

Only Classical Parameterised States have Optimal Measurements under Least Squares Loss

Wilfred Salmon^{1,2,3}, Sergii Strelchuk¹, and David R. M. Arvidsson-Shukur²

¹DAMTP, Centre for Mathematical Sciences, University of Cambridge, Cambridge CB30WA, UK

²Hitachi Cambridge Laboratory, J. J. Thomson Avenue, CB3 0HE, Cambridge, UK

³was29@cam.ac.uk

Measurements of quantum states form a key component in quantum-information processing. It is therefore an important task to compare measurements and furthermore decide if a measurement strategy is optimal. Entropic quantities, such as the quantum Fisher information, capture asymptotic optimality but not optimality with finite resources. We introduce a framework that allows one to conclusively establish if a measurement is optimal in the non-asymptotic regime. Our method relies on the fundamental property of expected errors of estimators, known as risk, and it does not involve optimisation over entropic quantities. The framework applies to finite sample sizes and lack of prior knowledge, as well as to the asymptotic and Bayesian settings. We prove a no-go theorem that shows that only classical states admit optimal measurements under the most common choice of error measurement: least squares. We further consider the less restrictive notion of an approximately optimal measurement and give sufficient conditions for such measurements to exist. Finally, we generalise the notion of when an estimator is inadmissible (i.e. strictly worse than an alternative), and provide two sufficient conditions for a measurement to be inadmissible.

1 Introduction

Extracting information from systems requires performing a measurement. The outcome of the measurement is then used to infer information about some property of the system of interest. Parameter estimation is the well-established study of different strategies of measurement and the subsequent inference of some number of parameters of a system. In particular, when the underlying system is quantum then understanding the non-classical features of quantum measurement becomes of paramount importance and forms the central object of study in quantum metrology.

The asymptotic regime of such problems, corresponding to an infinite number of measurements, has been

extensively studied [1, 2]. This regime is also known as local estimation, as one only considers infinitesimal changes in parameter values. The quantum Fisher information [1, 3] has emerged as the standard metric of asymptotic information for a parametrised quantum state. Its operational significance is emphasised by the quantum Cramér-Rao bound [1, 3] where its inverse lower-bounds the variance of locally unbiased estimators. However, asymptotically optimal measurements, as characterised by the quantum Fisher information, may be sub-optimal in finite-resource settings (see below). Additionally, in practice, parameter estimation is a global problem; one must consider parameters across the whole parameter space. Global estimation is often studied within the Bayesian regime, where one has some *a priori* belief about the distribution of the parameter. Optimal measurement choices in the Bayesian regime are well understood [4, 5], however their optimality relies on an accurate choice of prior. To the best of our knowledge, a thorough analysis of measurement optimality in the most general setting of global parameter estimation without prior knowledge is missing.

In this paper, we provide such an analysis. Our methodology relies on the fundamental property of expected errors of estimators, known as risk, defined below. We establish a fundamental result of quantum-information processing: *only* classical parameterised states [1], those that are a classical mixture of some parameter-independent pure states, admit optimal measurements. We go on to propose three ways in which a measurement can be approximately optimal, and we prove that if a state is close to being classical, there exists an approximately optimal measurement. Finally, we consider when one measurement dominates (is strictly better than) another measurement. We present two “bad” classes of measurements—those that extract no information from a system and those that can be fine-grained. We show that these measurements are always dominated by suitable different measurements, which we construct.

arXiv:2205.14142v2 [quant-ph] 3 May 2023

2 Technical Background

We begin by reviewing classical parameter estimation [6]. As an example, suppose one wishes to conduct an experiment to estimate $q \in [0, 1]$, some unknown proportion of a population satisfying some property. We sample some subset of n members of the population and observe a realisation of a binomially distributed random variable $X \sim B(n, q)$. We do not, *a priori*, know X 's distribution, only that it will be binomially distributed according to the parameter q . Based off a realisation x of X , one gives an estimate of q , most commonly x/n . In the general setting, one wishes to estimate some parameter θ , which could be any of the elements of the parameter space Θ . We observe the outcome of some random variable X , whose distribution is unknown to us, but is contained in the set $\{P_\theta : \theta \in \Theta\}$. An estimator of θ is then a function $\hat{\theta}(x)$, where $\hat{\theta}(x)$ encodes our estimate of the underlying parameter θ in the case that X takes the value x .

To discuss the merit of estimators, one introduces a loss function L , where $L(\theta_1, \theta_2)$ quantifies how bad a guess of θ_1 if the true parameter were θ_2 . Common examples include least squares $L(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|^2$, the Kullback-Leibler divergence [7] and the Bhattacharyya distance [8]. The loss function induces a risk function for an estimator:

$$R(\hat{\theta}, \theta) = \mathbb{E}_{X \sim P_\theta} [L(\hat{\theta}(X), \theta)]. \quad (1)$$

$R(\hat{\theta}, \theta)$ measures how badly one expects an estimator to perform for a fixed value of the underlying parameter θ . An estimator is ‘‘good’’ at the point θ if it has a low risk there. We say that $\hat{\theta}$ dominates $\tilde{\theta}$ if $\hat{\theta}$ is always at least as good as $\tilde{\theta}$ and sometimes better, i.e.,

$$\begin{aligned} \forall \theta \in \Theta, R(\hat{\theta}, \theta) &\leq R(\tilde{\theta}, \theta). \\ \text{and } \exists \theta \in \Theta \text{ s.t. } R(\hat{\theta}, \theta) &< R(\tilde{\theta}, \theta). \end{aligned} \quad (2)$$

If $\tilde{\theta}$ is dominated by some estimator $\hat{\theta}$ one says that it is inadmissible, otherwise it is said to be admissible. Admissibility is a very weak condition; constant estimators—always guessing θ_0 regardless of the outcome of the experiment—are often admissible [9], since they have zero risk at θ_0 . However, admissibility is certainly desirable, or, by definition, one can choose a nominally ‘‘better’’ estimator.

In the quantum setting [10, 11], our parameter is encoded in some quantum state $\rho(\theta)$. Often, one may have n copies of the same state $\sigma(\theta)$ in which case $\rho(\theta) = \sigma(\theta)^{\otimes n}$. We choose some (parameter independent) generalised measurement $M = \{M_k\}$, where each M_k is a linear operator satisfying $M_k \geq 0$, and $\sum_k M_k = \mathbb{1}$. This gives rise to a probability distribution parametrised by θ :

$$p_k^{(M)}(\theta) = \text{Tr}[\rho(\theta)M_k]. \quad (3)$$

Thus, after measurement, the quantum experiment reduces to a classical parameter-estimation problem and one can pick an estimator $\hat{\theta}_M(k)$ using the probability distribution in Eq. (3). Hence, the objects of interest in quantum parameter estimation are pairs $(M, \hat{\theta}_M)$, where M is a generalised measurement and $\hat{\theta}_M$ is an estimator based off the outcome of M . Note that an experimenter may choose a POVM without any particular estimator $\hat{\theta}_M$ in mind. However when the experimental data is eventually used to estimate the parameter, the way in which the data is processed exactly corresponds to picking an estimator. Often, when it is clear from context, we will omit the M label of the estimator. We can compare the performance of estimators based on different measurements. In particular, for two pairs $(M, \hat{\theta}_M)$ and $(F, \tilde{\theta}_F)$ we write $\hat{\theta}_M \leq \tilde{\theta}_F$ to mean that $\hat{\theta}_M$ is always at least as good as $\tilde{\theta}_F$, i.e. $\forall \theta \in \Theta, R(\hat{\theta}_M, \theta) \leq R(\tilde{\theta}_F, \theta)$. We say that $\hat{\theta}_M$ dominates $\tilde{\theta}_F$, written $\hat{\theta}_M < \tilde{\theta}_F$, if $\hat{\theta}_M \leq \tilde{\theta}_F$ and $\hat{\theta}_M$ is sometimes better, i.e. $\exists \theta \in \Theta$ s.t. $R(\hat{\theta}_M, \theta) < R(\tilde{\theta}_F, \theta)$, in close analogy to equation (2).

Before we introduce our notion of measurement optimality, we briefly discuss how the commonly used notion, based on the (quantum) Fisher information $\mathcal{F}(\theta)$ ¹, is, in general, unsatisfactory. $\mathcal{F}(\theta)$ is an entropic quantity that characterises the performance of estimation protocols in the asymptotic limit: asymptotic risk is lower-bounded by $\text{Tr}(\mathcal{F}(\theta)^{-1})$. We now give two examples where $\mathcal{F}(\theta)$ does not capture optimality. First, consider estimating the mean of three or more normally distributed variables. The most common choice of estimator, the maximum likelihood estimator, has risk everywhere equal to $\text{Tr}(\mathcal{F}(\theta)^{-1})$. Thus, based on the Fisher information, it is asymptotically optimal. However, the James-Stein estimator [12] always has strictly smaller risk (it dominates the MLE). Secondly, suppose one can query a unitary $U(\theta) = \text{diag}\{1, e^{i\theta}\}$ a number N times to estimate θ . A strategy that maximises $\mathcal{F}(\theta)$, applies the unitary N times to the $|+\rangle$ state and measures in the $\{|\pm\rangle\}$ basis [13]. However, in this scenario, one only receives one bit of information so that the risk of $\hat{\theta}$ does not scale with N . In the optimal case, (e.g. the quantum phase estimation algorithm [14]) the risk scales as $1/N$.

