Optimal estimation of one-parameter quantum channels

Mohan Sarovar and G J Milburn

Centre for Quantum Computer Technology and School of Physical Sciences,
The University of Queensland, St Lucia, QLD 4072, Australia

E-mail: mohan@physics.uq.edu.au and milburn@physics.uq.edu.au

Abstract
We explore the task of optimal quantum channel identification and in particular,
the estimation of a general one-parameter quantum process. We derive new
characterizations of optimality and apply the results to several examples
including the qubit depolarizing channel and the harmonic oscillator damping
channel. We also discuss the geometry of the problem and illustrate the
usefulness of using entanglement in process estimation.

PACS numbers: 03.65.Ta, 03.67.−a

1. Introduction

Recent experimental progress in quantum information processing has highlighted the
importance of quantum control and, in particular, the task of system identification. For
example, it is essential for the verification of quantum gates to be able to effectively identify
quantum processes. This identification might require a full process tomography, but quite
often an estimation of a number of free parameters may be sufficient to determine the proper
operation of a gate. In such cases, it is obviously beneficial to use optimal schemes to estimate
the free parameters. With that motivation, we consider the problem of optimally estimating
a quantum process with one free parameter. Figure 1 represents a schematic of the problem.
The goal is to identify an input state and a measurement scheme that will permit one to gain
the most informaiton about the free parameter. We will make these notions more precise in
what follows.

The field of quantum parameter estimation is concerned with the methods of estimating—
especially optimally estimating—properties of quantum states or processes. The field has a
fairly short but rich history, beginning with the pioneering work of Helstrom [1] and Holevo [2]
and has recently seen a rise in interest, especially from the quantum information community.

The task of estimating quantum states has an enormous literature dedicated to it (a small
sample is [3–7], and for a recent review see [8]). There are several notions of optimality in this
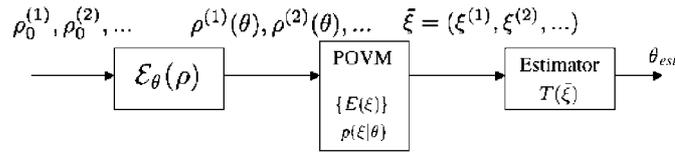
Figure 1. The estimation procedure.

scenario and some of them highlight the essential connection between the optimal estimation problem and the geometry of the space of quantum states [3, 9]; the statistical distinguishability of states induces a unique metric (the *Fisher metric*) on the space of quantum states [3, 10]. This is a pleasing state of affairs, which although not of much practical use (it cannot be used to identify optimal estimation strategies except in some special cases [11]) is tremendously edifying.

In contrast, parameter estimation for quantum processes is an area that is far less developed. The estimation of unitary quantum processes has been examined by a number of authors over the past few decades [1, 2, 4, 11, 12]; in these treatments, the system evolves unitarily and the parameters to be estimated are unknown parameters of this unitary. However, the general form of the problem, where the evolution is a general quantum operation, has received considerably less attention. Some recent treatments of special cases of this more general problem are given in [13–15].

One way to attack the problem of optimal quantum process estimation is to treat it as an optimal state estimation problem where the states under consideration are restricted to the parametrized (by the parameters of the process) set of output states of the quantum process. This is a practical point of view which recognizes that the only operational access to the parameters of the process is through probe input states that are measured at the output of the process (see figure 1). This approach breaks the process parameter estimation into two parts: the choice of an optimal input state and the choice of an optimal estimation scheme for the parametric family of output states $\{\mathcal{E}_\theta(\rho_0)\}_\theta$. Such an approach has been taken in the past (e.g. [13]), but with the dynamics of the process (or equivalently, the dependence of the output parametric family on the parameters of the process) being represented by rather abstract superoperators such as the *symmetric logarithmic derivative* [3]. The disadvantage of using such a representation is that explicit expressions for such superoperators are often difficult to calculate.

The aim of this paper is to use the same approach to the estimation of quantum processes, but to use a more common representation for them. In particular, we assume that the Kraus representation (operator sum decomposition) of the quantum process is given, with the parameters to be estimated being free parameters of the Kraus operators. We consider the simplest quantum process estimation problem in this general setting: the estimation of a one-parameter, trace-preserving quantum operation and investigate the advantages and disadvantages of using the Kraus representation to describe the process. We assume that the input state remains fixed and optimize over the estimation scheme to arrive at a measure of optimality. The measure is non-unique and non-constructive (it does not allow one to deduce the optimal POVM), primarily due to the non-uniqueness of the Kraus representation. However, for a special family of quantum channels we show that it can be used to test the optimality of an estimation scheme. We also discuss the geometry of the problem and apply the results to several examples, including one which illustrates the value of using entanglement to increase the statistical distinguishability of quantum processes.

The paper is organized as follows: section 2 covers some preliminary concepts and sets the notation. Section 3 derives the optimality conditions for the estimation scheme and discusses them. Then we consider several examples in section 4 and finally conclude in section 5.

2. Preliminaries

2.1. Quantum operations

Given two Hilbert spaces, $\mathcal{H}_1$ and $\mathcal{H}_2$, with dimensions d_1 and d_2 respectively, a quantum operation $\mathcal{E}: \mathcal{H}_1^T \to \mathcal{H}_2^T$ is a *completely positive* linear map between trace-class operators in $\mathcal{H}_1$ and $\mathcal{H}_2$($\mathcal{H}_i^T$ is the space of trace-class operators acting on Hilbert space $\mathcal{H}_i$; $\dim \mathcal{H}_i^T = d_i^2$). These maps represent the most general transformations between two density operators in quantum mechanics. The complete positivity of the map is a physically motivated and highly restrictive condition; however, there is a well-known representation of such maps originally due to Choi [16] and then popularized by Kraus [17], called the *operator sum decomposition* or the *Kraus decomposition*. This decomposition represents the action of the map $\mathcal{E}: \mathcal{H}_1^T \to \mathcal{H}_2^T$ as

$$\tilde{\rho} \equiv \mathcal{E}(\rho) = \sum_k \Upsilon_k \rho \Upsilon_k^\dagger, \tag{1}$$

where $\rho \in \mathcal{H}_1^T$ and $\tilde{\rho} \in \mathcal{H}_2^T$. The operators $\Upsilon_k: \mathcal{H}_1 \to \mathcal{H}_2$ are called the *Kraus operators* of the decomposition. The sum in this decomposition can run over an infinite set, but in general $\mathcal{E}$ can be represented using at most $d_1 d_2$ Kraus operators [18]. In this paper we will restrict our attention to *trace-preserving* quantum operations between isomorphic Hilbert spaces (i.e. $d_1 = d_2 = d$). The trace preservation implies that the Kraus operators satisfy a normalization condition:

$$\sum_k \Upsilon_k^\dagger \Upsilon_k = I_d, \tag{2}$$

where I_d is the d-dimensional identity operator. Of course, for a closed quantum system undergoing unitary evolution, there is only one Kraus operator in the decomposition and it is unitary, i.e. $\mathcal{E}(\rho) = V \rho V^\dagger$ with $V^\dagger V = I_d$.

An important property of this decomposition is the non-uniqueness of the Kraus operators. A new set of Kraus operators for the same quantum operation can be derived by an arbitrary unitary remixing of the original set. That is, if $\{\Upsilon_k\}_1^N$ are the Kraus operators in a decomposition of $\mathcal{E}$, then the set $\{\Omega_j\}_1^N$:

$$\Omega_j = \sum_{k=1}^N u_{jk} \Upsilon_k, \tag{3}$$

where u_{jk} are the elements of a unitary matrix, is the Kraus operators for a different, but equivalent, operator sum decomposition of $\mathcal{E}$. This non-uniqueness will become important when we describe the estimation of quantum operations in terms of their Kraus operators; in particular, a specific decomposition, called the *canonical decomposition*, will be important. The canonical decomposition is a distinguished Kraus decomposition that can be constructed *with respect to any given input state* with the following property: $\mathcal{E}(\rho) = \sum_k \Upsilon_k \rho \Upsilon_k^\dagger$ with $\operatorname{tr}\left(\Upsilon_k^\dagger \Upsilon_j \rho\right) = \delta_{jk} p_k$. Trace preservation implies $\sum_k p_k = 1$. The explicit construction of these Kraus operators is detailed in [19]; however, for our purposes the important thing to note is their dependence on the input state—if the input state changes, the Kraus operators in the canonical decomposition for a given quantum process will change.

