

Richard D. Gill[†]

Dept. of Mathematics, University of Utrecht, Box 80010, 3508 TA Utrecht, Netherlands

Serge Massar[‡]

Service de Physique Théorique, CP 225, Université Libre de Bruxelles, B1050 Brussels, Belgium

We consider the problem of estimating the state of a large but finite number N of identical quantum systems. As N becomes large the problem simplifies dramatically. The only relevant measure of the quality of estimation becomes the mean quadratic error matrix. Here we present a bound on this quantity: a new quantum Cramér-Rao inequality. The new bound expresses succinctly how in the quantum case one can trade information about one parameter for information about another. The bound holds for arbitrary measurements on pure states, but only for separable measurements on mixed states—a striking example of non-locality without entanglement for mixed but not for pure states. Cramér-Rao bounds are generally only derived for unbiased estimators. Here we give a version of our bound for biased estimators, and a simple asymptotic version for large N . Finally we prove that when the unknown state belongs to a two dimensional Hilbert space our quantum Cramér-Rao bound can always be attained and we provide an explicit measurement strategy that attains it. Thus we have a complete solution to the problem of estimating as efficiently as possible the unknown state of a large ensemble of qubits in the same pure state. The same is true for qubits in the same mixed state if one restricts oneself to separable measurements, but non-separable measurements allow dramatic increase of efficiency. Exactly how much increase is possible is a major open problem.

03.67.a and 02.50.-r

I INTRODUCTION

One of the central problems of quantum measurement theory is the estimation of an unknown quantum state. Originally only of theoretical interest, this problem is becoming of increasing practical importance. Indeed there are now several beautiful experimental realizations of quantum state reconstruction in such diverse systems as quantum optics[1], molecular states[2], trapped ions[3] and atoms in motion[4].

The theoretical work which is the basis for these experiments is concerned with devising measurement strategies that are simple to realize experimentally and which allow an unambiguous reconstruction of the quantum state. The best known such technique is quantum state tomography[5], adapted in [6] for the case of finite dimensional Hilbert spaces. However other techniques are also available, see [7] for a recent discussion in the case of finite dimensional Hilbert spaces. However, all these works suppose that the measurements are perfect and that any operator can be measured with infinite precision. But in general the quality of the reconstruction will be limited by experimental error[8] or by finite statistics. The present work is devoted to studying this latter aspect when the unknown state belongs to a finite dimensional Hilbert space.

Thus the setting of the problem is that we may dispose of a finite number N of copies of an unknown quantum state ρ (pure or mixed). Our task is to determine ρ as well as possible. This is by now a classical problem[9][10].

A common approach is first to specify a cost function which numerically quantifies the deviation of the estimate from the true state. One then tries to devise a measurement and estimation strategy which minimizes the mean cost. Since the mean cost typically depends on the unknown state itself, one typically averages over all possible states to arrive at a single number expressing the quality of the estimation. However optimal strategies have only been found in some simple highly symmetric cases (the covariant measurements of [10], see also [11][12]).

However when the number of copies N becomes large one can hope that the problem becomes simpler so that one might be able to find the optimal strategies in this limit. The reason for this is that in the large N limit the estimation problem ceases to be a ‘global’ problem and becomes ‘local’. Indeed for small N the estimated state will often be very different from the true state. Hence the optimal measurement strategy must take into account the behaviour of the cost function for large estimation errors. On the other hand in the limit of an infinite number of copies any two

[†]Electronic address: gill@math.uu.nl; URL: <http://www.math.uu.nl/people/gill>; also affiliated to EURANDOM, Netherlands

[‡]Electronic address: smassar@ulb.ac.be

states can be distinguished with certainty. So the relevant question to ask about the estimation strategy is at what *rate* it distinguishes neighboring states. And in that case we are only concerned with the behaviour of the estimator and of the cost function very close to the true value.

To formulate the problem with precision, let us suppose that the unknown state $\rho(\theta)$ depends on a vector of p unknown real parameters $\theta = (\theta_1, \dots, \theta_p)$. For instance the θ_i could correspond to various settings or physical properties of the apparatus that produces the state ρ . After carrying out a measurement on the N copies of ρ , one will guess what is θ . Call $\hat{\theta}^N = (\hat{\theta}_1^N, \dots, \hat{\theta}_p^N)$ the guessed value. For a good estimation strategy we expect the mean quadratic error (m.q.e.) to decrease as $1/N$:

$$\mathbb{E}_\theta \left((\hat{\theta}_i^N - \theta_i)(\hat{\theta}_j^N - \theta_j) \right) \equiv V_{ij}^N(\theta) \simeq \frac{W_{ij}(\theta)}{N} \quad (1)$$

where the scaled m.q.e. matrix $W(\theta) = (W_{ij}(\theta)) \simeq NV^N(\theta)$ does not depend on N . \mathbb{E}_θ denotes the mean taken over repetitions of the measurement with the value of θ fixed.

Consider now a smooth cost function $f(\hat{\theta}, \theta)$, which measures how much the estimated value $\hat{\theta}$ differs from the true value θ of the parameter. f will have a minimum at $\hat{\theta} = \theta$, hence can be expanded as

$$f(\hat{\theta}, \theta) = f_0(\theta) + \sum_{ij} C_{ij}(\theta)(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) + O(\|\hat{\theta} - \theta\|^3) \quad (2)$$

where $C(\theta) = (C_{ij}(\theta))$ is a nonnegative matrix. Thus for a reasonable estimation strategy the mean value of the cost will decrease as

$$\mathbb{E}_\theta(f(\hat{\theta}^N, \theta)) = f_0(\theta) + N^{-1} \sum_{ij} C_{ij}(\theta)W_{ij}(\theta) + o(N^{-1}) \quad (3)$$

since we expect the expectation value of higher order terms in $\hat{\theta} - \theta$ to decrease faster than $1/N$. The problem has become local: only the quadratic cost matrix $C(\theta)$ and the scaled mean quadratic error matrix $W(\theta)$ at θ intervene. The essential question about state estimation for large ensembles is therefore *what scaled m.q.e. matrices $W(\theta)$ are attainable through arbitrary measurement and estimation procedures?* In particular, what does the boundary of this set of attainable m.q.e. matrices look like?

In the case when the parameter θ is one-dimensional ($p = 1$) the problem has been solved: a bound on the variance of unbiased estimators—the quantum Cramér-Rao bound—was given in [9], and a strategy for attaining the bound in the large N limit was proposed in [15]. This justifies taking the bound to induce a ‘distinguishability metric’ on the space of states [13][14]. In the case of a multidimensional parameter however, though different bounds for the matrix W have been established, in general they are not tight [9][16][10].

In this paper we present a new bound for W in the multiparameter case which is inspired by the discussion in [15]. This bound expresses in a natural way how one can trade information about one parameter for information about another. The interest of this new bound depends on the precise problem one is considering:

- When $\rho(\theta) = |\psi(\theta)\rangle\langle\psi(\theta)|$ is a pure state belonging to a 2 dimensional Hilbert space, the bound is sharp: it provides a necessary and sufficient condition that W must satisfy in order to be attainable. Furthermore, the bound can be attained by carrying out separate measurements on each particle. This completely solves the problem of estimating the state of a large ensemble of spin 1/2 particles (qubits) in the same pure state.
- When $\rho(\theta)$ is a pure state belonging to a Hilbert space of dimension d larger than 2, then our bound on W applies, but it is not sharp.
- When the unknown state is mixed and belongs to a 2 dimensional Hilbert space, and if one restricts oneself to measurements that act separately on each particle, then our bound applies and is sharp.
- When the unknown state is mixed and belongs to a Hilbert space of dimension $d > 2$, and if one restricts oneself to measurements that act separately on each particle, then our bound applies but is not sharp.
- If the unknown state is mixed and one allows collective measurements, then our bound is not necessarily satisfied.

This last point is surprising and points to a fundamental difference between measuring pure states and mixed states. Indeed it is known that carrying out measurements on several identical copies of the same pure state can generally be done better with collective measurements on the different copies[17][11]. This is known as ‘non-locality without entanglement’[18]. The first point shows that in the limit of a large number of copies, pure states of spin 1/2 do not exhibit non-locality without entanglement. On the other hand the last point shows that in the limit of a large number of copies mixed states of spin 1/2 continue to exhibit non locality without entanglement.