3 Optimal Measurements

Below, we extend the aforementioned framework for estimator comparisons to include quantum measurements. Here, we define what we argue is the only pos-

¹For the ‘‘single parameter’’ case $\Theta \subseteq \mathbb{R}$, the quantum Fisher information $\mathcal{F}(\theta)$ is defined as $\text{Tr}(\rho(\theta)\Lambda(\theta)^2)$. Here, $\Lambda(\theta)$ is the symmetric logarithmic derivative, defined implicitly by the equation $2\partial_\theta \rho(\theta) = \Lambda(\theta)\rho(\theta) + \rho(\theta)\Lambda(\theta)$. For the ‘‘multiparameter case’’ of $\Theta \subseteq \mathbb{R}^N$ with $N > 1$ there are multiple definitions of the quantum Fisher information; for a review see [1].

sible such extension, unless further modelling assumptions are made. We begin by extending the notion of admissibility of estimators to measurements. We say that a generalised measurement M is at least as good as a generalised measurement F , written $M \preceq F$ if

$$\forall \hat{\theta}_F \exists \tilde{\theta}_M \text{ s.t. } \tilde{\theta}_M \leq \hat{\theta}_F. \quad (4)$$

If $M \preceq F$, then if one swaps an experiment's measurement from F to M , then whatever estimation strategy $\hat{\theta}_F$ was in use before, one can pick a new one $\hat{\theta}_M$ that is at least as good as $\hat{\theta}_F$ regardless of the true underlying parameter. If, however, $M \not\preceq F$, then one cannot do so—by swapping to M , there will be at least one estimation strategy $\hat{\theta}_F$ such that whatever new strategy $\hat{\theta}_M$ one picks there are some values of the underlying parameter where one would expect to do worse. Thus, our definition is the natural and only way one can define M as being at least as good as F without further assumptions on the estimation problem. We say that M dominates F , written $M \prec F$, if $M \preceq F$, but $F \not\preceq M$. In analogue with the classical case, we say that a measurement M is admissible if no other measurement dominates it. Clearly, as in the classical case, admissibility of a measurement is strongly desirable.

There is then a natural definition of an optimal measurement: we say that M is optimal if for all other possible measurements F , $M \preceq F$. This is the strongest and most general definition of optimality one could make—we emphasise that any other definition must make some additional modelling assumptions (e.g., some *a priori* belief about the unknown parameter). Nonetheless, and somewhat surprisingly, we show that optimal measurements exist for a family of parameter estimation problems.

Theorem 3.1. *Let $\Theta \subseteq \mathbb{R}^N$ and let L be a choice of loss function that is convex in its first argument. Suppose that $\rho(\theta) = \sum_i p_i(\theta) |i\rangle\langle i|$, for some probability distribution $\{p_i(\theta)\}$ and a fixed, θ -independent, basis $\{|i\rangle\}$, then the measurement $M = \{|i\rangle\langle i|\}$ is optimal.*

Proof. Suppose $F = \{F_k\}$ is some other measurement with an estimator $\hat{\theta}_F$. Let $m_{k,i} = \text{Tr}(F_k |i\rangle\langle i|)$. Note that $m_{k,i} \geq 0$, $\sum_k m_{k,i} = 1$ and $\sum_i p_i(\theta) m_{k,i} = p_k^{(F)}(\theta)$. Define an estimator using M by $\hat{\theta}_M(i) = \sum_k m_{k,i} \hat{\theta}_F(k)$. Then

$$R(\hat{\theta}_M, \theta) = \sum_i p_i(\theta) L\left(\sum_k m_{k,i} \tilde{\theta}_F(k), \theta\right), \quad (5)$$

$$\leq \sum_{i,k} p_i(\theta) m_{k,i} L(\tilde{\theta}_F(k), \theta) = R(\tilde{\theta}_F, \theta). \quad (6)$$

Consequently, by swapping $\tilde{\theta}_F$ to $\hat{\theta}_M$ we can only decrease the risk and thus $M \preceq F$. \square

Note that the parameter may be a vector ($N > 1$), so that Theorem 3.1 holds in the multi-parameter

setting. States where the eigenbasis of the quantum state may be expressed in a parameter independent way, are called classical parametrised states [1]. More precisely, a parametrised quantum state $\rho(\theta)$ is called classical if there exists a θ -independent eigenbasis $\{|i\rangle\}$ and probabilities $\{p_i(\theta)\}$ such that $\rho(\theta) = \sum_i p_i(\theta) |i\rangle\langle i|$. For example, classical states are relevant for estimating temperature, where the system is in a (thermal) state:

$$\rho(\beta) = \frac{e^{-\beta H}}{\text{Tr}(e^{-\beta H})}. \quad (7)$$

Theorem 3.1 shows that measuring the energy of the system is always an optimal strategy, regardless of the ensuing choice of estimator.

As a second example, we consider estimating the strength of depolarizing noise acting on a fixed quantum state. Suppose that depolarising noise acts on some d -dimensional pure state $|\psi\rangle$ with probability² $p \in [0, 1]$ to produce the mixed state:

$$\rho(p) = (1 - p) |\psi\rangle\langle\psi| + p \frac{\mathbb{1}}{d}. \quad (8)$$

Theorem 3.1 shows that measuring in any basis containing $|\psi\rangle$ is optimal for estimating p .

It is natural to ask whether there exist any parametrised quantum states that are not classical, but still admit an optimal measurement. We give a partial converse to Theorem 3.1, in the case that the loss function is least squares.

Theorem 3.2. *Suppose $\rho(\theta)$ is some parameter estimation problem, where $\Theta \subseteq \mathbb{R}^N$ is convex. Unless $\rho(\theta)$ is classical, there is no optimal measurement M under least-squares loss.*

The proof of this Theorem is given in Appendix B. However, here we provide a sketch of the proof.

Proof Sketch. We assume the existence of an optimal measurement M and deduce that $\rho(\theta)$ must be classical. We first restrict our analysis to the single-parameter ($N = 1$) case. We fix $\theta_1, \theta_2 \in \Theta$ and consider the restricted parameter space $\{\theta_1, \theta_2\} \subseteq \Theta$. We then consider the Bayesian scenario of being given the states $\rho(\theta_1)$ with probability p and $\rho(\theta_2)$ with probability $1 - p$. We show that if an optimal measurement exists for this Bayesian parameter-estimation problem, then $\rho(\theta_1)$ and $\rho(\theta_2)$ commute. However, we also show that M must be optimal for any such Bayesian problem. Thus, we deduce that if an optimal measurement exists on Θ , all $\{\rho(\theta) : \theta \in \Theta\}$ must commute. We use this to simultaneously diagonalise the $\{\rho(\theta)\}$ and thus we deduce that ρ is classical. By restricting to lines in higher dimension parameter spaces, we

²The full range of p for which this is a valid state is $p \in [0, d/(d-1)]$. However, for $p > 1$ the state does not represent realistic noise models.

generalise the single parameter case to the multiparameter case. \square

Thus, for $\rho(\theta)$ non-classical, there is no optimal measurement strategy. Whatever measurement M one picks, there is always a measurement estimator pair $(F, \hat{\theta}_F)$ that one cannot guarantee to do at least as well as. Some trade-off or additional modelling assumption is required to decide on the best measurement strategy.

To develop physical intuition for Theorems 3.1 and 3.2, it is useful to consider distinguishability of states. For σ and ν quantum states, the optimal measurement to distinguish them [10] is with respect to the eigenbasis of $\sigma - \nu$. If $\rho(\theta)$ is a classical parametrised state, then for any $\theta_1, \theta_2 \in \Theta$ the eigenbasis of $\rho(\theta_1) - \rho(\theta_2)$ is constant and thus the same measurement is optimal for distinguishing all possible pairs of states. Since it is pairwise optimal, it is intuitive that this measurement is then ‘‘globally’’ optimal for parameter estimation. For non-classical states, different measurements are better at distinguishing different parts of the parameter space. This effect is best demonstrated by the ‘‘most non-classical’’ parametrised state: a pure state. For $\theta \in [0, 2\pi)$, the Mach-Zehnder state [11] is

$$|\psi(\theta)\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i\theta}|1\rangle). \quad (9)$$

Consider a measurement M of the state defined by measuring with respect to the basis $\{|+\rangle, |-\rangle\}$. M saturates the quantum Fisher information (for all θ) [11] and is thus a good candidate for an optimal measurement. It is easy to check that the outcome probability distribution is

$$p_+(\theta) = \cos^2(\theta/2), \quad p_-(\theta) = \sin^2(\theta/2). \quad (10)$$

