)}{\partial \theta^l} + O(t),\\ \frac{\partial \phi^l}{\partial t} &= v^l.\end{aligned}$$

Substituting these back into (2.36) gives (2.34). Similarly, for $t \to 0$,

$$p_k(\theta + tv) = p_k(\theta) + O(t), \tag{2.37}$$

$$\frac{dp_k(\theta + tv)}{dt} = \sum_l p_k^{(l)} v^l + O(t), \quad p_k^{(l)} = \frac{\partial p_k(\theta)}{\partial \theta^l} \tag{2.38}$$

$$\frac{d|w_k(\theta + tv)\rangle}{dt} = \sum_l |w_k^{(l)}\rangle v^l + O(t), \quad |w_k^{(l)}\rangle = \frac{\partial |w_k(\theta)\rangle}{\partial \theta^l}. \tag{2.39}$$

Substituting (2.37) – (2.39) into (2.22) gives

$$\begin{aligned}\tilde{\lambda}_t &= \sum_{k, p_k \neq 0} \frac{\sum_l p_k^{(l)} v^l + O(t)}{p_k + O(t)} |w_k\rangle\langle w_k|\\ &+ \sum_{j \neq k, p_j + p_k > 0} 2\frac{p_j - p_k + O(t)}{p_j + p_k + O(t)} \left(\sum_l v^l \left\langle w_j^{(l)} | w_k \right\rangle + O(t) \right) |w_j\rangle\langle w_k|\\ &= \sum_l v^l \left(\sum_{k, p_k \neq 0} \frac{p_k^{(l)}}{p_k} |w_k\rangle\langle w_k| + \sum_{j \neq k, p_j + p_k > 0} 2\frac{p_j - p_k}{p_j + p_k} \left\langle w_j^{(l)} | w_k \right\rangle |w_j\rangle\langle w_k| \right) + O(t).\end{aligned}$$

Thus $\tilde{\lambda}_t$ has the form (2.35).

Theorem 2.4 *For multi-parameter channels,*

$$H_\theta \leq C_\Upsilon(\theta), \tag{2.40}$$

with equality if and only if

$$\left\langle w_j^{(l)} \middle| w_k \right\rangle = 0, \quad \forall j, k, l \quad \text{with} \quad p_j, p_k > 0, \tag{2.41}$$

$$\text{where} \quad \left| w_j^{(l)} \right\rangle = \frac{\partial}{\partial \theta^l} \middle| w_j \right\rangle.$$

Proof. Equation (2.40) is equivalent to

$$v^T H_\theta v \leq v^T C_\Upsilon(\theta) v, \quad \text{for all } v \in \mathbb{R}^p. \tag{2.42}$$

For given θ and v in $\mathbb{R}^p$, consider the set of one-parameter channels

$$\rho_0 \mapsto \sum_k \Upsilon_k(\theta + tv)\rho_0 \Upsilon_k^\dagger(\theta + tv), \quad t \in \mathbb{R}. \tag{2.43}$$

From Theorem 2.1 it is known that $H_t \leq C_\Upsilon(t)$, i.e.

$$\mathrm{tr}\left\{\tilde{\lambda}_t \rho_{\theta+tv} \tilde{\lambda}_t\right\} \leq 4 \sum_{l=1}^{d} \mathrm{tr}\left\{\frac{d}{dt}\Upsilon_l(\theta+tv)\rho_0 \frac{d}{dt}\Upsilon_l(\theta+tv)^\dagger\right\}.$$

Using (2.34) and (2.35) and evaluating at $t = 0$ gives

$$\sum_{m,n} v^m v^n \mathrm{tr}\left\{\tilde{\lambda}_\theta^{(m)} \rho_\theta \tilde{\lambda}_\theta^{(n)}\right\} \leq 4 \sum_{m,n,l} v^m v^n \mathrm{tr}\left\{\Upsilon_l(\theta)^{(m)} \rho_0 \Upsilon_l(\theta)^{(n)\dagger}\right\}.$$

This is equivalent to (2.42). Since this holds for all $v \in \mathbb{R}^p$, (2.40) holds.

Equality in (2.40) is equivalent to

$$v^T H_\theta v = v^T C_\Upsilon(\theta) v, \tag{2.44}$$

for all $v \in \mathbb{R}^p$. It follows that (2.44) holds for all $v \in \mathbb{R}^p$ if and only if, for one-parameter channels of the form (2.43) for given θ and $v \in \mathbb{R}^p$, $H_t|_{t=0} = C_\Upsilon(t)|_{t=0}$. From Theorem 2.1, this holds if and only if the channel (2.43) satisfies (2.26) at the point $t = 0$. This condition is equal to

$$\left(\frac{d}{dt}\langle w_j|\right)|w_k\rangle\Big|_{t=0} = 0, \quad \forall j,k \quad \text{with} \quad p_j, p_k > 0.$$

Using (2.39), this condition can be rewritten as

$$\sum_{l=1}^{m} v^l \left\langle w_j^{(l)} \middle| w_k \right\rangle = 0 \quad \forall j,k \quad \text{with} \quad p_j, p_k > 0. \tag{2.45}$$

Condition (2.45) holds for all v if and only if (2.41) is satisfied.

Lemma 2.3 *For channels, with output states for which $p_j(\theta) > 0$ for all j and θ, equality holds in (2.40) if and only if the channel is quasi-classical.*

Proof. This follows from (2.41) and the same analysis as in Lemma 2.1.

Lemma 2.4 *For unitary channels, equality holds in (2.40) if and only if*

$$\mathrm{tr}\left\{U_\theta \rho_0 \frac{\partial U_\theta}{\partial \theta^l}^\dagger\right\} = 0, \quad \forall l.$$

Proof. This follows from (2.41) and the same analysis as in Lemma 2.2.

Example 2.2 *There exist channels which are neither quasi-classical or unitary for which equality holds in (2.40). The channel with an arbitrary pure input state and output states*

$$\rho_\theta = f(\theta)^2 |w_1(\theta)\rangle\langle w_1(\theta)| + (1 - f(\theta)^2)|w_2(\theta)\rangle\langle w_2(\theta)|,$$

where $f(\theta)$ and $g(\theta)$ are real functions of θ with $0 \leq f(\theta), g(\theta) \leq 1$ and

$$|w_1(\theta)\rangle = (g(\theta), \sqrt{1 - g(\theta)^2}, 0)^T, \quad |w_2(\theta)\rangle = (0, 0, 1)^T,$$

satisfies (2.41) and hence achieves equality in (2.40).

Theorem 2.5 *For multi-parameter channels,*

$$F_\theta^M \leq C_\Upsilon(\theta), \tag{2.46}$$

with equality if and only if (2.41) holds and there exists a POVM satisfying

$$M_m^{1/2} \lambda^j \rho^{1/2} = \xi_m^j M_m^{1/2} \rho^{1/2}, \quad \xi_m^j \in \mathbb{R}, \quad \forall j, m. \tag{2.47}$$

Proof. This follows from Theorems 1.1, 1.2 and 2.4.

Theorem 2.6 *For multi-parameter channels,*

$$H_\theta \leq C_E(\theta), \tag{2.48}$$

with equality if and only if the set of output states satisfies (2.41) and a fixed unitary matrix $U = [u_{jk}]$ exists such that the Kraus operators E_j are related to the canonical Kraus operators Υ_k by

$$E_j(\theta) = \sum_k u_{jk} \Upsilon_k(\theta). \tag{2.49}$$

Proof. Inequality (2.48) follows from (2.31) and the same analysis used in the proof of Theorem 2.4 with Υ_k replaced by E_k.

Equality holds in (2.48) if and only if, for the set of channels (2.43), $H_t|_{t=0} = C_E(t)|_{t=0}$ for all v. From Theorem 2.3 this is satisfied if and only if the output states of the channel satisfy (2.26) at $t = 0$ and the Kraus operators E_j are related to the canonical Kraus operators Υ_k by

$$E_j(\theta + tv)|_{t=0} = \sum_k u_{jk}(\theta + tv)\Upsilon_k(\theta + tv)\ |_{t=0}\,,$$

where

$$\sum_{jk} p_j \left| \frac{du_{jk}}{dt}\right|_{t=0}\right|^2 = 0. \tag{2.50}$$

From the proof of Theorem 2.4 it can be seen that for channels of the form (2.43), satisfying (2.26) at $t = 0$ is equivalent to satisfying (2.41). Condition (2.50) can be rewritten as

$$\sum_{jk} p_j \left| \sum_l \frac{\partial u_{jk}}{\partial \theta^l} v^l \right|^2 = 0.$$

This is satisfied for all v if and only if a unitary matrix $U = [u_{jk}]$ exists satisfying (2.49) that does not depend on θ.

Theorem 2.7 *For multi-parameter channels,*

$$F_\theta^M \leq C_E(\theta), \tag{2.51}$$

with equality if and only if the set of output states satisfies (2.41), a fixed unitary matrix $U = [u_{jk}]$ exists such that the Kraus operators E_j are related to the canonical Kraus operators Υ_k by (2.49) and there exists a POVM satisfying (2.47).

Proof. This follows from Theorems 1.1, 1.2 and 2.6.

Chapter 3

The bound of Sarovar and Milburn as a metric on the space of quantum states

3.1 Introduction

Various statistical notions can be expressed in differential-geometric terms (Amari and Nagaoka, 2000). This area is sometimes known as 'information geometry'. Of special importance is Fisher information, which is the unique monotone metric on the space of probability measures (Morozova and Čencov, 1990). However, there is no unique monotone metric on the space of quantum states (Petz and Sudár, 1996). (Definitions of monotonicity and invariance were given below (1.50).)

The following theorem of Morozova and Čencov (1990) is of great interest.

Theorem 3.1 *A Riemannian metric is invariant if and only if at every density matrix*

$$\rho = \sum_j p_j |j\rangle\langle j|,$$

the squared length of any tangent vector A is of the form

$$C \sum_i \frac{1}{p_i} |A_{ii}|^2 + 2 \sum_{j<k} c(p_j, p_k) |A_{jk}|^2, \quad A_{jk} = \langle j|A|k\rangle, \tag{3.1}$$

where C is a constant, $c(\alpha x, \alpha y) = \alpha^{-1} c(x, y)$ and $c(x, y) = c(y, x)$.

This result was augmented by the following theorem of Petz and Sudár (1996).

Theorem 3.2 *A Riemannian metric on the space of quantum states is monotone if and only if at every density matrix*

$$\rho = \sum_j p_j |j\rangle\langle j|,$$

the squared length of any tangent vector A is of the form (3.1) and the function $f(t) = 1/c(t,1)$ is operator monotone. (A function $f(t)$ is operator monotone if for self-adjoint $n \times n$ matrices A and B, with $A \leq B$, $f(A) \leq f(B)$, (Bengtsson and Życzkowski, 2006, Section 12.1).)

For parametric families of states, put $A = d\rho/d\theta$. In this case (3.1) becomes

$$C \sum_i \frac{1}{p_i} \left| \left(\frac{d\rho}{d\theta} \right)_{ii} \right|^2 + 2 \sum_{j<k} c(p_j, p_k) \left| \left(\frac{d\rho}{d\theta} \right)_{jk} \right|^2 . \tag{3.2}$$

For the SLD, KMB and RLD quantum informations (Petz and Sudár, 1996), $C = 1$ and

$$\begin{aligned} c_{SLD}(x,y) &= \frac{2}{x+y} \\ c_{KMB}(x,y) &= \frac{\ln x - \ln y}{x-y} \\ c_{RLD}(x,y) &= \frac{1}{2}\left(\frac{1}{x} + \frac{1}{y}\right). \end{aligned}$$

For a more thorough background to the theory of metrics on the space of quantum states see (Bengtsson and Życzkowski, 2006, Chapter 14).

The *Symmetric Logarithmic Derivative (SLD)*, *Kubo-Mori Bogoliubov (KMB)* and *Right Logarithmic Derivative (RLD)* metrics (see Section 1.9) are the most frequently encountered monotone metrics in recent literature. The SLD quantum information is the minimum monotone metric on the space of quantum states (Petz and Sudár, 1996). It has been used widely in the estimation of states (Helstrom, 1967, 1976, Holevo, 1982, Hayashi, 2005) and quantum channels (Fujiwara, 2001, 2002, 2004, Fujiwara and Imai, 2003, Ballester, 2004a,b). For one-parameter families of states, the SLD quantum information is equal to the maximum attainable Fisher information (Braunstein and Caves, 1994). The SLD quantum information is related to the bures distance,

$$b^2(\rho, \sigma) = 1 - \mathrm{tr}\{\sqrt{\rho^{1/2}\sigma\rho^{1/2}}\}, \tag{3.3}$$

in the following way (Hayashi, 2006b, (6.23))

$$H_\theta^S = 8 \lim_{\epsilon \to 0} \frac{b^2(\rho_\theta, \rho_{\theta+\epsilon})}{\epsilon^2}. \tag{3.4}$$

The bures distance is a quantum analogue of the Hellinger distance

$$d_2^2(p\|q) = 1 - \sum_{i=1}^{k} \sqrt{p_i}\sqrt{q_i}, \tag{3.5}$$

where $p = (p_1, \ldots, p_k)$ and $q = (q_1, \ldots, q_k)$. The result (3.4) is interesting since, given a probability distribution $p_\theta = \{p_i(\theta)\}$, the 'classical' Fisher information is related to the Hellinger distance by

$$F_\theta = 8 \lim_{\epsilon \to 0} \frac{d_2^2(p_\theta \| p_{\theta+\epsilon})}{\epsilon^2}. \tag{3.6}$$

The KMB quantum information is equal to the limit of the quantum relative entropy $D(\rho\|\sigma) = \mathrm{tr}(\rho(\ln\rho - \ln\sigma))$ (Hayashi, 2002). That is,

$$H_\theta^K = \lim_{\epsilon \to 0} \frac{2D(\rho_\theta \| \rho_{\theta+\epsilon})}{\epsilon^2}. \tag{3.7}$$

This is analogous to the fact that the 'classical' Fisher information is the limit of the 'classical' relative entropy $D(p\|q) = \sum_{i=1}^{k} p_i \ln(p_i/q_i)$, where $p = (p_1, \ldots, p_k)$, $q = (q_1, \ldots, q_k)$. That is, given a probability distribution $p_\theta = \{p_i(\theta)\}$,

$$F_\theta = \lim_{\epsilon \to 0} \frac{2D(p_\theta \| p_{\theta+\epsilon})}{\epsilon^2}. \tag{3.8}$$

The RLD quantum information is the maximal monotone metric on the space of quantum states (Petz and Sudár, 1996). It has also been used in estimation theory (Fujiwara, 1994).

In Chapter 2 it was shown that Sarovar and Milburn's bound $C_\Upsilon(\theta)$ for one-parameter channels could be generalized to a Riemannian metric on Θ. In this chapter $C_\Upsilon(\theta)$ will be referred to as the *SM quantum information*. It seems natural to look at the properties of $C_\Upsilon(\theta)$. Is it is well-defined? Is it useful?

In this chapter it is shown that the SM quantum information is not a well-defined metric, since different choices of phase of the eigenvectors lead to different metrics. A new metric C_L is defined from C_Υ. Properties of C_L are investigated and it is seen that it is invariant but not monotone.

3.2 Analysis of the SM quantum information

The SM quantum information for the family of states

$$\rho_\theta = \sum_{k=1}^{d} p_k(\theta) |w_k(\theta)\rangle\langle w_k(\theta)| \tag{3.9}$$

was shown in Proposition 2.1 to be equal to

$$\begin{aligned} C_\Upsilon &= \sum_i \frac{1}{p_i}\left(\frac{dp_i}{d\theta}\right)^2 + 4\sum_{j<k}(p_j+p_k)|\langle w_j'|w_k\rangle|^2 \\ &+ 4\sum_i p_i|\langle w_i'|w_i\rangle|^2. \end{aligned} \tag{3.10}$$

This can be rewritten as

$$\begin{aligned} C_\Upsilon &= \sum_i \frac{1}{p_i}\left(\frac{dp_i}{d\theta}\right)^2 + 4\sum_{j<k}\frac{p_j+p_k}{(p_j-p_k)^2}\left|\left\langle w_j\left|\frac{d\rho}{d\theta}\right|w_k\right\rangle\right|^2 \\ &+ 4\sum_i p_i(\theta)|\langle w_i'|w_i\rangle|^2. \end{aligned} \tag{3.11}$$

It can be seen that $C_\Upsilon(\theta)$ is not of the form (3.2), and hence is neither invariant nor monotone. The SM quantum information $C_\Upsilon(\theta)$ for a family of states is defined in terms of its eigenvectors and eigenvalues by (3.10). The eigenvectors of a state are unique up to a change of phase. It turns out that different choices of phase for the eigenvectors lead to different metrics.

Example 3.1 *Consider the set of 2-dimensional states*

$$\rho_{r,\theta,\phi} = \frac{1}{2}\begin{pmatrix} 1+r\cos\theta & r\sin\theta e^{-i\phi} \\ r\sin\theta e^{i\phi} & 1-r\cos\theta \end{pmatrix}, \tag{3.12}$$

with $0 \le r \le 1$, $0 \le \theta \le \pi$ and $0 \le \phi \le 2\pi$. Any qubit, mixed or pure, can be written in the form (3.12) with specific values of r, θ and ϕ. Each state (3.12) has spectral decomposition

$$\begin{aligned} \rho_{r,\theta,\phi} &= \frac{1+r}{2}\left|w_1(\theta,\phi)\right\rangle\left\langle w_1(\theta,\phi)\right| + \frac{1-r}{2}\left|w_2(\theta,\phi)\right\rangle\left\langle w_2(\theta,\phi)\right|, \\ |w_1(\theta,\phi)\rangle &= (\cos(\theta/2)e^{-i\phi/2}, \sin(\theta/2)e^{i\phi/2})^T, \\ |w_2(\theta,\phi)\rangle &= (\sin(\theta/2)e^{-i\phi/2}, -\cos(\theta/2)e^{i\phi/2})^T. \end{aligned}$$

The SM quantum information for the family of states ρ_θ calculated from the above eigenvalues and eigenvectors is

$$C_\Upsilon(r,\theta,\phi) = \begin{pmatrix} \dfrac{1}{1-r^2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Changing the eigenvectors by the phase shift $e^{-i\phi/2}$, i.e. $|w_k(\theta,\phi)\rangle \mapsto e^{-i\phi/2}|w_k(\theta,\phi)\rangle$, leaves the density matrix unchanged but the SM quantum information calculated from the eigenvalues and shifted eigenvectors becomes

$$C_\Upsilon(r,\theta,\phi) = \begin{pmatrix} \frac{1}{1-r^2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2+2r\cos\theta \end{pmatrix}.$$

Hence the SM quantum information is not a well-defined metric.