Now we describe an important interpretation of quantum operations. By the postulates of quantum mechanics, a closed quantum system will undergo unitary evolution. Any departure from such an evolution is caused by coupling of the system to additional degrees of freedom (usually termed the environment). Therefore, any non-unitary quantum operation is the effective description of the evolution of a quantum system that is coupled to an environment. This leads us to think of a quantum operation, acting on a system of interest, as a unitary map of the system plus some environment (which combined form a closed system) after which the environment is traced out [18]. That is,

$$\mathcal{E}(\rho) = \text{tr}_{\text{env}}[U(\rho \otimes \rho_{\text{env}})U^\dagger], \tag{4}$$

where ρ is a density operator for the system and ρ_{env} is a density operator for the environment. U is a unitary operator acting on both the system and the environment. This is sometimes referred to as the *dilation* of the quantum operation. Now, assuming that $\{|e_k\rangle\}$ is a complete orthonormal basis for the state space of the environment and the initial environment state is $\rho_{\text{env}} = |e_0\rangle\langle e_0|$, then the Kraus decomposition of $\mathcal{E}$ can be written explicitly by performing the trace in this basis:

$$\begin{aligned}
\mathcal{E}(\rho) &= \text{tr}_{\text{env}}[U(\rho \otimes \rho_{\text{env}})U^\dagger] \\
&= \sum_k \langle e_k|U[\rho \otimes |e_0\rangle\langle e_0|]U^\dagger|e_k\rangle \\
&= \sum_k \Upsilon_k \rho \Upsilon_k^\dagger,
\end{aligned} \tag{5}$$

where $\Upsilon_k = \langle e_k|U|e_0\rangle$ is an operator acting only on the system subspace. The unitary freedom in the choice of the Kraus operators is exactly the same as the unitary freedom in choosing the environmental basis states in which to perform the trace. The assumption of a pure initial state for the environment is not very restrictive because we can always choose the environment to be large enough so that this condition is met. The other assumption, that the initial state is separable, is a much more subtle one and a full discussion of it is out of the scope of this paper. It has been extensively discussed in the literature, and we refer the interested reader to the recent treatment in [20].

2.2. Optimal estimation

We are concerned with optimally estimating a one-parameter quantum operation, $\mathcal{E}_\theta$. We assume that we have available an operator sum decomposition of the operation, but that the Kraus operators have a free *real, continuous* parameter, θ, which is to be estimated:

$$\rho(\theta) \equiv \mathcal{E}_\theta(\rho_0) = \sum_k \Upsilon_k(\theta)\rho_0 \Upsilon_k^\dagger(\theta), \tag{6}$$

with $\sum_k \Upsilon_k(\theta)^\dagger \Upsilon_k(\theta) = I_d$. Here, $\rho_0 \in \mathcal{H}^T (\dim \mathcal{H}^T = d^2)$ is the input state to the operation over which we have complete control. A schematic of the estimation process is given in figure 1. In each run of the experiment, the input state ρ_0 is fed into the quantum operation and a generalized measurement is performed on the output. The generalized measurement is described by non-negative, Hermitian operators $E(\xi)$, or POVMs, which satisfy the completeness property: $\int d\xi\, E(\xi) = I_d$. ξ labels the result/s of the measurement and can be univariate or multivariate as well as continuous or discrete (in the discrete case, the completeness integral becomes a sum). Given such a description of the measurement, the probability density for the measurement result, conditioned on a given state is given by $p(\xi|\theta) = \text{tr}(E(\xi)\rho(\theta))$. We will assume that there are N independent runs of this experiment after which the parameter θ is estimated using the results of N independent,

separable measurements on N outputs from the channel. That is, $\theta_{\text{est}} = \theta_{\text{est}}(\xi_1, \xi_2, \ldots, \xi_N)$. The estimator T is any function of the N measurement outcomes, and it attempts to reconstruct θ from these outcomes. We will assume that the estimator, T, is unbiased. That is, $E_\theta\{\theta_{\text{est}}\} = \theta$ where $E_\theta\{\cdot\}$ is the expectation value with respect to the probability distribution for θ. This is a mild assumption that will not affect the essential results. An example of an estimator (which is unbiased also) is the *sample mean*: $\bar{m} = 1/N \sum_{i=1}^{N} x_i$, which is an estimate of the true mean of the probability distribution that x_i are drawn from.

This is a very general setting in which to describe parameter estimation. We can account for scenarios such as entanglement-assisted estimation by simply changing the definition of the quantum operation to be a tensor product of two or more operations. We will see an example of this in section 4. We are seeking the optimal scheme for estimating the free parameter, θ. Loosely, this amounts to specifying an input state, a measurement scheme and an estimator that will permit one to gain the most information about θ, and we will now proceed by making this notion of optimality more precise.

Firstly, the deviation of our estimate from the actual value of the parameter can be measured by

$$\delta\theta \equiv \theta_{\text{est}} - \theta. \tag{7}$$

It is natural to consider the estimation scheme that minimizes the variance of this estimation error as the optimal one. That is, we want to minimize $\langle(\delta\theta)^2\rangle$. Braunstein and Caves consider a similar problem in [3], and we will follow their treatment in order to find the optimal scheme.

As mentioned in the introduction, this optimization problem can be split into two subproblems: (i) the choice of an optimal input state ρ_0 and (ii) the choice of an optimal scheme (choice of $\{E(\xi)\}$ and T) to estimate from the one-parameter family of the output states $\rho(\theta) = \mathcal{E}(\rho_0)$. In this paper, we will not consider the first subproblem; we will assume that the input state is fixed and focus on the optimization of the estimation scheme. The second subproblem, the optimization of the estimation scheme, can be further broken down into two steps: first, a minimization of the error variance over estimators T for a given quantum measurement and second, a minimization over all quantum measurements. The optimization over the estimators is an entirely classical one; it is well known in the statistical inference literature and results in the famous Cramér–Rao bound [21]:

$$\langle(\delta\theta)^2\rangle \geqslant \frac{1}{NF(\theta)}, \tag{8}$$

where N is the number of measurement results used in the estimation and $F(\theta)$ is the *Fisher information*:

$$
\begin{aligned}
F(\theta) &\equiv \int \mathrm{d}\xi \; p(\xi|\theta) \left(\frac{\partial \ln p(\xi|\theta)}{\partial\theta}\right)^2 \\
&= \int \mathrm{d}\xi \frac{1}{p(\xi|\theta)} \left(\frac{\partial p(\xi|\theta)}{\partial\theta}\right)^2.
\end{aligned} \tag{9}
$$

The Fisher information represents the amount of information about θ contained in the measurement result ξ. The dependence of this quantity on the choice of quantum measurement is clear from the fact that $p(\xi|\theta) = \mathrm{tr}(E(\xi)\rho(\theta))$. Strictly, this form of the bound is only valid for unbiased estimators. As we will only consider such estimators, we will not state the more general form of the bound here.

The Cramér–Rao bound effectively takes the estimator out of the picture. It says that for a given input state and measurement (i.e. for a given $p(\xi|\theta)$), the variance is lower-bounded by the quantity on the right-hand side of equation (8). Thus, subproblem (ii)—the estimation

on the parametric family of output states—simplifies to finding the best measurement: the one that minimizes this lower bound, or equivalently, maximizes the Fisher information. As an aside, under mild regularity conditions on $p(\xi|\theta)$ the lower bound in equation (8) is an asymptotically achievable one, that is, there exist estimators that can attain this bound as $N \rightarrow \infty$, and an example is the maximum likelihood estimator [22]. The estimators that achieve this bound have been extensively studied in the field of statistical inference; therefore, we will not consider them here but will rather concentrate on the quantum aspects of the problem.

Given the Cramér–Rao bound, the next step of the optimization becomes a maximization of equation (9) over all possible quantum measurements. We express this maximization, and the result, by

$$F^*(\theta) \equiv \max_{\{E(\xi)\}} F(\theta). \tag{10}$$

2.2.1. A geometric perspective. Before treating this maximization, we examine the process estimation problem from a geometric perspective. The one-parameter family formed by the output of the process for a fixed input state, $\{\rho(\theta)\}_\theta$, defines a curve in density operator space which is parametrized by the continuous, real parameter θ. The curve is itself a manifold defined by $\rho_0 \equiv |\psi_0\rangle\langle\psi_0|$ and $\mathcal{E}_\theta$, and the advantage of regarding members of $\{\rho(\theta)\}_\theta$ as outputs of a quantum process represented by its Kraus decomposition is that we can define a natural local coordinate patch at each point on the curve[1]. That is, $\rho(\theta) = \sum_k \Upsilon_k(\theta)|\psi_0\rangle\langle\psi_0|\Upsilon_k^\dagger(\theta) = \sum_k |e_k(\theta)\rangle\langle e_k(\theta)|$, where $|e_k(\theta)\rangle \equiv \Upsilon_k(\theta)|\psi_0\rangle$ are unnormalized vectors. Now if we exclusively use the canonical decomposition for the process, this can be rewritten as an eigendecomposition of $\rho(\theta)$: $\sum_k p_k(\theta)|f_k(\theta)\rangle\langle f_k(\theta)|$, where $|f_k(\theta)\rangle = \frac{1}{\sqrt{p_k(\theta)}}|e_k(\theta)\rangle$ and $\langle f_j(\theta)|f_k(\theta)\rangle = \delta_{jk}$. The set $\{|f_k(\theta)\rangle\}$ can be considered a local orthonormal coordinate basis at the point θ on the curve. A point to note is that while we write the eigendecomposition as a sum over k, the number of Kraus operators in the canonical decomposition, it does not mean that $\rho(\theta)$ has the same number of eigenvectors as the number of Kraus operators are there in the Kraus decomposition. $\rho(\theta)$ can have at most d eigenvectors, where d is the dimension of the system, while there is no limit on the number of Kraus operators. In the canonical decomposition, $\Upsilon_k(\theta)|\psi_0\rangle = 0$ for some values of k, and these terms will drop out in the eigendecomposition sum.