Consider a different measurement F , defined by measuring with respect to the basis

$$\begin{aligned} |e_1\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + e^{i\pi/4}|1\rangle), \\ |e_2\rangle &= \frac{1}{\sqrt{2}}(e^{-i\pi/4}|0\rangle - |1\rangle). \end{aligned} \quad (11)$$

Measuring with respect to F gives rise to a probability distribution of

$$\begin{aligned} p_1(\theta) &= \cos^2\left[\frac{1}{2}\left(\theta - \frac{\pi}{4}\right)\right], \\ p_2(\theta) &= \sin^2\left[\frac{1}{2}\left(\theta - \frac{\pi}{4}\right)\right]. \end{aligned} \quad (12)$$

Define an estimator $\hat{\theta}_F$ by $\hat{\theta}_F(1) = \pi/4$, $\hat{\theta}_F(2) = \pi/2$. Note in particular that $\hat{\theta}_F$ has zero risk at $\pi/4$ and thus we have a non-constant estimator that has zero risk at $\pi/4$. If M were optimal there would exist

some estimator $\hat{\theta}_M$ such that $\hat{\theta}_M \leq \hat{\theta}_F$. But then we need $\hat{\theta}_M$ to have zero risk at $\pi/4$, but from equation (10), we see this is only possible with the constant estimator $\hat{\theta}_M(+)=\hat{\theta}_M(-)=\pi/4$. However, in this case $R(\hat{\theta}_F, \pi/2) < R(\hat{\theta}_M, \pi/2)$ and we deduce that $M \not\leq F$. Thus M is certainly not optimal.

Note that M and F both have maximal Fisher information and thus the measurements both have the same asymptotic, local, estimation ability. However, within global estimation they are *not* interchangeable; M cannot discriminate between $[0, \pi)$ and $[\pi, 2\pi)$ whereas F cannot discriminate between $[\pi/4, 5\pi/4)$ and $[5\pi/4, 9\pi/4)$. Thus, given a large number of copies it would be a poor strategy to measure all of the states with M or all of the states with F as one could not distinguish θ from $\theta + \pi$. Thus in practice, one would use a collection of different measurements; see for example [15].

The above argument generalises to other pure states. Suppose one has some (generically multiparameter) parametrised pure state $|\psi(\theta)\rangle$. Then for any fixed θ_* , letting F_{θ_*} be a measurement in some basis containing $|\psi(\theta_*)\rangle$, one can clearly construct an estimator with zero risk at θ_* . However, in general, the only estimators $\hat{\theta}_{F_{\theta_*}}$ with zero risk at some $\theta \neq \theta_*$ will be constant. Thus we expect none of the $\{F_{\theta_*} | \theta_* \in \Theta\}$ to be optimal and thus expect no optimal measurement for $|\psi(\theta)\rangle$.

4 Approximately Optimal Measurements

Since the previous section shows that optimal measurements (under least-squares loss) only exist for the very restrictive class of classical parametrised states, it is natural to ask whether measurements can be approximately optimal. There are three clear candidates for defining when a measurement is approximately optimal:

- (i) A measurement M is ϵ -additively optimal if, for any other measurement-estimator pair $(F, \hat{\theta}_F)$ there exists an estimator $\hat{\theta}_M$ such that for all $\theta \in \Theta$,

$$R(\hat{\theta}_M, \theta) \leq R(\hat{\theta}_F, \theta) + \epsilon. \quad (13)$$

- (ii) A measurement M is η -multiplicatively optimal if, for any other measurement-estimator pair $(F, \hat{\theta}_F)$ there exists an estimator $\hat{\theta}_M$ such that for all $\theta \in \Theta$,

$$R(\hat{\theta}_M, \theta) \leq (1 + \eta)R(\hat{\theta}_F, \theta). \quad (14)$$

- (iii) A measurement is δ -locally optimal if, for any other measurement-estimator pair $(F, \hat{\theta}_F)$ there exists an estimator $\hat{\theta}_M$ and subset $S \subseteq \Theta$ such

that S has volume (measure) less than or equal to δ and for all θ in $\Theta \setminus S$,

$$R(\hat{\theta}_M, \theta) \leq R(\hat{\theta}_F, \theta). \quad (15)$$

We can combine the third definition with either of the first two. For example, a measurement is ϵ -additively and δ -locally optimal if, for any other measurement-estimator pair $(F, \hat{\theta}_F)$ there exists an estimator $\hat{\theta}_M$ and subset $S \subseteq \Theta$ such that S has volume (measure) less than or equal to δ and for all θ in $\Theta \setminus S$,

$$R(\hat{\theta}_M, \theta) \leq R(\hat{\theta}_F, \theta) + \epsilon. \quad (16)$$

For each of the three definitions of approximate optimality, there is a corresponding notion of “closeness” of quantum states such that if $\rho(\theta)$ is close to being classical, there is an approximately optimal measurement—approximately classical implies an approximately optimal measurement. We briefly review each of the results here, the first two of which are proved in Appendix C. The proofs are made up of a series of technical inequalities.

We start with additive optimality, considering closeness in trace norm: for any matrix A , define its trace norm as

$$\|A\|_1 = \text{Tr}(\sqrt{A^\dagger A}). \quad (17)$$

We denote $\sqrt{A^\dagger A}$ as $|A|$.

We can now state the first result:

Lemma 4.1. *Suppose $\rho(\theta)$ is some parametrised state and $\sigma(\theta)$ is some classical state. Suppose further that Θ has finite diameter $d = \sup_{\theta, \phi} L(\theta, \phi)$. If, for all θ , we have that $\|\rho(\theta) - \sigma(\theta)\|_1 \leq \epsilon$, then there is a $2d\epsilon$ -additively optimal measurement M .*

For the multiplicative error, we need a different notion of closeness, namely the maximum relative entropy [16]: for quantum states ρ and σ , we define

$$D_{\max}(\rho|\sigma) = \inf\{\gamma : \rho \leq e^\gamma \sigma\}. \quad (18)$$

We can then state the multiplicative result:

Lemma 4.2. *Suppose $\rho(\theta)$ is some parametrised state and $\sigma(\theta)$ is some classical state. If, for all θ , we have that $\exp\{D_{\max}[\rho(\theta)|\sigma(\theta)] + D_{\max}[\sigma(\theta)|\rho(\theta)]\} \leq 1 + \eta$, then there is a η -multiplicatively optimal measurement M .*

The final closeness result is for local optimality. For a subset of real vectors $Y \subseteq \mathbb{R}^p$, denote its volume (Lebesgue measure) by $|Y|$.

Lemma 4.3. *Suppose $\rho(\theta)$ is some parametrised state, which is classical on some subset $\Gamma \subseteq \Theta$. Then there is a $(|\Theta| - |\Gamma|)$ -locally optimal measurement M .*

Proof. Take the optimal measurement for $\rho|_\Gamma$. This clearly has the desired property. \square

Note that it is not expected that being close to a classical state is necessary for an approximately optimal measurement to exist. For example, pure states $|\psi(\theta)\rangle$ where the quantum Fisher information can be saturated are expected to have an approximately optimal measurement for a large number of copies $|\Psi(\theta)\rangle = |\psi(\theta)\rangle^{\otimes n}$, $n \gg 1$. However, a large number of copies of a pure state is still a pure state and thus such states are far from being classical.

5 Admissibility of Measurements

Our definition of when a measurement is better than another ($M \prec F$) naturally gave rise to a definition of admissibility—a measurement M is admissible if no other measurement dominates it. In this section we investigate two classes of intuitively “bad” measurements and demonstrate their inadmissibility under some mild assumptions.

To state these assumptions, we must introduce a specific class of loss functions—Bregman divergences [17]. Specifically, for a convex set $\Theta \subseteq \mathbb{R}^p$ and any real-valued continuously differentiable, strictly convex function $g(\theta)$, the Bregman divergence associated to g is given by

$$B(\theta_1, \theta_2) = g(\theta_1) - g(\theta_2) - [\nabla g(\theta_2)]^T (\theta_1 - \theta_2). \quad (19)$$

A Bregman divergence is any function that can be written in such a fashion. For example, least squares is a Bregman divergence with $g(\theta) = \|\theta\|^2$ and the Kullback-Liebler divergence is a Bregman divergence with $g(\theta) = \sum_i \theta_i \log(\theta_i)$. Bregman divergences may be viewed as generalisations of the least-squares loss function. Appendix D reviews two of their properties that we require for our results.