3.3 A new metric

The C_L *quantum information* for the family of states (3.9) will be defined as

$$C_L = C_\Upsilon - 4\sum_i p_i |\langle w_i'|w_i\rangle|^2. \tag{3.13}$$

Thus

$$C_L = \sum_i \frac{1}{p_i}\left(\frac{dp_i}{d\theta}\right)^2 + 4\sum_{j<k}(p_j+p_k)|\langle w_j'|w_k\rangle|^2 \tag{3.14}$$

$$= \sum_i \frac{1}{p_i}\left(\frac{dp_i}{d\theta}\right)^2 + 4\sum_{j<k}\frac{p_j+p_k}{(p_j-p_k)^2}\left|\left\langle w_j\left|\frac{d\rho}{d\theta}\right|w_k\right\rangle\right|^2. \tag{3.15}$$

Remark 3.1 *Unlike the RLD and KMB quantum informations, the C_L quantum information can be defined for families of pure states. For pure states, $C_L(\rho_\theta) = H(\rho_\theta)$ (see (3.27)).*

The C_L quantum information is of the form (3.2) with $C = 1$ and

$$c_L(p_j,p_k) = 2\frac{p_j+p_k}{(p_j-p_k)^2}. \tag{3.16}$$

This function is symmetric and $c_L(\alpha x,\alpha y) = \alpha^{-1}c_L(x,y)$. Hence, C_L is invariant. Thus it does not suffer the same defect as C_Υ. The C_L quantum information provides each parameterized family $\{\rho_\theta : \theta \in \Theta\}$ with a unique Riemannian metric on Θ.

For a metric to be monotone, it must be of the form (3.1) and the function $f(t)$ associated with the metric must be monotone and satisfy $f(t) = tf(t^{-1})$.

The functions associated with the SLD, KMB and RLD quantum informations are

$$\begin{aligned} f_{SLD}(t) &= \frac{1+t}{2} \\ f_{KMB}(t) &= \frac{t-1}{\log t} \\ f_{RLD}(t) &= \frac{2t}{1+t}. \end{aligned}$$

Calculation shows that the function associated with C_L is

$$f_{C_L}(t) = \frac{(t-1)^2}{2(1+t)}. \tag{3.17}$$

If f is a monotone function then $f(0) \leq f(t_1) \leq f(t_2)$ whenever $0 \leq t_1 \leq t_2$. The function $f_{C_L}(t)$ satisfies $f_{C_L}(t) = tf_{c_L}(t^{-1})$ but is not monotone, as $f_{C_L}(0) > f_{C_L}(1)$. Hence, C_L is an invariant but not monotone Riemannian metric.

Example 3.2 *The depolarizing channel,(1.33), acts on* 3*-dimensional states in the following way*

$$\rho_0 \mapsto (1-\epsilon)\rho_0 + \frac{\epsilon}{3}\mathbb{I}_3, \qquad 0 \leq \epsilon \leq 1. \tag{3.18}$$

Consider the one-parameter set of 3*-dimensional mixed states*

$$\begin{aligned} \rho_\theta &= (1-2\delta)|w_1\rangle\langle w_1| + \delta|w_2\rangle\langle w_2| + \delta|w_3\rangle\langle w_3|, \\ |w_1\rangle &= (1,0,0)^T, \\ |w_2\rangle &= (0,\cos\theta,\sin\theta)^T, \\ |w_3\rangle &= (0,-\sin\theta,\cos\theta)^T, \end{aligned}$$

where δ is fixed. The $C_L(\theta)$ quantum information of this family of states is 8δ. Under the action of the depolarizing channel the set of output states is

$$\begin{aligned} \mathcal{E}(\rho_\theta) &= \left((1-\epsilon)(1-2\delta) + \frac{\epsilon}{3}\right)|w_1\rangle\langle w_1| \\ &+ \left((1-\epsilon)\delta + \frac{\epsilon}{3}\right)|w_2\rangle\langle w_2| + \left((1-\epsilon)\delta + \frac{\epsilon}{3}\right)|w_3\rangle\langle w_3| \end{aligned}$$

with $|w_i\rangle$ unchanged. The C_L quantum information for the family of states $\mathcal{E}(\rho_\theta)$ is $8\delta + 8\epsilon(1/3-\delta)$. Now

$$C_L(\mathcal{E}(\rho_\theta)) - C_L(\rho_\theta) = 8\epsilon(1/3-\delta). \tag{3.19}$$

For $\epsilon > 0$ and $\delta < 1/3$, C_L has increased under the action of a TP-CP map, thus demonstrating the non-monotonicity of C_L.

3.4 Ordering of C_L, C_Υ and H

Theorem 3.3 *Given a parameterised quantum model* $\{\rho_\theta = \sum_{k=1}^{d} p_k(\theta)|w_k(\theta)\rangle\langle w_k(\theta)| : \theta \in \mathbb{R}^p\}$,

$$H_\theta \leq C_L(\theta) \leq C_\Upsilon(\theta), \tag{3.20}$$

where the multi-parameter versions of H_θ, $C_L(\theta)$ *and* $C_\Upsilon(\theta)$ *are defined in (3.44), (3.35) and (3.28) respectively. Equality holds in* $H_\theta \leq C_L(\theta)$ *for families of states* ρ_θ *if and only if*

$$\left\langle w_j^{(m)} \middle| w_k \right\rangle = 0, \quad \forall m, \quad j \neq k, p_j, p_k > 0. \tag{3.21}$$

Equality holds in $C_L(\theta) \leq C_\Upsilon(\theta)$ *for families of states* ρ_θ *if and only if*

$$\left\langle w_i^{(m)} \middle| w_i \right\rangle = 0, \quad \forall m, i, \quad p_i > 0. \tag{3.22}$$

A proof of Theorem 3.3 will be given first for the one-parameter case and then for the general case.

3.4.1 One-parameter case

Lemma 3.1

$$C_L(\theta) \leq C_\Upsilon(\theta), \tag{3.23}$$

with equality if and only if

$$\langle w_i'|w_i\rangle = 0, \quad \forall i, \quad p_i > 0. \tag{3.24}$$

Proof. This follows from the definition of C_L, (3.13), and the fact that $|\langle w_i'|w_i\rangle|$ is non-negative.

Lemma 3.2

$$H_\theta \leq C_L(\theta), \tag{3.25}$$

with equality if and only if

$$\langle w_j'|w_k\rangle = 0, \quad \forall j \neq k, p_j, p_k > 0. \tag{3.26}$$

Proof. Proposition 2.2 showed that

$$H = \sum_k \frac{1}{p_k}\left(\frac{dp_k}{d\theta}\right)^2 + \sum_{j<k} 4\frac{(p_j - p_k)^2}{p_j + p_k}|\langle w_j'|w_k\rangle|^2,$$

and hence

$$\begin{aligned} C_L - H &= 16\sum_{j<k}\frac{p_j p_k}{p_j + p_k}|\langle w_j'|w_k\rangle|^2 \\ &= 8\sum_{j\neq k}\frac{p_j p_k}{p_j + p_k}|\langle w_j'|w_k\rangle|^2, \end{aligned} \tag{3.27}$$

since $|\langle w_j'|w_k\rangle|^2$ is symmetric with respect to j and k (2.15). The right hand side of (3.27) is non-negative, and equal to zero if and only if (3.26) holds.

3.4.2 The multi-parameter case

Proposition 3.1 *In the multi-parameter case the SM quantum information is the matrix with entries*

$$\begin{aligned} (C_\Upsilon)_{kl} &= \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial\theta^k}\right)\left(\frac{\partial p_i}{\partial\theta^l}\right) + 4\Re\sum_{i<j}(p_i + p_j)\left\langle w_i^{(k)}\middle|w_j\right\rangle\left\langle w_j\middle|w_i^{(l)}\right\rangle \\ &+ 4\sum_i p_i\left\langle w_i^{(k)}\middle|w_i\right\rangle\left\langle w_i\middle|w_i^{(l)}\right\rangle. \end{aligned} \tag{3.28}$$

Proof. The multi-parameter version of C_Υ was defined, (2.33), as the matrix with entries

$$(C_\Upsilon)_{kl} = 4\sum_i \Re\mathrm{tr}\left\{\Upsilon_i^{(k)}\rho_0\Upsilon_i^{(l)\dagger}\right\}, \qquad \Upsilon_i^{(k)} = \frac{\partial}{\partial\theta^k}\Upsilon_i.$$

Using (2.8) and (2.9)

$$\begin{aligned} \Re\mathrm{tr}\left\{\Upsilon_i^{(k)}\rho_0\Upsilon_i^{(l)\dagger}\right\} &= \Re\Bigg[\frac{1}{4p_i}\frac{\partial p_i}{\partial\theta^k}\frac{\partial p_i}{\partial\theta^l} + p_i\left\langle w_i^{(l)}\middle|w_i^{(k)}\right\rangle \\ &+ \frac{1}{2}\left(\frac{\partial p_i}{\partial\theta^l}\left\langle w_i\middle|w_i^{(k)}\right\rangle + \frac{\partial p_i}{\partial\theta^k}\left\langle w_i^{(l)}\middle|w_i\right\rangle\right)\Bigg]. \end{aligned} \tag{3.29}$$

The contributions of the final two terms on the right hand side of (3.29) are zero since they are purely imaginary (see below (2.12)). Thus,

$$(C_\Upsilon)_{kl} = \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial\theta^k}\right)\left(\frac{\partial p_i}{\partial\theta^l}\right) + 4\Re\sum_i p_i\left\langle w_i^{(l)}\middle|w_i^{(k)}\right\rangle. \tag{3.30}$$

Inserting the identity $\mathbb{I}_d = \sum_j |w_j\rangle\langle w_j|$ into the second term on the right hand side of (3.29) gives

$$\begin{aligned}\sum_i p_i \left\langle w_i^{(l)} \middle| w_i^{(k)} \right\rangle &= \sum_{i\neq j} p_i \left\langle w_i^{(l)} \middle| w_j \right\rangle \left\langle w_j \middle| w_i^{(k)} \right\rangle \\ &+ \sum_i p_i \left\langle w_i^{(l)} \middle| w_i \right\rangle \left\langle w_i \middle| w_i^{(k)} \right\rangle. \end{aligned} \tag{3.31}$$

The first term on the right hand side of (3.31) can be written as

$$\begin{aligned}\sum_{i\neq j} p_i \left\langle w_i^{(l)} \middle| w_j \right\rangle \left\langle w_j \middle| w_i^{(k)} \right\rangle &= \sum_{i<j} p_i \left\langle w_i^{(l)} \middle| w_j \right\rangle \left\langle w_j \middle| w_i^{(k)} \right\rangle \\ &+ \sum_{i>j} p_i \left\langle w_i^{(l)} \middle| w_j \right\rangle \left\langle w_j \middle| w_i^{(k)} \right\rangle. \end{aligned} \tag{3.32}$$

Swapping the indices, i and j, on the second term on the right hand side of (3.32) gives

$$\begin{aligned}\Re \sum_{i>j} p_i \left\langle w_i^{(l)} \middle| w_j \right\rangle \left\langle w_j \middle| w_i^{(k)} \right\rangle &= \Re \sum_{i<j} p_j \left\langle w_j^{(l)} \middle| w_i \right\rangle \left\langle w_i \middle| w_j^{(k)} \right\rangle \\ &= \Re \sum_{i<j} p_j \left\langle w_j \middle| w_i^{(l)} \right\rangle \left\langle w_i^{(k)} \middle| w_j \right\rangle, \end{aligned} \tag{3.33}$$

using (2.14). From (3.31), (3.32) and (3.33) it follows that

$$\begin{aligned}\Re \sum_i p_i \left\langle w_i^{(k)} \middle| w_i^{(l)} \right\rangle &= \Re \sum_{i<j} (p_i + p_j) \left\langle w_i^{(k)} \middle| w_j \right\rangle \left\langle w_j \middle| w_i^{(l)} \right\rangle \\ &+ \Re \sum_i p_i \left\langle w_i^{(k)} \middle| w_i \right\rangle \left\langle w_i \middle| w_i^{(l)} \right\rangle. \end{aligned} \tag{3.34}$$

The required result follows from (3.30) and (3.34).

The multivariate version of C_L will be defined as the matrix with entries

$$\begin{aligned}(C_L)_{kl} &= (C_\Upsilon)_{kl} - 4\Re \sum_i p_i \left\langle w_i^{(k)} \middle| w_i \right\rangle \left\langle w_i \middle| w_i^{(l)} \right\rangle \qquad (3.35) \\ &= \sum_i \frac{1}{p_i} \left(\frac{\partial p_i}{\partial \theta^j}\right) \left(\frac{\partial p_i}{\partial \theta^k}\right) + 4\Re \sum_{i<j} (p_i + p_j) \left\langle w_i^{(k)} \middle| w_j \right\rangle \left\langle w_j \middle| w_i^{(l)} \right\rangle. \end{aligned}$$

Lemma 3.3

$$C_L(\theta) \leq C_\Upsilon(\theta). \tag{3.36}$$

with equality if and only if (3.22) holds.

Proof. Equation (3.36) is equivalent to

$$v^T C_L(\theta) v \leq v^T C_\Upsilon(\theta) v, \tag{3.37}$$

for all $v \in \mathbb{R}^p$. For given θ and v in $\mathbb{R}^p$, consider the set of one-parameter states

$$\rho_{\theta+tv} = \sum_{k=1}^{d} p_k(\theta+tv)|w_k(\theta+tv)\rangle\langle w_k(\theta+tv)|, \qquad t \in \mathbb{R}.$$

It was shown in the proof of Proposition 2.3 that

$$\begin{aligned}
\frac{d}{dt}p_k(\theta+tv) &= \sum_l \frac{\partial p_k(\theta)}{\partial \theta^l} v^l + O(t), \quad t \to 0, && (3.38)\\
\frac{d}{dt}|w_k(\theta+tv)\rangle &= \sum_l \left| w_k(\theta)^{(l)} \right\rangle v^l + O(t), \quad t \to 0, && (3.39)\\
\left| w_k(\theta)^{(l)} \right\rangle &= \frac{\partial}{\partial \theta^l}|w_k(\theta)\rangle,
\end{aligned}$$

where v^l is the lth component of the vector v. From Lemma 3.1 it is known that $C_L(t) \leq C_\Upsilon(t)$, i.e.

$$\begin{aligned}
\sum_i \frac{1}{p_i(\theta+tv)}\left(\frac{dp_i}{dt}\right)^2 &+ 4\sum_{j<k}(p_j(\theta+tv)+p_k(\theta+tv))\left|\left\langle \frac{dw_j}{dt}\middle| w_k\right\rangle\right|^2 \\
\leq \sum_i \frac{1}{p_i(\theta+tv)}\left(\frac{dp_i}{dt}\right)^2 &+ 4\sum_{j<k}(p_j(\theta+tv)+p_k(\theta+tv))\left|\left\langle \frac{dw_j}{dt}\middle| w_k\right\rangle\right|^2 \\
&+ 4\sum_i p_i(\theta)\left|\left\langle \frac{dw_i}{dt}\middle| w_i\right\rangle\right|^2. \qquad (3.40)
\end{aligned}$$

Using (3.38) and (3.39) and evaluating at $t = 0$ gives

$$\begin{aligned}
&\sum_{m,n} v^m v^n \left(\sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta^m}\right)\left(\frac{\partial p_i}{\partial \theta^n}\right) + 4\sum_{i<j}(p_i+p_j)\left\langle w_i^{(m)}\middle| w_j\right\rangle\left\langle w_j\middle| w_i^{(n)}\right\rangle \right) \\
&\leq \sum_{r,s} v^r v^s \left(\sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta^r}\right)\left(\frac{\partial p_i}{\partial \theta^s}\right) + 4\sum_{i<j}(p_i+p_j)\left\langle w_i^{(r)}\middle| w_j\right\rangle\left\langle w_j\middle| w_i^{(s)}\right\rangle \right) \\
&\qquad + 4\sum_i p_i \left\langle w_i^{(m)}\middle| w_i\right\rangle\left\langle w_i\middle| w_i^{(n)}\right\rangle \Bigg).
\end{aligned}$$

This can be rewritten as

$$\sum_{m,n} v^m v^n C_L(\theta)_{mn} \leq \sum_{r,s} v^r v^s C_\Upsilon(\theta)_{rs}.$$

This is equivalent to (3.37). Since this holds for all v in $\mathbb{R}^p$, (3.36) holds.

Equality in (3.36) is equivalent to

$$v^T C_L(\theta) v = v^T C_\Upsilon(\theta) v, \tag{3.41}$$

for all $v \in \mathbb{R}^p$. From the proof of Lemma 3.3 it is seen that (3.41) is satisfied for all $v \in \mathbb{R}^p$ if and only if, for one-parameter families of states $\rho_{\theta+tv}$, $C_L(t)|_{t=0} = C_\Upsilon(t)|_{t=0}$. From Lemma 3.1 this is possible if and only if the channel satisfies (3.24) at the point $t = 0$. This condition is equal to

$$\sum_i p_i(t) \left| \left\langle \frac{dw_i}{dt} \middle| w_i \right\rangle \right|^2 \Bigg|_{t=0} = 0. \tag{3.42}$$

Using (3.39), this condition can be rewritten as

$$\sum_{l=1}^{m} v^m v^n \sum_i p_i \left\langle w_i^{(m)} \middle| w_i \right\rangle \left\langle w_i \middle| w_i^{(n)} \right\rangle = 0, \quad \forall m, n. \tag{3.43}$$

Condition (3.43) holds for all v if and only if (3.22) is satisfied.