The Fisher information can be used to define a Riemannian metric on this curve (submanifold) that measures the statistical distinguishability of neighbouring one-parameter quantum operations given the fixed input state $|\psi_0\rangle$. To see this, we go back to the definitions above and note that the Cramér–Rao bound (8) is a lower bound on the variance in the error when estimating the parameter θ. Thus, it is a lower bound on the error in reliably distinguishing between two neighbouring quantum operations: $\mathcal{E}_\theta$ and $\mathcal{E}_{\theta+\mathrm{d}\theta}$. Therefore, as in [3], we can consider it a *distinguishability metric* along the curve of one-parameter quantum operations defined by $|\psi_0\rangle$ and $\mathcal{E}_\theta$. More formally, let us establish $\min[\sqrt{N}\langle(\delta\theta)^2\rangle^{1/2}]$ to be a measure of statistical deviation. $\sqrt{N}$ removes the improvement in estimation due to increased sampling, and the minimization is over measurement schemes to ensure that we are considering the most discriminating scheme. A statistical measure of distinguishability should be proportional to the inverse of this deviation measure, i.e. the more the deviation, the less distinguishable neighbouring operations become. Thus, we can define a distinguishability

---

[1] Here, we have assumed that the input state is pure. We will see below that the optimal input state will always be a pure one, and therefore this assumption is justified.

metric along the curve as

$$ds^2 = \frac{d\theta^2}{\min[N\langle(\delta\theta)^2\rangle]}. \tag{11}$$

$(ds/d\theta)^2$ is well known as the *statistical distance* [3], and using equation (8), we can rewrite it in terms of the Fisher information as

$$\left(\frac{ds}{d\theta}\right)^2 = \max_{\{E(\xi)\}} F(\theta) = F^*(\theta). \tag{12}$$

This is exactly the maximization we are considering for optimal estimation. Note that it is only over the measurement POVMs ($\{E(\xi)\}$) because this is the statistical distance over a curve defined by a particular input state. A caveat is required here; when we refer to $F^*(\theta)$ as being the metric on the curve, this is strictly only true if it can be shown that the bound set by $F^*(\theta)$ can be achieved. This question of achievability will be important in the following.

3. The optimization

As outlined in section 2.2, the procedure of finding the best estimation scheme can be phrased as a sequence of optimizations. The first optimization, which is entirely classical, results in the Cramér–Rao bound, and in this section we shall examine the quantum aspects of the problem.

3.1. Optimal estimation on the output family

The optimal quantum measurement scheme is the set of POVMs that maximize the Fisher information, for a fixed input state. Using the definition of the Fisher information, equation (9), and the fact that $p(\xi|\theta) = \text{tr}(E(\xi)\rho(\theta))$, we get

$$F^*(\theta) = \max_{\{E(\xi)\}} \int d\xi \, \frac{(\text{tr}[E(\xi)\rho'(\theta)])^2}{\text{tr}[E(\xi)\rho(\theta)]}, \tag{13}$$

where $\rho(\theta) \equiv \mathcal{E}_\theta(\rho_0) = \sum_k \Upsilon_k(\theta)\rho_0\Upsilon_k^\dagger(\theta)$ and $\rho'(\theta) = \partial\rho(\theta)/\partial\theta$. Let $\{\Upsilon_k\}$ be the Kraus operators for an arbitrary Kraus decomposition of $\mathcal{E}$. Now, the next logical step would be to replace $\rho(\theta)$ and $\rho'(\theta)$ by their definitions in terms of the Kraus operators that define the quantum operation. However, this makes the maximization of (13) difficult due to the introduction of the Kraus decomposition sum in the numerator. Instead, we will take a step back and use the dilation of the quantum operation.

As mentioned in section 2.1, a quantum operation can be thought of as a unitary map of the system plus some environment after which the environment is traced out. Given this, we will label our system A and the environment B, and define

$$\rho_A(\theta) = \mathcal{E}_\theta\big(\rho_A^0\big) = \sum_k \Upsilon_k(\theta)\rho_A^0\Upsilon_k^\dagger(\theta)$$
$$= \text{tr}_B\{U(\theta)[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|]U^\dagger(\theta)\}, \tag{14}$$

where $U(\theta)$ is some unitary operator acting on systems A and B, and ρ_A^0 is the input state on subsystem A. The mapping between $U(\theta)$ and $\{\Upsilon_k(\theta)\}$ is not unique because of the freedom in choosing the environment basis states, and we will return to this point shortly. Also,

$$\rho_A'(\theta) = \text{tr}_B\left\{U'(\theta)\big[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\big]U^\dagger(\theta) + U(\theta)\big[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\big]U^{\dagger\prime}(\theta)\right\}$$
$$= \text{tr}_B\{\Omega(\theta) + \Omega^\dagger(\theta)\}, \tag{15}$$

where $U'(\theta) = \partial U(\theta)/\partial\theta$ and $\Omega(\theta) = U'(\theta)\big[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\big]U^\dagger(\theta)$.

Now we return to the problem of maximizing (13). Substituting (14) and (15), we get (in the following, we will use $E_A(\xi) \otimes I_B$ and $E_A(\xi) I_B$ interchangeably to denote the same operator)

$$\int d\xi \frac{(\mathrm{tr}_A\{E_A(\xi) \, \mathrm{tr}_B\{\Omega(\theta) + \Omega^\dagger(\theta)\}\})^2}{\mathrm{tr}_A\left\{E_A(\xi) \, \mathrm{tr}_B\left\{U(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]U^\dagger(\theta)\right\}\right\}}$$

$$= \int d\xi \frac{(\mathrm{tr}_A\{\mathrm{tr}_B\{(E_A(\xi)I_B)(\Omega(\theta) + \Omega^\dagger(\theta))\}\})^2}{\mathrm{tr}_A\left\{\mathrm{tr}_B\left\{(E_A(\xi)I_B)U(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]U^\dagger(\theta)\right\}\right\}}$$

$$= \int d\xi \frac{(\mathrm{tr}\{(E_A(\xi)I_B)\Omega(\theta) + (E_A(\xi)I_B)\Omega^\dagger(\theta)\})^2}{\mathrm{tr}\left\{(E_A(\xi)I_B)U(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]U^\dagger(\theta)\right\}}$$

$$= 4 \int d\xi \frac{(\mathrm{Re}\,\mathrm{tr}\{(E_A(\xi)I_B)\Omega(\theta)\})^2}{\mathrm{tr}\left\{(E_A(\xi)I_B)U(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]U^\dagger(\theta)\right\}}$$

$$\leqslant 4 \int d\xi \frac{|\mathrm{tr}\{(E_A(\xi)I_B)\Omega(\theta)\}|^2}{\mathrm{tr}\left\{(E_A(\xi)I_B)U(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]U^\dagger(\theta)\right\}}$$

$$= 4 \int d\xi \frac{\left|\mathrm{tr}\left\{(E_A(\xi) \otimes I_B)U'(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]U^\dagger(\theta)\right\}\right|^2}{\mathrm{tr}\left\{(E_A(\xi) \otimes I_B)U(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]U^\dagger(\theta)\right\}}.$$

We proceed by applying the Cauchy–Schwarz inequality: $|\mathrm{tr}(O^\dagger P)|^2 \leqslant \mathrm{tr}(O^\dagger O)\,\mathrm{tr}(P^\dagger P)$, with equality when $O = \lambda P$ for some constant λ. We will apply this inequality to the numerator, where $O^\dagger = (E_A(\xi)I_B)^{1/2}U'(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]^{1/2}$ and $P = \left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]^{1/2}U^\dagger(\theta)(E_A(\xi)I_B)^{1/2}$, to get

$$F(\theta) \leqslant 4 \int d\xi \frac{\mathrm{tr}\left\{E_A(\xi)I_B)U(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]U^\dagger(\theta)\right\}}{\mathrm{tr}\left\{(E_A(\xi)I_B)U(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]U^\dagger(\theta)\right\}}$$

$$\times \mathrm{tr}\left\{(E_A(\xi)I_B)U'(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]U^{\dagger'}(\theta)\right\}$$

$$= 4 \int d\xi \, \mathrm{tr}\left\{(E_A(\xi) \otimes I_B)U'(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]U^{\dagger'}(\theta)\right\}$$

$$= 4\,\mathrm{tr}\left\{U'(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]U^{\dagger'}(\theta)\right\}$$

$$= 4\,\mathrm{tr}_A\,\mathrm{tr}_B\left\{U'(\theta)\left[\rho_A^0 \otimes |e_0\rangle_B\langle e_0|\right]U^{\dagger'}(\theta)\right\}$$

$$= 4\,\mathrm{tr}_A\left\{\sum_k \Upsilon_k'(\theta)\rho_A^0 \Upsilon_k^{\dagger'}(\theta)\right\}$$

$$= 4\,\mathrm{tr}\left\{\sum_k \Upsilon_k^{\dagger'}(\theta)\Upsilon_k'(\theta)\rho_0\right\}, \tag{16}$$

where we have dropped all subscripts in the last line because all operators are in subsystem A, and the completeness relation $\int d\xi\, E(\xi) = I_d$ has been used.