The first class of measurements we discuss are those that can be fine-grained to extract more information from a system. Suppose that one has some parameter estimation problem $\rho(\theta)$ and use some measurement F where one knows the post measurement states. That is to say that $F = \{F_i^\dagger F_i\}$ and, in the case of outcome i , the post measurement state is

$$\rho_i = \frac{F_i \rho F_i^\dagger}{\text{Tr}(F_i \rho F_i^\dagger)}. \quad (20)$$

Intuitively, if some ρ_i still depends on θ , then one can, in the case of outcome i , perform another measurement on the system to extract more information about the unknown parameter—refining our measurement. Recall that ρ might be many copies of some density matrix: $\rho = \sigma^{\otimes n}$, so that outcome i could be the result of many measurements of individual density matrices. More precisely, we say that F is refineable if there exists an outcome (WLOG the first) as well as $\theta_1, \theta_2 \in \Theta$ such that $p_1^{(F)}(\theta_1), p_1^{(F)}(\theta_2) > 0$ and

$\rho_1(\theta_1) \neq \rho_1(\theta_2)$ ³. In Appendix D, we prove the following Lemma, that shows that such measurements are indeed inadmissible:

Lemma 5.1. *Let $\Theta \subseteq \mathbb{R}^N$, L be some Bregman divergence loss function on Θ and $\rho(\theta)$ be some non-constant parametrised state. Suppose that $F = \{F_i^\dagger F_i\}$ is a refineable measurement, then F is inadmissible.*

Proof Sketch. Constructing an M such that $M \preceq F$ is relatively straightforward—in the case of the first outcome, our state still depends on the parameter θ , so one can measure it again. We are always free to ignore the result of this subsequent measurement, and thus it is clear that $M \preceq F$. The proof that $F \not\preceq M$ is somewhat technical and involves the framework of (classical) Bayesian estimation [6], which is outlined in Appendix A. \square

The second case of inadmissibility that we consider is when a measurement does not extract information from a system. For example, one could do a measurement with a single outcome or measure in a mutually unbiased basis to the basis our parameter is encoded in. Formally, this corresponds to the case where the measurement outcome probabilities are independent of the underlying parameter. Again, in Appendix D, we prove the following Lemma:

Lemma 5.2. *Let $\Theta \subseteq \mathbb{R}^N$, L be some Bregman divergence loss function on Θ and $\rho(\theta)$ be some non-constant parametrised state. Suppose that $F = \{F_k\}$ is a measurement whose outcome probabilities are independent of θ , then F is inadmissible under least-squares loss.*

Proof Sketch. There are two main ideas in this proof. First, we show that the only admissible estimators for F are constant estimators (described above) and thus any measurement M satisfies $M \preceq F$. Second, we find an M dominating F using the Bayesian framework. \square

Note both results Lemmas show that the identity measurement (i.e. doing nothing and just guessing) is admissible iff. the state ρ is constant. Thus no matter how weakly a system varies, there is always a measurement that extracts useful information from it.

6 Conclusion

In Conclusion, we have defined the most general way in which the performance of quantum measurements for parameter estimation can be compared. As we have argued, any other comparison would require

³The first of these conditions is necessary to ensure that ρ_1 is well-defined at θ_1 and θ_2 .

additional modelling assumptions. Remarkably, we demonstrated a class of parametrised states—classical ones—that admit optimal measurements, even at this level of generality. Under some more strict assumptions, we have shown that these are the only such states.

Since only a very restrictive class of parametrised states have optimal measurements, we proposed several criteria for when a measurement may be considered approximately optimal. A further direction of research that appears interesting is characterising when a state has an approximately optimal measurement—we have given a selection of sufficient conditions but, as argued, they are not necessary. In particular, understanding approximately optimal measurements in the asymptotic regime could provide a new perspective on well established ideas, such as the quantum Fisher information.

Finally, we have demonstrated two inadmissible classes of measurements—those that can be refined or those that do not extract information from the quantum state of interest. Another possible further direction of research would be to attempt to characterise the admissible estimators for a given parametrised quantum state.

Acknowledgments

The authors wish to thank N. Mertig, J. Smith and C. Long for their useful discussions and comments. W.S. was supported by the EPSRC and Hitachi. S.S. acknowledges support from the Royal Society University Research Fellowship. D.R.M.A.-S. was supported by the Lars Hierta’s Memorial Foundation, and Girton College.

References

- [1] F. Albarelli et al. “A perspective on multiparameter quantum metrology: From theoretical tools to applications in quantum imaging”. In: *Physics Letters A* 384.12 (2020), p. 126311. DOI: <https://doi.org/10.1016/j.physleta.2020.126311>.
- [2] Wojciech Górecki and Rafał Demkowicz-Dobrzański. “Multiparameter quantum metrology in the Heisenberg limit regime: Many-repetition scenario versus full optimization”. In: *Phys. Rev. A* 106 (1 July 2022), p. 012424. DOI: 10.1103/PhysRevA.106.012424.
- [3] Samuel L. Braunstein and Carlton M. Caves. “Statistical distance and the geometry of quantum states”. In: *Phys. Rev. Lett.* 72 (22 May 1994), pp. 3439–3443. DOI: 10.1103/PhysRevLett.72.3439.

- [4] Rafał Demkowicz-Dobrzański, Wojciech Górecki, and Mădălin Guță. “Multi-parameter estimation beyond quantum Fisher information”. In: *Journal of Physics A: Mathematical and Theoretical* 53.36 (Aug. 2020), p. 363001. DOI: 10.1088/1751-8121/ab8ef3.
- [5] Christopher A. Fuchs and Rüdiger Schack. “Priors in Quantum Bayesian Inference”. In: *AIP Conference Proceedings* 1101.1 (2009), pp. 255–259. DOI: 10.1063/1.3109948.
- [6] Eric L. Lehmann and George Casella. *Theory of Point Estimation*. Springer, New York, NY, 1998. DOI: <https://doi.org/10.1007/b98854>.
- [7] S. Kullback and R. A. Leibler. “On Information and Sufficiency”. In: *The Annals of Mathematical Statistics* 22.1 (Mar. 1951), pp. 79–86. DOI: 10.1214/aoms/1177729694.
- [8] A. Bhattacharyya. “On a Measure of Divergence between Two Multinomial Populations”. In: *Sankhyā: The Indian Journal of Statistics (1933-1960)* 7.4 (1946), pp. 401–406.
- [9] Larry Wasserman. *All of Statistics*. Springer, New York, NY, 2004, p. 202. DOI: <https://doi.org/10.1007/978-0-387-21736-9>.
- [10] Carl W. Helstrom. “Quantum detection and estimation theory”. In: *Journal of Statistical Physics* 1 (June 1969), pp. 231–252. DOI: <https://doi.org/10.1007/BF01007479>.
- [11] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. “Advances in quantum metrology”. In: *Nature Photonics* 5.4 (2011), pp. 222–229. DOI: 10.1038/nphoton.2011.35.
- [12] Charles Stein and W. James. “Estimation with Quadratic Loss”. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1 (1961), pp. 361–379.
- [13] Hwang Lee, Pieter Kok, and Jonathan P. Dowling. “A quantum Rosetta stone for interferometry”. In: *Journal of Modern Optics* 49.14-15 (2002), pp. 2325–2338. DOI: 10.1080/0950034021000011536.
- [14] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. DOI: 10.1017/CBO9780511976667.
- [15] Joseph G. Smith, Crispin H. W. Barnes, and David R. M. Arvidsson-Shukur. “Iterative quantum-phase-estimation protocol for shallow circuits”. In: *Phys. Rev. A* 106 (6 Dec. 2022), p. 062615. DOI: 10.1103/PhysRevA.106.062615.
- [16] M. Mosonyi and T. Ogawa. “Quantum Hypothesis Testing and the Operational Interpretation of the Quantum Rényi Relative Entropies”. In: *Communications in Mathematical Physics* 334.3 (2015). DOI: 10.1007/s00220-014-2248-x.
- [17] A. Banerjee, Xin Guo, and Hui Wang. “On the optimality of conditional expectation as a Bregman predictor”. In: *IEEE Transactions on Information Theory* 51.7 (2005), pp. 2664–2669. DOI: 10.1109/TIT.2005.850145.

A Bayesian estimation

In this section we introduce Bayesian estimation, which we will need for the proofs of the results from the main text.

In a Bayes setting, one has some *a priori* belief about our underlying parameter [9]. For example, one may believe that it is likely to be centred around some value, and correspondingly very unlikely to be far away from it. Formally, this corresponds to a probability distribution $\pi(\theta)$ on our parameter space. A common choice is a normally distributed prior $\mathcal{N}(\mu, \Sigma)$; to encode an underlying expectation of our parameter μ along with decay of probability away from μ encoded by Σ . Since one has some probability distribution on our parameter space, one can now talk about the risk of an estimator averaged across different possible values of the parameter rather than at a specific parameter value. That is, one defines the Bayes risk corresponding to π as

$$R_\pi(\hat{\theta}) = \mathbb{E}_{\theta \sim \pi(\theta)}[R(\hat{\theta}, \theta)] = \int d\theta \pi(\theta) R(\hat{\theta}, \theta). \quad (21)$$

An estimator $\hat{\theta}^B$ is said to be Bayes if it minimises the Bayes risk, that is for any other estimator $\hat{\theta}$, $R_\pi(\hat{\theta}^B) \leq R_\pi(\hat{\theta})$. We note the following fact about Bayesian estimators.