To describe our bound on W , we first consider for simplicity the case of a pure state of spin 1/2 particles. Suppose the unknown state is a spin 1/2 known to be in a pure state, and the state is known to be almost pointing in the $+z$ direction:

$$|\psi(\theta_1, \theta_2)\rangle \simeq |\uparrow_z\rangle + \frac{1}{2}(\theta_1 + i\theta_2)|\downarrow_z\rangle \quad (4)$$

where we have written an expression valid to first order in θ_1, θ_2 . Suppose we carry out a measurement of the operator σ_x . We obtain the outcome $\pm x$ with probability $p(\pm x) = (1 \pm \theta_1)/2$. Thus the outcome of this measurement tells us about the value of θ_1 . Similarly we can carry out a measurement of σ_y . We obtain the outcome $\pm y$ with probability $p(\pm y) = (1 \pm \theta_2)/2$. The outcome of this measurement tells us about θ_2 . But the measurements σ_x and σ_y are incompatible, i.e., the operators do not commute and cannot be measured simultaneously. Thus if one obtains knowledge about θ_1 , it is at the expense of θ_2 . Indeed suppose one has N copies of the state ψ and one measures σ_x on N_1 copies and σ_y on $N_2 = N - N_1$ copies. Our estimator for θ_1 is the fraction of $+x$ outcomes minus the fraction of $-x$ outcomes. This estimator is unbiased. The resulting uncertainty (at the point $\theta_1 = \theta_2 = 0$) about θ_1 is then $E_\theta((\hat{\theta}_1 - \theta_1)^2) = \frac{1}{N_1}$. Similarly we can estimate θ_2 and the corresponding uncertainty is $E_\theta((\hat{\theta}_2 - \theta_2)^2) = \frac{1}{N_2}$. We can combine these two expressions in the following relation:

$$\frac{1}{E_\theta((\hat{\theta}_1 - \theta_1)^2)} + \frac{1}{E_\theta((\hat{\theta}_2 - \theta_2)^2)} = \frac{1}{V_{11}^N} + \frac{1}{V_{22}^N} = N \quad (5)$$

which expresses in a compact form how we can trade knowledge about θ_1 for knowledge about θ_2 . We shall show that it is impossible to do better than precisely (5) when one restricts attention to unbiased estimators based on arbitrary measurements, and asymptotically not possible to do better with any estimator whatsoever.

To generalize (5), we rewrite it in a more abstract form, and state it as an inequality. We use polar coordinates to parameterize the unknown state of the spin 1/2 particle: $|\psi\rangle = \cos \frac{\eta}{2} |\uparrow\rangle + \sin \frac{\eta}{2} e^{i\varphi} |\downarrow\rangle$. We introduce the tensor

$$H_{\eta\eta} = 1 \quad , \quad H_{\varphi\varphi} = \sin^2 \eta \quad , \quad H_{\eta\varphi} = 0 \quad (6)$$

which is simply the Euclidean metric on the sphere. Then the bound (5) can be reexpressed as

$$\text{tr } H^{-1}(V^N)^{-1} \leq N \quad (7)$$

where V^N is the m.q.e. matrix defined in (1).

For mixed states belonging to a 2 dimensional Hilbert space, (7) can be generalized as follows. Let us suppose that the state $\rho(\theta)$ depends on three unknown parameters. Then we can parameterize it by $\rho(\theta) = \frac{1}{2}(I + \sum_i \theta_i \sigma_i)$ where I is the identity matrix, σ_i are the Pauli matrices and the 3 parameters θ_i obey $\|\theta\|^2 = \sum_i \theta_i^2 \leq 1$. We now introduce the tensor

$$H_{ij}(\theta) = \delta_{ij} + \frac{\theta_i \theta_j}{1 - \|\theta\|^2} \quad (8)$$

which generalizes the tensor (6) to the case of mixed states. Then, upon restricting oneself to separable measurements, we will show that the m.q.e. matrix V^N must satisfy (exactly for unbiased estimators, and otherwise asymptotically)

$$\text{tr } H(\theta)^{-1} V^N(\theta)^{-1} \leq N \quad (9)$$

As an application of these results, the minimum of the cost function (3) in the case of spin 1/2 particles (for mixed states restricting oneself to separable measurement) is

$$\min E_{\theta}(f(\hat{\theta}, \theta)) = f_0(\theta) + \frac{\left(\text{tr} \sqrt{H(\theta)^{-1/2} C(\theta) H(\theta)^{-1/2}}\right)^2}{N} + o(1/N) \quad (10)$$

which is obtained simply by minimizing (3) subject to the constraints (7) or (9).

We can compare (10) with the exact results which are known in the case of covariant measurements on pure states of spin 1/2 particles [10][11]. In this problem one is given N spin 1/2 particles polarized along the direction Ω . The Ω are uniformly distributed on the sphere. One wants to devise a measurement and estimation strategy that minimize the mean value of the cost function $\cos^2 \omega/2$, where ω is the angle between the estimated direction $\hat{\Omega}$ and the true direction Ω . Expanding the cost function to second order in ω (to obtain the quadratic cost matrix C) and averaging (10) over the sphere, one finds

$$E(\cos^2 \omega/2) \geq 1 - \frac{1}{N} + o\left(\frac{1}{N}\right) \quad (11)$$

which in the limit for large N coincides with the results (exact for all N) of [10][11]. If the directions Ω are not uniformly distributed, then [10][11] do not apply, but (10) stays valid. However we cannot compare our results with the recent analysis of covariant measurements on mixed states[12] because we suppose separability of the measurement, whereas [12] does not.

Equations (7) and (9) have a simple generalization to the case of particles belonging to higher dimensional Hilbert spaces. But in these cases these bounds are no longer sharp.

In order to appreciate the above results, we must recall some results from classical statistical inference. This is the subject of the next section.

II CLASSICAL CRAMÉR-RAO BOUND

Consider a random variable X with probability density $p(x, \theta)$. The connection with the quantum problem is that we can view $p(x, \theta)$ as the probability density that a quantum measurement on the system yields outcome x given that the state was $\rho(\theta)$. We take a random sample of size N from the distribution and use it to estimate the value of each parameter θ_i . Call $\hat{\theta}_i^N$ the estimated value. The following results about the m.q.e. matrix of the estimator are well known:

1. Suppose that the estimator is unbiased, that is $E_{\theta}(\hat{\theta}^N - \theta) = 0$, where E_{θ} is the expectation value at fixed θ , i.e., the integral $\int dx p(x|\theta)$. Define its m.q.e. matrix $V^N(\theta)$ by

$$V_{ij}^N(\theta) = E_{\theta}((\hat{\theta}_i^N - \theta_i)(\hat{\theta}_j^N - \theta_j)). \quad (12)$$

Furthermore define the Fisher information matrix $I(\theta)$ by

$$\begin{aligned} I_{ij}(\theta) &= E_{\theta}(\partial_{\theta_i} \ln p(X|\theta) \partial_{\theta_j} \ln p(X|\theta)) \\ &= \int dx \frac{\partial_{\theta_i} p(x|\theta) \partial_{\theta_j} p(x|\theta)}{p(x|\theta)}. \end{aligned} \quad (13)$$

Then for any N , the following inequalities, known as the Cramér-Rao inequalities, hold[19][9]

$$V^N(\theta) \geq I(\theta)^{-1}/N \quad (14)$$

or equivalently

$$V^N(\theta)^{-1} \leq NI(\theta), \quad (15)$$

the inequality meaning that the difference of the two sides is a nonnegative matrix.