We have arrived at a bound on the Fisher information:

$$C_\Upsilon(\theta) \equiv 4\,\mathrm{tr}\left\{\sum_k \Upsilon_k^{\dagger'}(\theta)\Upsilon_k'(\theta)\rho_0\right\}. \tag{17}$$

This bound is equal to the maximum, $F^*(\theta)$, if it is achievable. To achieve the bound we need to saturate the two inequalities used in the derivation. The condition for meeting the first inequality is

$$\operatorname{Im} \operatorname{tr}\{(E_A(\xi) \otimes I_B)\Omega(\theta)\} = 0$$

$$\Longleftrightarrow \operatorname{Im} \operatorname{tr}_A \left\{ E(\xi) \sum_k \Upsilon'_k(\theta)\rho_0 \Upsilon_k^\dagger(\theta) \right\} = 0 \tag{18}$$

$$\Longleftrightarrow \operatorname{Im} \operatorname{tr} \left\{ \sum_k \Upsilon_k^\dagger(\theta) E(\xi) \Upsilon'_k(\theta)\rho_0 \right\} = 0 \quad \forall \xi.$$

The condition for meeting the Cauchy–Schwarz bound is

$$(E_A(\xi) \otimes I_B)^{\frac{1}{2}} U'(\theta)\big[\rho_A^0 \otimes |e_0\rangle_B \langle e_0|\big]^{\frac{1}{2}}$$

$$= \lambda_\xi(\theta)(E_A(\xi) \otimes I_B)^{\frac{1}{2}} U(\theta)\big[\rho_A^0 \otimes |e_0\rangle_B \langle e_0|\big]^{\frac{1}{2}} \quad \forall \xi,$$

where the constant $\lambda_\xi(\theta)$ can generally depend on ξ and θ.

We would like a condition is terms of the Kraus operators instead of the unitary U. To do this, multiply from the left by an identity in the form $I_A \otimes \sum_k |e_k\rangle_B \langle e_k|$ to get

$$\sum_k \big[E(\xi)^{\frac{1}{2}} \Upsilon'_k(\theta)\rho_0^{\frac{1}{2}}\big]_A \otimes |e_k\rangle_B \langle e_0| = \lambda_\xi(\theta) \sum_k \big[E(\xi)^{\frac{1}{2}} \Upsilon_k(\theta)\rho_0^{\frac{1}{2}}\big]_A \otimes |e_k\rangle_B \langle e_0| \quad \forall \xi$$

$$\Longleftrightarrow E(\xi)^{\frac{1}{2}} \Upsilon'_k(\theta)\rho_0^{\frac{1}{2}} = \lambda_\xi(\theta) E(\xi)^{\frac{1}{2}} \Upsilon_k(\theta)\rho_0^{\frac{1}{2}} \quad \forall \xi, k, \tag{19}$$

where the last step uses the orthogonality of $|e_k\rangle_B$. Equation (19) defines a series of conditions that the optimal measurement must satisfy. In general, k runs from 1 to d^2 where d is the dimension of the state space of ρ, and therefore we see simply from the number of constraining equations that the optimal measurement is severely restricted.

We can reduce these two conditions to one by substituting (19) into the statement for the first condition (18)

$$\operatorname{Im} \operatorname{tr} \left\{ \sum_k \Upsilon_k^\dagger(\theta) E(\xi) \Upsilon'_k(\theta)\rho_0 \right\} = 0 \quad \forall \xi$$

$$\Rightarrow \operatorname{Im} \operatorname{tr} \left\{ \sum_k \Upsilon_k^\dagger(\theta) E(\xi) \lambda_\xi(\theta) \Upsilon_k(\theta)\rho_0 \right\} = 0 \quad \forall \xi$$

$$\Rightarrow \operatorname{Im} \lambda_\xi(\theta) \operatorname{tr} \rho(\theta) E(\xi) = 0 \quad \forall \xi. \tag{20}$$

Now, because the trace in the last line is always real, this condition is met if and only if $\lambda_\xi(\theta)$ is real. Therefore, to summarize, the optimal measurement scheme must satisfy the conditions

$$E(\xi)^{\frac{1}{2}} \Upsilon'_k(\theta)\rho_0^{\frac{1}{2}} = \lambda_\xi(\theta) E(\xi)^{\frac{1}{2}} \Upsilon_k(\theta)\rho_0^{\frac{1}{2}} \quad \forall \xi, k, \tag{21}$$

where $\lambda_\xi(\theta)$ is a *real* number that can depend on ξ.

Condition (21) is a condition on the optimal measurement POVM and the optimal input state. However, although it defines the optimal strategy, it is not a constructive condition. Except for special cases (that we will outline below), it is difficult to define the optimal measurement in terms of $\{\Upsilon_k\}$ and ρ_0 from the above condition. We will say more about the satisfiability of these conditions and thus the achievability of the Fisher information bound below.

We conclude this subsection by noting that an immediate consequence of the form $C_\Upsilon(\theta)$ is that a pure input state, $\rho_0 \equiv |\psi_0\rangle\langle\psi_0|$, is optimal; this follows from the linearity of trace, the concavity of density operators and the positivity of the operator $\Pi \equiv \sum_k \Upsilon_k^{\dagger\prime}(\theta)\Upsilon'_k(\theta)$.

3.2. Uniqueness and achievability

We derived a bound on the Fisher information above, and here we will discuss the uniqueness and achievability of this bound. These are both important questions because the achievability makes the bound meaningful (and means that $C_\Upsilon(\theta)$ can be viewed as a metric on the curve formed by the parametrized output family) and the uniqueness makes it useful as a characterization of optimality.

The Fisher information bound we have in equation (17) is non-unique because the Kraus operators of the quantum operation are not unique. For each choice of Kraus decomposition, the value of $C_\Upsilon(\theta)$ provides a possibly different upper bound to the Fisher information. So a natural question is: how does changing the Kraus decomposition modify the bound?

Let both $\{\Upsilon_k(\theta)\}$ and $\{\Omega_k(\theta)\}$ be the Kraus operator sets for $\mathcal{E}_\theta$. Then from section 2.1 we know that these two sets are related by a unitary transformation

$$\Omega_k(\theta) = \sum_j u_{jk}(\theta)\Upsilon_j(\theta),\tag{22}$$

where $u_{jk}(\theta)$ are the elements of a unitary matrix, which crucially can also depend on θ. Now, the Fisher information bound $C_\Omega(\theta)$ for the Kraus operator choice $\{\Omega_k(\theta)\}$ in terms of the Kraus operators $\{\Upsilon_k(\theta)\}$ is

$$
\begin{aligned}
C_\Omega(\theta) &= 4\,\mathrm{tr}\left\{\sum_k \Omega_k^{\dagger\prime}(\theta)\Omega_k'(\theta)\rho_0\right\}\\
&= 4\sum_k \mathrm{tr}\left[\left(\sum_j u_{kj}^{*\prime}\Upsilon_j^\dagger + \sum_j u_{kj}^*\Upsilon_j^{\dagger\prime}\right)\left(\sum_l u_{kl}'\Upsilon_l + \sum_l u_{kl}\Upsilon_l'\right)\rho_0\right]\\
&= 4\left(\sum_{jkl} u_{kj}^* u_{kl}\,\mathrm{tr}\left[\Upsilon_j^{\dagger\prime}\Upsilon_l'\rho_0\right] + \sum_{jkl} u_{kj}^{*\prime} u_{kl}'\,\mathrm{tr}\left[\Upsilon_j^\dagger\Upsilon_l\rho_0\right]\right.\\
&\quad\left. + \sum_{jkl} u_{kj}^{*\prime} u_{kl}\,\mathrm{tr}\left[\Upsilon_j^\dagger\Upsilon_l'\rho_0\right] + \sum_{jkl} u_{kj}^* u_{kl}'\,\mathrm{tr}\left[\Upsilon_j^{\dagger\prime}\Upsilon_l\rho_0\right]\right).
\end{aligned}\tag{23}
$$

From this expression it is clear that this bound, C_Ω, can be made as large as desired by appropriately choosing the unitary matrix $[u_{ij}(\theta)]$. In particular, it can be made to diverge by choosing $[u_{ij}(\theta)]$ that is discontinuous in θ. Therefore, the sensible thing to consider is the minimum value of this bound, that is, the minimum of equation (17) with respect to a choice of the Kraus operators.