Lemma A.1. *A unique Bayes estimator $\hat{\theta}^B$ is admissible*

Proof. Suppose $\hat{\theta}$ dominates $\hat{\theta}^B$, then by definition $R(\hat{\theta}, \theta) \leq R(\hat{\theta}^B, \theta)$ and $\hat{\theta} \neq \hat{\theta}^B$. But then taking expectation with respect to θ , we see that $R_\pi(\hat{\theta}) \leq R_\pi(\hat{\theta}^B)$. But this contradicts that $\hat{\theta}^B$ is the unique minimiser of the Bayes risk. \square

In the quantum case, where one has a state $\rho(\theta)$, one must minimise the Bayes risk over estimators based off a fixed measurement and then additionally minimise this across all possible measurements. Thus one searches for a measurement-estimator pair which minimises the Bayes risk. In such a case, both the measurement and estimator are described as Bayes.

In the single parameter case ($\Theta \subseteq \mathbb{R}$), we also introduce some notation: for any measurement-estimator pair $(F, \hat{\theta}_F)$, we define

$$\Lambda = \sum_i F_i \hat{\theta}_F(i), \quad \Lambda_2 = \sum_i F_i \hat{\theta}_F(i)^2, \quad (22)$$

Where F_i are the measurement operators of F .

For any prior $\pi(\theta)$ on Θ , we define two operators

$$\bar{\rho} = \int d\theta \pi(\theta) \rho(\theta), \quad \bar{\rho}' = \int d\theta \pi(\theta) \theta \rho(\theta). \quad (23)$$

B Proof of Theorem 3.2

In this section, we prove that only classical parametrised quantum states have optimal measurements. We specialise to the case that our parameter θ can be expressed as a real vector - $\Theta \subseteq \mathbb{R}^N$ and that we use least-squares loss - where $L(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|^2$.

We now present the proof of Theorem 3.2 as a series of Lemmas. We first restrict to the single parameter ($N = 1$) case. We fix $\theta_1, \theta_2 \in \Theta$ and consider the restricted parameter space $\{\theta_1, \theta_2\} \subseteq \Theta$ and show that $\rho(\theta_1)$ and $\rho(\theta_2)$ commute. This is achieved by considering the Bayesian measurements associated with all possible priors on $\{\theta_1, \theta_2\}$. This then shows that if an optimal measurement exists on Θ , all $\{\rho(\theta) : \theta \in \Theta\}$ must commute. We use this to simultaneously diagonalise the $\{\rho(\theta)\}$ and thus deduce ρ is classical. By restricting to lines in higher dimension parameter spaces, we use the single parameter case to prove the multiparameter case.

Lemma B.1. *Suppose $\rho(\theta)$ is some parametrised state with an optimal measurement M . Then M is a Bayes measurement for any prior $\pi(\theta)$ on Θ .*

Proof. Suppose we have some Bayes estimator $\hat{\theta}_F^B$. Then since M is optimal, $M \preceq F$, and thus we can find an estimator $\hat{\theta}_M$ satisfying $\hat{\theta}_M \leq \hat{\theta}_F^B$. But then, by definition, $\hat{\theta}_M$ must also be Bayes. Thus for any prior, we can always measure with M and pick a Bayes estimator based on the resulting probability distribution. \square

Lemma B.2. *Suppose $\rho(\theta)$ is some single parameter state ($\Theta \subseteq \mathbb{R}$) under least-squares loss. Suppose further we have some prior $\pi(\theta)$ on θ where $\bar{\rho}$ has full rank. Then a measurement $F = \{F_i\}$, F is Bayesian iff. the following hold*

(i) F is the projective measurement onto Λ 's eigenspaces, or a fine-graining of it.

(ii) $\{\Lambda, \bar{\rho}\} = 2\bar{\rho}'$

Proof. Note that this lemma, and most of this proof, is presented in [4] as a sufficient condition. We prove that the conditions are also necessary.

Fix some measurement-estimator pair $(F, \hat{\theta}_F)$. Note that we may write the Bayesian risk in terms of Λ and Λ_2 (in the case of least-squares loss):

$$R_\pi(\hat{\theta}_F) = \int d\theta \sum_i \pi(\theta) \text{Tr}(F_i \rho(\theta)) (\hat{\theta}_F(i) - \theta)^2, \quad (24)$$

$$= \text{Tr}(\Lambda_2 \bar{\rho}) - 2\text{Tr}(\Lambda \bar{\rho}') + \int d\theta \pi(\theta) \theta^2. \quad (25)$$

We note that

$$0 \leq \sum_i (\hat{\theta}(i)\mathbb{1} - \Lambda)^\dagger F_i (\hat{\theta}(i)\mathbb{1} - \Lambda), \quad (26)$$

$$= \Lambda_2 - \Lambda^2, \quad (27)$$

and thus $\text{Tr}(\Lambda_2 \bar{\rho}) \geq \text{Tr}(\Lambda^2 \bar{\rho})$. Since we want to minimise the Bayes risk, we now check when this inequality is saturated. Note that for any positive semi-definite operator A ,

$$\text{Tr}(\bar{\rho}A) = 0 \Leftrightarrow \text{Tr}((\sqrt{\bar{\rho}}\sqrt{A})^\dagger \sqrt{\bar{\rho}}\sqrt{A}) = 0, \quad (28)$$

$$\Leftrightarrow \sqrt{\bar{\rho}}\sqrt{A} = 0, \quad (29)$$

$$\Leftrightarrow A = 0, \quad (30)$$

where we used the assumption that $\bar{\rho}$ is full rank, and hence $\sqrt{\bar{\rho}}$, is invertible. Thus we saturate the inequality iff. $\sum_i (\hat{\theta}(i) - \Lambda)^\dagger F_i (\hat{\theta}(i) - \Lambda) = 0$. But since this is a sum of positive semi-definite operators, this can only happen if, for each i , $(\hat{\theta}(i) - \Lambda)^\dagger F_i (\hat{\theta}(i) - \Lambda) = 0$. If $\hat{\theta}(i)$ is not an eigenvalue of Λ , then $F_i = 0$. Otherwise, F_i must only have support on the $\hat{\theta}(i)$ eigenspace of Λ . Since $\Lambda = \sum F_i \hat{\theta}(i)$, condition (i) follows.

On saturation of the inequality, we see that finding a Bayes measurement reduces to minimising the quantity

$$\mathcal{C} = \text{Tr}(\Lambda^2 \bar{\rho}) - 2\text{Tr}(\Lambda \bar{\rho}'), \quad (31)$$

over Hermitian operators Λ . By differentiating with respect to Λ and setting the derivative to zero, we see that there is a unique minimum and that it satisfies the equation

$$\frac{1}{2}\{\Lambda, \bar{\rho}\} = \bar{\rho}'. \quad (32)$$

□

Lemma B.3. *Suppose $\rho(\theta)$ is some single parameter state ($\Theta \subseteq \mathbb{R}$) with an optimal measurement M under least-squares loss. Fix $\theta_1, \theta_2 \in \Theta$ distinct. If $\rho(\theta_1), \rho(\theta_2)$ both have full rank, then there exist simultaneously diagonalisable Hermitian maps Γ, Γ' satisfying*

$$\{\Gamma, \rho_2\} = \rho_1, \quad \{\Gamma', \rho_1\} = \rho_2. \quad (33)$$

Proof. For notational convenience, let $\rho_i = \rho(\theta_i)$ for $i = 1, 2$. For $p \in (0, 1)$ fix a prior on Θ where we are given ρ_1 with probability p and ρ_2 with probability $1 - p$. Note that for any $p \in (0, 1)$, $\bar{\rho}$ has full rank. By Lemma B.1 we see that M must be Bayesian for any value of p - let $\hat{\theta}_M^p$ be a Bayes estimator for a fixed value of p . Then $\Lambda = \sum_i M_i \hat{\theta}_M^p(i)$ and M must satisfy condition (i) of Lemma B.2. Then, as a function of p we may expand $\Lambda = \sum_k \mu_k(p) |k\rangle\langle k|$ - in an eigenbasis that does not depend on p .

By condition (ii) of Lemma B.2, for any value of $p \in (0, 1)$ we know that Λ satisfies

$$\frac{1}{2}\{\Lambda, p\rho_1 + (1-p)\rho_2\} = p\theta_1\rho_1 + (1-p)\theta_2\rho_2. \quad (34)$$

As well as being invertible, the smallest eigenvalue of $\bar{\rho}$, λ_{\min} is bounded below by the minimum of the eigenvalues of ρ_1 and ρ_2 . Consider equation (34) in the limit of $p \rightarrow 0$. Let Γ be defined implicitly as the solution to the equation

$$\{\Gamma, \rho_2\} = \rho_1. \quad (35)$$

Note since ρ_1 is invertible, by expanding equation (35) in the eigenbasis of ρ_2 , we see that Γ is well defined and is furthermore Hermitian.