2. The hypothesis of unbiased estimators is very restrictive since most estimators will be biased. Happily it is possible to relax this condition. Here are just two of the many results available:

- (a) First of all, if one is interested in averaging the mean cost over possible values of θ with respect to a given prior distribution $\lambda(\theta)$, then there is a Bayesian version of the Cramér-Rao inequality, the van Trees inequality [20][21]. In the multivariate case, upon giving oneself a quadratic cost function determined by a matrix $C(\theta)$, one can derive the inequality

$$\int d\theta \lambda(\theta) \text{tr } C(\theta) V^N(\theta) \geq \frac{\int d\theta \lambda(\theta) \text{tr } C(\theta) I^{-1}(\theta)}{N} - \frac{\alpha}{N^2} \quad (16)$$

where α is a positive number that depends on $C(\theta)$, $I(\theta)$, and $\lambda(\theta)$, but is independent of N .

- (b) The second approach makes no reference to any prior distribution for θ , but only holds in the limit N tending to infinity and lays a mild restriction on the estimators considered. Specifically, if the probability distribution of $\sqrt{N}(\hat{\theta}^N - \theta)$ converges uniformly in θ towards a distribution depending continuously on θ , say of a random vector Z , then the limiting scaled m.q.e. matrix $W(\theta)$ defined by $W_{ij}(\theta) = E_\theta(Z_i Z_j)$ obeys $W \geq I^{-1}$.

3. Furthermore in the limit of arbitrarily large samples one can attain the Cramér-Rao bound. This is proven by explicitly constructing an estimator that attains the bound in the extended senses 2a) (apart from the $1/N^2$ term) or 2b) just indicated: the maximum likelihood estimator (m.l.e.).

Modern statistical theory contains many other results having the same flavour as point 2 above, namely that the Cramér-Rao bound holds in an approximate sense for large N , without the restriction to biased estimators. Result 2a) applies to a larger class of estimators than 2b), but only gives a result on the average behaviour over different values of θ . On the other hand combining results 3) and 2b) tells us that the maximum likelihood estimator is for large N an optimal estimator for each value of θ separately. The reason why in 2b) additional regularity is demanded is because of the phenomenon of super-efficiency (see [22] for a recent discussion) whereby an estimator can have mean quadratic error of smaller order than $1/N$ at isolated points. Modern statistical theory (see again [22] or [23]) has concentrated on the more difficult problem of obtaining non-Bayesian results (i.e., pointwise rather than average) making much use of the technical tool of ‘local asymptotic normality’. A major challenge in the quantum case is to obtain a result of type 2b) when this technique is definitely not available.

III QUANTUM CRAMÉR-RAO BOUND

In this paper we show that similar results to 1, 2a, 2b, 3 can be obtained when one must estimate the state of an unknown quantum system $\rho(\theta)$ of which one possesses N copies. This problem is most simply addressed, following [14], by decomposing it into a first (quantum) step in which one carries out a measurement on $\rho^N = \rho \otimes \dots \otimes \rho$ and a second (classical) step in which one uses the result of the measurement to estimate the value of the parameters θ .

The most general way to describe the measurement is by a positive operator-valued measurement (POVM) $M = (M_\xi)$ whose elements satisfy $M_\xi \geq 0$, $\sum_\xi M_\xi = I$. (For simplicity we take the outcomes of the POVM to be discrete.) The generalization to an arbitrary outcome space is just a question of translating into measure-theoretic language.)

Quantum mechanics tells us the probability to obtain outcome ξ given state $\rho(\theta)$:

$$p(\xi|\theta) = \text{tr } \rho^N(\theta) M_\xi. \quad (17)$$

From the outcome ξ of the measurement one must guess what are the values of the p parameters θ_i . Call $\hat{\theta}^N$ the estimated value of the parameter vector. We want to obtain bounds on the m.q.e. matrix $V^N(\theta)$ of the estimator $\hat{\theta}^N$ when the true parameter value is θ , thus $V_{ij}^N(\theta) = E_\theta(\hat{\theta}_i^N - \theta_i)(\hat{\theta}_j^N - \theta_j)$. To proceed we temporarily make the simplifying assumption that the estimators are unbiased, $E_\theta \hat{\theta}^N = \theta$. Then we can apply the classical Cramér-Rao inequality to the probability distribution $p(\xi|\theta)$ to obtain:

$$V^N \geq I^N(M, \theta)^{-1} \quad (18)$$

or

$$(V^N)^{-1} \leq I^N(M, \theta) \quad (19)$$

where the Fisher information matrix I^N for the measurement M is defined by

$$\begin{aligned} I_{ij}^N(M, \theta) &= \sum_{\xi} \frac{\partial_i p(\xi|\theta) \partial_j p(\xi|\theta)}{p(\xi|\theta)} \\ &= \sum_{\xi} \frac{\text{tr}(\rho_{,i}^N M_{\xi}) \text{tr}(\rho_{,j}^N M_{\xi})}{\text{tr}(M_{\xi} \rho^N)} \end{aligned} \quad (20)$$

with $\rho_{,i}^N = \partial_{\theta_i} \rho^N$.

These expressions suggest the following questions:

1. is there a simple bound for the m.q.e. V^N of unbiased estimators $\hat{\theta}^N$, or equivalently for the Fisher information $I^N(M, \theta)$?
2. is the bound also valid for sufficiently well behaved but possibly biased estimators—at least in the limit of large N ?
3. can this bound be attained—at least in the limit of a large number of copies N ?

Most of the work on this subject has been devoted to answering the question 1). We now recall what is known about these questions.

Suppose first the parameter θ is one-dimensional, $p = 1$. The symmetric logarithmic derivative (s.l.d.) λ_{θ} of ρ is the Hermitian matrix defined implicitly by

$$\rho_{,\theta} = \frac{\lambda_{\theta} \rho + \rho \lambda_{\theta}}{2}. \quad (21)$$

In a basis where ρ is diagonal, $\rho = \sum_k p_k |k\rangle \langle k|$, this can be inverted to yield

$$(\lambda_{\theta})_{kl} = (\rho_{,\theta})_{kl} \frac{2}{p_k + p_l}. \quad (22)$$

Then we have the bound

$$I_{\theta\theta}^N(M, \theta) \leq N \text{tr} \rho \lambda_{\theta} \lambda_{\theta}. \quad (23)$$

Furthermore it was suggested in [15] how to adapt the classical m.l.e. so as to attain, in the limit of large N , the bound (23).

In the multiparameter case the bound based on the s.l.d. can be generalized in a natural way. Define the s.l.d. along direction θ_i by

$$\rho_{,i} = \frac{\lambda_i \rho + \rho \lambda_i}{2}, \quad (24)$$

and Helstrom's quantum information matrix H by

$$H_{ij} = \text{tr} \rho \frac{\lambda_i \lambda_j + \lambda_j \lambda_i}{2}. \quad (25)$$

(This is the same matrix that was introduced for spin 1/2 particles for a particular choice of parameters in (6) and (8)). Then one can prove the bound [9],

$$I^N(M, \theta) \leq NH(\theta). \quad (26)$$

(This can be deduced directly from (23) as proven in [14]. Indeed since (23) holds for each path in parameter space, it implies the matrix equation (26)).

However this bound is in general not achievable. Another bound has been proposed based on an asymmetric logarithmic derivative (a.l.d.) [16] which in some cases is better than (26). Holevo [10] has proposed yet another bound that is stronger than both the s.l.d. and the a.l.d. bound, but this bound is not explicit: it requires a further minimization. As far as we know no general achievable bound is known in the multiparameter case.

The difficulty in obtaining a simple bound in the multiparameter case is that there are many inequivalent ways in which one can minimize the m.q.e. matrix V_{ij}^N . That is, in order to build a good estimator one must make a choice of what one wants to estimate, and according to this choice the measurement strategy followed will be different. Hence a bound in the form of a matrix inequality like (26) cannot be expected to be tight.

IV RESULTS

In this paper we obtain answers to the three questions raised above in the multiparameter case. Our results are summarized in this section.

We first discuss point 1), that is bounds on the Fisher information. We shall show the following:

Theorem I: When $\rho(\theta) = |\psi(\theta)\rangle\langle\psi(\theta)|$ is a pure state, then the Fisher information $I^N(M, \theta)$ defined in (20) must satisfy the following relation

$$\text{tr } H^{-1}(\theta) I^N(M, \theta) \leq (d-1)N \quad (27)$$

where H^{-1} is the inverse of the quantum information matrix defined in (25) and d is the dimension of the Hilbert space to which $\rho(\theta)$ belongs. Note that this inequality (27) is invariant under change of parameterization $\theta \rightarrow \theta'(\theta)$.