Proposition 3.2 *In the multi-parameter case the SLD quantum information is the matrix with entries*

$$(H)_{kl} = \sum_i \frac{1}{p_i} \left(\frac{\partial p_i}{\partial \theta^k} \right) \left(\frac{\partial p_i}{\partial \theta^l} \right) + 4\Re \sum_{i<j} \frac{(p_i - p_j)^2}{p_i + p_j} \left\langle w_i^{(k)} \middle| w_j \right\rangle \left\langle w_j \middle| w_i^{(l)} \right\rangle. \tag{3.44}$$

Proof. In the multi-parameter case a particular choice of SLD with respect to the parameter θ^k is

$$\tilde{\lambda}^k = \sum_{i, p_i \neq 0} \frac{1}{p_i} \frac{\partial p_i}{\partial \theta^k} \left| w_i \right\rangle \left\langle w_i \right| + \sum_{i \neq j, p_i + p_j > 0} 2 \frac{p_i - p_j}{p_i + p_j} \left\langle w_i^{(k)} \middle| w_j \right\rangle \left| w_i \right\rangle \left\langle w_j \right|. \tag{3.45}$$

Proposition 3.2 follows almost identically to the one-parameter case (see proof of Proposition 2.2).

Lemma 3.4

$$H_\theta \leq C_L(\theta), \tag{3.46}$$

with equality if and only if (3.21) holds.

Proof. This follows from Lemma 3.2 in the same way as Lemma 3.3 follows from Lemma 3.1.

3.5 C_L as the minimum of C_Υ

Example 3.1 showed that for C_Υ, different choices of eigenvectors of ρ_θ can result in completely different metrics. Since C_Υ is an upper bound on Fisher information, it seems sensible to choose the minimum among possible values of C_Υ. It will now be investigated whether there exists a choice of eigenvectors such that $C_\Upsilon = C_L$.

3.5.1 One-parameter case

Consider a family of states $\rho_\theta = \sum_i p_i(\theta)|w_i(\theta)\rangle\langle w_i(\theta)|$. A phase change of the eigenvectors $|w_1(\theta)\rangle, \ldots, |w_d(\theta)\rangle$ sends these vectors to $|v_1(\theta)\rangle, \ldots, |v_d(\theta)\rangle$, where $|v_j(\theta)\rangle = e^{i\alpha_j(\theta)}|w_j(\theta)\rangle$ for some real-valued functions $\alpha_1, \ldots, \alpha_d$. The density matrix ρ_θ is unchanged. Now

$$\frac{d}{d\theta}|v_k(\theta)\rangle = i\frac{d\alpha_k}{d\theta}e^{i\alpha_k(\theta)}|w_k(\theta)\rangle + e^{i\alpha_k(\theta)}\frac{d}{d\theta}|w_k(\theta)\rangle$$

and hence

$$\langle v_k'|v_k\rangle = -i\frac{d\alpha_k}{d\theta} + \langle w_k'|w_k\rangle.$$

Choosing

$$\alpha_k(\theta) = -i\int_{\theta_0}^{\theta}\langle w_k'(\phi)|w_k(\phi)\rangle d\phi,$$

(3.24) is satisfied. (Since $\langle w_k'|w_k\rangle$ is purely imaginary, α_k is real.) Thus in the one-parameter case C_L is the minimum among C_Υ.

3.5.2 Multi-parameter case

A phase change of the eigenvectors $|w_1(\theta)\rangle, \ldots, |w_d(\theta)\rangle$ sends these vectors to $|v_1(\theta)\rangle, \ldots, |v_d(\theta)\rangle$, where $|v_j(\theta)\rangle = e^{i\alpha_j(\theta)}|w_j(\theta)\rangle$ for some real-valued functions $\alpha_1, \ldots, \alpha_d$. In this case $\theta = (\theta^1, \ldots, \theta^p)$. Equality holds in (3.36) if and

only if (3.22) holds. Now,

$$\frac{\partial}{\partial\theta^m}|v_j(\theta)\rangle = i\frac{\partial\alpha_j}{\partial\theta^m}e^{i\alpha_j(\theta)}|w_j(\theta)\rangle + e^{i\alpha_j(\theta)}\frac{\partial}{\partial\theta^m}|w_k(\theta)\rangle$$

and hence

$$\left\langle v_j^{(m)}\middle|v_j\right\rangle = -i\frac{\partial\alpha_j}{\partial\theta^m} + \left\langle w_j^{(m)}\middle|w_j\right\rangle.$$

This is zero if and only if

$$i\frac{\partial\alpha_j}{\partial\theta^m} = \left\langle\frac{\partial w_j}{\partial\theta^m}\middle|w_j\right\rangle \in i\mathbb{R} \qquad \forall j,m.$$

This is solvable for $\alpha_1,\ldots,\alpha_d$ if and only if

$$\frac{\partial^2\alpha_j}{\partial\theta^k\partial\theta^l} = \frac{\partial^2\alpha_j}{\partial\theta^l\partial\theta^k} \quad \forall j,k,l.$$

This is equivalent to

$$\frac{\partial}{\partial\theta^k}\left\langle\frac{\partial w_j}{\partial\theta^l}\middle|w_j\right\rangle = \frac{\partial}{\partial\theta^l}\left\langle\frac{\partial w_j}{\partial\theta^k}\middle|w_j\right\rangle \quad \forall j,k,l,$$

which is equivalent to

$$\left\langle\frac{\partial^2 w_j}{\partial\theta^k\partial\theta^l}\middle|w_j\right\rangle + \left\langle\frac{\partial w_j}{\partial\theta^l}\middle|\frac{\partial w_j}{\partial\theta^k}\right\rangle = \left\langle\frac{\partial^2 w_j}{\partial\theta^l\partial\theta^k}\middle|w_j\right\rangle + \left\langle\frac{\partial w_j}{\partial\theta^k}\middle|\frac{\partial w_j}{\partial\theta^l}\right\rangle \quad \forall j,k,l.$$

Since $|w_j\rangle$ is assumed to be continuously differentiable,

$$\left\langle\frac{\partial^2 w_j}{\partial\theta^k\partial\theta^l}\middle|w_j\right\rangle = \left\langle\frac{\partial^2 w_j}{\partial\theta^l\partial\theta^k}\middle|w_j\right\rangle \quad \forall j,k,l,$$

and hence it is required that

$$\left\langle\frac{\partial w_j}{\partial\theta^l}\middle|\frac{\partial w_j}{\partial\theta^k}\right\rangle = \left\langle\frac{\partial w_j}{\partial\theta^k}\middle|\frac{\partial w_j}{\partial\theta^l}\right\rangle \qquad \forall j,k,l.$$

This is satisfied if and only if

$$\left\langle\frac{\partial w_j}{\partial\theta^l}\middle|\frac{\partial w_j}{\partial\theta^k}\right\rangle \in \mathbb{R} \qquad \forall j,k,l, \tag{3.47}$$

which, in general, does not hold. Hence, for multi-parameter families of states, C_L is not generally the minimum among C_Υ.

Example 3.3 *For the family of states given in Example 3.1,*

$$\begin{aligned}\left\langle \frac{\partial w_1}{\partial \theta} \middle| \frac{\partial w_1}{\partial \phi} \right\rangle &= \frac{i}{2}\sin(\theta/2)\cos(\theta/2), \\ \left\langle \frac{\partial w_2}{\partial \theta} \middle| \frac{\partial w_2}{\partial \phi} \right\rangle &= \frac{-i}{2}\sin(\theta/2)\cos(\theta/2).\end{aligned}$$

Since (3.47) is not satisfied, C_L is not the minimum among C_Υ, for this family of states.

3.6 Relationship between C_L and SLD information of mixtures

For a general family of states $\rho_\theta = \sum_i p_i(\theta)|w_i(\theta)\rangle\langle w_i(\theta)|$, C_L was defined in (3.35) as the matrix with entries

$$\begin{aligned}(C_L)_{kl} &= (C_\Upsilon)_{kl} - 4\sum_i p_i \left\langle w_i^{(k)} \middle| w_i \right\rangle \left\langle w_i \middle| w_i^{(l)} \right\rangle \\ &= \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta^k}\right)\left(\frac{\partial p_i}{\partial \theta^l}\right) \\ &+ 4\sum_i p_i \left(\Re \left\langle w_i^{(k)} \middle| w_i^{(l)} \right\rangle - \left\langle w_i^{(k)} \middle| w_i \right\rangle \left\langle w_i \middle| w_i^{(l)} \right\rangle \right) \qquad (3.48)\end{aligned}$$

by (3.30). It is not difficult to show that the SLD quantum information for $\rho_i(\theta) = |w_i(\theta)\rangle\langle w_i(\theta)|$ is the matrix with entries

$$(H(\rho_i))_{kl} = 4\Re \left\langle w_i^{(k)} \middle| w_i^{(l)} \right\rangle - \left\langle w_i^{(k)} \middle| w_i \right\rangle \left\langle w_i \middle| w_i^{(l)} \right\rangle. \qquad (3.49)$$

Thus,

$$C_L\left(\sum_i p_i \rho_i\right) = F_\theta(p) + \sum_i p_i H(\rho_i), \qquad (3.50)$$

where $F_\theta(p)$ is the Fisher information matrix for $p = (p_1, \ldots, p_d)$, which has entries

$$(F_\theta(p))_{kl} = \sum_i \frac{1}{p_i}\left(\frac{\partial p_i}{\partial \theta^k}\right)\left(\frac{\partial p_i}{\partial \theta^l}\right). \qquad (3.51)$$

The result (3.50) states that the C_L quantum information is equal to the classical Fisher information of the probability distribution $\{p_1, \ldots, p_d\}$ plus

a weighted sum of the SLD quantum informations of the pure states $\rho_i(\theta)$ of which the state ρ_θ is a convex mixture. From (3.48) it can be seen that for pure states, for which there is only one non-zero p_i, $C_L = H$, and so

$$C_L\left(\sum_i p_i\rho_i\right) = F_\theta(p) + \sum_i p_i C_L(\rho_i). \tag{3.52}$$

Note that (3.50) and (3.52) are analogous to (7.4) of Amari (1982): Given random variables X and Y depending on θ with

$$f(x,y;\theta) = g(x;\theta)h(y|x;\theta), \tag{3.53}$$

$$F_\theta^{X,Y} = F_\theta^X + E_X[F_\theta^Y|x], \tag{3.54}$$

where

$$\begin{aligned} F_\theta^{X,Y} &= \int\int f(x,y;\theta)\left(\frac{d\log f(x,y;\theta)}{d\theta}\right)^2 dxdy \\ F_\theta^X &= \int g(x;\theta)\left(\frac{d\log g(x;\theta)}{d\theta}\right)^2 dx \\ E_X[F_\theta^Y|x] &= \int g(x;\theta)\left(\int h(y|x;\theta)\left(\frac{d\log h(y|x;\theta)}{d\theta}\right)^2 dy\right) dx. \end{aligned}$$

Chapter 4

Simultaneous estimation of several commuting quantum unitary channels

4.1 Introduction

The situation in which there are n non-identical commuting channels which are 'dependent' (having the same parameter but different forms) will be considered. This chapter introduces the idea of estimation of different commuting unitary channels simultaneously, as opposed to estimating them separately. Using the SLD quantum information as a measure of performance, it will be shown that this can give considerable improvement over estimating the channels individually.

4.1.1 Estimation of unitary channels

Estimation of an unknown or partially unknown unitary channel has received a lot of attention recently, see Rudolph and Grover (2003), Ji *et al.* (2008), Acín *et al.* (2001), Bagan *et al.* (2004a,b), Ballester (2004a,b), de Martini *et al.* (2003), Fujiwara (2002), Hayashi (2006a). Almost every quantum information protocol assumes perfect knowledge of a quantum channel. In practice, knowledge will be imperfect; hence estimation of quantum channels has to precede most other quantum information schemes, and its optimization is of fundamental importance.

It will be assumed that the unitary channel comes from a parametric family of channels. When estimating a parameter θ in a one-parameter model, the SLD quantum information H_θ will be used as a measure of performance.

When H_θ is attainable, i.e. there exists an M such that $F_\theta^M = H_\theta$, the following result is of importance: As the number of observations $N \to \infty$, using the POVM M and an unbiased maximum likelihood estimator,

$$NE[(\hat{\theta} - \theta)^2] \to \frac{1}{H_\theta} \tag{4.1}$$

(Van der Vaart, 1998, p. 63). When $\dim \theta > 1$, the performance of estimation will be quantified using the trace of the SLD quantum information ($\mathrm{tr}\{H_\theta\}$). When $\dim \theta > 1$, the SLD quantum informations for different parametric families of states may be incomparable. That is, given two families of states $\rho_\theta^{(1)}$ and $\rho_\theta^{(2)}$, with SLD quantum informations $H_\theta^{(1)}$ and $H_\theta^{(2)}$, it may be that $H_\theta^{(1)} \not\geq H_\theta^{(2)}$ and $H_\theta^{(1)} \not\leq H_\theta^{(2)}$. The quantity $\mathrm{tr}\{H_\theta\}$ is useful since (Ballester, 2004a)

(i) it treats the parameters $\theta^1, \ldots, \theta^p$ with equal importance,

(ii) if $\mathrm{tr}\{H_\theta^{(1)}\} \geq \mathrm{tr}\{H_\theta^{(2)}\}$ then $H_\theta^{(1)} \not< H_\theta^{(2)}$.

The output state will be measured using POVMs which satisfy (1.69) (possibly using an adaptive measurement), and an estimate of θ and hence U_θ will be obtained using the maximum likelihood estimator.

Previous work in estimation (see Section 1.11.1) has looked at the case where there are n copies of some U_θ. In this chapter a more general problem is considered: given n channels which are not identical, is it better to estimate each of them individually or is it possible to improve on this by using the channels in parallel, as in the case of n identical channels? It may be that in practice, more commonly, there are n channels which are different (but functionally dependent) than n channels which are identical.

In this chapter the performance of estimation will be considered as a function of N, the number of times each of the n channels is used. It will be assumed that each channel can be used only once on each input state.

4.2 Simplifying Matsumoto's equality condition

The following result will simplify later calculations.

It was mentioned in Theorem 1.3 that for pure state models $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$, there exists a POVM and estimator such that equality holds in the quantum Cramér-Rao inequality, (1.79), at $\theta = \theta_0$ if and only if

$$\Im\langle l_j(\theta)|l_k(\theta)\rangle = 0, \quad \forall j, k, \tag{4.2}$$

where $|l_j(\theta)\rangle = \lambda_\theta^j|\psi_\theta\rangle$ (Matsumoto, 1997, Fujiwara, 2002, Matsumoto, 2002). An equivalent condition which is simpler to check, and will be used in this chapter, is given in the following lemma.

Lemma 4.1 *For families of pure states $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$, equality holds in the quantum Cramér-Rao inequality, (1.79), at $\theta = \theta_0$ if and only if*

$$\Im\langle\psi_\theta^{(j)}|\psi_\theta^{(k)}\rangle = 0 \quad \forall j,k, \tag{4.3}$$

where $|\psi_\theta^{(j)}\rangle = \partial|\psi_\theta\rangle/\partial\theta^j$.

Proof. For pure states, equality holds in (1.79) at $\theta = \theta_0$ if and only if (4.2) is satisfied. Now $|l_j(\theta)\rangle = \lambda_\theta^j|\psi_\theta\rangle$ is independent of the choice of λ_θ^j (Fujiwara, 2002, Appendix A, before (7)). A possible choice is

$$\lambda_\theta^j = 2\partial\rho_\theta/\partial\theta^j = 2(|\psi_\theta^{(j)}\rangle\langle\psi_\theta| + |\psi_\theta\rangle\langle\psi_\theta^{(j)}|). \tag{4.4}$$

A little algebra gives

$$\langle l_j(\theta)|l_k(\theta)\rangle = 4(\langle\psi_\theta^{(j)}|\psi_\theta^{(k)}\rangle + \langle\psi_\theta^{(j)}|\psi_\theta\rangle\langle\psi_\theta^{(k)}|\psi_\theta\rangle). \tag{4.5}$$

The second term is always real, since $\langle\psi_\theta^{(l)}|\psi\rangle$ is purely imaginary for all l (see below (2.12)). Thus condition (4.2) is equivalent to condition (4.3).

Remark 4.1 *Although $\langle\psi_\theta^{(j)}|\psi_\theta^{(k)}\rangle$ depends on the choice of phase of $|\psi_\theta\rangle$, $\Im\langle\psi_\theta^{(j)}|\psi_\theta^{(k)}\rangle$ does not.*

Lemma 4.2 *If $|x_1\rangle, \ldots, |x_n\rangle \in \mathbb{C}^d$ such that*

(i) $|x_1\rangle, \ldots, |x_n\rangle$ are $\mathbb{R}$-linearly independent,

(ii) $\langle x_j|x_k\rangle \in \mathbb{R}$ for all $j,k = 1,\ldots,n$,

then $n \leq d$.

Proof. Suppose that $\exists\ \alpha_1, \ldots, \alpha_n \in \mathbb{C}$ such that

$$\sum_{j=1}^n \alpha_j|x_j\rangle = 0.$$

Putting $\alpha_j = a_j + ib_j$, where $a_j, b_j \in \mathbb{R}$ for $j = 1,\ldots,n$, gives

$$\sum_{j=1}^n (a_j + ib_j)|x_j\rangle = 0,$$

and so

$$\sum_{j=1}^{n}(a_j + ib_j)\langle x_k|x_j\rangle = 0, \quad \text{for all } k. \tag{4.6}$$

From condition (ii),

$$\sum_{j=1}^{n} a_j\langle x_k|x_j\rangle = 0, \text{ for all } k.$$

Thus

$$\sum_{j,k=1}^{n} a_j a_k\langle x_k|x_j\rangle = 0,$$

and so

$$\sum_{j=1}^{n} a_j|x_j\rangle = 0.$$

Since by (i) $|x_1\rangle, \ldots, |x_n\rangle$ are $\mathbb{R}$-linearly independent, $a_j = 0$ for $j = 1, \ldots, n$. Similarly (4.6) gives $b_j = 0$ for $j = 1, \ldots, n$. Thus $\alpha_j = 0$ for $j = 1, \ldots, n$. Consequently, $|x_1\rangle, \ldots, |x_n\rangle$ are $\mathbb{C}$-linearly independent. Therefore, if $|x_1\rangle, \ldots, |x_n\rangle \in \mathbb{C}^d$ satisfy (i) and (ii), then $n \leq d$.