For some ‘‘remainder’’ matrix R , we expand Λ as $\Lambda = \theta_2\mathbb{1} + 2p(\theta_1 - \theta_2)\Gamma + R$. Substituting this ansatz into equation (34), we see that R must satisfy the equation

$$\frac{1}{2}\{R, \bar{\rho}\} = p^2(\theta_2 - \theta_1)\{\Gamma, \rho_1 - \rho_2\}. \quad (36)$$

Expanding in an eigenbasis of $\bar{\rho}$ and recalling that all the eigenvalues of $\bar{\rho}$ are bounded below by λ_{\min} (which is independent of p), we see that $R = \mathcal{O}(p^2)$. Consider an eigenstate $|\psi\rangle$ of Λ with eigenvalue $\mu(p)$, then note that

$$\left\| \Gamma |\psi\rangle - \frac{\mu(p) - \theta_2}{2p(\theta_1 - \theta_2)} |\psi\rangle \right\| = \mathcal{O}(p) \quad (37)$$

Taking the limit as $p \rightarrow 0$, we see that $|\psi\rangle$ must be an eigenstate of Γ . But, by symmetry, in the limit $p \rightarrow 1$, we may define Γ' by the equation

$$\{\Gamma', \rho_1\} = \rho_2, \quad (38)$$

and $|\psi\rangle$ must also be an eigenstate of Γ' . Thus Γ and Γ' must be simultaneously diagonalisable in Λ 's eigenbasis. □

Lemma B.4. *Suppose A, B are positive definite and Γ, Γ' are simultaneously diagonalisable Hermitian maps satisfying*

$$\{\Gamma, A\} = B, \quad \{\Gamma', B\} = A. \quad (39)$$

Then A and B commute.

Proof. Let Γ and Γ' be simultaneously diagonalisable in the orthonormal basis $\{|i\rangle\}_{i=1}^n$. Let $\Gamma|i\rangle = \lambda_i|i\rangle$, $\Gamma'|i\rangle = \mu_i|i\rangle$, $\langle i|A|j\rangle = A_{ij}$ and $\langle i|B|j\rangle = B_{ij}$, for $i, j = 1, \dots, n$. By positive definiteness, note that $A_{ii}, B_{ii} > 0$ for any i . Rewriting equations (39) in components of this basis, we reach the set of equations

$$(\lambda_i + \lambda_j)A_{ij} = B_{ij}, \quad (\mu_i + \mu_j)B_{ij} = A_{ij}. \quad (40)$$

Setting $i = j$ we deduce that

$$\lambda_i = B_{ii}/2A_{ii}, \quad \mu_i = A_{ii}/2B_{ii}. \quad (41)$$

Note equation (40) shows $A_{ij} = 0 \Leftrightarrow B_{ij} = 0$. Suppose $A_{ij} \neq 0$, then multiplying the two individual equations in (40) and dividing by $A_{ij}B_{ij}$ we see that

$$\begin{aligned} (A_{ii}/2B_{ii} + A_{jj}/2B_{jj})(B_{ii}/2A_{ii} + B_{jj}/2A_{jj}) &= 1 \\ \Leftrightarrow \frac{A_{ii}B_{jj}}{B_{ii}A_{jj}} + \frac{B_{ii}A_{jj}}{A_{ii}B_{jj}} &= 2. \end{aligned} \quad (42)$$

But $x + 1/x = 2$ iff. $x = 1$ and thus we deduce

$$\frac{A_{ii}}{B_{ii}} = \frac{A_{jj}}{B_{jj}} \quad (43)$$

Consider a graph G on n nodes, with an edge between i and j iff. $A_{ij} \neq 0$. If G is connected, then for any nodes i, j in G there is path between them. Assuming connectedness, we apply the result of equation (43) along the path, we deduce that $A_{ii}B_{jj} = B_{ii}A_{jj}$ for every $i, j = 1, \dots, n$. But then summing over j , we deduce that $A_{ii} = \frac{\text{Tr}(A)}{\text{Tr}(B)} B_{ii}$. Substituting this into equations (40) and (41), we see that A and B are proportional and thus commute. If G is not connected, then we can apply the above procedure to each connected component of G . The result is that A and B are block diagonal, with the diagonal matrix entries proportional and thus A and B commute. \square

Lemma B.5. *Suppose $\rho(\theta)$ is some single parameter state ($\Theta \subseteq \mathbb{R}$) with an optimal measurement M under least-squares loss. Fix $\theta_1, \theta_2 \in \Theta$ distinct, then $\rho(\theta_1), \rho(\theta_2)$ commute (regardless of their rank).*

Proof. Again, for notational convenience, let $\rho_i = \rho(\theta_i)$ for $i = 1, 2$. Redefine our parameter space $\Theta = \{\theta_1, \theta_2\}$, but still allow estimators to take any real value. Note, by definition, that M must also be optimal for this parameter estimation problem.

Suppose that $U = \ker(\rho_2) \cap \ker(\rho_1) \neq 0$. Let U^\perp be the orthogonal compliment of U and let Π_{U^\perp} be the orthogonal projection matrix onto U^\perp . Then note that replacing M with the measurement $\{\Pi_{U^\perp} M_i \Pi_{U^\perp}, \mathbb{1} - \Pi_{U^\perp}\}$ does not change any of the measurement probabilities when measuring ρ_1 or ρ_2 . Thus we may WLOG restrict to U^\perp , which still has an optimal measurement for this restricted parameter estimation problem.

Suppose that ρ_2 is not full rank, i.e. it has some non-trivial kernel $K \leq \mathcal{H}$. Take the measurement E of the two orthogonal projectors Π_K and Π_{K^\perp} , along with an estimator $\hat{\theta}_E(K) = \theta_1, \hat{\theta}_E(K^\perp) = \theta_2$ so that

$$R(\hat{\theta}_E, \theta_2) = 0, \quad (44)$$

$$R(\hat{\theta}_E, \theta_1) = (\theta_1 - \theta_2)^2 (1 - \text{Tr}(\Pi_K \rho_1)). \quad (45)$$

Since M is optimal, we have that $M \preceq E$. Then, in particular, there must be an estimator $\tilde{\theta}_M$ satisfying $\tilde{\theta}_M \leq \hat{\theta}_E$. In particular, $\tilde{\theta}_M$ must have zero risk at θ_2 and thus if $\text{Tr}(\rho_2 M_i) \neq 0$, then we must have $\tilde{\theta}_M(i) = \theta_2$. Let $I = \{i \mid \text{Tr}(\rho_2 M_i) = 0\}$, then, by the above, $\tilde{\theta}_M$ must satisfy

$$R(\tilde{\theta}_M, \theta_1) \geq (\theta_1 - \theta_2)^2 \left(1 - \sum_{i \in I} \text{Tr}(\rho_1 M_i) \right), \quad (46)$$

with equality iff. for every $i \in I, \tilde{\theta}_M(i) = \theta_1$. WLOG Assume that this holds, so that we saturate equation (46). If $\text{Tr}(\rho_2 M_i) = 0$, then, as $M_i \geq 0$, every

eigenvector of M_i with non-zero eigenvalue must lie in K and thus $\sum_{i \in I} M_i \leq \Pi_K$. Thus, $R(\tilde{\theta}_M, \theta_1) \geq R(\hat{\theta}_E, \theta_2)$ with equality iff. $\rho_1 (\Pi_K - \sum_{i \in I} M_i) = 0$. Since ρ_1 does not kill any vector in K (as we restricted to U^\perp) we get equality iff. $\sum_{i \in I} M_i = \Pi_K$. Thus for $i \notin I$ and $|k\rangle \in K, M_i |k\rangle = 0$.

Now, for $p \in (0, 1)$ fix a prior on Θ where we are given ρ_1 with probability p and ρ_2 with probability $1 - p$. Taking Λ corresponding to M by Lemma B.1, we see that Λ decomposes as $\Lambda = \Lambda_K + \Lambda_{K^\perp}$, where $\Lambda_K = \Pi_K \Lambda_K \Pi_K$ and $\Lambda_{K^\perp} = \Pi_{K^\perp} \Lambda_{K^\perp} \Pi_{K^\perp}$. Furthermore, in order for our estimator to be Bayes, we must always guess θ_1 in the case of a K outcome i.e. $\Lambda_K = \theta_1 \Pi_K$. Take the inner product of condition (ii) of lemma B.2 with $|k\rangle \in K$ and $|\ell\rangle \in K^\perp$. This gives

$$\langle \ell | \Lambda_{K^\perp} \rho_1 | k \rangle = \theta_1 \langle \ell | \rho_1 | k \rangle. \quad (47)$$

Note that as $p \rightarrow 0$, we must have $\Lambda_{K^\perp} \rightarrow \Pi_{K^\perp} \theta_2$, as our estimates must approach θ_2 for our estimator to be Bayes. Thus the only way for equation (47) to be satisfied for all $p \in (0, 1)$ is for $\langle \ell | \rho_1 | k \rangle = 0$. Thus ρ_1 fixes K and we may decompose it as $\rho_1 = \Pi_K \rho_1 \Pi_K + \Pi_{K^\perp} \rho_1 \Pi_{K^\perp}$. Thus equation (34) becomes

$$\begin{aligned} & \frac{1}{2} \{ \Lambda_{K^\perp} + p \Pi_{K^\perp} \rho_1 \Pi_{K^\perp} + (1 - p) \Pi_{K^\perp} \rho_2 \Pi_{K^\perp} \} \\ & = p \theta_1 \Pi_{K^\perp} \rho_1 \Pi_{K^\perp} + (1 - p) \theta_2 \Pi_{K^\perp} \rho_2 \Pi_{K^\perp}. \end{aligned} \quad (48)$$