The second issue to be addressed is the achievability of the bound. Let the Kraus decomposition be fixed; then to show the attainability of the bound, we must show that there always exists some POVM $\{E(\xi)\}$ that can meet the optimality conditions of equation (21). This attainability is a subtle task, as was pointed out by Barndorff-Nielsen and Gill in [5]. They show that in general, the optimal choice of a POVM is dependent on the actual value of the unknown parameter θ, which means that in practice, an adaptive strategy that narrows in on the value of θ has to be used [5]. Only in some special cases [10] does one strategy achieve the bound uniformly over θ. We will refer to an estimation strategy that is optimal at some value of θ, but possibly not at other values of θ as a *locally optimal* strategy.

Therefore, to give the bound $C_\Upsilon(\theta)$ a unique meaning, we would ideally like to identify the Kraus decomposition that minimizes the bound and show that the optimal measurement conditions containing operators from this decomposition can be met by some POVM. That

is, we want to minimize $C_\Upsilon(\theta)$ over the Kraus operators for a channel, while stipulating that the optimality conditions (21) are met (locally, or globally). At this stage we do not have a method of performing this optimization for the general case, but in the next section we carry it out for a special family of quantum channels.

3.3. A special case: the quasi-classical process

In this section, we identify a special case where the issues of uniqueness (or equivalently, minimality) and attainability of the bound in equation (17) can be settled. For this case, we are able to show that the appropriate Kraus operators to use in calculating $C_\Upsilon(\theta)$ are the ones forming the canonical Kraus decomposition induced by the fixed input state.

Consider a one-parameter quantum process $\mathcal{E}_\theta$ and input state $\rho_0 = |\psi_0\rangle\langle\psi_0|$ with canonical decomposition $\mathcal{E}_\theta(\rho_0) = \sum_k \Upsilon_k(\theta)\rho_0\Upsilon_k^\dagger(\theta)$. As mentioned in section 2.1, this decomposition is characterized by the orthogonality of the Kraus operators according to an inner product based on the particular state ρ_0: $\mathcal{E}_\theta(\rho_0) = \sum_k \Upsilon_k(\theta)\rho_0\Upsilon_k^\dagger(\theta)$ with $\mathrm{tr}\left(\Upsilon_k^\dagger(\theta)\Upsilon_j(\theta)\rho_0\right) = \delta_{jk}p_k$. Now, assume an additional orthogonality constraint on the canonical decomposition: $\mathrm{tr}\left(\Upsilon_j^\dagger(\theta)\Upsilon_k'(\theta)\rho_0\right) = \mu_k(\theta)\delta_{jk}, \mu_k \in \mathbb{R}$. In terms of the local basis on the curve, this constraint can be written as $\langle e_j(\theta)|\partial_\theta e_k(\theta)\rangle = \mu_k(\theta)\delta_{jk}$, which implies $\langle f_j(\theta)|\partial_\theta f_k(\theta)\rangle = \tilde{\mu}_k(\theta)\delta_{jk}$ where $\tilde{\mu}_k \in \mathbb{R}$ also. Note that this additional constraint specifies a *quasi-classical model* where the eigenbasis of the output density operator remains the same for all θ, and it is only the eigenvalues that change along the curve, i.e. a locally orthogonal basis can also be considered a globally orthogonal basis along the curve. The output density operators $\{\rho(\theta)\}_\theta$ form a commuting parametric family. In the following, we prove the attainability and uniqueness of the Fisher information bound for this special case.

Theorem 1. *Attainability: for the quasi-classical model, the optimality conditions*

$$E(\xi)^{\frac{1}{2}}\Upsilon_k'(\theta)\rho_0^{\frac{1}{2}} = \lambda_\xi(\theta)E(\xi)^{\frac{1}{2}}\Upsilon_k(\theta)\rho_0^{\frac{1}{2}} \quad \forall\xi, k, \tag{24}$$

where $\{\Upsilon_k\}$ are members of the canonical decomposition, can be met with a locally optimal strategy.

Proof. We will prove this by explicitly constructing the POVM that meets these conditions. Consider the optimality conditions when the input state is a pure state $|\psi_0\rangle$:

$$\begin{aligned}
&E_j\Upsilon_k(\theta)'|\psi_0\rangle = \lambda_j(\theta)E_j\Upsilon_k(\theta)|\psi_0\rangle \\
&E_j|\partial_\theta e_k\rangle = \lambda_j(\theta)E_j|e_k\rangle \\
&E_j\left(\frac{p_k'(\theta)}{2\sqrt{p_k(\theta)}}|f_k(\theta)\rangle + \sqrt{p_k(\theta)}|\partial_\theta f_k(\theta)\rangle\right) = \lambda_j(\theta)\sqrt{p_k(\theta)}E_j|f_k(\theta)\rangle
\end{aligned} \tag{25}$$

for all j, k (we are now assuming that the POVM has a discrete number of elements and are thus using a discrete index j). Now consider the choice $E_j = |f_j(\theta)\rangle\langle f_j(\theta)|$ for the POVM; the completeness condition on POVMs is automatically satisfied because $\{|f_j\rangle\}$ are eigenstates of a Hermitian operator and therefore span the space. Given this choice of POVM, the optimality conditions become

$$\left(\frac{p_k'}{2\sqrt{p_k}} + \sqrt{p_k}\tilde{\mu}_k - \lambda_j\sqrt{p_k}\right)|f_k\rangle\delta_{jk} = 0, \tag{26}$$

where we have suppressed θ because all quantities depend on it. This condition can be satisfied for all j, k by the choice $\lambda_k = \tilde{\mu}_k + p_k'/2p_k = \mu_k/p_k$, and thus in this special case we can construct the optimal POVM. However, note that this choice, $E_j = |f_j(\theta)\rangle\langle f_j(\theta)|$, presumes

knowledge of θ and therefore can only be implemented adaptively [5]. As mentioned above, given the set $\{|f_j(\tilde{\theta})\rangle\}$ for some $\tilde{\theta}$, the additional constraint, $\langle f_j(\theta)|\partial_\theta f_k(\theta)\rangle = \tilde{\mu}_k(\theta)\delta_{jk}$, ensures that this set remains *orthogonal* for all θ. Despite this, the POVM has to be adapted during the estimation because the normalization of the elements of the set varies with θ. $\square$

Theorem 2. *Uniqueness: for the quasi-classical model, the minimum of the Fisher information bound of equation (17) over the valid Kraus decompositions is achieved by the canonical decomposition.*

Proof. Firstly, by the preceding theorem, we know that for the quasi-classical model the optimality conditions can be satisfied for the canonical Kraus operators, that is, some POVM set $\{E(\xi)\}$ can be found such that conditions (21) are satisfied for the canonical Kraus operators, and thus the bound is achievable. Now, let C_Υ denote the Fisher information bound when the canonical Kraus decomposition is used. We will show the value of the bound using any other Kraus decomposition, i.e. C_Ω is larger than C_Υ. Consider C_Ω as given in equation (23). Using the identities $\sum_k u_{kj}^* u_{kl} = \delta_{jl}$ and $\mathrm{tr}\left(\Upsilon_k^\dagger(\theta)\Upsilon_j(\theta)\rho_0\right) = \delta_{jk}p_k$, we can rewrite this expression as

$$C_\Omega(\theta) = 4\left(\sum_j \mathrm{tr}\,\Upsilon_j^{\dagger\prime}\Upsilon_j'\rho_0 + \sum_{jk} |u_{kj}'|^2 p_j\right.$$
$$\left. + \sum_{jkl} u_{kj}^{*\prime} u_{kl}\,\mathrm{tr}\left[\Upsilon_j^\dagger\Upsilon_l'\rho_0\right] + \sum_{jkl} u_{kj}^* u_{kl}'\,\mathrm{tr}\left[\Upsilon_j^{\dagger\prime}\Upsilon_l\rho_0\right]\right). \tag{27}$$