Theorem 4.1 *For a d-dimensional non-degenerate family of pure states $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$, $\theta = (\theta^1, \ldots, \theta^p)$, H_θ is attainable only if $p \leq d - 1$.*

Proof. The vectors $\{|l_j(\theta)\rangle\}$, where $|l_j(\theta)\rangle = \lambda_\theta^j|\psi_\theta\rangle$ are $\mathbb{R}$-linearly independent (due to the nondegeneracy of the parameterization $\theta \mapsto \rho_\theta$) (Fujiwara, 2002, Appendix A). Since $\langle l_j(\theta)|\psi_\theta\rangle = \mathrm{tr}\{\rho_\theta\lambda_\theta^j\} = 0$ for all j, the vectors $\{|\psi_\theta\rangle, |l_1(\theta)\rangle, \ldots, |l_p(\theta)\rangle\}$ are also $\mathbb{R}$-linearly independent. From (4.2) it is seen that H_θ is attainable if and only if the set of vectors $\{|\psi_\theta\rangle, |l_1(\theta)\rangle, \ldots, |l_p(\theta)\rangle\}$ satisfy conditions (i) and (ii) in Lemma 4.2. It follows from Lemma 4.2 that H_θ is attainable only if $p \leq d - 1$.

Remark 4.2 *As any unitary channel can be specified by $d^2 - 1$ parameters, Theorem 4.1 shows the importance of enlarging the Hilbert space to estimate a completely unknown $U \in SU(d)$, i.e. letting $\mathbb{I}_d \otimes U$ act on a state $|\phi\rangle \in \mathbb{C}^{d^2}$. In this case it is possible to have a maximum of $d^2 - 1$ parameters such that H_θ is attainable.*

For the channels considered in this chapter an extension of the form $\mathbb{I}_d \otimes \mathcal{E}$ does not increase the maximum attainable Fisher information.

4.3 A 2-dimensional family of non-identical channels

Consider the following set of 2-dimensional channels, which are all functions of the parameter θ,

$$U_\theta^1 = \begin{pmatrix} 1 & 0 \\ 0 & e^{if_1(\theta)} \end{pmatrix}, \ldots, \quad U_\theta^n = \begin{pmatrix} 1 & 0 \\ 0 & e^{if_n(\theta)} \end{pmatrix}, \tag{4.7}$$

where $0 \leq \theta \leq q$, for some q, and $f_j : \mathbb{R} \to \mathbb{R}$. The following conditions are imposed on the functions f_j:

(a) $\dfrac{df_j(\theta)}{d\theta} > 0$,

(b) $0 \leq \sum_j f_j(\theta) \leq \pi$,

for all j and θ.

Remark 4.3 *Throughout this chapter similar restrictions will be given on the unitary matrices to be estimated. Condition (a) means that as θ is increased the angle through which states are rotated is also increased, though the amount by which the phase increases varies from unitary to unitary; condition (b) can be thought of as having some prior information about the phases to be estimated, possibly through a knowledge of the experimental arrangements.*

The SLD quantum informations of the schemes

(i) letting each of the n channels act on identical copies of $|\psi_x\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle)$, i.e.

$$|\psi_x\rangle \mapsto U_\theta^j |\psi_x\rangle,$$

(ii) arranging all n of the channels in parallel and using the entangled input state $|\psi\rangle = 1/\sqrt{2}(|00\cdots0\rangle + |11\cdots1\rangle) \in \mathbb{C}^{2^n}$, i.e.

$$|\psi\rangle \mapsto (U_\theta^1 \otimes \cdots \otimes U_\theta^n)|\psi\rangle, \tag{4.8}$$

will be compared. If U_θ^j acts on the state $|\psi_x\rangle$, the output state is $1/\sqrt{2}(|0\rangle + e^{if_j(\theta)}|1\rangle)$. This gives $H_\theta^j = (df_j(\theta)/d\theta)^2$, which is attainable by measuring in x, i.e. using the POVM $M^x = \{M_0 = |\psi_x\rangle\langle\psi_x|, \mathbb{I} - M_0\}$. Thus for the n channnels

$$H_\theta^{(i)} = \sum_{j=1}^{n} \left(\frac{df_j(\theta)}{d\theta} \right)^2 \tag{4.9}$$

and is attainable. An estimate $\hat{\theta}^{(i)}$ is obtained using the maximum likelihood estimator.

Now the n-partite input state $|\psi\rangle$ will be considered. The output state is $1/\sqrt{2}(|00\cdots0\rangle + e^{i\sum_{j=1}^{n} f_j(\theta)}|11\cdots1\rangle)$. Computation gives

$$H_\theta^{(ii)} = \left(\sum_{j=1}^{n} \frac{df_j(\theta)}{d\theta}\right)^2, \tag{4.10}$$

which is attainable using the POVM $M = \{M_0 = |\psi\rangle\langle\psi|, \mathbb{I} - M_0\}$. Because of conditions (a) and (b), θ can be identified. An estimate $\hat{\theta}^{(ii)}$ is obtained using the maximum likelihood estimator.

The SLD quantum informations (4.9) and (4.10) may look similar, but they are not. The position of the bracket makes a considerable difference. From condition (a) on the functions f_j, (4.10) is considerably larger than (4.9). For example, in the case when $f_j(\theta) = \theta$ for all j, the SLD quantum informations are Nn and Nn^2, respectively.

A consequence of this is that the asymptotic limit of the mean square error is considerably smaller using approach (ii). The asymptotic limits of the mean square errors for approaches (i) and (ii) are, respectively,

$$NE[(\hat{\theta}^{(i)} - \theta)^2] \to \frac{1}{\sum_{j=1}^{n} \left(\frac{df_j(\theta)}{d\theta}\right)^2}, \tag{4.11}$$

$$NE[(\hat{\theta}^{(ii)} - \theta)^2] \to \frac{1}{\left(\sum_{j=1}^{n} \frac{df_j(\theta)}{d\theta}\right)^2}. \tag{4.12}$$

4.3.1 Sequential method

Here it will be shown that, without using entanglement, it is possible to obtain the same SLD quantum information for the set of non-identical channels (4.7), as was obtained in approach (ii). A third scheme for estimating the set of channels (4.7) will be introduced, which will be referred to as the *sequential scheme*. The sequential scheme makes no use of entanglement.

(iii) The channels (4.7) are each used once on the same separable input state $|\psi_x\rangle$, i.e.

$$\begin{aligned} |\psi_x\rangle &\mapsto U_\theta^n \cdots U_\theta^2 U_\theta^1 |\psi_x\rangle \\ &= \frac{1}{\sqrt{2}}(|0\rangle + e^{i\sum_{j=1}^{n} f_j(\theta)}|1\rangle). \end{aligned} \tag{4.13}$$

Calculation gives

$$H_\theta^{(iii)} = \left(\sum_j \frac{df_j(\theta)}{d\theta} \right)^2 \tag{4.14}$$

and is attainable by measuring in x. An estimate $\hat{\theta}^{(iii)}$ is obtained using the maximum likelihood estimator. The SLD quantum information obtained in approach (iii) is equal to that obtained in approach (ii), thus will have the same asymptotic limit on the mean square error, (4.12).

4.4 A more general family of one-parameter channels

Often physicists are interested in unitary channels parameterised as $V_\theta = \exp(i\theta H)$, where H is an observable related to the energy in a system, known as the *Hamiltonian*. This seemingly simple channel has many examples in interferometry and measurement of small forces. (For more on channels of this type see Giovannetti *et al.*, 2006, and the references therein.) Consider n d-dimensional unitary channels parameterised as

$$U_\theta^j = \exp(i f_j(\theta) H), \quad 1 \leq j \leq n, \tag{4.15}$$

where $0 \leq \theta \leq q$, for some q, $f_j : \mathbb{R} \to \mathbb{R}$ for all j. The following conditions are imposed on the functions f_j:

(a) $\dfrac{df_j(\theta)}{d\theta} > 0$,

(b) $0 \leq \sum_j f_j(\theta) \leq \pi$,

for all j and θ. The problem of finding the optimal input state will not be considered. The SLD quantum informations for

(i) letting each of the n channels act on identical copies of some $|\psi_0\rangle$, i.e.

$$|\psi_0\rangle \mapsto U_\theta^j |\psi_0\rangle,$$

(ii) letting each of the n channels act on the same separable state $|\psi_0\rangle$, i.e.

$$|\psi_0\rangle \mapsto U_\theta^n \cdots U_\theta^2 U_\theta^1 |\psi_0\rangle$$

will be compared. The SLD quantum informations for (i) and (ii) are, respectively,

$$H_\theta^{(i)} = 4\sum_{j=1}^{n}\left(\frac{df_j(\theta)}{d\theta}\right)^2[\langle\phi_0|H^2|\phi_0\rangle - \langle\phi_0|H|\phi_0\rangle^2], \tag{4.16}$$

$$H_\theta^{(ii)} = 4\left(\sum_{j=1}^{n}\frac{df_j(\theta)}{d\theta}\right)^2[\langle\phi_0|H^2|\phi_0\rangle - \langle\phi_0|H|\phi_0\rangle^2]. \tag{4.17}$$

Because of condition (a) the SLD quantum information of (ii), given by (4.17), is considerably larger than that of (i), given by (4.16). These results hold for all choices of input state $|\phi_0\rangle$.

4.5 A d-dimensional family of non-identical channels

The situtation of having n 'dependent' d-dimensional commuting channels will be considered. These will be parameterised in a similar way to that used by Ballester (2004a). Ballester (2004a) looked at commuting unitary channels. Any commuting unitary channel can be specified using $d-1$ parameters, i.e. by a parameter $\theta = (\theta_1, \ldots, \theta_{d-1})$. Given a set of $d \times d$ matrices t_k, $k = 1, \ldots, d-1$, satisfying

(i) $t_k = t_k^\dagger$,

(ii) $\mathrm{tr}\{t_k\} = 0$,

(iii) $\mathrm{tr}\{t_k t_l\} = \delta_{kl}$,

(iv) $t_k t_l = t_l t_k$,

Ballester parameterised the set of commuting unitary channels as

$$U_\theta = \exp\left(i\sum_{k=1}^{d-1}\theta_k t_k\right). \tag{4.18}$$

Since, from (iv), t_k and t_l commute, they share a basis $\{|w_k\rangle\}$, which is assumed to be known. Consequently, any t_m can be written as

$$t_m = \sum_{i=1}^{d} c_{mi}|w_i\rangle\langle w_i|. \tag{4.19}$$

From condition (i) it follows that $c_{mi} \in \mathbb{R}$ for all m, i. Conditions (ii) and (iii) give

$$\sum_{i=1}^{d} c_{mi} = 0, \tag{4.20}$$

$$\sum_{i=1}^{d} c_{mi}c_{ni} = \delta_{mn}. \tag{4.21}$$

Ballester showed that there is no advantage in extending U_θ and using a maximally entangled input state. The maximum value of $\mathrm{tr}\{H_\theta\}$ can be attained using the separable state

$$|\psi_{sep}\rangle = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} |w_k\rangle. \tag{4.22}$$

Consider the set of channels

$$U_\theta^j = \exp\left(i \sum_{k=1}^{d-1} f_j(\theta_k) t_k \right), \quad 1 \leq j \leq n, \tag{4.23}$$

where $0 \leq \theta \leq q$, for some q, $f_j : \mathbb{R} \to \mathbb{R}$ and $f_j(\theta_0) = 0$ for all j. All n channels depend on the parameter $\theta = (\theta_1, \ldots, \theta_{d-1})$, and each channel depends on every component of θ. The following conditions are imposed on the functions f_j:

(a) $\dfrac{df_j(\theta)}{d\theta} > 0$

(b) $0 \leq \sum_j f_j(\theta_k) \leq \pi$,

for all j, k and θ. The traces of the SLD quantum information for

(i) letting each of the n channels act on identical copies of $|\psi_{sep}\rangle$ given in (4.22), i.e.

$$|\psi_{sep}\rangle \mapsto U_\theta^j |\psi_{sep}\rangle,$$

(ii) letting each of the n channels act on the same separable state $|\psi_{sep}\rangle$, i.e.

$$|\psi_{sep}\rangle \mapsto U_\theta^n \cdots U_\theta^2 U_\theta^1 |\psi_{sep}\rangle$$

will be compared.

Proposition 4.1 *The traces of the SLD quantum informations for (i) and (ii) are, respectively,*

$$\mathrm{tr}\{H_\theta^{(i)}\} = \frac{4}{d}\sum_{i=1}^{d-1}\sum_{j=1}^{n}\left(\frac{\partial f_j(\theta_i)}{\partial\theta_i}\right)^2, \tag{4.24}$$

$$\mathrm{tr}\{H_\theta^{(ii)}\} = \frac{4}{d}\sum_{i=1}^{d-1}\left(\sum_{j=1}^{n}\frac{\partial f_j(\theta_i)}{\partial\theta_i}\right)^2. \tag{4.25}$$

From condition (a), the trace of the SLD quantum information of (ii), given by (4.25), is considerably larger than that of (i), given by (4.24).

Proof. A proof will be given for (4.25); the proof for (4.24) is very similar. The jth unitary channel will be denoted by U_θ^j. As the U_θ^j commute,

$$\prod_{j=1}^{n} U_\theta^j = \exp\left\{i\sum_{k=1}^{d-1} g_k(\theta_k)t_k\right\}, \quad g_k(\theta_k) = \sum_{j=1}^{n} f_j(\theta_k). \tag{4.26}$$

Using each of the n channels on the single input state (4.22) gives the output state

$$|\psi_\theta\rangle = \left(\prod_{j=1}^{n} U_\theta^j\right)|\psi_{sep}\rangle = \exp\left\{i\sum_{k=1}^{d-1} g_k(\theta_k)t_k\right\}|\psi_{sep}\rangle. \tag{4.27}$$

An arbitrary diagonal element of H_θ is equal to

$$\begin{aligned}
(H_\theta^{(ii)})_{mm} &= 4\left[\langle\psi_\theta^{(m)}|\psi_\theta^{(m)}\rangle - |\langle\psi_\theta^{(m)}|\psi_\theta\rangle|^2\right], \quad |\psi_\theta^{(m)}\rangle = \partial|\psi_\theta\rangle/\partial\theta^m, \\
&= 4\left(\frac{\partial g_m(\theta_m)}{\partial\theta_m}\right)^2\left[\langle\psi_\theta|t_m t_m|\psi_\theta\rangle - |\langle\psi_\theta|t_m|\psi_\theta\rangle|^2\right] \\
&= 4\left(\frac{\partial g_m(\theta_m)}{\partial\theta_m}\right)^2\left[\langle\psi_{sep}|t_m t_m|\psi_{sep}\rangle - |\langle\psi_{sep}|t_m|\psi_{sep}\rangle|^2\right] \\
&= 4\left(\frac{\partial g_m(\theta_m)}{\partial\theta_m}\right)^2\left[\frac{1}{d}\sum_{k=1}^{d} c_{mk}^2 - \left|\frac{1}{d}\sum_{k=1}^{d} c_{mk}\right|^2\right] \text{ by (4.19)} \\
&= \frac{4}{d}\left(\frac{\partial g_m(\theta_m)}{\partial\theta_m}\right)^2 \text{ by (4.20) and (4.21).}
\end{aligned}$$

Thus

$$\begin{aligned}\mathrm{tr}\{H_\theta^{(ii)}\} &= \frac{4}{d}\sum_{m=1}^{d-1}\left(\frac{\partial g_m(\theta_m)}{\partial\theta_m}\right)^2 \\ &= \frac{4}{d}\sum_{m=1}^{d-1}\left(\sum_{j=1}^{N}\frac{\partial f_j(\theta_m)}{\partial\theta_m}\right)^2.\end{aligned}$$

Proposition 4.2 *The SLD quantum information (4.25) is attainable.*

Proof. The set of output states is given by (4.27). Now

$$\begin{aligned}\langle\psi_\theta^{(m)}|\psi_\theta^{(n)}\rangle &= \left(\frac{\partial g_m(\theta_m)}{\partial\theta_m}\right)\left(\frac{\partial g_n(\theta_n)}{\partial\theta_n}\right)\langle\psi_\theta|t_m t_n|\psi_\theta\rangle, \quad |\psi_\theta^{(m)}\rangle = \partial|\psi_\theta\rangle/\partial\theta^m, \\ &= \left(\frac{\partial g_m(\theta_m)}{\partial\theta_m}\right)\left(\frac{\partial g_n(\theta_n)}{\partial\theta_n}\right)\langle\psi_{sep}|t_m t_n|\psi_{sep}\rangle \\ &= \frac{\delta_{mn}}{d}\left(\frac{\partial g_m(\theta_m)}{\partial\theta_m}\right)^2.\end{aligned}$$

which is always real. Thus (4.3) is satisfied, and consequently H_θ is attainable.

Chapter 5

An iterative phase estimation algorithm

5.1 Introduction

This chapter considers phase estimation, which is of fundamental importance to quantum information and quantum computation. Phase estimation is related to some very important problems such as estimating eigenvalues (Wei and Nori, 2004, Aspuru-Guzik *et al.*, 2005, Wang *et al.*, 2008, 2009), precision measurement of length and optical properties, and clock synchronization (de Burgh and Bartlett, 2005). (The work in this chapter has been published in O'Loan (2010).)

Consider a unitary matrix U_θ depending on an unknown parameter θ for which one of its eigenvectors $|u\rangle$ is completely known; furthermore U_θ acts on $|u\rangle$ by $U_\theta|u\rangle = e^{i2\pi\theta}|u\rangle$, where $\theta \in [0, 1)$. The task of phase estimation is to estimate the eigenvalue $e^{i2\pi\theta}$, and consequently θ, as accurately as possible. This chapter considers phase estimation of a unitary matrix with known eigenvectors, which acts on a 2-dimensional Hilbert space. In particular, unitary matrices of the form

$$U_\theta = \begin{pmatrix} 1 & 0 \\ 0 & e^{i2\pi\theta} \end{pmatrix}, \tag{5.1}$$

are considered, where $\theta \in [0, 1)$. The angle θ will be thought of as a point on a circle of unit circumference, and confidence intervals for θ as arcs on a circle of unit circumference, known as confidence arcs. The distance between the point θ and an estimate $\hat{\theta}$, will be defined as

$$|\hat{\theta} - \theta|_1 = \min\left((\hat{\theta} - \theta)_{\mathrm{mod}\,1}, (\theta - \hat{\theta})_{\mathrm{mod}\,1}\right). \tag{5.2}$$

The performance of phase estimation schemes will be quantified in terms of the expected fidelity $\langle F(U_{\hat{\theta}}, U_\theta)\rangle$. The cost function

$$1 - \langle F(U_{\hat{\theta}}, U_\theta)\rangle = 1 - \frac{\left\langle |\mathrm{tr}\{U_{\hat{\theta}}^{-1} U_\theta\}|^2 \right\rangle}{d^2} \tag{5.3}$$

will be used, and its asymptotic scaling analysed as a function of n — the number of times that U_θ is used.