This result immediately gives an inequality for the mean quadratic error matrix of *unbiased* estimators $\hat{\theta}^N$ by invoking the classical Cramér-Rao inequality in order to replace $I^N(M, \theta)$ by the inverse of the m.q.e. $V^N(\theta)$:

$$\text{tr } H^{-1}(\theta) (NV^N(\theta))^{-1} \leq d-1. \quad (28)$$

Theorem II: When $\rho(\theta)$ is a mixed state, and if the measurement M consists of separate measurements on each particle, then the Fisher information also satisfies (27). Hence for separable measurements on a mixed state, the m.q.e. matrix of an unbiased estimator satisfies (28).

Theorem III (non additivity of optimal Fisher information): In the case of mixed states, it is in general possible to devise a collective measurement for which the Fisher information does not satisfy the inequality (27).

The second part of the paper consists in proving that the constraint (28) also holds for biased estimators under suitable additional conditions. We give two forms of this generalized form of (28) corresponding to the two forms 2a) and 2b) of the generalized classical Cramér-Rao inequality.

Consider N copies of a state $\rho(\theta)$. If ρ is pure we can make either collective or separable measurements. If ρ is mixed we restrict ourselves to separable measurements (since Theorem III shows that in this case collective measurements can beat (27)). Based on the outcome of the measurement we estimate the value of the parameter vector θ . Call $\hat{\theta}$ the estimator, and denote by $V^N = V^N(\theta)$ its m.q.e. matrix when the true value of the parameter is θ .

We shall prove the following generalization of result of type 2b) concerning the behaviour of the mean quadratic error matrix as N tends to infinity:

Theorem IV: Suppose that the scaled m.q.e. $NV^N(\theta)$ has the limit $W(\theta)$ as $N \rightarrow \infty$. Suppose that the convergence is uniform in θ and that W is continuous at the point $\theta = \theta^0$. Furthermore we suppose that H and its derivatives are bounded in a neighbourhood of this point. Then we shall prove in section VI that $W(\theta^0)$ must satisfy

$$\text{tr } H^{-1}(\theta^0)W^{-1}(\theta^0) \leq (d-1). \quad (29)$$

This result gives a bound on the mean value of a quadratic cost function C as N tends to infinity. Indeed using a Lagrange multiplier to impose the condition (29), the minimum cost is readily found to be

$$\lim_{N \rightarrow \infty} N \text{tr } C(\theta^0) V^N(\theta^0) \geq \left(\text{tr } \sqrt{H^{-\frac{1}{2}}(\theta^0) C(\theta^0) H^{-\frac{1}{2}}(\theta^0)} \right)^2. \quad (30)$$

In terms of a cost function, it is also possible to prove a Bayesian version of the Cramér-Rao inequality which is the analogue of the classical result 2a):

Theorem V: Suppose that one is given a quadratic cost function $C(\theta)$ and a prior distribution $\lambda(\theta)$ for the parameters θ . If C , λ and H are sufficiently smooth functions of θ (continuity of the first derivatives is sufficient), while λ is zero outside a compact region with smooth boundary, then

$$\int d\theta \lambda(\theta) \text{tr } C(\theta) V^N(\theta) \geq \frac{1}{N} \int d\theta \lambda(\theta) \text{tr} \left(\sqrt{H^{-\frac{1}{2}}(\theta) C(\theta) H^{-\frac{1}{2}}(\theta)} \right)^2 - \frac{\alpha}{N^2} \quad (31)$$

where α is a constant independent of N but which depends on C , λ and H .

Theorems I, II, IV and V put bounds on the m.q.e. matrix of an estimator of an unknown state $\rho(\theta)$ (for mixed states, under the restriction that the measurement is separable). The third part of this article is devoted to showing that in the case of spin 1/2 systems ($d = 2$) these bounds can be attained. We first show that at any point θ^0 we can attain equality in (27).

Theorem VI: Suppose one has N spin 1/2 particles in an unknown (possibly mixed) state $\rho(\theta)$. Fix any point θ^0 . Give yourself a matrix G^0 satisfying $\text{tr } H^{-1}(\theta^0) G^0 \leq 1$. We call G^0 the target scaled information matrix. Then there exists a measurement M^{θ^0} (depending on the choice of θ^0) acting on each spin separately such that $I^N(M^{\theta^0}, \theta^0) = N G^0$. This measurement is described in detail in section VII A.

For large N we can also approximately attain equality at all points θ simultaneously:

Theorem VII: Suppose one has N spin 1/2 particles in an unknown pure state $|\psi(\theta)\rangle$ or suppose that one has N spin 1/2 particles in an unknown mixed state $\rho(\theta)$. In the latter case we also require that the state never be pure, i.e. $\text{tr } \rho(\theta)^2 < 1$ for all θ .

Give oneself a smooth positive matrix function $G(\theta)$ satisfying $\text{tr } H^{-1}(\theta) G(\theta) \leq 1$ for all θ , the target scaled information for each possible value of θ . Define the corresponding target scaled m.q.e. matrix $W(\theta) = G(\theta)^{-1}$. Suppose that $W(\theta)$ is non singular (i.e. $G(\theta)$ never has a zero eigenvalue).

Then there exists a measurement M acting on each spin separately, and a corresponding estimator $\hat{\theta}$, whose m.q.e. matrix $V^N(\theta)$ satisfies

$$V^N(\theta) = \frac{W(\theta)}{N} + o(1/N) \quad (32)$$

for all values of θ simultaneously. For this estimation strategy $\sqrt{N}(\hat{\theta} - \theta)$ converges in distribution towards $N(0, W)$, the normal distribution with mean zero and covariance W . The measurement M and estimation strategy is described in detail in section VII B.

It is interesting to note that the measurement strategy which satisfies (32) is an adaptive one. That is one first carries out a measurement on a small fraction of the particles. This gives a preliminary estimate of the quantum state which allows a fine tuning of the measurements that are carried out on the remaining particles. This is to be contrasted to previously proposed state estimation strategies in the case of finite dimensional Hilbert spaces [7][6] in which the same measurement is carried out on all the particles. The necessity of an adaptive measurement strategy if one wants to minimise the m.q.e. was pointed out in [15].

When the unknown state belongs to a Hilbert space of dimension $d > 2$, then the bound (27) cannot be attained in general. Indeed we shall show in section V F that for $d > 2$, neither (26) nor (27) implies the other.

In this section we prove Theorems I, II, III. That is we prove (27) for general measurements in the case of pure states and for separate measurements on each particle in the case of mixed states.

A Preliminary results

The first step in proving (27) is to show that one can restrict oneself to POVM's whose elements are proportional to one dimensional projectors. Indeed any POVM can always be refined to yield a POVM whose elements are proportional to one dimensional projectors. We call such a measurement *exhaustive*. This yields a refined probability distribution $(p(\xi, \theta))$. It is well known that under such refining of the probability distribution, the Fisher information can only increase[24].

The second step in proving (27) consists in increasing the number of parameters. Suppose that $\rho(\theta)$ depends on p parameters θ_i , $i = 1, \dots, p$. If $\rho = |\psi(\theta)\rangle\langle\psi(\theta)|$ is a pure state, then $p \leq 2d - 2$ (since $|\psi(\theta)\rangle$ is normalized and defined up to a phase). If ρ is a mixed state, then Hermiticity and the condition $\text{tr } \rho = 1$ impose that $p \leq d^2 - 1$. Suppose that $p < \nu$ is less then the maximum number of possible parameters ($\nu = 2d - 2$ or $\nu = d^2 - 1$ according to whether the state is pure or mixed). Then one can always increase the number of parameters up to the maximum. Indeed let us suppose that to the p parameters, one adds independent parameters $\theta_{i'}$, $i' = p + 1, \dots, \nu$. We can now consider the quantum information matrix \tilde{H} , and Fisher information matrix \tilde{I} , for the completed set of parameters. We shall show below that

$$\text{tr } H^{-1}(\theta) I^N(M, \theta) \leq \text{tr } \tilde{H}^{-1}(\theta) \tilde{I}^N(M, \theta). \quad (33)$$

Therefore it will be sufficient to prove (27) in the case when there are ν parameters.