Note that the first term is simply the Fisher information bound for the canonical Kraus decomposition, C_Υ, and the second term is always positive. Hence, we see that the question of whether $C_\Omega \geqslant C_\Upsilon$ depends on the sign of the third and fourth terms, i.e. we are interested in the sign of

$$G(\theta) \equiv \sum_{jkl} u_{kj}^{*\prime} u_{kl}\,\mathrm{tr}\left[\Upsilon_j^\dagger\Upsilon_l'\rho_0\right] + \sum_{jkl} u_{kj}^* u_{kl}'\,\mathrm{tr}\left[\Upsilon_j^{\dagger\prime}\Upsilon_l\rho_0\right]. \tag{28}$$

To determine this, use the optimality conditions of equation (21) to get rid of the derivatives within the trace. Explicitly, insert a resolution of identity of the form $\int \mathrm{d}\xi\,E(\xi)$:

$$G(\theta) \equiv \sum_{jkl} u_{kj}^{*\prime} u_{kl}\,\mathrm{tr}\left[\Upsilon_j^\dagger \int \mathrm{d}\xi\,E(\xi)\Upsilon_l'\rho_0\right] + \sum_{jkl} u_{kj}^* u_{kl}'\,\mathrm{tr}\left[\Upsilon_j^{\dagger\prime} \int \mathrm{d}\xi\,E(\xi)\Upsilon_l\rho_0\right]$$
$$= \int \mathrm{d}\xi\,\lambda_\xi(\theta) \sum_{jl} \mathrm{tr}\left[\Upsilon_j^\dagger E(\xi)\Upsilon_l\rho_0\right] \sum_k (u_{kj}^{*\prime} u_{kl} + u_{kj}^* u_{kl}')$$
$$= 0, \tag{29}$$

where the second line follows from the optimality conditions equation (21) and the third line follows from taking the derivative of the orthonormality condition on the elements of the unitary matrix $[u_{ij}(\theta)]$. Therefore,

$$C_\Omega(\theta) = 4\left(\sum_j \mathrm{tr}\,\Upsilon_j^{\dagger\prime}\Upsilon_j'\rho_0 + \sum_{jk} |u_{kj}'|^2 p_j\right) \geqslant C_\Upsilon(\theta) \quad \forall\theta,$$

and the minimum of equation (17) for the quasi-classical model is achieved by the canonical Kraus decomposition. $\square$

For this quasi-classical model, we can truly say that $C_\Upsilon(\theta) = F^*(\theta)$ (where $\{\Upsilon_k\}$ form the canonical decomposition) and hence we have a measure of the statistical distance along the curve formed by the one-parameter family of output states. In fact, we can express this metric on the curve explicitly in terms of the local coordinate system set up by the Kraus decomposition:

$$\left(\frac{ds}{d\theta}\right)^2 = F^*(\theta) = 4 \sum_k \langle \partial_\theta e_k(\theta) | \partial_\theta e_k(\theta) \rangle$$

$$= \sum_k \frac{(p_k(\theta)')^2}{p_k(\theta)} + 4 \sum_k p_k(\theta) |\langle f_k(\theta) | \partial_\theta f_k(\theta) \rangle|^2. \tag{30}$$

This result has operational significance. It means that if the input state and quantum channel are such that the output family is a mutually commuting one, then the optimal estimation scheme can be identified and the statistical distinguishability be calculated easily.

4. Examples

In this section we will consider several examples to illustrate the results of the previous sections.

4.1. Qubit depolarization channel

The qubit depolarization channel is defined as

$$\rho(p) = \mathcal{D}(\rho) = p\frac{\mathbb{I}}{2} + (1-p)\rho$$

$$= (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z), \tag{31}$$

where ρ is a density matrix in a Hilbert space of dimension 2 and X, Y, Z are the Pauli matrices. This channel can be best understood by examining its action on the Bloch sphere representation of a qubit; it has the effect of uniformly shrinking the Bloch sphere towards the centre. The parameter to estimate is the rate of this shrinking, parametrized by $0 \leqslant p \leqslant 1$.

From the definition of the channel, it is clear that $\{\rho(p)\}_p$ forms a commuting family. Thus, we can use the canonical decomposition of the channel to analyse its estimation.

The unitary invariance of this channel (spherical symmetry in the Bloch sphere picture) implies that all pure state inputs will perform identically when it comes to estimation performance. Therefore, choose $|\psi_0\rangle = |0\rangle$, the +1 eigenstate of Z. The canonical decomposition of the channel with respect to this initial state is

$$\Upsilon_1(p) = i\sqrt{\frac{p}{6}}X + \sqrt{\frac{p}{6}}Y \qquad \Upsilon_2(p) = \frac{q}{\sqrt{q + \frac{p}{3}}}\mathbb{I} + \frac{\frac{p}{3}}{\sqrt{q + \frac{p}{3}}}Z$$

$$\Upsilon_3(p) = \sqrt{\frac{p}{6}}X + i\sqrt{\frac{p}{6}}Y \qquad \Upsilon_4(p) = -\sqrt{\frac{q\frac{p}{3}}{\frac{p}{3} + q}}\mathbb{I} + \sqrt{\frac{q\frac{p}{3}}{\frac{p}{3} + q}}Z, \tag{32}$$

where $q = 1 - p$. The optimality conditions of equation (21) are easily seen to be satisfied by projective measurements onto the Z-basis, i.e. the POVM $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$. The statistical distance for this channel with the input state $|\psi_0\rangle = |0\rangle$ is given by $F^*(p) = \frac{6}{p(9-6p)}$, which can be achieved uniformly by the estimation strategy of measuring each output in the Z-basis. Note that this bound diverges at $p = 0$, but not at $p = 1$. This is because at $p = 1$ we still cannot distinguish perfectly between the action of the three Paulis with one qubit.
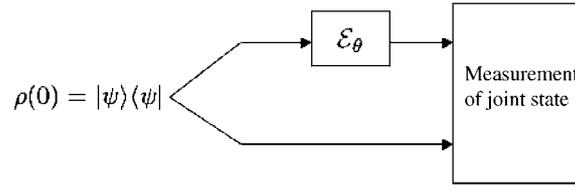
Figure 2. Estimating quantum channels using entanglement. If $\mathcal{E}_\theta$ is a quantum operation acting on operators in a Hilbert space of dimension d, then $\dim(|\psi\rangle)$ is at least d^2.

It is a well-known fact that using entanglement can improve estimation [13, 23]. To compare the performance of a scheme that uses entanglement to the one that does not, we can compare the statistical distinguishability for the two cases. Consider the estimation of the depolarizing channel using a maximally entangled state as the input into the channel $\mathbb{I} \otimes \mathcal{D}$; see figure 2. This is a common setup for estimating channels because it can be shown that the output state completely characterizes the channel [24]. The canonical decomposition of the channel $\mathbb{I} \otimes \mathcal{D}$ with respect to the maximally entangled input state $|\psi_0\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ consists of the Kraus operators:

$$
\begin{aligned}
\Upsilon_1(p) &= \sqrt{1-p}\, \mathbb{I} \otimes \mathbb{I} & \Upsilon_2(p) &= \sqrt{p/3}\, \mathbb{I} \otimes X \\
\Upsilon_3(p) &= \sqrt{p/3}\, \mathbb{I} \otimes Y & \Upsilon_4(p) &= \sqrt{p/3}\, \mathbb{I} \otimes Z.
\end{aligned}
\tag{33}
$$

Given this, it is easy to show that the optimality conditions of equation (21) can be satisfied by a POVM formed by projectors onto the Bell basis[2]. Note that if the singlet state $|\phi^-\rangle$ is used as the input state, then the measurement scheme need only discriminate between the singlet and triplet subspaces to be optimal. The statistical distance in this case is $F_e^*(p) = \frac{1}{p(1-p)}$. Figure 3 plots the two values of statistical distance and clearly shows the improved distinguishability of the parameter (uniformly across p) that the entanglement-assisted scheme affords.

This example illustrates the effect of using entanglement for estimation; it can have the advantage of increasing the statistical distinguishability of channels. It also illustrates the ability of this formalism to treat entangled input states and non-local measurements. Such scenarios simply change the definition of the channel to a suitable tensor product of single party channels, while the statistical distance and optimality conditions retain their form.

4.2. Estimation of pure T_2 qubit dephasing time

There has been considerable interest recently in accurately estimating single qubit T_2 relaxation times for various quantum computing architectures [25–27]. Such estimations have also been commonplace in the NMR community for several decades now. We can use the formalism developed above to investigate the schemes used for estimating T_2.