For a simple phase estimation approach where U_θ is used once on n identical copies of some input state (see Section 5.1.1), $1 - \langle F\rangle = O(1/n)$. This rate at which $1 - \langle F\rangle$ approaches zero is known as the *standard quantum limit* (de Burgh and Bartlett, 2005).

However, it has been shown (Hayashi, 2006a, Kahn, 2007, Imai and Fujiwara, 2007) that it is possible to obtain $1 - \langle F\rangle = O(1/n^2)$. This rate at which $1 - \langle F\rangle$ approaches zero is known as the *Heisenberg limit* (Giovannetti *et al.*, 2004), and cannot be beaten (Kahn, 2007). These methods require n copies of U_θ and entangled states.

It has further been shown that it is possible to achieve the Heisenberg limit without entanglement, and with only a single copy of U_θ (see Section 5.1.5). Estimation schemes of this type require a rotation gate capable of performing arbitrary rotations to perfect precision.

Kitaev (1996) sketched an iterative phase estimation method which requires only a single copy of U_θ and basic measurements: no extra rotation gate is needed. For this method $1 - \langle F\rangle = O((\log n/n)^2$, which is within a logarithmic factor of the Heisenberg limit. However, as will be shown in this chapter, attempts to give a detailed account for such a scheme have been unsuccessful. This chapter seeks to give a correct detailed phase estimation scheme similar to that of Kitaev (1996), which requires only a single copy of U_θ and basic measurements.

A selection of different phase estimation schemes will now be given.

5.1.1 Simple approach

A very simple method of phase estimation is to let U_θ act on the input state $|\psi_x\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle)$; the output state is $|\psi_\theta\rangle = 1/\sqrt{2}(|0\rangle + e^{i2\pi\theta}|1\rangle)$. After measuring in x, outcome 0 is observed with probability $p(0;\theta) = (1+\cos(2\pi\theta))/2$. Performing N measurements gives an estimate $\cos(2\pi\hat{\theta}) = 2N_{x=0}/N - 1$ of $\cos(2\pi\theta)$, where $N_{x=0}$ is the number of times outcome 0 is observed. After measuring in y, outcome 0 is observed with probability $p(0;\theta) = (1+\sin(2\pi\theta))/2$. Performing N measurements gives an estimate $\sin(2\pi\hat{\theta}) = 2N_{y=0}/N - 1$ of $\sin(2\pi\theta)$, where $N_{y=0}$ is the number of times

outcome 0 is observed. From estimates of $\cos(2\pi\theta)$ and $\sin(2\pi\theta)$ an estimate of θ can be obtained.

5.1.2 Kitaev's procedure

The first l-stage iterative phase estimation procedure was given by Kitaev (1996). (The number of stages l is chosen beforehand, and will be a compromise between the precision desired and experimental resources and limitations.) At the kth stage of Kitaev's procedure, U_θ acts 2^{k-1} times on a qubit, which is then measured. The experimenter performs some multiple of $\log(l/\epsilon)$ measurements of $(2^{k-1}\theta)_{\mathrm{mod}\,1}$. This ensures that it is possible to 'localize each of the numbers $2^{k-1}\theta$ in one of the 8 intervals $[(s-1)/8, (s+1)/8]$ $(s = 0, \ldots, 7)$ with error probability $\leq \epsilon/l$'. Using this information, an algorithm — which is not given — gives an estimate $\hat{\theta}$ satisfying

$$\Pr\left(\left(\hat{\theta} - 1/2^{l+2}, \hat{\theta} + 1/2^{l+2}\right) \ni \theta\right) \geq 1 - \epsilon. \tag{5.4}$$

5.1.3 The scheme of Rudolph and Grover

Rudolph and Grover (2003) looked at the problem of transmitting a reference frame from Alice to Bob, which is linked to estimation of an unknown $U \in SU(2)$, parametrized by three parameters α, θ, ϕ. The scheme of Rudolph and Grover involves estimating the parameters α, θ, ϕ individually using the following l-stage iterative procedure. The parameter $\theta \in [0, 1)$ is thought of in terms of an infinite binary expansion $\theta = w_1 w_2 \ldots w_l \ldots$. At the kth stage a qubit is sent back and forth between Alice and Bob in such a way that, when Bob finally measures it, he observes outcome 0 with probability $p_k(0;\theta) = (1 + \cos(2^k \pi\theta))/2$.

This is repeated a minimum of $N = 32\log_2(2l/\epsilon)$ times (Rudolph and Grover, 2003), which ensures that Bob's estimate $\hat{p}_k(1;\theta)$ of $p_k(1;\theta)$ satisfies

$$\Pr\left((\hat{p}_k - 1/4, \hat{p}_k + 1/4) \ni p_k\right) \geq 1 - \epsilon/l. \tag{5.5}$$

It is assumed that if $|\hat{p}_k - p_k| \leq 1/4$, then Bob can estimate the kth bit of θ correctly. If this is so, then from (5.5), the probability that Bob estimates the kth bit of θ correctly is at least $1 - \epsilon/l$, and the probability that he estimates all of the binary digits of θ correctly is at least $1 - \epsilon$. After l stages, an estimate $\hat{\theta} = \hat{w}_1 \hat{w}_2 \ldots \hat{w}_l$ is obtained, satisfying

$$\Pr\left(\left(\hat{\theta} - 1/2^{l}, \hat{\theta} + 1/2^{l}\right) \ni \theta\right) \geq 1 - \epsilon. \tag{5.6}$$

A similar scheme is then used to estimate the parameters α and ϕ. The method of Rudolph and Grover has been used by de Burgh and Bartlett (2005) for the problem of clock synchronization.

5.1.4 The procedure of Ji *et al.*

Ji *et al.* (2008) highlighted two errors with the method of Rudolph and Grover:

(i) knowing $|\hat{\theta} - \theta|_1 \leq 1/2^m$ does not give the first m bits of the binary expansion of θ – consider $\theta = 0.49$, $\hat{\theta} = 0.5$ and $m = 1$,

(ii) the method is problematic (in the sense explained in section 5.2) for θ close to $1/2$.

Ji *et al.* gave the following l-stage procedure. In the first stage, the experimenter lets U_θ act on $|\psi_x\rangle$ and then measures in x; outcome 0 is observed with probability $p(0;\theta) = (1 + \cos(2\pi\theta))/2$. The state $U_\theta|\psi_x\rangle$ is measured N times (N is some multiple of $\log(l/\epsilon)$), which gives an estimate $\hat{\theta}$ satisfying

$$\Pr\left(\left(\hat{\theta} - 1/12, \hat{\theta} + 1/12\right) \ni \theta\right) \geq 1 - \epsilon/l. \tag{5.7}$$

Having obtained an estimate $\hat{\theta}$,

1) if $\hat{\theta} \in [0, 5/12)$, define $r_1 = 2$ and $\nu_1 = 0$,

2) if $\hat{\theta} \in [5/12, 7/12)$, define $r_1 = 3$ and $\nu_1 = 1$,

3) if $\hat{\theta} \in [7/12, 1]$, define $r_1 = 2$ and $\nu_1 = 1$.

At the kth stage the experimenter lets U_θ act $r_1 r_2 \ldots r_{k-1}$ times on $|\psi_x\rangle$. After measuring $U_\theta^{r_1 r_2 \ldots r_{k-1}}|\psi_x\rangle$ N times, $(r_1 r_2 \ldots r_{k-1}\theta)_{\mathrm{mod}\,1}$ is estimated and r_k and ν_k are obtained in a similar way to r_1 and ν_1. After l stages, values are obtained for $(r_1, \ldots, r_l, \nu_1, \ldots, \nu_l)$. The final estimate of θ is

$$\hat{\theta} = \sum_{i=1}^{l} \frac{\nu_i}{\prod_{j=1}^{i} r_j}. \tag{5.8}$$

5.1.5 The method of Dobšíček *et al.*

A popular iterative estimation method is to take θ to have a binary expansion of given length l plus some small remainder, that is $\theta = w_1 w_2 \ldots w_l + \Delta$. The binary digits $w_1, \ldots, w_l$ are estimated one at a time with a single measurement. This has been done by Childs *et al.* (2000), Dobšíček *et al.* (2007), Knill *et al.* (2007). The method will be reviewed as described by Dobšíček *et al.* (2007).

At the kth stage the experimenter lets $U_\theta^{2^{l-k+1}}$ act on one of two qubits. The other qubit is acted on by a Z-rotation gate $e^{i\alpha_k \sigma_z}$ before being measured

— where $\alpha_0 = 0$ and α_k for $k = 2, \ldots, l$ depend on the results from the previous $k-1$ stages. From this measurement, an estimate $\hat{w}_{l-k+1}$ is obtained of the $(l-k+1)$th binary digit. After l stages an estimate $\hat{\theta} = \hat{w}_1 \hat{w}_2 \ldots \hat{w}_l$ is obtained of θ which satisfies

$$\Pr\left(\left(\hat{\theta} - 1/2^{l+1}, \hat{\theta} + 1/2^{l+1}\right) \ni \theta\right) \geq 0.81. \tag{5.9}$$

The probability that the final interval contains θ can be increased to $1-\epsilon$ by either (a) increasing the number of rounds to $l' = l + \log(2 + 1/(2\epsilon))$ or (b) using $O(\log^2(1/\epsilon))$ extra measurements of the first few binary digits (Dobšíček *et al.*, 2007). The method of Dobšíček *et al.* has recently been carried out on experimental data by Liu *et al.* (2007). Similar work has also been done by Higgins *et al.* (2007).

5.2 Problems

There is nothing wrong with Kitaev's method of iterative estimation. However, he does not give an algorithm for

(i) choosing which of the intervals contains $(2^{k-1}\theta)_{\mathrm{mod}\,1}$ with probability $1 - \epsilon/l$,

(ii) reconstructing θ given confidence intervals for $(2^{k-1}\theta)_{\mathrm{mod}\,1}$.

As will be seen in this section, there are gaps in the methods of Rudolph and Grover, and Ji *et al.* for (i). There are two main gaps in the method of Rudolph and Grover, which will now be explained. Firstly, $p_k(0;\theta) = (1 + \cos(2^k \pi\theta))/2$ is a multimodal function of θ. For example, $\theta = 3/4$ and $\theta = 1/4$ give the same value of $p_1(0;\theta)$, even though they differ in the first binary digit. To overcome this, an estimate of $\sin(2\pi\theta)$ is needed as well. This however is a trivial point and is easily overcome.

Secondly, if $\theta = 1/2 \pm \delta$, where δ is small, a large number of measurements is required to determine the first bit of θ correctly with high probability. If a mistake is made then, for the final estimate $\hat{\theta}$, $|\hat{\theta} - \theta|_1 \geq \delta$. This problem, which occurs for θ close to $1/2$, was pointed out by Ji *et al.* (2008).

A similar problem also occurs for $\theta = 0 \pm \delta$. Because of this, difficulties will be encountered in estimating the kth bit of θ whenever $(2^{k-1}\theta)_{\mathrm{mod}\,1} \approx 0$, $(2^{k-1}\theta)_{\mathrm{mod}\,1} \approx 1$ or $(2^{k-1}\theta)_{\mathrm{mod}\,1} \approx 1/2$. However, it may also be possible to overcome this problem using extra rotation gates in these cases.

There are also gaps in the method of Ji *et al.* (2008). Firstly, like Rudolph and Grover, they overlook the fact that $p_1(0;\theta) = (1+\cos(2\pi\theta))/2$ is bimodal. Secondly, the accuracy of their final estimate relies on the assumption that

if $|\hat{\theta} - \theta|_1 \leq 1/12$ and $\hat{\theta} \in [0, 5/12)$, then $\theta \in [0, 1/2)$. This is not true – consider $\theta = -1/12 \notin [0, 1/2)$. Similarly, they assume that if $|\hat{\theta} - \theta|_1 \leq 1/12$ and $\hat{\theta} \in [7/12, 1)$, then $\theta \in [1/2, 1)$, which again is not true – consider $\theta = 1/12 \notin [1/2, 1)$. Again problems will be encountered at the kth stage if $(r_1 \cdots r_{k-1}\theta)_{\mathrm{mod}\,1} \approx 0$ or $(r_1 \cdots r_{k-1}\theta)_{\mathrm{mod}\,1} \approx 1$.

5.3 An iterative estimation algorithm

This section contains a new method of phase estimation. Firstly, an iterative algorithm is given for going from confidence arcs for $\theta, (2\theta)_{\mathrm{mod}\,1}, (4\theta)_{\mathrm{mod}\,1}, \ldots, (2^{l-1}\theta)_{\mathrm{mod}\,1}$, of length $1/3$ and coverage probability at least $1 - \epsilon/l$, to a confidence arc for θ of length $1/(2^{l-1} \times 3)$ and coverage probability at least $1 - \epsilon$. Secondly, a method is given for obtaining a confidence arc for $(2^{k-1}\theta)_{\mathrm{mod}\,1}$, of length $1/3$ and coverage probability at least $1 - \epsilon/l$. Thirdly, one of Bernstein's inequalities is used to calculate the number of measurements needed at each stage. Finally, it is shown that it is possible to choose a value of ϵ such that $1 - \langle F(U_{\hat{\theta}}, U_\theta)\rangle = O((\log n/n)^2)$.

5.3.1 The iterative algorithm

First an intuitive approach is given using examples. For computational simplicity, confidence arcs of length 0.3 and coverage probability 1 will be considered. L_k and J_k will denote confidence arcs for $(2^{k-1}\theta)_{\mathrm{mod}\,1}$ and $2^{k-1}\theta$ respectively, of length 0.3 and coverage probability 1. (In the more general algorithm L_k and J_k will have length $1/3$ and coverage probability at least $1 - \epsilon/l$.) For the examples, $l = 3$.

Example 1
Suppose that after doing some measurements of U_θ, U_θ^2 and U_θ^4 it is found that

$$L_1 = [0.6, 0.9] \ni \theta \tag{5.10}$$
$$L_2 = [0.3, 0.6] \ni (2\theta)_{\mathrm{mod}\,1} \tag{5.11}$$
$$L_3 = [0.8, 1.1] \ni (4\theta)_{\mathrm{mod}\,1}. \tag{5.12}$$

It follows from (5.10) that

$$2L_1 = [1.2, 1.8] \ni 2\theta. \tag{5.13}$$

Using (5.11) and (5.13), it follows that

$$J_2 = [1.3, 1.6] \ni 2\theta. \tag{5.14}$$

From (5.14) it is known that

$$2J_2 = [2.6, 3.2] \ni 4\theta. \tag{5.15}$$

Using (5.12) and (5.15) gives

$$J_3 = [2.8, 3.1] \ni 4\theta. \tag{5.16}$$

Using confidence arcs (5.10), (5.11) and (5.12) for θ, $(2\theta)_{\mathrm{mod}\,1}$ and $(4\theta)_{\mathrm{mod}\,1}$ respectively, of length 0.3 and coverage probability 1, a confidence arc (5.16) has been derived for 4θ of length 0.3 and coverage probability 1. This gives a confidence arc for θ of length $0.3/2^{3-1} = 0.075$ and coverage probability 1, namely

$$(1/4)J_3 = [0.7, 0.775] \ni \theta. \tag{5.17}$$

Remember that confidence arcs on a circle are being considered. On the circle the arc $[1.2, 1.8]$ is equivalent to the arc $[0.2, 0.8]$, as are $[2.2, 2.8]$, $[3.2, 3.8]$ .... Similarly, $[2.6, 3.2]$ is equivalent to $[0.6, 1.2]$.

The symbol $\subset_1$ will be used to signify that a confidence arc on the circle is a subset of another confidence arc on the circle. Similarly, the symbol $\in_1$ will be used to signify that a point is contained within an arc on the circle, e.g. $0.3 \in_1 [1.2, 1.8]$. The previous example was rather simple in that $[0.3, 0.6] \subset_1 [1.2, 1.8]$ and $[0.8, 1.1] \subset_1 [2.6, 3.2]$.

Consider the following example for which, $L_{k+1} \not\subset_1 2J_k$. (Note that $L_1 = J_1$.)

Example 2
Suppose that after doing some measurements of U_θ, U_θ^2 and U_θ^4 it is found that

$$\begin{aligned} L_1 &= [0.1, 0.4] \ni \theta & (5.18)\\ L_2 &= [0.7, 1.0] \ni (2\theta)_{\mathrm{mod}\,1} & (5.19)\\ L_3 &= [0.9, 1.2] \ni (4\theta)_{\mathrm{mod}\,1}. & (5.20) \end{aligned}$$

It follows from (5.18) that

$$2J_1 = [0.2, 0.8] \ni 2\theta. \tag{5.21}$$

Now $L_2 \not\subset_1 2J_1$. From (5.19) and (5.21) it follows that

$$[0.7, 0.8] \ni 2\theta. \tag{5.22}$$

However, for simplicity, the confidence arcs J_k will be kept of equal length (in this example 0.3, in the more general algorithm 1/3). There is no unique

way to do this. A convenient way is to keep $J_k \subset 2J_{k-1}$ and J_k of length 0.3. Thus for this example the upper bound for 2θ remains as 0.8 and the lower bound is chosen to be $0.8 - 0.3 = 0.5$. This gives

$$J_2 = [0.5, 0.8] \ni 2\theta. \tag{5.23}$$

From (5.23) it follows that

$$2J_2 = [1.0, 1.6] \ni 4\theta. \tag{5.24}$$

Now, again $L_3 \not\subset_1 2J_2$. To keep $J_k \subset_1 2J_{k-1}$ and J_k of length 0.3, the lower bound remains as 1.0 and the upper bound becomes $1.0 + 0.3 = 1.3$,

$$J_3 = [1.0, 1.3] \ni 4\theta. \tag{5.25}$$

A confidence arc for $2^{3-1}\theta$ has been found of length 0.3 and coverage probability 1. This gives a confidence arc for θ of length $0.3/2^{3-1} = 0.075$ and coverage probability 1, namely

$$(1/4)J_3 = [0.25, 0.325] \ni \theta. \tag{5.26}$$

General Algorithm

The general algorithm will now be presented. Confidence arcs are now of length 1/3 rather than 0.3, and coverage probability 1,

$$L_k = [x(k), x(k) + 1/3], \qquad x(k) \in [0, 1) \tag{5.27}$$

$$J_k = [z(k), z(k) + 1/3]. \tag{5.28}$$

As in the examples, $2J_k$ and L_{k+1} are used to find a confidence arc J_{k+1}, with $J_{k+1} \subset 2J_k$. For $J_{k+1} \subset 2J_k$ it is required that $z(k+1) \in [2z(k), 2z(k)+1/3]$. Assuming that $J_k \ni 2^{k-1}\theta$ and $L_{k+1} \ni (2^k\theta)_{\mathrm{mod}\,1}$, there are three possibilities. For each possibility a figure is given (showing, for simplicity, a line instead of an arc), with a small vertical line representing the choice of the lower boundary $z(k+1)$ of J_{k+1}. Note that $J_1 = L_1$.