To prove (33), fix a particular point θ^0 . At this point we have the derivative $\rho_{,i}$ and s.l.d. λ_i of ρ for $i = 1, \dots, p$. Introduce a set of Hermitian matrices $\lambda_{i'}$ with $\text{tr } \rho(\theta^0) \lambda_{i'} = 0$, for $i' = p + 1, \dots, \nu$, such that

$$\text{tr } \rho(\theta^0) \frac{\lambda_i \lambda_{i'} + \lambda_{i'} \lambda_i}{2} = 0 \quad , \quad i = 1, \dots, p \quad , \quad i' = p + 1, \dots, \nu. \quad (34)$$

This is always possible because we can view (34) as a scalar product between λ_i and $\lambda_{i'}$ and a Gram-Schmidt orthogonalization procedure will then yield the matrices $\lambda_{i'}$. Now define matrices $\rho_{,i'}$ by $\rho_{,i'} = (\rho(\theta^0) \lambda_{i'} + \lambda_{i'} \rho(\theta^0))/2$ and define additional parameters $\theta_{i'}$ satisfying, at θ^0 , $\partial_{\theta_{i'}} \rho = \rho_{,i'}$. The point of this construction is that because of (34), the quantum information matrix \tilde{H} is block diagonal with the first block equal to H . Let $\tilde{I}(M)$ be the Fisher information matrix for the enlarged set of parameters (but the same measurement). Then $\text{tr } \tilde{H}^{-1} \tilde{I}(M) = \text{tr } (\tilde{H}^{-1})_{11} \tilde{I}_{11}(M) + \text{tr } (\tilde{H}^{-1})_{22} \tilde{I}_{22}(M)$ where the indices 11 and 22 denote the blocks of these matrices corresponding to the original and the new parameters. But both terms are nonnegative since all matrices involved are nonnegative, and $(\tilde{H}^{-1})_{11} = H^{-1}$, so we obtain (33) at θ_0 and for the particular parameters just introduced. But since the right hand side of (33) is invariant under reparametrization, it is true for any parametrization, and at any θ .

B Pure states

To proceed we shall consider a POVM whose elements are proportional to one dimensional projectors and calculate explicitly the left hand side of (27) in the case where the number of parameters is the maximum $p = 2d - 2$ in a basis where H is diagonal.

We fix a point θ^0 . At this point we chose a basis such that

$$\rho(\theta^0) = |1\rangle\langle 1| . \quad (35)$$

Hence the density matrix of the N copies is

$$\rho^N = |1\rangle\langle 1| \otimes \dots \otimes |1\rangle\langle 1| . \quad (36)$$

Consider now the $2d - 2$ Hermitian operators

$$\begin{aligned}\rho_{,k+} &= |1\rangle\langle k| + |k\rangle\langle 1| \quad , \quad 1 < k \leq d \quad , \\ \rho_{,k-} &= i|1\rangle\langle k| - i|k\rangle\langle 1| \quad , \quad 1 < k \leq d \quad .\end{aligned}\tag{37}$$

We choose a parameterisation such that in the vicinity of θ^0 , it has the form $\rho = \rho(\theta^0) + \sum_{k,\pm} (\theta_{k\pm} - \theta_{k\pm}^0) \rho_{,k\pm}$ with the unknown parameters $\theta_{k\pm}$, $k = 2, \dots, d$. With this parametrisation the derivatives of ρ^N are

$$\rho_{,k\pm}^N = \rho_{,k\pm} \otimes \rho \dots \otimes \rho + \dots + \rho \otimes \dots \otimes \rho_{,k\pm} \quad .\tag{38}$$

One then calculates the s.l.d. of ρ and hence the quantum information matrix H . One verifies that in this basis H is diagonal:

$$H_{k\pm, k'\pm'} = 4\delta_{kk'}\delta_{\pm\pm'} \quad .\tag{39}$$

Consider any POVM whose elements are proportional to one dimensional projectors

$$M_\xi = |\psi_\xi\rangle\langle\psi_\xi| \quad , \quad |\psi_\xi\rangle = \sum_{k_1=1}^d \dots \sum_{k_N=1}^d a_{\xi k_1 \dots k_N} |k_1 \dots k_N\rangle \quad .\tag{40}$$

The completeness relation $\sum_\xi M_\xi = I$ takes the form

$$\sum_\xi a_{\xi k_1 \dots k_N}^* a_{\xi k'_1 \dots k'_N} = \delta_{k_1 k'_1} \dots \delta_{k_N k'_N} \quad .\tag{41}$$

To proceed we need the following formulae:

$$\text{tr } \rho(\theta^0) M_\xi = |a_{\xi 1 \dots 1}|^2\tag{42}$$

and

$$\text{tr } \rho(\theta^0)_{,k+} M_\xi = \sum_{p=1}^N (a_{\xi 1 \dots 1}^* a_{\xi 1 \dots k_p = k \dots 1} + a_{\xi 1 \dots k_p = k \dots 1}^* a_{\xi 1 \dots 1})\tag{43}$$

and similarly for $\text{tr } \rho(\theta^0)_{,k-} M_\xi$. Thus we obtain

$$(\text{tr } \rho(\theta^0)_{,k+} M_\xi)^2 + (\text{tr } \rho(\theta^0)_{,k-} M_\xi)^2 = \sum_{p=1}^N 4|a_{\xi 1 \dots 1}|^2 |a_{\xi 1 \dots k_p = k \dots 1}|^2 \quad .\tag{44}$$

Putting everything together yields

$$\begin{aligned}\text{tr } H^{-1} I(M) &= \sum_\xi \frac{1}{\text{tr } \rho(\theta^0) M_\xi} \frac{1}{4} \sum_{k=2}^d \sum_{\pm} (\text{tr } \rho(\theta^0)_{,k+} M_\xi)^2 + (\text{tr } \rho(\theta^0)_{,k-} M_\xi)^2 \\ &= \sum_{k=2}^d \sum_{p=1}^N \sum_\xi |a_{\xi 1 \dots k_p = k \dots 1}|^2 \\ &= N(d-1)\end{aligned}\tag{45}$$

which proves that equality holds in (27) for arbitrary exhaustive measurements in the case of pure states.

Deriving (27) for mixed states is more complicated than for pure, and we shall proceed in two steps. First we shall consider the case of one mixed state ($N = 1$) and show that equality in (27) holds in this case for arbitrary exhaustive measurements. Then we shall consider the case of an arbitrary number N of mixed states.

We first diagonalize ρ at a point θ^0 : $\rho(\theta^0) = \sum_{k=1}^d p_k |k\rangle \langle k|$. We now introduce the following complete set of Hermitian traceless matrices:

$$\begin{aligned} \rho_{,kl+} &= |k\rangle \langle l| + |l\rangle \langle k| \quad , \quad k < l \quad , \\ \rho_{,kl-} &= i |k\rangle \langle l| - i |l\rangle \langle k| \quad , \quad k < l \quad , \\ \rho_{,m} &= \sum_{k=1}^d c_{mk} |k\rangle \langle k| \quad , \quad m = 1, \dots, d-1 \end{aligned} \quad (46)$$

where the coefficients c_{mk} obey

$$\begin{aligned} \sum_k c_{mk} &= 0 \quad , \\ \sum_k \frac{1}{p_k} c_{m'k} c_{mk} &= \delta_{m'm} . \end{aligned} \quad (47)$$

Let us denote the matrices $\rho_{,kl\pm}$ and $\rho_{,m}$ collectively as $\rho_{,i}$. (They constitute a set of generators of $\text{su}(d)$).