If we restrict the dynamics to be purely dephasing, we can model the single qubit channel as

$$
\frac{d\rho}{dt} = \gamma(Z\rho Z - \rho),
\tag{34}
$$

---

[2] The Bell states are four maximally entangled states of two qubits: $|\psi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, $|\psi^-\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$, $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$, $|\phi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. These four states span the Hilbert space of two qubits and are therefore called the *Bell basis*. The symmetric subspace of two-qubit Hilbert space is spanned by the *triplet* states $|\psi^\pm\rangle$ and $|\phi^+\rangle$, and the anti-symmetric subspace is spanned by the *singlet* state $|\phi^-\rangle$.
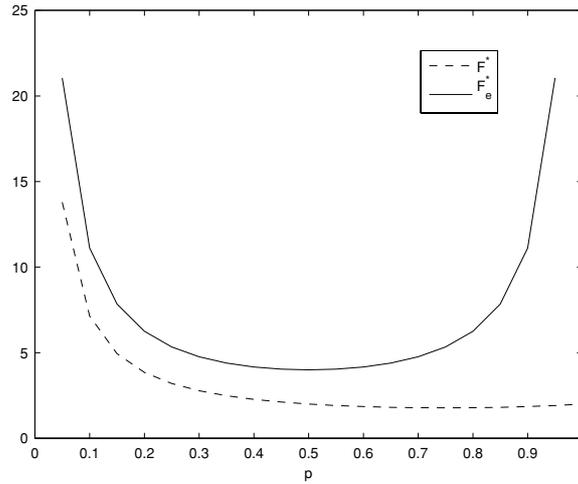
Figure 3. Statistical distinguishability of the depolarizing channel with and without the use of entanglement during estimation. The values of F^* at $p = 0$ and F_e^* at $p = 0, 1$ are not plotted because the quantities diverge at those points.

where γ is the dephasing rate and the parameter we are trying to estimate. A Kraus decomposition for this process is

$$\rho(\theta) = \mathcal{D}_Z(\rho) = \frac{1 + e^{-2\theta}}{2} \rho + \frac{1 - e^{-2\theta}}{2} Z\rho Z, \tag{35}$$

where $\theta = \gamma t$ is a simple transformation of the parameter we want to estimate. In the following, we will denote the single-qubit equal superposition states by $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. Note that these are eigenstates of the operator X and thus will be collectively referred to as the X-basis.

The standard techniques for estimating T_2 times are based on the *spin echo* [28] pulse sequence which is essentially the preparation of a $|+\rangle$ or $|-\rangle$ initial state and then a measurement in the X-basis after the channel has acted. This pulse sequence actually has added features designed to nullify bulk sample inhomogeneities, but the basic idea is as mentioned. It is easy to check that given this input state, equation (35) is the canonical decomposition for this channel and also that the output family $\{\rho(\theta)\}_\theta$ is a mutually commuting one. Thus, we can use the operators in the canonical decomposition to determine the optimality of this scheme by checking the optimality conditions (21), which turn out to be

$$(e^{-2\theta} + \lambda_j(\theta)[1 + e^{-2\theta}])E_j^{1/2}|+\rangle = 0$$
$$(e^{-2\theta} - \lambda_j(\theta)(1 - e^{-2\theta}))E_j^{1/2}|-\rangle = 0 \tag{36}$$

for all j, where we are using a discrete number of POVMs indexed by j. These two conditions can be met by a measurement in the X-basis, that is, with a POVM $\{E_+ = |+\rangle\langle+|, E_- = |-\rangle\langle-|\}$. The choices required for λ are $\lambda_+ = -1/(e^{2\theta} + 1)$ and $\lambda_- = 1/(e^{2\theta} - 1)$. So we see that the standard spin echo technique of estimating single qubit pure dephasing time, T_2, is indeed an optimal one for an X-basis input state.

Again, it is possible to show that using entanglement helps in the estimation of the parameter θ for this channel. However, a channel extension of the form $\mathbb{I} \otimes \mathcal{D}_Z$ does not increase the statistical distance; instead, a channel extension of the form $\mathcal{D}_Z \otimes \mathcal{D}_Z$ must be used with a maximally entangled state.

4.3. The random shift channel

The *random shift* channel is defined by the master equation

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \gamma \left(U_\beta \rho U_\beta^\dagger - \rho \right), \tag{37}$$

where γ is a real, positive constant and U_β is a unitary operator: $U = \mathrm{e}^{-\mathrm{i}\beta H}$ for some continuous spectrum Hermitian operator H and some real number β. A Kraus decomposition for this channel is

$$\Lambda_k(\theta) = \frac{\theta^{k/2}}{\sqrt{k!}} \, \mathrm{e}^{-\theta/2} U_\beta^k, \qquad k = 0, 1, 2, \ldots, \tag{38}$$

where $\theta = \gamma t$. These equations describe a channel that delivers a Poisson distributed number of unitary displacements (or unitary 'kicks') by β to the input state. The average number of kicks in a time t is given by $\theta = \gamma t$ and is the parameter we are estimating.

The way to optimally estimate the unitary version of this channel $\rho \to U_\beta \rho U_\beta^\dagger$ is to use the fact that H is a generator of translations in some basis [7, 11]. That is, if $|x\rangle$ is an eigenstate of an operator conjugate to H, then $U|x\rangle = \mathrm{e}^{-\mathrm{i}\beta H}|x\rangle = |x + \beta\rangle$. Then the optimal scheme is to input a fiducial state $|x_0\rangle$ and to use a POVM that is formed from projectors onto translated versions of this state $\{|x\rangle : |x\rangle = U_\beta |x_0\rangle, \beta \in \mathbb{R}\}$ ($\langle x|x'\rangle = \delta(x - x')$). For example, if $H = \hat{p}$, the momentum operator (and hence U is a spatial translation), then we would choose E and ρ_0 to be projectors onto position eigenstates.

Since U_β is simply a representation of an Abelian group and $U_\beta U_{\beta'} = U_{\beta+\beta'}$, we would expect the optimal scheme for estimating the random shift channel to be the same as that for estimating the unitary shift channel. Note that when the input state is a fiducial state $|x_0\rangle$, which is translated by U_β, the Kraus operators given by equation (38) form the canonical decomposition. Additionally, the model is quasi-classical because the output family is a commuting one. Therefore, we can check the optimality of using the unitary channel estimation scheme for estimating the random shift channel by examining the optimality conditions with the canonical Kraus operators. $\Lambda_k(\theta)' = \left(-\frac{1}{2} + \frac{k}{2\theta} \right) \Lambda_k(\theta)$, and when the input state is $|x_0\rangle$, the optimality conditions become

$$\left(-\frac{1}{2} + \frac{k}{2\theta} - \lambda_\xi(\theta) \right) E^{1/2}(\xi) \Lambda_k(\theta)|x_0\rangle = 0 \qquad \forall \xi, k. \tag{39}$$

Now, choosing $E(\xi) \equiv E(x) = |x\rangle\langle x|$, we get

$$\left(-\frac{1}{2} + \frac{k}{2\theta} - \lambda_x \right) \delta(x - (x_0 + k\beta))|x\rangle = 0 \qquad \forall x, k. \tag{40}$$

The left-hand side is zero except when $x = x_0 + k\beta$, and in that case we can choose $\lambda_x = \frac{1}{2} - \frac{x - x_0}{2\theta\beta}$, so that the left-hand side goes to zero in all cases. Therefore, as in the unitary case, using a POVM formed of projectors onto shifted states is optimal when the input is an element of this same set. The statistical distance for this estimation scheme is $F^*(\theta) = 1/\theta$.
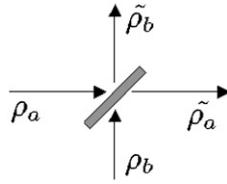
Figure 4. The beam splitter interpretation of the damping channel.

4.4. The damping channel

As our final example of a one-parameter quantum process, we consider the harmonic oscillator damping channel (DC). This is a quantum process described by the master equation

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \gamma \left(\hat{a}\rho\hat{a}^{\dagger} - \frac{1}{2}(\hat{a}^{\dagger}\hat{a}\rho + \rho\hat{a}^{\dagger}\hat{a}) \right), \qquad (41)$$

where $\hat{a}^{\dagger}$ and $\hat{a}$ are creation and annihilation operators for a harmonic oscillator mode, and γ is a real, positive constant. This channel describes the effects of random photon loss. An operator sum decomposition for this process can be obtained by expanding the above master equation as a Dyson series and solving. This yields the following Kraus operators:

$$\Delta_k(\theta) = \frac{(1 - \mathrm{e}^{-\theta})^{k/2}}{\sqrt{k!}} \, \mathrm{e}^{-\frac{\theta}{2}\hat{a}^{\dagger}\hat{a}}\hat{a}^{k}, \qquad k = 0, 1, 2, \ldots, \qquad (42)$$

where $\theta = \gamma t$ is the parameter to be estimated for this channel. Note that the state space of $\rho(\theta)$ is infinite dimensional and there is also an infinite number of Kraus operators.