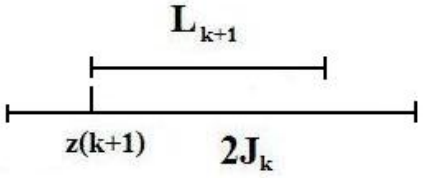


Figure 5.1: Situation (i)

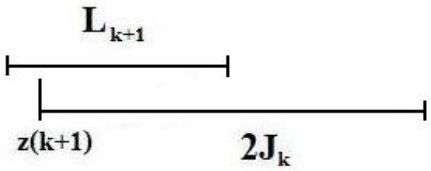


Figure 5.2: Situation (ii)

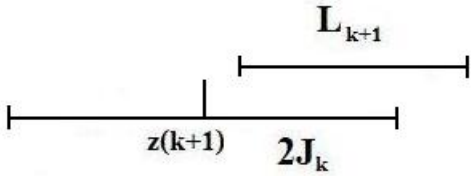


Figure 5.3: Situation (iii)

(i) The simplest possibility is that $L_{k+1} \subset_1 2J_k$. This occurs when

$$(x(k+1) - 2z(k))_{\mathrm{mod}\,1} \in [0, 1/3).$$

In this case the lower boundary of J_{k+1} is taken to be

$$z(k+1) = 2z(k) + (x(k+1) - 2z(k))_{\mathrm{mod}\,1}.$$

(ii) Another possibility is that $x(k+1) \notin_1 2J_k$ but $x(k+1) + 1/3 \in_1 2J_k$. This occurs when

$$(x(k+1) - 2z(k))_{\mathrm{mod}\,1} \in [2/3, 1).$$

In this case the lower boundary of J_{k+1} is taken to be

$$z(k+1) = 2z(k).$$

(iii) The final possibility is that $x(k+1) \in_1 2J_k$ but $x(k+1) + 1/3 \notin_1 2J_k$. This occurs when

$$(x(k+1) - 2z(k))_{\mathrm{mod}\,1} \in [1/3, 2/3).$$

In this case the lower boundary of J_{k+1} is taken to be

$$z(k+1) = 2z(k) + \frac{1}{3}.$$

This iterative scheme gives the confidence arc $J_l = [z(l), z(l) + 1/3]$ for $2^{l-1}\theta$ of length $1/3$ and coverage probability 1. This gives a confidence arc

for θ of length $1/(2^{l-1}\times 3)$, and coverage probability 1, namely $(1/2^{l-1})J_l = [z(l)/2^{l-1}, (z(l)+1/3)/2^{l-1}]$. The centre of this interval, modulo 1, is taken as the final estimate $\hat{\theta}$ of θ, i.e.

$$\hat{\theta} = \left(\frac{z(l)+1/6}{2^{l-1}}\right)_{\mathrm{mod}\,1}.$$

The final confidence arc for θ of length $1/(2^{l-1}\times 3)$ contains θ if $L_k \ni (2^{k-1}\theta)_{\mathrm{mod}\,1}$, for every $k = 1,\ldots,l$. If, for every $k = 1,\ldots,l$, L_k has coverage probability at least $1-\epsilon/l$, the coverage probability of the final confidence arc is at least $1-\epsilon$.

5.3.2 Finding L_k

The following function will be used:

$$\mathrm{atan2}(x,y) = \begin{cases} \arctan(y/x) & x>0, \\ \arctan(y/x)+\pi & x<0, \quad y\geq 0, \\ \arctan(y/x)-\pi & x<0, \quad y<0, \\ \pi/2 & x=0, \quad y>0, \\ -\pi/2 & x=0, \quad y<0, \\ \text{undefined} & x=0, \quad y=0. \end{cases}$$

Here, details are given for calculating confidence arcs L_k for $(2^{k-1}\theta)_{\mathrm{mod}\,1}$ of length $1/3$ and coverage probability at least $1-\epsilon/l$. First it will be shown how to compute a confidence arc of length $1/3$, then, how to make the coverage probability at least $1-\epsilon/l$.

The problem of finding a confidence arc for θ will be considered. The analysis is exactly the same as for $(2^{k-1}\theta)_{\mathrm{mod}\,1}$, except that in the latter case the experimenter lets U_θ act 2^{k-1} times on the same $|\psi_x\rangle$.

The experimenter lets U_θ act on $|\psi_x\rangle$ and then measures in x. Outcome 0 is observed with probability $p_x(0;\theta) = (1+\cos(2\pi\theta))/2$. The state $U_\theta|\psi_x\rangle$ is measured in x a total of N times and outcome 0 is observed $N_{x=0}$ times. This gives an estimate $2N_{x=0}/N - 1$ of $\cos(2\pi\theta)$.

The experimenter lets U_θ act on $|\psi_x\rangle$ and measures in y. Outcome 0 is observed with probability $p_y(0;\theta) = (1+\sin(2\pi\theta))/2$. The state $U_\theta|\psi_x\rangle$ is measured in y a total of N times and outcome 0 is observed $N_{y=0}$ times. This gives an estimate $2N_{y=0}/N - 1$ of $\sin(2\pi\theta)$. Estimates of $\sin(2\pi\theta)$ and $\cos(2\pi\theta)$ give the estimate

$$\hat{\theta}_1 = \frac{1}{2\pi}\left(\mathrm{atan2}\left(\frac{2N_{y=0}}{N}-1, \frac{2N_{x=0}}{N}-1\right)\right)_{\mathrm{mod}\,2\pi} \tag{5.29}$$

of θ. The confidence arc is

$$L_1 = ((\hat{\theta}_1 - 1/6)_{\mathrm{mod}\,1}, (\hat{\theta}_1 - 1/6)_{\mathrm{mod}\,1} + 1/3). \tag{5.30}$$

More generally, an estimate $(2^{k-1}\hat{\theta}_k)_{\mathrm{mod}\,1}$ of $(2^{k-1}\theta)_{\mathrm{mod}\,1}$ gives the confidence arc

$$L_k = \left(\left((2^{k-1}\hat{\theta}_k)_{\mathrm{mod}\,1} - 1/6\right)_{\mathrm{mod}\,1}, \left((2^{k-1}\hat{\theta}_k)_{\mathrm{mod}\,1} - 1/6\right)_{\mathrm{mod}\,1} + 1/3 \right). \tag{5.31}$$

It is necessary to find the accuracy needed for the estimates of $p_x(0;\theta)$ and $p_y(0;\theta)$ to ensure that $|\hat{\theta} - \theta|_1 \leq 1/6$, and hence $L_1 \ni \theta$.

Put $x = \cos(2\pi\theta)$, $y = \sin(2\pi\theta)$, $x_0 = 2N_{x=0}/N - 1$, $y_0 = 2N_{y=0}/N - 1$ and $\phi(x,y) = \mathrm{atan}(y,x)$. Define

$$|\hat{\phi} - \phi|_{2\pi} = \min\left((\hat{\phi} - \phi)_{\mathrm{mod}\,2\pi}, (\phi - \hat{\phi})_{\mathrm{mod}\,2\pi}\right). \tag{5.32}$$

Given that

$$|x - x_0| \leq \alpha, \tag{5.33}$$
$$|y - y_0| \leq \alpha, \tag{5.34}$$

an upper bound is sought on $|\phi(x,y) - \phi(x_0,y_0)|_{2\pi}$. This will be done in steps.
(i)

$$\begin{aligned} |\phi(x,y) - \phi(x_0,y_0)|_{2\pi} &= |[\phi(x,y) - \phi(x,y_0)] \\ &+ [\phi(x,y_0) - \phi(x_0,y_0)]|_{2\pi} \\ &\leq |\phi(x,y) - \phi(x,y_0)|_{2\pi} \\ &+ |\phi(x,y_0) - \phi(x_0,y_0)|_{2\pi}. \end{aligned} \tag{5.35}$$

Put

$$\begin{aligned} \psi_1 &= |\phi(x,y) - \phi(x,y_0)|_{2\pi} \\ \psi_2 &= |\phi(x,y_0) - \phi(x_0,y_0)|_{2\pi}. \end{aligned}$$

(ii) Consider the triangle T_1 given by the points $(0,0), (x,y)$ and (x,y_0), with y_0 satisfying (5.34). The angle at the point $(0,0)$ is ψ_1, and is opposite a side of length $|y - y_0|$. The angle, say ψ_A, at the point (x,y_0) will be opposite a side of length 1. Using the sine rule for T_1 gives

$$\frac{\sin\psi_1}{|y - y_0|} = \frac{\sin\psi_A}{1}. \tag{5.36}$$

For any triangle the angles add up to π. The largest angle will be opposite the longest side. For any angle, ψ^* say, not opposite the longest side, $\psi^* \in [0, \pi/2]$. If $\alpha \leq 1/2$ then from (5.34) $|y - y_0| \leq 1$ and so $\psi_1 \in [0, \pi/2]$. Thus $\psi_1 \leftrightarrow \sin\psi_1$, and hence $\psi_1 = \arcsin\beta$, with $\beta = |y - y_0| \times \sin\psi_A$. As $\psi_A \in [0, \pi]$, consequently $\sin\psi_A \in [0, 1]$, and using (5.34) it follows that $\beta \in [0, \alpha]$. Since arcsin is a monotone function on $[0, \alpha]$, it follows that

$$\psi_1 \leq \arcsin(\alpha). \tag{5.37}$$

(iii) Consider the triangle T_2 given by the points $(0, 0), (x, y_0)$ and (x_0, y_0), with x_0 and y_0 satisfying (5.33) and (5.34) respectively. The angle at the point $(0, 0)$ is ψ_2 and is opposite a side of length $|x - x_0|$. The angle, say ψ_B, at the point (x_0, y_0) is opposite a side of length r, where

$$\begin{aligned} r &= \sqrt{x^2 + y_0^2} \\ &\geq \min_{\Delta} \sqrt{x^2 + (y + \Delta)^2}, \quad \Delta \in [-\alpha, \alpha] \\ &= \min_{\Delta} \sqrt{x^2 + y^2 + 2y\Delta + \Delta^2} \\ &= \min_{\Delta} \sqrt{1 + 2y\Delta + \Delta^2} \\ &\geq \min_{\Delta} \sqrt{1 - 2|\Delta| + |\Delta|^2} \\ &= \min_{\Delta} \quad 1 - |\Delta| \\ &= 1 - \alpha. \end{aligned} \tag{5.38}$$

Using the sine rule for T_2 gives

$$\frac{\sin\psi_2}{|x - x_0|} = \frac{\sin\psi_B}{r}. \tag{5.39}$$

If $\alpha \leq 1/2$ then $\alpha \leq 1 - \alpha$ and so from (5.33) and (5.38), $|x - x_0| \leq r$. It follows that $\psi_2 \in [0, \pi/2]$ and so $\psi_2 \leftrightarrow \sin\psi_2$. Using (5.33), (5.38), (5.39) and monoticity of arcsin on $[0, 1]$ gives

$$\psi_2 \leq \arcsin\left(\frac{\alpha}{1 - \alpha}\right). \tag{5.40}$$

Theorem 5.1 *Given (5.33) and (5.34) for $\alpha \leq 1/2$,*

$$|\phi(x, y) - \phi(x_0, y_0)|_{2\pi} \leq \arcsin(\alpha) + \arcsin\left(\frac{\alpha}{1 - \alpha}\right). \tag{5.41}$$

Proof. This follows from (5.35),(5.37) and (5.40).

For the iterative algorithm it is required that $|\hat{\theta} - \theta|_1 \leq 1/6$, which is equivalent to

$$|\phi(x, y) - \phi(x_0, y_0)|_{2\pi} \leq \frac{\pi}{3}. \tag{5.42}$$

If $\alpha = 0.3794$ then (5.42) holds, and (5.33) and (5.34) are equivalent to

$$\begin{aligned} |N_{x=0}/N - p_x(0;\theta)| &\leq 0.1897, & (5.43)\\ |N_{y=0}/N - p_y(0;\theta)| &\leq 0.1897. & (5.44) \end{aligned}$$

It follows that if

$$\Pr\Big(|N_{x=0}/N - p_x(0;\theta)| \leq 0.1897\Big) \geq \sqrt{1 - \epsilon/l} \tag{5.45}$$

and

$$\Pr\Big(|N_{y=0}/N - p_y(0;\theta)| \leq 0.1897\Big) \geq \sqrt{1 - \epsilon/l}, \tag{5.46}$$

then

$$\Pr\left(L_1 \ni \theta\right) \geq 1 - \epsilon/l. \tag{5.47}$$

An analogous result holds for $L_k, k = 2, \ldots, l$. In Section 5.3.3 it is shown that if $N = 24.437 \log(4l/\epsilon)$ then (5.45) and (5.46) hold.

5.3.3 Number of measurements needed

The following Bernstein inequality (Hazewinkel, 2002) will be used:

Theorem 5.2 *If the equations*

$$E[Y_j] = 0, \quad E[Y_j^2] = b_j, \quad j = 1, \ldots, n,$$

hold for the independent random variables $Y_1, \ldots, Y_n$ *with*

$$E[|Y_j|^l] \leq \frac{b_j}{2} H^{l-2} l! \tag{5.48}$$

(where $l > 2$ *and* H *is a constant independent of* j*), then the following inequality holds for the sum* $S_n = \sum_{j=1}^n Y_j$*:*

$$\Pr(|S_n| > r) \leq 2 \exp\left(-\frac{r^2}{2(B_n + Hr)}\right), \tag{5.49}$$

where $B_n = \sum_{j=1}^n b_j$*.*

The observed measurement outcomes from a single measurement in x have distribution and moments

$$X_j \sim \text{Bin}(1, p), \quad E[X_j] = p, \quad E[X_j^2] = p(1-p),$$

where $p = (1 + \cos(2\pi\theta))/2$. Put $b_j = p(1-p)$ for $j = 1, \ldots, N$ and consider the random variable $R_j = X_j - p$, which has moments

$$E[R_j] = 0, \quad E[R_j^2] = p(1-p) = b_j.$$

Now, for $l > 2$,

$$\begin{aligned} E[|R_j|^l] &= p|1-p|^l + (1-p)|0-p|^l \\ &\leq p(1-p)^2 + (1-p)p^2 \\ &= p(1-p) \\ &= b_j. \end{aligned} \tag{5.50}$$

Thus, comparing (5.50) with (5.48), $H = 1$ is a suitable choice. Substituting $B_N = \sum_{j=1}^N b_j = Np(1-p)$ and $S_N = \sum_{j=1}^N R_j = N_{x=0} - Np$ into (5.49) gives

$$\Pr(|N_{x=1} - Np| > r) \leq 2\exp\left(-\frac{r^2}{2(Np(1-p)+r)}\right).$$

Putting $r = N\delta$, gives

$$\begin{aligned} \Pr(|N_{x=1}/N - p| > \delta) &\leq 2\exp\left(-\frac{N\delta^2}{2(p(1-p)+\delta)}\right) \\ &\leq 2\exp\left(-\frac{N\delta^2}{2(1/4+\delta)}\right). \end{aligned} \tag{5.51}$$

The inequality $\Pr(|N_{x=0}/N - p| < \delta) \geq \sqrt{1-\epsilon/l}$, is equivalent to the inequality $\Pr(|N_{x=0}/N - p| > \delta) \leq 1 - \sqrt{1-\epsilon/l}$, which holds if $\Pr(|N_{x=0}/N - p| > \delta) \leq \epsilon/(2l)$. Substituting $\delta = 0.1897$ into (5.51), it can be found that (5.45) holds if

$$N = 24.437 \ln(4l/\epsilon) \tag{5.52}$$

measurements in x are performed at each stage. The analysis is exactly the same for measurements in y, and so a total number of

$$N_{tot} = 48.874 \ln(4l/\epsilon) \tag{5.53}$$

measurements are required at each stage. This ensures that (5.45) and (5.46) hold, and consequently (5.47) holds.

5.3.4 The behaviour of the fidelity

The behaviour of $1-\langle F(U_{\hat{\theta}}, U_\theta)\rangle$ will be analysed as a function of the number n of times U_θ is used. As in Rudolph and Grover (2003), the worst–case value of $1-\langle F(U_{\hat{\theta}}, U_\theta)\rangle$ will be sought. That is, if the final confidence arc does not contain θ then $\hat{\theta} = (\theta + 1/2)_{\mathrm{mod}\,1}$, and if it does then θ lies on the boundary of the confidence arc, i.e. $|\hat{\theta} - \theta|_1 = 1/(2^l \times 3)$. This gives

$$\begin{aligned} 1 - \langle F(U_{\hat{\theta}}, U_\theta)\rangle &\leq 1 - \left((1-\epsilon)\frac{1+\cos(2\pi/(2^l\times 3))}{2} + \epsilon \times 0\right) \\ &\approx \epsilon + \frac{\pi^2}{2^{2l}\times 9} - \frac{\epsilon\pi^2}{2^{2l}\times 9}. \end{aligned}$$

If $\epsilon = 1/2^{2l}$, then $1 - \langle F(U_{\hat{\theta}}, U_\theta)\rangle = O(1/2^{2l})$. This requires a total of

$$N_{tot} = 48.874 \log(4l \times 2^{2l}) \tag{5.54}$$

measurements at each stage. The number of times U_θ is used is $n = N_{tot}(2^l - 1)$, and so $1/2^l \approx N_{tot}/n$. The number of measurements, (5.54), made at each stage is $O(l)$; noticing that $\log n$ is also $O(l)$, it follows that

$$1 - \langle F(U_{\hat{\theta}}, U_\theta)\rangle = O\left(\left(\frac{\log n}{n}\right)^2\right). \tag{5.55}$$

5.4 Simulations

The analysis in Section 5.3.4 concentrated on optimizing the worst-case asymptotic scaling of $1 - \langle F\rangle$ with respect to n. Here a more pragmatic line will be taken. Of interest is the minimum number of measurements needed such that the final confidence arc contains θ a satisfactory proportion of the time.