We choose a parameterization such that in the vicinity of θ^0 , it has the form $\rho = \rho(\theta^0) + \sum_i (\theta_i - \theta_i^0) \rho_{,i}$. One then calculates the s.l.d. of ρ and from this the quantum information matrix H . One verifies that in this basis H is diagonal:

$$\begin{aligned} H_{kl\pm, k'l'\pm'} &= \frac{4}{p_k + p_l} \delta_{kk'} \delta_{ll'} \delta_{\pm\pm'} \quad , \\ H_{kl\pm, m} &= 0 \quad , \\ H_{m, m'} &= \delta_{m'm} . \end{aligned} \quad (48)$$

Consider any POVM whose elements are proportional to one dimensional projectors

$$\begin{aligned} M_\xi &= |\psi_\xi\rangle \langle \psi_\xi| \quad , \\ |\psi_\xi\rangle &= \sum_k a_{\xi k} |k\rangle . \end{aligned} \quad (49)$$

The l.h.s. of (27) can now be written as

$$\text{tr } H^{-1} I(M) = \sum_\xi \frac{1}{\langle \psi_\xi | \rho | \psi_\xi \rangle} \left(\sum_{k < l} \sum_{\pm} \frac{p_k + p_l}{4} \langle \psi_\xi | \rho_{,kl\pm} | \psi_\xi \rangle^2 + \sum_m \langle \psi_\xi | \rho_{,m} | \psi_\xi \rangle^2 \right) . \quad (50)$$

Using the following expressions

$$\begin{aligned} \langle \psi_\xi | \rho_{,m} | \psi_\xi \rangle &= \sum_k |a_{\xi k}|^2 c_{mk} \quad , \\ \langle \psi_\xi | \rho_{,kl+} | \psi_\xi \rangle^2 + \langle \psi_\xi | \rho_{,kl-} | \psi_\xi \rangle^2 &= 4 |a_{\xi k}|^2 |a_{\xi l}|^2 \end{aligned} \quad (51)$$

one obtains

$$\begin{aligned}
\text{tr } H^{-1}I(M) &= \sum_{\xi} \frac{1}{\langle \psi_{\xi} | \rho | \psi_{\xi} \rangle} \left(\sum_{k < l} (p_k + p_l) |a_{\xi k}|^2 |a_{\xi l}|^2 + \sum_m \left(\sum_k |a_{\xi k}|^2 c_{mk} \right)^2 \right) \\
&= \sum_{\xi} \frac{1}{\langle \psi_{\xi} | \rho | \psi_{\xi} \rangle} \left(\sum_{k \neq l} p_k |a_{\xi k}|^2 |a_{\xi l}|^2 + \sum_k \sum_l |a_{\xi k}|^2 |a_{\xi l}|^2 \sum_m c_{mk} c_{ml} \right). \tag{52}
\end{aligned}$$

We now use the following relation

$$\sum_m c_{mk} c_{ml} = \delta_{kl} p_k - p_k p_l \tag{53}$$

which is derived from (47) as follows: define $v_{mk} = c_{mk}/\sqrt{p_k}$ ($m = 1, \dots, d-1$) and $v_{dk} = \sqrt{p_k}$. Then (47) can be rewritten as $\sum_k v_{mk} v_{m'k} = \delta_{mm'}$. The vectors v_{mk} therefore are a complete orthonormal basis of R^d , hence they obey $\sum_m v_{mk} v_{mk'} = \delta_{kk'}$. Reexpressing in terms of c_{mk} yields (53). Inserting it in (52) we obtain

$$\begin{aligned}
\text{tr } H^{-1}I(M) &= \sum_{\xi} \frac{1}{\langle \psi_{\xi} | \rho | \psi_{\xi} \rangle} \left(\sum_k \sum_l p_k (1 - p_l) |a_{\xi k}|^2 |a_{\xi l}|^2 \right) \\
&= \sum_k (1 - p_k) \sum_{\xi} |a_{\xi k}|^2 = \sum_{\xi} \text{tr } (I - \rho) M_{\xi} \\
&= d - 1 \tag{54}
\end{aligned}$$

as announced.

Note that this has demonstrated that equality holds in (27) whenever $N = 1$, $p = d^2 - 1$, and the POVM is exhaustive. It follows from the classical properties of the Fisher information that equality also holds for arbitrary N whenever the POVM can be considered as a sequence of N separate exhaustive measurements on each copy of the system. It also holds if the n 'th measurement is chosen at random depending on the outcomes of the previous measurements.

D Separable measurements on N mixed states

We shall now prove that if we possess N identical mixed states of spin 1/2 particles, and carry out separable measurements, then

$$\text{tr } H^{-1}I(M) \leq N(d-1). \tag{55}$$

We recall that a separable measurement is one that can be carried out sequentially on separate particles, where the measurement on one particle at any stage (and indeed which particle to measure: one is allowed to measure particles several times) can depend arbitrarily on the outcomes so far, see [17] for a discussion. It is therefore more general than the case considered at the end of the previous subsection where the measurement on the n th particle could only depend on the measurements carried out on the $n-1$ previous particles.

If a POVM is separable, then its elements M_{ξ} can be decomposed into a sum of terms proportional to projectors onto unentangled states

$$\begin{aligned}
M_{\xi} &= \sum_i |\psi_{\xi i}\rangle \langle \psi_{\xi i}|, \\
|\psi_{\xi i}\rangle &= |\psi_{\xi i}^1\rangle \otimes \dots \otimes |\psi_{\xi i}^N\rangle. \tag{56}
\end{aligned}$$

We call measurements having such a representation *nonentangled*. (Note that there exist nonentangled POVMs which are not separable[18]).

By refining a separable measurement (which increases the Fisher information) one can restrict oneself to measurements whose POVM elements are proportional to projectors onto product states

$$M_\xi = |\psi_\xi\rangle\langle\psi_\xi| = |\psi_\xi^1\rangle\langle\psi_\xi^1| \otimes \dots \otimes |\psi_\xi^N\rangle\langle\psi_\xi^N|. \quad (57)$$

We now evaluate the l.h.s. of (55) for measurements of the form (57). First recall that the N unknown states have the form

$$\rho^N = \rho \otimes \dots \otimes \rho = \sum_{k_1=1}^d \dots \sum_{k_N=1}^d p_{k_1 \dots k_N} |k_1 \dots k_N\rangle \langle k_1 \dots k_N| \quad (58)$$

and the derivatives of ρ^N have the form

$$\rho_{,i}^N = \rho_{,i} \otimes \rho \dots \otimes \rho + \dots + \rho \otimes \dots \otimes \rho_{,i} = \sum_{p=1}^N \rho \otimes \dots \rho_{,i} \dots \otimes \rho \quad (59)$$

where in the second rewriting it is understood that $\rho_{,i}$ is at the p 'th position in the product.

Using the product form of measurement (57), one finds that

$$\begin{aligned} \langle\psi_\xi|\rho^N|\psi_\xi\rangle &= \langle\psi_\xi^1|\rho|\psi_\xi^1\rangle \dots \langle\psi_\xi^N|\rho|\psi_\xi^N\rangle \\ \langle\psi_\xi|\rho_{,i}^N|\psi_\xi\rangle &= \sum_{p=1}^N \langle\psi_\xi^1|\rho|\psi_\xi^1\rangle \dots \langle\psi_\xi^p|\rho_{,i}|\psi_\xi^p\rangle \dots \langle\psi_\xi^N|\rho|\psi_\xi^N\rangle \end{aligned} \quad (60)$$