This essentially reduces the Hilbert space to K^\perp , on which ρ_2 has full rank. We may then repeat the above to assume ρ_1 has full rank (note that, as we restricted to $U^\perp, \ker(\rho_1) \leq K^\perp$. By Lemmas B.3 and B.4, ρ_1 and ρ_2 commute on their joint support. But as they fix each other's kernels and are Hermitian, they must, therefore, fully commute. \square

Lemma B.6. *Suppose $\rho(\theta)$ is some single parameter state ($\Theta \subseteq \mathbb{R}$). If there is an optimal measurement M under least-squares loss, then $\rho(\theta)$ is a classical state problem.*

Proof. Fix θ_1 in Θ and decompose \mathcal{H} as a direct sum of $\rho(\theta_1)$'s (distinct) eigenspaces $\mathcal{H} = \oplus_j V_j$. If this is an eigenspace decomposition of $\rho(\theta)$ for every $\theta \in \Theta$, then by definition ρ is a classical state. Otherwise, take θ_2 such that some V_j is not an eigenspace of $\rho(\theta_2)$. But by Lemma B.5, $[\rho(\theta_1), \rho(\theta_2)] = 0$ and thus V_j must decompose a sum of eigenspaces of $\rho(\theta_2)$, $V_j = \oplus V'_j$. Take this new decomposition into eigenspaces. We may repeat this process, noting that it must terminate as each eigenspace must have dimension at least one, to see that $\rho(\theta)$ must be classical. \square

Theorem 3.2. *Suppose $\rho(\theta)$ is some parameter estimation problem, where $\Theta \subseteq \mathbb{R}^N$ is convex. Unless $\rho(\theta)$ is classical, there is no optimal measurement M under least-squares loss.*

Proof. Suppose we have a general $\rho(\theta)$ and some optimal measurement M . Fix $\theta_1, \theta_2 \in \Theta$ and let $T = [0, \|\theta_2 - \theta_1\|]$. By convexity, for $t \in T$, $\gamma(t) = \theta_1 + t(\theta_2 - \theta_1)/\|\theta_2 - \theta_1\|$ is in Θ . Thus $\rho(\gamma(t))$ gives a single parameter estimation problem - estimating t . We have constructed this problem such that the least-squares loss functions agree - that is for $s, t \in T$,

$$\|\gamma(t) - \gamma(s)\|^2 = |t - s|^2. \quad (49)$$

But then (by projecting any estimators $\hat{\theta}_M$ onto the line segment between θ_1 and θ_2) M must also be optimal for this problem, and thus, by Lemma B.6, $\rho(\theta_1)$ and $\rho(\theta_2)$ commute. But then, since θ_1 and θ_2 were arbitrary, by the same argument as Lemma B.6, we see that $\rho(\theta)$ must be classical. \square

We remark that the condition of convexity in Theorem 3.2 can be weakened. By Lemma B.2, we see that the optimal measurement must be a fine-graining of the projection onto ρ 's joint eigenspaces. But then for any two distinct sets on which ρ is classical, we see that it must be classical on the union of these sets too. Then the result holds for Θ a disjoint union of convex sets too. Moreover, if Θ is open, around any $\theta \in \Theta$ we can fit a convex set in which, ρ must be classical. But then by the same reasoning as before, we see that ρ must be classical on the whole of Θ . We do not prove this result in full detail, as most parameter spaces of interest are convex.

C Approximately Classical Implies an Approximately Optimal Measurement

In this section we provide the remaining two proofs for each of the ‘‘close’’ to classical implies ‘‘close’’ to optimal results from Section 4. It will be useful to slightly extend our notation for risk functions to include a label for the state we are considering, i.e. we write $R_\rho(\hat{\theta}_M, \theta)$.

We start with the first result:

Lemma 4.1. *Suppose $\rho(\theta)$ is some parametrised state and $\sigma(\theta)$ is some classical state. Suppose further that Θ has finite diameter $d = \sup_{\theta, \phi} L(\theta, \phi)$. If, for all θ , we have that $\|\rho(\theta) - \sigma(\theta)\|_1 \leq \epsilon$, then there is a $2d\epsilon$ -additively optimal measurement M .*

Proof. Since $\sigma(\theta)$ is classical, we can fix some optimal measurement M . Fix some measurement-estimator pair $(F, \hat{\theta}_F)$ and $\theta \in \Theta$. Then note that

$$\begin{aligned} & |R_\rho(\hat{\theta}_F, \theta) - R_\sigma(\hat{\theta}_F, \theta)| \\ &= \left| \sum_i \text{Tr}[F_i(\rho(\theta) - \sigma(\theta))]L(\hat{\theta}_F(i), \theta) \right|, \quad (50) \end{aligned}$$

$$\leq d \sum_i |\text{Tr}[F_i(\rho(\theta) - \sigma(\theta))]|. \quad (51)$$

Fix some $F_i \geq 0$. Diagonalising $\rho(\theta) - \sigma(\theta) = \sum_j \lambda_j |j\rangle\langle j|$, note that

$$|\text{Tr}[F_i(\rho(\theta) - \sigma(\theta))]| \leq \sum_j |\lambda_j| \langle j|F_i|j\rangle, \quad (52)$$

$$= \text{Tr}(F_i|\rho(\theta) - \sigma(\theta)|). \quad (53)$$

Substituting inequality (53) into (51), we see that

$$|R_\rho(\hat{\theta}_F, \theta) - R_\sigma(\hat{\theta}_F, \theta)| \leq d\epsilon \quad (54)$$

Since M is optimal on $\sigma(\theta)$ there exists $\hat{\theta}_M$ such that for all θ , $R_\sigma(\hat{\theta}_M, \theta) \leq R_\sigma(\hat{\theta}_F, \theta)$. Then applying (54) twice, we see

$$R_\rho(\hat{\theta}_M, \theta) - R_\rho(\hat{\theta}_F, \theta) \leq R_\sigma(\hat{\theta}_M, \theta) - R_\sigma(\hat{\theta}_F, \theta) + 2d\epsilon, \quad (55)$$

$$\leq 2d\epsilon. \quad (56)$$

\square

For the multiplicative error, we will make use of the following property of maximum relative entropy,

Lemma C.1. *For two states ρ, σ*

$$D_{\max}(\rho||\sigma) = \log \sup_{0 \leq M \leq \mathbb{1}} \frac{\text{Tr}(M\rho)}{\text{Tr}(M\sigma)}. \quad (57)$$

We omit the proof, it is given in [16].

Using this, we can prove the closeness result for multiplicative approximate optimality

Lemma 4.2. *Suppose $\rho(\theta)$ is some parametrised state and $\sigma(\theta)$ is some classical state. If, for all θ , we have that $\exp\{D_{\max}[\rho(\theta)||\sigma(\theta)] + D_{\max}[\sigma(\theta)||\rho(\theta)]\} \leq 1 + \eta$, then there is a η -multiplicatively optimal measurement M .*

Proof. Fix some measurement F , note that

$$\frac{R_\rho(\hat{\theta}_F, \theta)}{R_\sigma(\hat{\theta}_F, \theta)} = \frac{\sum_i \text{Tr}(\rho(\theta)F_i)L(\hat{\theta}_F(i), \theta)}{\sum_i \text{Tr}(\sigma(\theta)F_i)L(\hat{\theta}_F(i), \theta)}, \quad (58)$$

$$\leq \max_i \frac{\text{Tr}(\rho(\theta)F_i)}{\text{Tr}(\sigma(\theta)F_i)}, \quad (59)$$

$$\leq e^{D_{\max}(\rho||\sigma)}, \quad (60)$$

where we have used Lemma C.1. But then fixing an optimal measurement for σ and using the inequality (60) twice, the result follows. \square

D Admissibility of Measurements

The aim of this section is to prove that the two classes of measurements discussed in Section 5 are inadmissible.

To begin, we must prove a series of technical results about Bregman divergences and Bayesian estimation (see Appendix A). There are two properties of Bregman divergences that we will need, stated below in Lemma D.1. We will not prove them, instead referring the reader to [17].

Lemma D.1. *Let L a Bregman divergence. Then*

(i) L is strictly convex in its first argument.

(ii) For any prior π on Θ , the Bayes estimator $\hat{\theta}^B$ is unique and is given by the posterior mean $\hat{\theta}^B(x) = \mathbb{E}_\pi[\theta | \text{outcome } x]$.

Next we prove some technical lemmas to do with Bayesian estimation.