The iterative algorithm will now be investigated using simulations with the computer package MAPLE. A value for the parameter $\theta \in [0, 1)$ is given by a random variable with a uniform distribution. Measurement results can be simulated, since the number of times outcome 0 is observed has a Binomial distribution. For example, at the kth iterative stage, measuring in x, $N_{x=0} \sim \mathrm{Bin}(N, (1 + \cos(2^k\pi\theta))/2)$. From the simulated results of measurements in x and y for stages $1, \ldots, l$, an estimate of θ is obtained using the iterative algorithm given in section 5.3.1. It can then be checked whether the final confidence arc contains θ. This is done for $100,000$ randomly chosen values of θ, and the number of times the final interval contains θ is recorded.

For most recent iterative schemes the total number of iterations is reasonably small: 6 in Higgins *et al.* (2007) and 7 in Liu *et al.* (2007). Simulations were performed with the number of iterations varying between 6 and 9. Table 5.1 gives the number of times the final confidence arc contains the true value of θ.

	Number of iterative stages (l)			
N_{tot}	6	7	8	9
20	99,792	99,729	99,747	99,712
30	99,993	99,987	99,982	99,978
40	99,999	100,000	99,998	99,999
50	100,000	100,000	99,999	100,000

Table 5.1: Numbers of trials out of 100,000 with $|\hat{\theta} - \theta| \leq 1/(2^l \times 3)$.

It seems a waste to use $N_{tot} = 48.874 \log(2l \times 2^{2l})$ measurements at each stage, since the simulations suggest that for practical purposes it is sufficient to use fewer measurements – even as few as 20 or 30.

5.4.1 Estimating the coverage probability

Using the above simulations the coverage probability can be estimated, i.e. the probability that, using the iterative algorithm, the known true value θ is contained in the final confidence interval.

Suppose the true (unknown) coverage probability is p. For the ith trial put

$$\begin{aligned} W_i &= 1 \quad \text{if interval covers } \theta \\ &= 0 \quad \text{if not.} \end{aligned}$$

Then $W_1, \ldots, W_M$ are independent identically distributed Bernoulli random variables, i.e. $W_i \sim \text{Bin}(1, p)$. Thus

$$W_1 + \cdots + W_M \sim \text{Bin}(M, p).$$

If m out of M intervals cover θ then p is estimated by m/M. An approximate 95% confidence interval for p is

$$\frac{m}{M} \pm 1.96 \sqrt{\frac{\frac{m}{M}\left(1 - \frac{m}{M}\right)}{M}}.$$

The longest confidence interval (0.00066) is that for using 9 iterative stages and a total of 20 measurements at each stage. Using the half-length of this confidence interval, the confidence interval

$$\frac{m}{100,000} \pm 0.00033$$

can be computed from the results given in Table 5.1. It has coverage probability at least 95%.

5.5 The noisy case

It is known that when even a small amount of noise is present the performance of phase estimation schemes is greatly reduced (Huelga *et al.*, 1997, Shaji and Caves, 2007).

This section investigates the performance of the iterative estimation algorithm when depolarizing noise is present. The channel

$$\rho_0 \mapsto (1-r)U_\theta \rho_0 U_\theta^\dagger + \frac{r}{2}\mathbb{I}_2, \qquad 0 < r < 1, \tag{5.56}$$

is considered, where U_θ is the same as before, (5.1), and $\rho_0 = |\psi_x\rangle\langle\psi_x|$. (The channel (5.56) is identical to $U_\theta\rho_0 U_\theta^\dagger$ undergoing phase damping with $\lambda = r(2-r)$ (Nielsen and Chuang, 2000, p. 383).) Ji *et al.* (2008) gave the very interesting result that if $r > 0$, then the optimal asymptotic rate at which $1 - \langle F(U_{\hat{\theta}}, U_\theta)\rangle$ approaches zero is given by the standard quantum limit.

Defining n' as the maximum number of times the experimenter lets U_θ act on the same input state, Ji *et al.* (2008) argued that if $(1-r)^{n'}$ is close to 1, and thus $n'r << 1$, then it is still possible to estimate θ as before with the rate $O((\log n/n)^2)$.

The whole point of using an iterative scheme is that the distinguishability of θ from $\cos(n2\pi\theta)$, with $n >> 1$, is considerably greater than from $\cos(2\pi\theta)$.

To measure distinguishability, the quantity F_θ^M/m will be used, where m is the number of times U_θ acts on the same input state. This is because of interest is to maximize the distinguishability of θ per use of the channel.

If there is no noise, and the experimenter lets U_θ act m times on the input state and measures in x, then outcome 0 is observed with probability $p(0;\theta) = (1+\cos(m2\pi\theta))/2$ and 1 with probability $p(1;\theta) = 1 - p(0;\theta)$. The Fisher information from this measurement is $F_\theta^{M_x} = 4\pi^2 m^2$, which is equal to the SLD quantum information. Measuring in y gives the same Fisher information. Thus $F_\theta^{M_x}/m = F_\theta^{M_y}/m = 4\pi^2 m$. At the kth stage of the

iterative procedure, the experimenter lets U_θ act $m = 2^{k-1}$ times on the input state, and so $F_\theta^{M_x}/m = F_\theta^{M_y}/m = \pi^2 2^{k+1}$. Thus F_θ^M/m (where M is an arbitrary measurement in x or y) increases exponentially with k.

In the noisy case, when the experimenter lets U_θ act m times on the output state and then measures in x, outcome 0 is observed with probability $p(0;\theta) = (1 + (1-r)^m \cos(m2\pi\theta))/2$ and 1 with probability $p(1;\theta) = 1 - p(0;\theta)$. Measuring in y, outcome 0 is observed with probability $p(0;\theta) = (1 + (1-r)^m \sin(m2\pi\theta))/2$ and 1 with probability $p(1;\theta) = 1 - p(0;\theta)$. This gives

$$\begin{aligned} F_\theta^{M_x} &= \frac{4\pi^2 m^2 (1-r)^{2m} \sin^2(2m\pi\theta)}{1 - (1-r)^{2m} \cos^2(2m\pi\theta)} \\ F_\theta^{M_y} &= \frac{4\pi^2 m^2 (1-r)^{2m} \cos^2(2m\pi\theta)}{1 - (1-r)^{2m} \sin^2(2m\pi\theta)} \\ H_\theta &= 4\pi^2 m^2 (1-r)^{2m}. \end{aligned}$$

Notice that

$$F_\theta^{M_x} + F_\theta^{M_y} \approx H_\theta.$$

Thus measuring both in x and y, the average Fisher information from a single measurement M is approximately $H_\theta/2$.

The maximal value of F_θ^M/m, taken over m, will occur close to the maximal value of H_θ/m. When $r > 0$, H_θ/m, and hence F_θ^M/m, does not increase indefinitely with m. Instead it reaches its maximum at

$$m = -\frac{1}{2\log(1-r)}, \tag{5.57}$$

after which it decreases. When r is small, this maximum is obtained at

$$m \approx \frac{1}{2r}. \tag{5.58}$$

Thus in the noisy case the number of iterative stages that should be performed is limited by the amount of noise. The number of stages that can be performed, for small r, such that H_θ/m, and hence F_θ^M/m, increases at each stage is approximately $l \approx -\log_2 r$.

Figures 5.4 – 5.8 give H_θ/m at the kth iterative stage. It can be seen that H_θ/m increases up to $k = -\log_2 r$, decreases slightly near $k = -\log_2 r + 1$ and falls rapidly for $k > -\log_2 r + 1$.

Tables 5.2 – 5.7 contain the results of simulations, for magnitudes of noise $r = 2^{-4}, 2^{-5}, \ldots, 2^{-8}$ and total number of iterative stages $l = 4, \ldots, 9$ – the number of measurements at each stage is fixed. Consider the diagonals of

Tables 5.2 – 5.7, from $r = 2^{-4}$, $l = 4$ to $r = 2^{-8}$, $l = 8$. This corresponds to the experimenter performing $l = -\log_2 r$ iterative stages, which involves going up to the iterative stage at which F_θ^M/m is maximized. Similarly, the diagonal from $r = 2^{-4}$, $l = 5$ to $r = 2^{-8}$, $l = 9$ corresponds to the experimenter performing $l = -\log_2 r + 1$ iterative stages etc. It is interesting to note that when $l > -\log_2 r$, there is a significant decrease in the number of confidence intervals containing θ, and when $l > -\log_2 r + 1$, an even greater decrease in the number of confidence intervals containing θ. For example, using 30 measurements at each stage, if the experimenter performs $l = -\log_2 r$ iterative stages then the final confidence interval contains θ approximately 98% of the time; if the experimenter increases to $l = -\log_2 r +$ 1 iterative stages, then the final confidence interval contains the true value of θ approximately 89% of the time. If the experimenter increases to $l = -\log_2 r + 2$ iterative stages, then approximately 61% of the time the final confidence interval contains θ – a considerable drop in performance. It can be seen from Table 5.7, for which 200 measurements are performed at each stage, that this drop in performance does not just occur when performing relatively small numbers of measurements at each stage.

It is interesting to see that the drop off in performance, in terms of the coverage probability – which can be calculated from Tables 5.2 – 5.7, occurs at the same point as the drop in performance as measured by H_θ/m, and consequently F_θ/m – seen in Figures 5.4 – 5.8.

Since F_θ^M/m starts to decrease after $l = -\log_2 r$ iterative stages, it makes no sense to choose $l > -\log_2 r$. The simulations also suggest that it is safer to do no more than $l = -\log_2 r$ iterative stages. This is equivalent to letting U_θ act no more than $n' = 1/(2r)$ times on the same input state. Thus for a given level of noise the experimenter can let U_θ act on an input state more times than n' satisfying $n'r << 1$ (though the $O((\log n/n)^2)$ rate may not be kept). A sensible suggestion is, more generally, that for the channel (5.56) the optimum number of iterative stages, where at the kth stage U_θ is used 2^{k-1} times, is $l = \lfloor -\log_2 r \rfloor$.

A related question was considered in Rubin and Kaushik (2007), where the 'stopping point', was N the number of entangled photons to be included in the NOON input states. Rubin and Kaushik found that the optimal precision in measurement occurred for $N = 1.279/L$, where L is the magnitude of loss (analogous to the point, $n' = 1/(2r)$, at which F_θ^M/m is maximized).

If $l = -\log_2 r$ iterative stages are performed and the final confidence interval contains θ, this corresponds to a precision $|\hat{\theta} - \theta|_1 \leq r/3$. If the experimenter desires greater precision in his final estimate than $|\hat{\theta} - \theta|_1 \leq r/3$, then it seems sensible for him to perform more measurements at the final iterative stage.

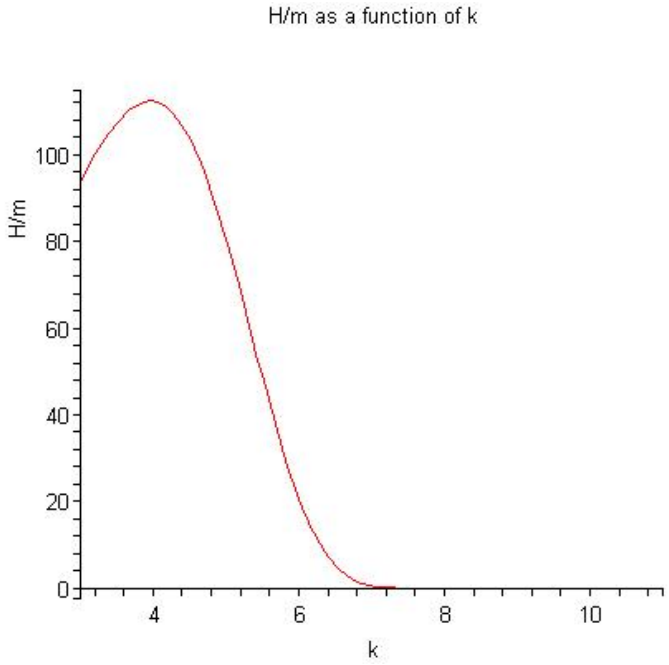


Figure 5.4: H_θ/m at the kth iterative stage, with $r = 2^{-4}$.

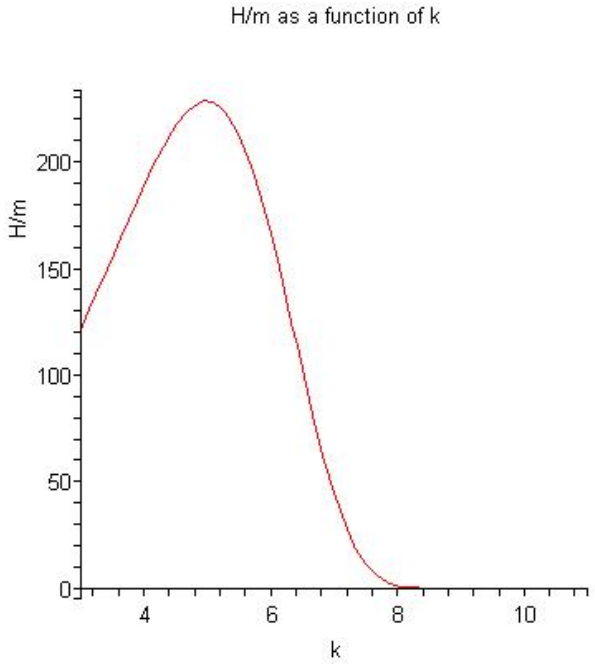


Figure 5.5: H_θ/m at the kth iterative stage, with $r = 2^{-5}$.

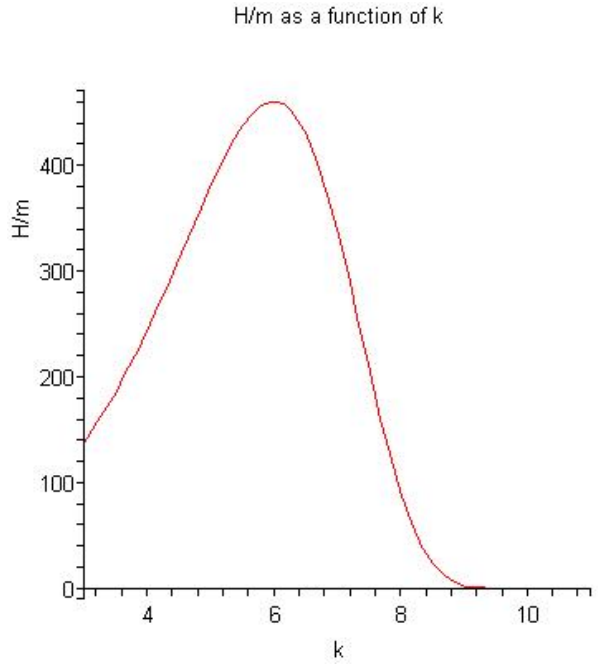


Figure 5.6: H_θ/m at the kth iterative stage, with $r = 2^{-6}$.

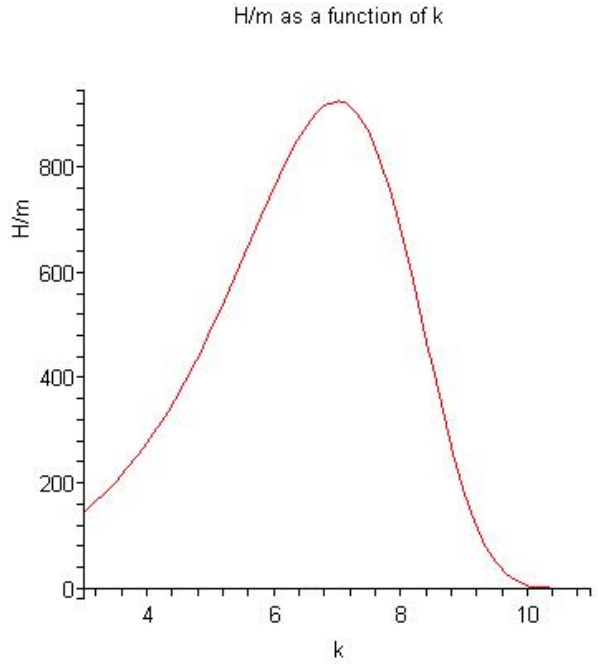


Figure 5.7: H_θ/m at the kth iterative stage, with $r = 2^{-7}$.

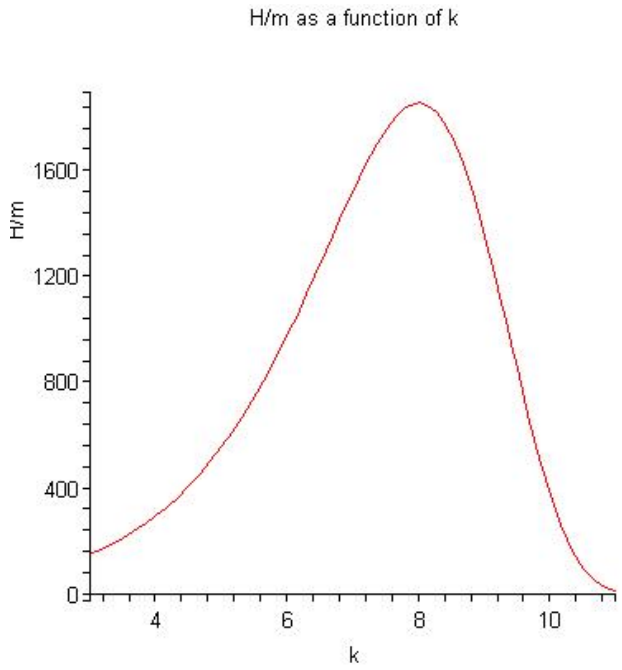


Figure 5.8: H_θ/m at the kth iterative stage, with $r = 2^{-8}$.

	Number of iterative stages (l)					
r	4	5	6	7	8	9
2^{-4}	94,601	78,896	48,831	24,429	11,514	5,430
2^{-5}	98,804	94,854	79,207	49,625	24,891	11,738
2^{-6}	99,608	98,728	94,840	79,428	50,121	24,887
2^{-7}	99,779	99,571	98,768	94,917	79,544	50,130
2^{-8}	99,823	99,719	99,571	98,764	94,715	79,745

Table 5.2: Numbers of trials out of 100,000 with $|\hat{\theta} - \theta|_1 \leq 1/(2^l \times 3)$, with $N_{tot} = 20$.