One interpretation of this quantum operation is that it describes the transformation of a state when combined with the vacuum at a beam splitter (see figure 4). That is, the state of mode a after the beam splitter is given by

$$\tilde{\rho}_a = \mathrm{tr}_b \, U(\phi)(\rho_a \otimes |0\rangle_b\langle 0|)U^{\dagger}(\phi),$$

where $U(\phi) = \exp(-\phi(\hat{a}^{\dagger}\hat{b} - \hat{a}\hat{b}^{\dagger}))$ is the beam splitter unitary transformation with $\hat{a}$ and $\hat{b}$ being the annihilation operators for modes a and b, respectively. Evaluating this trace gives the same CP map as the damping channel with $\mathrm{e}^{-\theta}$ replaced by $\cos^2\phi$, the intensity transmittance of the beam splitter. Therefore, our estimation task is equivalent to the estimation of the transmittance of a beam splitter.

A common method for probing such a channel would be with a Fock (photon number) state $|\psi_0\rangle = |N\rangle$, where $N = 0, 1, 2, \ldots$ is the number of photons in the mode. Similarly, common measurement techniques at the channel output would be photodetection, heterodyne or homodyne measurements.

To determine the optimal measurement strategy when the photon number states are used as input, we again confirm that firstly, the Kraus decomposition defined by equation (42) is indeed the canonical one when number states are used and, secondly, that the channel is quasi-classical with such input states. Therefore, we can decide on the optimal POVM by looking at the optimality conditions equation (21). $\Delta'_k(\theta) = \left(\frac{k\mathrm{e}^{-\theta}}{2(1-\mathrm{e}^{-\theta})} - \frac{1}{2}\hat{a}^{\dagger}\hat{a} \right)\Delta_k(\theta)$, and hence the conditions become

$$\frac{k\,\mathrm{e}^{-\theta}}{2(1 - \mathrm{e}^{-\theta})} E^{1/2}(\xi)\Delta_k(\theta)|N\rangle - \frac{1}{2}E^{1/2}(\xi)\hat{a}^{\dagger}\hat{a}\Delta_k(\theta)|N\rangle = \lambda_\xi(\theta)E^{1/2}(\xi)\Delta_k(\theta)|N\rangle, \qquad (43)$$

for all ξ and k. We can simplify this by applying all operators except $E(\xi)$ to the input state. Note that for $k > N$ the application of $\Delta_k(\theta)$ yields zero and the condition is trivially satisfied. For $k \leqslant N$ we have the condition that for all ξ,

$$\left(\frac{k\,e^{-\theta}}{2(1-e^{-\theta})} - \frac{N-k}{2} - \lambda_\xi(\theta) \right) E^{1/2}(\xi)|N-k\rangle = 0. \tag{44}$$

To satisfy this we can choose $E(\xi) \equiv E(M) = |M\rangle\langle M|$, a number state projector, which corresponds to photodetection. This choice requires $\lambda_M = \frac{Ne^{-\theta}-M}{2(1-e^{-\theta})}$. Therefore, the best strategy for estimating the damping channel when the Fock states are used as input is to perform photodetection at the output. The statistical distance for this scenario is $F_N^*(\theta) = \frac{N}{e^\theta-1}$, and hence larger N at the input makes the process more distinguishable.

4.5. A comment on estimation

In all the examples considered above, except the random shift channel, we are trying to estimate a continuous parameter with experiments that have discrete outcomes. This may seem peculiar, but the situation is clarified by the observation that the parameter is always a continuous function of the probabilities of the discrete outcomes (or rather, the probabilities are functions of the parameter). Therefore, from the point of view of the estimator, the problem is the same as estimating the parameter of a probability distribution from independent experiments that sample that distribution. This is a well-known estimation problem in classical estimation theory, and a *maximum likelihood estimator* [22] would be a practical estimator that would also achieve the Fisher information bound asymptotically.

5. Conclusion

We have investigated the problem of optimally estimating a general one-parameter quantum process. We have attempted to obtain characterizations of estimation optimality in terms of a common representation of quantum processes, the Kraus decomposition. We derived a bound on estimation accuracy and conditions of optimality when the input state is fixed; however, the non-uniqueness of the Kraus decomposition causes this bound to be non-unique. It also makes proving the achievability of the bound difficult. However, we have shown that in the special case of a quasi-classical channel, the issues of uniqueness and attainability can be settled, and in this special case, the characterization of optimal estimation (with a fixed input state) we derived is useful in determining the statistical distinguishability of quantum processes.

Representing a quantum process in terms of its Kraus decomposition has the advantage that it is often easy to do; however, in view of the above treatment, this representation is difficult for characterizing optimal estimation strategies because of its non-uniqueness. The immediate direction in which this work could be extended is to investigate the possibility of settling the issues of attainability and uniqueness for a general quantum channel. In particular, explicit expressions for the optimal POVM from the optimality conditions equation (21), for a fixed set of Kraus operators, would be extremely useful.

Acknowledgments

acknowledge useful discussions with Duncan Mortimer, Michael Nielsen, Charles Hill and Carl Caves.

References

[1] Helstrom C V 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
[2] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland)
[3] Braunstein S and Caves C 1994 Statistical distance and the geometry of quantum states *Phys. Rev. Lett.* **72** 3439
[4] Fujiwara A and Nagaoka H 1995 Quantum Fisher metric and estimation for pure state models *Phys. Lett.* A **201** 119
[5] Barndorff-Nielsen O E and Gill R D 2000 Fisher information in quantum statistics *J. Phys. A: Math. Gen.* **30** 4481
[6] Gill R D and Massar S 2000 State estimation for large ensembles *Phys. Rev.* A **61** 042312
[7] Holevo A S 2004 Asymptotic estimation of shift parameter of a quantum state *Prob. Theory Appl.* **49** 207
[8] Paris M and Rehacek J (ed) 2004 *Quantum State Estimation* (*Lecture Notes in Physics,* vol 649) (Berlin: Springer)
[9] Wootters W K 1981 Statistical distance and Hilbert space *Phys. Rev.* D **23** 357
[10] Barndorff-Nielsen O E, Gill R D and Jupp P E 2003 On quantum statistical inference *J. R. Stat. Soc.* B **65** 775
[11] Braunstein S, Caves C and Milburn G J 1994 Generalized uncertainty relations: theory, examples, and Lorentz invariance *Ann. Phys.* **247** 135
[12] Ballester M A 2004 Estimation of unitary quantum operations *Phys. Rev.* A **69** 022303
[13] Fujiwara A 2001 Quantum channel identification problem *Phys. Rev.* A **63** 042304
[14] Fujiwara A and Hiroshi I 2003 Quantum parameter estimation of a generalized Pauli channel *J. Phys. A: Math. Gen.* **36** 8093
[15] Sasaki M, Ban M and Barnett S M 2002 Optimal parameter estimation of depolarizing channel *Phys. Rev.* A **66** 022308
[16] Choi M D 1975 Completely positive linear maps on complex matricies *Linear Algebra Appl.* **10** 285
[17] Kraus K 1983 *States, Effects, and Operations: Fundamental Notions of Quantum Theory* (*Lecture Notes in Physics,* vol 190) (Berlin: Springer)
[18] Nielsen M A and Chuang I L 2001 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[19] Nielsen M A, Caves C M, Schumacher B and Barnum H 1998 Information-theoretic approach to quantum error correction and reversible measurement *Proc. R. Soc.* A **454** 257–486
[20] Benatti F and Floreanini R 2005 Open quantum dynamics: complete positivity and entanglement *Int. J. Mod. Phys.* B **19** 3063
[21] Cover T M and Thomas J A 1991 *Elements of Information Theory* (New York: Wiley-Interscience)
[22] Nahi N E 1969 *Estimation Theory and Applications* (New York: Wiley)
[23] D'Ariano G M, Lo Presti P and Paris M G A 2001 Using entanglement improves the precision of quantum measurements *Phys. Rev. Lett.* **87** 270404
[24] Arrighi P and Patricot C 2004 On quantum operations as quantum states *Ann. Phys.* **311** 26
[25] Tyryshkin A M, Lyon S A, Astashkin A V and Raitsimring A M 2003 Electron spin relaxation times of phosphorus donors in silico *Phys. Rev.* B **68** 193207
[26] Khaetskii A V, Loss D and Glazman L 2002 Electron spin decoherence in quantum dots due to interaction with nuclei *Phys. Rev. Lett.* **88** 186802
[27] Vion D, Aassime A, Cottet A, Joyez P, Pothier H, Urbina C, Esteve D and Devoret M H 2002 Manipulating the quantum state of an electrical circuit *Science* **296** 886
[28] Hahn E 1950 Spin echoes *Phys. Rev.* **80** 580