Lemma D.2. *Let $\Theta \subset \mathbb{R}^p$ and L be a Bregman divergence. Suppose that $\{p_i(\theta)\}$ is a distribution depending on θ and that there exist $\theta_1, \theta_2 \in \Theta$ such that $p_1(\theta_1)p_2(\theta_2) \neq p_2(\theta_1)p_1(\theta_2)$. Then, there exists an admissible estimator $\hat{\theta}$ where $\hat{\theta}(1) \neq \hat{\theta}(2)$.*

Proof. Define a Prior on Θ of uniformly random choice of θ_1 or θ_2 , that is to say,

$$\mathbb{P}(Y = \theta) = \begin{cases} 1/2, & \theta \in \{\theta_1, \theta_2\}, \\ 0, & \text{o.w.} \end{cases} \quad (61)$$

In this case, by Lemma D.1, we know that

$$\hat{\theta}^B(i) = \mathbb{E}_\pi[\theta | i], \quad (62)$$

$$= \mathbb{P}(\theta_1 | i)\theta_1 + \mathbb{P}(\theta_2 | i)\theta_2. \quad (63)$$

Since this is a convex combination of θ_1 and θ_2 which must be distinct to satisfy the conditions of the Lemma, we deduce that $\hat{\theta}^B(1) = \hat{\theta}^B(2)$ iff. $\mathbb{P}(\theta_1 | 1) = \mathbb{P}(\theta_1 | 2)$. But

$$\mathbb{P}(\theta_1 | i) = \frac{p_i(\theta_1)}{p_i(\theta_1) + p_i(\theta_2)}. \quad (64)$$

So

$$\mathbb{P}(\theta_1 | 1) = \mathbb{P}(\theta_1 | 2) \Leftrightarrow \frac{p_1(\theta_1)}{p_1(\theta_1) + p_1(\theta_2)} = \frac{p_2(\theta_1)}{p_2(\theta_1) + p_2(\theta_2)}, \quad (65)$$

$$\Leftrightarrow p_1(\theta_1)p_2(\theta_2) = p_2(\theta_1)p_1(\theta_2). \quad (66)$$

Thus by the assumptions of the Lemma, $\hat{\theta}^B(1) \neq \hat{\theta}^B(2)$. But by Lemma A.1, $\hat{\theta}^B$ is admissible. \square

Lemma D.3. *Let L be a Bregman divergence. Suppose that $\{p_i(\theta)\}$ is a distribution depending on θ and that $\hat{\theta}, \tilde{\theta}$ are two distinct estimators with the same risk function. Then $\hat{\theta}$ and $\tilde{\theta}$ are inadmissible unless they are equal with probability 1.*

Proof. Let $\hat{\theta}'(i) = [\hat{\theta}(i) + \tilde{\theta}(i)]/2$. By strict convexity of L in its first argument and since $\hat{\theta}, \tilde{\theta}$ are distinct, the result follows. \square

Lemma D.4. *Let L be a Bregman divergence and $\{p_i\}$ be some distribution independent of some parameter θ . Then an estimator $\hat{\theta}$ is admissible iff. it is constant with probability 1.*

Proof. Consider the constant estimator $\theta_0 = \sum_i p_i \hat{\theta}(i)$, which is well defined by the assumption that $p_k^{(F)}$ is constant. Then

$$R(\hat{\theta}, \theta) = \sum_i p_i L(\hat{\theta}(i), \theta) \quad (67)$$

$$\geq L(\theta_0, \theta) = R(\theta_0, \theta). \quad (68)$$

By lemma D.3, $\hat{\theta}$ is admissible only if it equals θ_0 with probability 1. \square

We can now prove the two Lemmas from Section 5. First, we deal with refineable measurements. Recall their definition: suppose that we have some parameter estimation problem $\rho(\theta)$ and a measurement $F = \{F_i^\dagger F_i\}_{i=1}^I$. In the case of outcome i the post measurement state is

$$\rho_i = \frac{F_i \rho F_i^\dagger}{\text{Tr}(F_i \rho F_i^\dagger)}. \quad (69)$$

Recall that a measurement $F = \{F_i^\dagger F_i\}_{i=1}^I$ is called refineable if one of the possible post measurement states (WLOG the first) still depends on θ . That is there are $\theta_1, \theta_2 \in \Theta$ such that

(i) $\rho_1(\theta_1) \neq \rho_1(\theta_2)$.

(ii) $p_1(\theta_1), p_1(\theta_2) \neq 0$.

The second of these conditions is to ensure that outcome one is possible at θ_1 and θ_2 . Otherwise, the post measurement state, as defined above, is not well-defined. We can now prove the first result on inadmissibility from the main text.

Lemma 5.1. *Let $\Theta \subseteq \mathbb{R}^N$, L be some Bregman divergence loss function on Θ and $\rho(\theta)$ be some non-constant parametrised state. Suppose that $F = \{F_i^\dagger F_i\}$ is a refineable measurement, then F is inadmissible.*

Proof. Take θ_1, θ_2 satisfying the conditions above. Since $\rho_1(\theta_1) \neq \rho_1(\theta_2)$, by the Helstrom Bound [10], there is some measurement $M = M_1, M_2$ where $p_1^{(M)}(\theta_1) \neq p_1^{(M)}(\theta_2)$. Consider the measurement MF corresponding to measuring first with F and in the case of outcome 1, measuring again but now using M . That is $MF = \{F_1^\dagger M_1^\dagger M_1 F_1, F_1^\dagger M_2^\dagger M_2 F_1\} \cup \{F_i^\dagger F_i\}_{i=2}^I$. We label the outcomes of this measurement as $(1, 1), (2, 1)$ and i for $i = 2, \dots, I$ in the obvious way. Note that for any estimator $(F, \hat{\theta})$ we can construct an estimator $(MF, \hat{\theta}')$ by $\hat{\theta}'(1, 1) = \hat{\theta}'(1, 2) = \hat{\theta}(1)$ and for $i \geq 2$, $\hat{\theta}'(i) = \hat{\theta}(i)$. By construction $R(\hat{\theta}, \theta) \equiv R(\hat{\theta}', \theta)$ and thus $MF \preceq F$. It remains to show $F \not\preceq MF$, which we show by contradiction.

Suppose we find an MF -admissible estimator $\hat{\theta}$ where $\hat{\theta}(1, 1) \neq \hat{\theta}(1, 2)$. If there is an estimator $(F, \tilde{\theta})$ such that $(F, \tilde{\theta}) \leq (MF, \hat{\theta})$, then by the above construction

$(MF, \tilde{\theta}') \leq (MF, \hat{\theta})$. But by construction, $\tilde{\theta}' \neq \hat{\theta}$, and so by lemma D.3 this contradicts admissibility of $\hat{\theta}$ and hence $F \not\prec MF$. Thus the problem reduces to finding such an estimator $\hat{\theta}$.

Note that by lemma D.2 it is sufficient to show $p_{(1,1)}^{(MF)}(\theta_1)p_{(1,2)}^{(MF)}(\theta_2) \neq p_{(1,2)}^{(MF)}(\theta_1)p_{(1,1)}^{(MF)}(\theta_2)$. But by expanding probabilities and dividing through by $p_1^{(F)}(\theta_1)p_1^{(F)}(\theta_2)$, we see that

$$p_{(1,1)}^{(MF)}(\theta_1)p_{(1,2)}^{(MF)}(\theta_2) = p_{(1,2)}^{(MF)}(\theta_1)p_{(1,1)}^{(MF)}(\theta_2), \quad (70)$$

$$\Leftrightarrow p_1^{(M)}(\theta_1)p_2^{(M)}(\theta_2) = p_2^{(M)}(\theta_1)p_1^{(M)}(\theta_2), \quad (71)$$

$$\Leftrightarrow p_1^{(M)}(\theta_1) = p_1^{(M)}(\theta_2). \quad (72)$$

But by the definition of M the final equality (72) cannot hold. Thus we can apply lemma D.2 and a desired $\hat{\theta}$ exists. \square

Next, we consider the case when a measurement extracts no information from a system.

Lemma 5.2. *Let $\Theta \subseteq \mathbb{R}^N$, L be some Bregman divergence loss function on Θ and $\rho(\theta)$ be some non-constant parametrised state. Suppose that $F = \{F_k\}$ is a measurement whose outcome probabilities are independent of θ , then F is inadmissible under least-squares loss.*

Proof. Since ρ is not constant, take $\theta_1, \theta_2 \in \Theta$ such that $\rho(\theta_1) \neq \rho(\theta_2)$. Again, by the Helstrom bound, [10], there is some measurement $M = \{M_1, M_2\}$ where $p_1^{(M)}(\theta_1) \neq p_1^{(M)}(\theta_2)$. Then, note that

$$p_1^{(M)}(\theta_1)p_2^{(M)}(\theta_2) = p_2^{(M)}(\theta_1)p_1^{(M)}(\theta_2) \quad (73)$$

$$\Leftrightarrow p_1^{(M)}(\theta_1) = p_1^{(M)}(\theta_2). \quad (74)$$

So by lemma D.2 there exists a non-constant admissible estimator $\hat{\theta}$. Applying Lemma D.4, we see $M \preceq F$ but $F \not\prec M$. \square