	Number of iterative stages (l)					
r	4	5	6	7	8	9
2^{-4}	98,290	88,340	60,423	32,445	16,059	8,042
2^{-5}	99,804	98,408	88,537	61,293	32,756	16,460
2^{-6}	99,967	99,807	98,430	88,708	61,148	32,595
2^{-7}	99,985	99,955	99,802	98,476	88,895	61,699
2^{-8}	99,988	99,977	99,962	99,812	98,467	88,864

Table 5.3: Numbers of trials out of 100,000 with $|\hat{\theta} - \theta|_1 \leq 1/(2^l \times 3)$, with $N_{tot} = 30$.

r	Number of iterative stages (l)					
	4	5	6	7	8	9
2^{-4}	99,336	91,501	63,433	33,429	16,130	7,940
2^{-5}	99,972	99,349	91,753	64,469	34,079	16,262
2^{-6}	99,999	99,962	99,391	92,139	64,768	33,861
2^{-7}	99,999	99,993	99,960	99,388	92,190	64,744
2^{-8}	99,998	99,999	99,997	99,957	99,371	92,287

Table 5.4: Numbers of trials out of 100,000 with $|\hat{\theta} - \theta|_1 \leq 1/(2^l \times 3)$, with $N_{tot} = 40$.

r	Number of iterative stages (l)					
	4	5	6	7	8	9
2^{-4}	99,741	94,475	68,789	37,529	18,793	9,308
2^{-5}	99,991	99,738	94,644	69,626	37,976	19,139
2^{-6}	99,999	99,993	99,759	94,909	70,021	38,232
2^{-7}	100,000	99,998	99,995	99,770	95,030	70,402
2^{-8}	100,000	100,000	99,999	99,995	99,790	94,983

Table 5.5: Numbers of trials out of 100,000 with $|\hat{\theta} - \theta|_1 \leq 1/(2^l \times 3)$, with $N_{tot} = 50$.

r	Number of iterative stages (l)					
	4	5	6	7	8	9
2^{-4}	99,993	98,780	78,515	43,641	21,599	10,983
2^{-5}	100,000	99,994	98,924	79,739	44,374	22,150
2^{-6}	100,000	100,000	99,999	98,904	79,899	44,762
2^{-7}	100,000	100,000	100,000	99,997	98,966	80,004
2^{-8}	100,000	100,000	100,000	100,000	99,998	98,989

Table 5.6: Numbers of trials out of 100,000 with $|\hat{\theta} - \theta|_1 \leq 1/(2^l \times 3)$, with $N_{tot} = 100$.

	Number of iterative stages (l)					
r	4	5	6	7	8	9
2^{-4}	100,000	99,907	87,516	50,576	25,592	12,868
2^{-5}	100,000	100,000	99,937	88,393	51,414	26,064
2^{-6}	100,000	100,000	100,000	99,946	88,754	52,022
2^{-7}	100,000	100,000	100,000	100,000	99,953	88,932
2^{-8}	100,000	100,000	100,000	100,000	100,000	99,938

Table 5.7: Numbers of trials out of 100,000 with $|\hat{\theta} - \theta|_1 \leq 1/(2^l \times 3)$, with $N_{tot} = 200$.

Appendix A

Notation

notation	definition
$\|\psi\rangle$	finite dimensional complex column vector of unit length (see (1.1)).
$\langle\psi\|$	dual of $\|\psi\rangle$ (see (1.2)).
$\langle\psi\|\phi\rangle$	inner product of $\|\psi\rangle$ and $\|\phi\rangle$ (see (1.3)).
$\|\phi\rangle\langle\psi\|$	outer product of $\|\psi\rangle$ and $\|\phi\rangle$ (see (1.6)).
$\|0\rangle$	$(1,0)^T$ (T denotes transpose).
$\|1\rangle$	$(0,1)^T$.
z^*	complex conjugate of z.
ρ	density matrix (see Section 1.3).
$\mathbb{I}$	identity matrix.
$\mathbb{I}_d$	$d \times d$ identity matrix.
$\sigma_x, \sigma_y, \sigma_z$	Pauli matrices (1.15).
$\sigma_1, \sigma_2, \sigma_3$	Pauli matrices.
M	POVM (see Section 1.4).

notation	definition
M_m	element of POVM M corresponding to outcome m.
$\mathcal{H}$	Hilbert space.
$\mathcal{H}_{A,B}$	$\mathcal{H}_A \otimes \mathcal{H}_B$ an extended Hilbert space.
$A^\dagger$	Hermitian transpose of A (see Section 1.2).
$\|\phi^A\phi^B\rangle$	$\|\phi^A\rangle \otimes \|\phi^B\rangle$.
$S(\mathcal{H})$	set of states on the Hilbert space $\mathcal{H}$.
ρ_A	reduced state on $S(\mathcal{H}_A)$ (see Section 1.5.1).
E_k	Kraus operator (see Section 1.7).
U	unitary matrix (any matrix satisfying $UU^\dagger = U^\dagger U = \mathbb{I}$).
$\mathcal{E}$	quantum channel (see Section 1.7).
F_θ	Fisher information (See Section 1.38).
F_θ^M	Fisher information from single measurement using M (See section 1.38).
$\rightsquigarrow$	converges in distribution to.
$\langle F \rangle$	expectation of F.
ρ_θ	parameterized family of states.
λ_{SLD}, λ, λ_θ	SLD quantum score (see (1.52)).
$C_E(\theta)$	Sarovar and Milburn's bound based on arbitrary set of Kraus operators $E = \{E_k\}$ (see 2.3).
Υ_k	canonical Kraus operator (see before (2.4)).
$C_\Upsilon(\theta)$	Sarovar and Milburn's bound based on canonical Kraus operators $\{\Upsilon_k\}$.
$C_L(\theta)$	metric derived from C_Υ (see (3.15)).
H_θ^S, H_θ	SLD quantum information (see Section 1.9).
$F(\hat{U}, U)$	fidelity between $\hat{U}$ and U (see (1.90)).

Index

Bibliography

Acín, A., Jané, E. and Vidal, G. (2001). Optimal estimation of quantum dynamics, *Phys. Rev. A* **64**(5): 050302.

Amari, S. (1982). Geometrical theory of asymptotic ancillarity and conditional inference, *Biometrika* **69**(1): 1–17.

Amari, S. and Nagaoka, H. (2000). *Methods of Information Geometry*, Oxford: Oxford University Press.

Aspuru-Guzik, A., Dutoi, A. D., Love, P. and Head-Gordon, M. (2005). Simulated Quantum Computation of Molecular Energies, *Science* **309**(5741): 1704–1707.

Bagan, E., Baig, M. and Muñoz-Tapia, R. (2004a). Entanglement-assisted alignment of reference frames using a dense covariant coding, *Phys. Rev. A* **69**(5): 050303.

Bagan, E., Baig, M. and Muñoz-Tapia, R. (2004b). Quantum reverse engineering and reference-frame alignment without nonlocal correlations, *Phys. Rev. A* **70**(3): 030301.

Bagan, E., Ballester, M., Gill, R., Monras, A. and Muñoz Tapia, R. (2006). Optimal full estimation of qubit mixed states, *Phys. Rev. A* **73**(3): 032301.

Ballester, M. A. (2004a). Entanglement is not very useful for estimating multiple phases, *Phys. Rev. A* **70**(3): 032310.

Ballester, M. A. (2004b). Estimation of unitary quantum operations, *Phys. Rev. A* **69**(2): 022303.

Ballester, M. A. (2005). *Estimation of Quantum States and Operations*, Ph.D. thesis, University of Utrecht.

Barndorff-Nielsen, O. E. and Gill, R. D. (2000). Fisher information in quantum statistics, *J. Phys. A* **33**(24): 4481–4490.

Barndorff-Nielsen, O. E., Gill, R. D. and Jupp, P. E. (2003). On quantum statistical inference, *J. R. Statist. Soc. B* **65**(4): 775–804.

Bengtsson, I. and Życzkowski, K. (2006). *Geometry of Quantum States: An Introduction to Quantum Entanglement*, Cambridge: Cambridge University Press.

Bennett, C. H., DiVincenzo, D. P., Fuchs, C. A., Mor, T., Rains, E., Shor, P. W., Smolin, J. A. and Wootters, W. K. (1999). Quantum nonlocality without entanglement, *Phys. Rev. A* **59**(2): 1070–1091.

Braunstein, S. L. and Caves, C. M. (1994). Statistical distance and the geometry of quantum states, *Phys. Rev. Lett.* **72**(22): 3439–3443.

Busch, P., Grabowski, M. and Lahti, P. J. (1995). *Operational Quantum Physics*, Springer.

Childs, A., Preskill, J. and Renes, J. (2000). Quantum information and precision measurement, *J. Modern Optics* **47**(22): 155–176.

Chiribella, G., D'Ariano, G. M., Perinotti, P. and Sacchi, M. F. (2004). Efficient use of quantum resources for the transmission of a reference frame, *Phys. Rev. Lett.* **93**(18): 180503.

Choi, M. (1975). Completely positive linear maps on complex matrices, *Linear Algebra and its Applications* **10**: 285–290.

Cochran, W. (1973). Experiments for nonlinear functions, *J. Amer. Statist. Assoc.* **68**: 771–781.

de Burgh, M. and Bartlett, S. D. (2005). Quantum methods for clock synchronization: Beating the standard quantum limit without entanglement, *Phys. Rev. A* **72**(4): 042301.

de Martini, F., Mazzei, A., Ricci, M. and D'Ariano, G. M. (2003). Exploiting quantum parallelism of entanglement for a complete experimental quantum characterization of a single-qubit device, *Phys. Rev. A* **67**(6): 062307.

Dobšíček, M., Johansson, G., Shumeiko, V. and Wendin, G. (2007). Arbitrary accuracy iterative quantum phase estimation algorithm using a single ancillary qubit: A two-qubit benchmark, *Phys. Rev. A* **76**(3): 030306.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics, *Philos. Trans. R. Soc. London* **A**(222): 309–369.

Fujiwara, A. (1994). Multi-parameter pure state estimation based on the right logarithmic derivative, *Univ. Tokyo, METR* **94**(9).

Fujiwara, A. (2001). Quantum channel identification problem, *Phys. Rev. A* **63**(4): 042304.

Fujiwara, A. (2002). Estimation of SU(2) operation and dense coding: An information geometric approach, *Phys. Rev. A* **65**(1): 012316.

Fujiwara, A. (2004). Estimation of a generalized amplitude-damping channel, *Phys. Rev. A* **70**(1): 012317.

Fujiwara, A. (2006). Strong consistency and asymptotic efficiency for adaptive quantum estimation problems, *J. Phys. A* **39**(40): 12489–12504.

Fujiwara, A. and Imai, H. (2003). Quantum parameter estimation of a generalized Pauli channel, *J. Phys. A* **36**(29): 8093–8103.

Fujiwara, A. and Imai, H. (2008). A fibre bundle over manifolds of quantum channels and its application to quantum statistics, *J. Phys. A* **41**(25): 255304.

Gill, R. D. (2001). Teleportation into quantum statistics, *J. Kor. Statist. Soc.* **30**: 291.

Gill, R. D. (2008). Conciliation of Bayes and Pointwise Quantum State Estimation, pp. 239-261 in Quantum Stochastics and Information: Statistics, Filtering and Control, Belavkin, V.P. and Guţă, M., eds., World Scientific.

Giovannetti, V., Lloyd, S. and Maccone, L. (2004). Quantum-enhanced measurements: beating the standard quantum limit, *Science* **306**(5700): 1330–1336.

Giovannetti, V., Lloyd, S. and Maccone, L. (2006). Quantum metrology, *Phys. Rev. Lett.* **96**(1): 010401.

Guţă, M. and Jencova, A. (2007). Local asymptotic normality in quantum statistics, *Communications in Mathematical Physics* **276**(2): 341–379.

Guţă, M. and Kahn, J. (2006). Local asymptotic normality for qubit states, *Phys. Rev. A* **73**(5): 052108.

Guţă, M., Janssens, B. and Kahn, J. (2007). Optimal estimation of qubit states with continuous time measurements, *Communications in Mathematical Physics* **277**(1): 127–160.

Hayashi, M. (2002). Two quantum analogues of Fisher information from a large deviation viewpoint of quantum estimation, *J. Phys. A* **35**(39): 7689–7727.

Hayashi, M. (2003a). presentations at MaPhySto and QUANTOP Workshop on Quantum Measurements and Quantum Stochastics, Aarhus, 2003, and Special Week on Quantum Statistics, Isaac Newton Institute for Mathematical Sciences, Cambridge, 2004.

Hayashi, M. (2003b). Quantum estimation and the quantum central limit theorem, *Bulletin of the Mathematical Society of Japan* **5**: 368–391.
URL: *Translated into English in quant-ph/0608198*

Hayashi, M. (2005). *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*, Singapore: World Scientific.

Hayashi, M. (2006a). Parallel treatment of estimation of SU(2) and phase estimation, *Phys. Lett. A* **354**: 183–189.

Hayashi, M. (2006b). *Quantum Information: an Introduction*, Springer.

Hayashi, M. and Matsumoto, K. (2004). Asymptotic performance of optimal state estimation in quantum two level system, *quant-ph/0411073* .

Hazewinkel, M. (2002). *Encyclopedia of Mathematics*, Springer online Reference works.
URL: *http://eom.springer.de/default.htm*

Helstrom, C. V. (1976). *Quantum Detection and Estimation*, New York: Academic.

Helstrom, C. W. (1967). Minimum mean-square error estimation in quantum statistics, *Phys. Lett.* **25A**: 101–102.

Higgins, B. L., Berry, D. W., Bartlett, S. D., Wiseman, H. M. and Pryde, G. J. (2007). Entanglement-free Heisenberg-limited phase estimation, *Nature* **450**: 393–396.

Holevo, A. S. (1982). *Probabilistic and Statistical Aspects of Quantum Theory*, Amsterdam: North-Holland.

Huelga, S. F., Macchiavello, C., Pellizzari, T., Ekert, A. K., Plenio, M. B. and Cirac, J. I. (1997). Improvement of frequency standards with quantum entanglement, *Phys. Rev. Lett.* **79**(20): 3865–3868.

Imai, H. and Fujiwara, A. (2007). Geometry of optimal estimation scheme for SU(d) channels, *J. Phys. A* **40**(16): 4391–4400.

Ji, Z., Wang, G., Duan, R., Feng, Y. and Ying, M. (2008). Parameter estimation of quantum channels, *IEEE Trans. Inf. Theory* **54**(11): 5172–5185.

Kahn, J. (2007). Fast rate estimation of a unitary operation in SU(d), *Phys. Rev. A* **75**(2): 022326.

Kahn, J. and Guţă, M. (2009). Local asymptotic normality for finite dimensional quantum systems, *Communications in Mathematical Physics* **289**(2): 597–652.

Kitaev, A. Y. (1996). Quantum measurements and the abelian stabilizer problem, *Electronic Colloquium on Computational Complexity* **3**(3): 1–22.

Knill, E., Ortiz, G. and Somma, R. D. (2007). Optimal quantum measurements of expectation values of observables, *Phys. Rev. A* **75**(1): 012328.

Kraus, K. (1983). *States, Effects, and Operations: Fundamental Notions of Quantum Theory (Lecture Notes in Physics)*, Vol. 190, Berlin; New York: Springer-Verlag.

Le Bellac, M. (2006). *A Short Introduction to Quantum Information and Quantum Computation*, Cambridge University Press, New York, NY, USA.

Liu, X., Luo, J. and Sun, X. (2007). Experimental realization of arbitrary accuracy iterative phase estimation algorithms on ensemble quantum computers, *Chin. Phys. Lett.* **24**(12): 3316–3319.

Luati, A. (2004). Maximum Fisher information in mixed state quantum systems, *Annals of Statistics* **32**(4): 1770.

Massar, S. (2000). Collective versus local measurements on two parallel or antiparallel spins, *Phys. Rev. A* **62**: 040101.

Matsumoto, K. (1997). A new approach to the Cramér-Rao-type bound of the pure-state model, *quant-ph/9711008v1* .

Matsumoto, K. (2002). A new approach to the Cramér-Rao-type bound of the pure-state model, *J. Phys. A* **35**(13): 3111–3123.

Morozova, E. A. and Čencov, N. N. (1990). Markov invariant geometry on state manifolds (in russian), *Itogi Nauki i Tekhniki* **36**: 69–102.

Nagaoka, H. (1988). An asymptotically efficient estimator for a one-dimensional parametric model of quantum statistical operators, *Proc. Int. Symp. on Inform. Theory* p. 198.

Nagaoka, H. (1989). On the parameter estimation problem for quantum statistical models, *Proc. 12th Symp. on Inform. Theory and its Appl.* pp. 577–82.

Nielsen, M. A. and Chuang, I. (2000). *Quantum Computation and Quantum Information*, Cambridge: Cambridge University Press.

O'Loan, C. J. (2007). Analysis of a convenient information bound for general quantum channels, *J. Phys. A* **40**(48): 14499–14513.

O'Loan, C. J. (2010). Iterative phase estimation, *J. Phys. A* **43**(1): 015301.

Petz, D. and Sudár, C. (1996). Geometries of quantum states, *J. Math. Phys.* **37**: 2662–2673.

Rubin, M. A. and Kaushik, S. (2007). Loss-induced limits to phase measurement precision with maximally entangled states, *Phys. Rev. A* **75**(5): 053805.

Rudolph, T. and Grover, L. (2003). Quantum communication complexity of establishing a shared reference frame, *Phys. Rev. Lett.* **91**(21): 217905.

Sarovar, M. and Milburn, G. J. (2006). Optimal estimation of one-parameter quantum channels, *J. Phys. A* **39**(26): 8487–8505.

Shaji, A. and Caves, C. M. (2007). Qubit metrology and decoherence, *Phys. Rev. A* **76**(3): 032111.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge: Cambridge University Press.

Wang, H. F., Wu, L. A., Liu, Y. X. and Nori, F. (2009). Measurement-based quantum phase estimation algorithm for finding eigenvalues of non-hermitian matrices, *arxiv:0906.2538, Phys. Rev. A, in press* .

Wang, X., You, J. and Nori, F. (2008). Simulated Quantum Computation of Molecular Energies, *Phys. Rev. A* **77**: 062339.

Wei, L. F. and Nori, F. (2004). Quantum phase estimation algorithms with delays: effects of dynamical phases, *J. Phys. A* **37**(16): 4607–4617.

Yuen, H. and Lax, M. (1973). Multiple-parameter quantum estimation and measurement of nonselfadjoint observables, *IEEE Trans. Inf. Theory* **19**(6): 740–750.