Inserting these expressions into the Fisher information matrix one finds

$$\begin{aligned} I_{ij}(M) &= \sum_{\xi} \frac{\langle\psi_\xi|\rho_{,i}^N|\psi_\xi\rangle \langle\psi_\xi|\rho_{,j}^N|\psi_\xi\rangle}{\langle\psi_\xi|\rho^N|\psi_\xi\rangle} \\ &= \sum_{\xi} \sum_{p \neq p'} \langle\psi_\xi^1|\rho|\psi_\xi^1\rangle \dots \langle\psi_\xi^p|\rho_{,i}|\psi_\xi^p\rangle \dots \langle\psi_\xi^{p'}|\rho_{,j}|\psi_\xi^{p'}\rangle \dots \langle\psi_\xi^N|\rho|\psi_\xi^N\rangle \\ &\quad + \sum_{\xi} \sum_p \langle\psi_\xi^1|\rho|\psi_\xi^1\rangle \dots \frac{\langle\psi_\xi^p|\rho_{,i}|\psi_\xi^p\rangle \langle\psi_\xi^p|\rho_{,j}|\psi_\xi^p\rangle}{\langle\psi_\xi^p|\rho|\psi_\xi^p\rangle} \dots \langle\psi_\xi^N|\rho|\psi_\xi^N\rangle \\ &= \sum_{\xi} \sum_p \langle\psi_\xi^1|\rho|\psi_\xi^1\rangle \dots \frac{\langle\psi_\xi^p|\rho_{,i}|\psi_\xi^p\rangle \langle\psi_\xi^p|\rho_{,j}|\psi_\xi^p\rangle}{\langle\psi_\xi^p|\rho|\psi_\xi^p\rangle} \dots \langle\psi_\xi^N|\rho|\psi_\xi^N\rangle, \end{aligned} \quad (61)$$

where we have used the fact that the first term in the second equality vanishes. Indeed it is equal to

$$\sum_{\xi} \sum_{p \neq p'} \langle\psi_\xi|\rho \otimes \dots \otimes \rho_{,i} \otimes \dots \otimes \rho_{,j} \otimes \dots \otimes \rho|\psi_\xi\rangle. \quad (62)$$

The sum over ξ can be carried out in (62) to yield the identity matrix and the resulting trace vanishes because $\text{tr} \rho \otimes \dots \otimes \rho_{,i} \otimes \dots \otimes \rho_{,j} \otimes \dots \otimes \rho = 0$.

We now insert (61) into $\text{tr} H^{-1} I(M)$. All the operations from (50) to (54) can be carried out exactly as in the previous subsection, and one arrives at the expression

$$\begin{aligned} \text{tr} H^{-1} I(M) &= \sum_p \sum_{\xi} \langle\psi_\xi|\rho \otimes \dots \otimes (I - \rho) \otimes \dots \otimes \rho|\psi_\xi\rangle \\ &= N(d-1) \end{aligned} \quad (63)$$

which is the sought for relation.

We now provide a counterexample showing that if one carries out a collective measurement on $N > 1$ mixed states one can violate (27). We take $N = 2$, and suppose the unknown states belong to a 2 dimensional Hilbert space. $\rho(\theta) = \frac{I}{2} + \sum_i \theta_i \sigma_i$. We take as reference point $\theta_i = 0$ corresponding to $\rho = \frac{I}{2}$. At this point $H_{ij}(\theta_i = 0) = \delta_{ij}$.

We consider as measurement on the two copies the following POVM

$$M = \left\{ \begin{array}{l} \frac{1}{2} |\uparrow_x \uparrow_x\rangle \langle \uparrow_x \uparrow_x|, \quad \frac{1}{2} |\downarrow_x \downarrow_x\rangle \langle \downarrow_x \downarrow_x|, \quad \frac{1}{2} |\uparrow_y \uparrow_y\rangle \langle \uparrow_y \uparrow_y|, \quad \frac{1}{2} |\downarrow_y \downarrow_y\rangle \langle \downarrow_y \downarrow_y|, \\ \frac{1}{2} |\uparrow_z \uparrow_z\rangle \langle \uparrow_z \uparrow_z|, \quad \frac{1}{2} |\downarrow_z \downarrow_z\rangle \langle \downarrow_z \downarrow_z|, \quad \frac{1}{2} |\uparrow_z \downarrow_z - \downarrow_z \uparrow_z\rangle \langle \uparrow_z \downarrow_z - \downarrow_z \uparrow_z| \end{array} \right\}. \quad (64)$$

This POVM cannot be realized by separate measurements on each particle since the last term projects onto an entangled state.

For this POVM one calculates that $I_{ij}(M, \theta_i = 0) = \delta_{ij}$. Hence the left hand side of (27) evaluates to $\sum_{ij} H_{ij}^{-1}(\theta_i = 0) I_{ij}(M, \theta_i = 0) = 3 > N(d-1) = 2$. This shows that the optimal Fisher information is non additive.

F Comparison with other Quantum Cramér-Rao bounds

An important question raised by the bound (27) is how it compares to other quantum Cramér-Rao bounds obtained in the literature. In this respect, our most important result is that (27) is both a necessary and sufficient condition that $I(M, \theta)$ must satisfy when the dimensionality of the system d equals 2 and the state is pure. This will be proven and discussed in detail in section VII.

When $d > 2$ (27) is not a sufficient condition that $I(M, \theta)$ must satisfy. To see this let us compare (27) with the bound derived by Helstrom based on the s.l.d. This bound is the matrix inequality $I^N(M, \theta) \leq NH(\theta)$, see (26). The comparison is most easily carried out by defining the matrix $F = \frac{1}{N} H^{-\frac{1}{2}} I^N H^{-\frac{1}{2}} = \sum_{i=1}^p \gamma_i f_i \otimes f_i$ where γ_i are the eigenvalues of F and f_i its eigenvectors. Helstrom's bound can be reexpressed as $\gamma_i \leq 1$ for all i , whereas the bound (27) states that $\sum_i \gamma_i \leq d-1$. From these expressions it results that the bound (27) is better than Helstrom's bound for $d = 2$. For $d > 2$ and $p \leq d-1$ Helstrom's bound is better than (27) as is seen by summing the inequalities $\gamma_i \leq 1$ to obtain $\sum_i \gamma_i \leq p$. For $p > d-1$, neither Helstrom's bound nor the bound (27) are better than the other.

Yuen and Lax[16] have proposed another matrix bound based on an asymmetric logarithmic derivative (a.l.d.). This bound is known to be worse than the bound based on the s.l.d. in the case of one parameter, but it can be better, for some loss functions, in the case of two or more parameters. We have however not been able to make a detailed comparison between the bound based on the a.l.d. and (27).

Although when $d > 2$, the bound (27) is not a sufficient condition it can be complemented by additional constraints based on partial traces of $H^{-1} I^N(M, \theta)$ which we now exhibit.

Consider a subset $i = 1, \dots, p'$ ($p' < p$) of the parameters. Let $\rho_{,i'}$ be the corresponding derivatives of $\rho(\theta)$. Let us define the effective dimension d' of the space in which these parameters act at the point θ^0 as follows. Let Π be a projector that commutes with $\rho(\theta^0)$ ($[\Pi, \rho(\theta^0)] = 0$) and such that $\rho_{,i'}$, $i' = 1, \dots, p'$ acts only within the eigenspace of Π (that is $\Pi \rho_{,i'} \Pi = \rho_{,i'}$). Then d' is the smallest dimension of the eigenspace of such a projector Π ($d' = \text{tr } \Pi$). To be more explicit, let us reexpress the definition of d' in coordinates. First we diagonalize $\rho(\theta^0) = \sum_k p_k |k\rangle \langle k|$. If some p_k are equal this can be done in many ways. The projector Π projects onto some of the eigenvectors of ρ : $\Pi = \sum_{k=1}^{d'} |k\rangle \langle k|$. Next we write the operators $\rho_{,i'}$ in this basis: $\rho_{,i'} = \sum_{k,l=1}^{d'} (\rho_{,i'})_{kl} |k\rangle \langle l|$ where the fact that the indices k, l go from one to d' expresses the fact that $\rho_{,i'}$ acts only within the eigenspace of Π . Finally we choose the smallest such d' .

We will show that

$$\sum_{i', j'=1}^{p'} H_{i' j'}^{-1} I_{i' j'}^N(M, \theta^0) \leq N(d' - 1). \quad (65)$$

Before proving this result let us illustrate it by an example. Consider an unknown pure state in d dimensions. In the neighbourhood of a particular point we can parameterize the state by

$$\psi = |1\rangle + (\theta_2 + i\eta_2) |2\rangle + \dots + (\theta_d + i\eta_d) |d\rangle \quad (66)$$

where the unknown parameters are θ_i and η_i , $i = 2, \dots, d$. There are thus $2d - 2$ parameters. At the point $\theta = \eta = 0$, H is diagonal in this parameterization: $H_{\theta_i\theta_j} = \delta_{ij}$, $H_{\eta_i\eta_j} = \delta_{ij}$, $H_{\theta_i\eta_j} = 0$. Hence (27) takes the form

$$\sum_i I_{\theta_i\theta_i}^N(M, \theta = \eta = 0) + I_{\eta_i\eta_i}^N(M, \theta = \eta = 0) \leq N(d - 1). \quad (67)$$

But using (65) we also find the constraints

$$I_{\theta_i\theta_i}^N(M, \theta = \eta = 0) + I_{\eta_i\eta_i}^N(M, \theta = \eta = 0) \leq N \quad , \quad i = 2, \dots, d \quad (68)$$

which are stronger than (67) since they must hold separately, but by summing them one obtains (67).

The proof of equation (65) proceeds as in section V. First we can restrict ourselves to POVM's whose elements are proportional to one dimensional projectors. Secondly, we can restrict ourselves to the subspace Π in evaluating (65). This follows from the inequality

$$\begin{aligned} I(M)_{i'j'} &= \sum_{\xi} \frac{\text{tr}(\rho_{i'} M_{\xi}) \text{tr}(\rho_{j'} M_{\xi})}{\text{tr}(\rho M_{\xi})} \\ &= \sum_{\xi} \frac{\text{tr}(\rho_{i'} \Pi M_{\xi} \Pi) \text{tr}(\rho_{j'} \Pi M_{\xi} \Pi)}{\text{tr}(\rho \Pi M_{\xi} \Pi) + \text{tr}(\rho(1 - \Pi) M_{\xi} (1 - \Pi))} \\ &\leq \sum_{\xi} \frac{\text{tr}(\rho_{i'} \Pi M_{\xi} \Pi) \text{tr}(\rho_{j'} \Pi M_{\xi} \Pi)}{\text{tr}(\rho \Pi M_{\xi} \Pi)}. \end{aligned} \quad (69)$$

Note that equality in (69) holds when the measurement consists of one dimensional projectors and when the POVM decomposes into the sum of two POVM's acting on the subspaces spanned by Π and $1 - \Pi$ separately (i.e., the POVM elements $M_{\xi} = |\psi_{\xi}\rangle \langle \psi_{\xi}|$ must commute with Π and $1 - \Pi$). Thirdly, we can increase the number of parameters from p' to $d'^2 - 1$. We then introduce exactly as in (46) a parameterization in which the $\rho_{i'}$ are particularly simple, but in place of (53) we use

$$\sum_{1 \leq m' \leq d'} c_{m'k'} c_{m'l'} = \delta_{k'l'} p_{k'} - \frac{p_{k'} p_{l'}}{\text{tr}(\Pi \rho)}. \quad (70)$$

After these preliminary steps the l.h.s. of (65) is calculated exactly as in subsections VB, VC and VD.

VI DROPPING THE CONDITION OF UNBIASED ESTIMATORS

A Quantum van Trees inequality

In the previous section we proved a bound on the m.q.e. of unbiased estimators $\hat{\theta}^N$ of N copies of the quantum system $\rho(\theta)$ (with the additional condition that if ρ is mixed the measurement should be separable). In this section we shall prove Theorems IV and V that under additional conditions it is possible to drop the hypothesis that the estimator is unbiased.

The starting point for the results in this section is a Bayesian form of the Cramér-Rao inequality, the van Trees inequality [20], and in particular the multivariate form of the van Trees inequality proven in [21]. Adapted to the problem of estimating a quantum state, this inequality takes the following form. Let $\hat{\theta}^N$ be an arbitrary estimator of the parameter θ based on a measurement M of the system $\rho^N(\theta)$. Suppose it has m.q.e. matrix $V^N(\theta)$, and Fisher information matrix $I^N(M, \theta)$. Let $\lambda(\theta)$ be a smooth density supported on a compact region (with smooth boundary) of the parameter space, and suppose λ vanishes on the boundary. By E_{λ} we denote expectation over a random parameter value Θ with the probability density $\lambda(\theta)$. Let $C(\theta)$ and $D(\theta)$ be two $p \times p$ matrix valued functions of θ , the former being symmetric and positive definite. Then the multivariate van Trees inequality reads

$$E_{\lambda} \text{tr} C(\Theta) V^N(\Theta) \geq \frac{(E_{\lambda} \text{tr} D(\Theta))^2}{E_{\lambda} \text{tr} C(\Theta)^{-1} D(\Theta) I^N(M, \Theta) D(\Theta)^{\top} + \tilde{I}(\lambda)} \quad (71)$$

where $^\top$ denotes the transpose of the matrix and

$$\tilde{\mathcal{I}}(\lambda) = \int d\theta \frac{1}{\lambda(\theta)} \sum_{ijkl} C_{ij}(\theta)^{-1} \partial_{\theta_k} \{D_{ik}(\theta) \lambda(\theta)\} \partial_{\theta_l} \{D_{jl}(\theta) \lambda(\theta)\}. \quad (72)$$

As a first application of this inequality we shall prove Theorem V, that is bound the minimum value averaged over θ of a quadratic cost function. Let $C(\theta)$ be the quadratic cost function. Consider the matrix $W_{opt}(\theta)$ that minimizes for each value of θ the cost $\text{tr} C(\theta) W(\theta)$ under the condition that $\text{tr} H(\theta)^{-1} W(\theta)^{-1} \leq d-1$. One easily finds that

$$W_{opt} = \frac{\text{tr} \sqrt{H^{-1/2} C H^{-1/2}}}{d-1} H^{-1/2} \sqrt{H^{1/2} C^{-1} H^{1/2}} H^{-1/2} \quad (73)$$

$$= \frac{\text{tr} \sqrt{C^{1/2} H^{-1} C^{1/2}}}{d-1} C^{-1/2} \sqrt{C^{1/2} H^{-1} C^{1/2}} C^{-1/2} \quad (74)$$

and that

$$\text{tr} C W_{opt} = \frac{\left(\text{tr} \sqrt{H^{-1/2} C H^{-1/2}}\right)^2}{d-1} = \frac{\left(\text{tr} \sqrt{C^{1/2} H^{-1} C^{1/2}}\right)^2}{d-1}. \quad (75)$$

We choose in (71) $D(\theta) = C(\theta) W_{opt}(\theta)$. Thus $\text{tr} D(\theta) = \text{tr} C(\theta) W_{opt}(\theta)$ is given by (75). Note that

$$D(\theta)^\top C(\theta)^{-1} D(\theta) = W_{opt}(\theta) C(\theta) W_{opt}(\theta) = \frac{\text{tr} C(\theta) W_{opt}(\theta)}{d-1} H(\theta)^{-1}. \quad (76)$$

Thus

$$\begin{aligned} \text{tr} D(\theta)^\top C(\theta)^{-1} D(\theta) I^N(M, \theta) &= \frac{\text{tr} C(\theta) W_{opt}(\theta)}{d-1} \text{tr} H(\theta)^{-1} I^N(M, \theta) \\ &\leq N \text{tr} C(\theta) W_{opt}(\theta). \end{aligned} \quad (77)$$

Inserting these expressions into (71) one obtains

$$\begin{aligned} \mathbb{E}_\lambda \text{tr} C(\Theta) V^N(\Theta) &\geq \frac{(\mathbb{E}_\lambda \text{tr} C(\Theta) W_{opt}(\Theta))^2}{N \mathbb{E}_\lambda \text{tr} C(\Theta) W_{opt}(\Theta) + \tilde{\mathcal{I}}(\lambda)} \\ &\geq \frac{\mathbb{E}_\lambda \text{tr} C(\Theta) W_{opt}(\Theta)^2}{N} - \frac{\alpha}{N^2} \end{aligned} \quad (78)$$

where

$$\alpha = \frac{\tilde{\mathcal{I}}(\lambda)}{\mathbb{E}_\lambda \text{tr} C(\Theta) W_{opt}(\Theta)} \quad (79)$$

is independent of N . This proves that upon averaging over θ it is impossible (for large N) to improve over the minimum cost (30).

B Asymptotic version of the Cramér-Rao inequality

We now prove Theorem IV, that is an asymptotic version of our main inequality (28) which is valid at every point θ and does not make the assumption of unbiased estimators. We must however slightly restrict the class of competing estimators since otherwise by the phenomenon of super-efficiency we can beat a given estimator at any specific value of the parameter, though we pay for this by bad behaviour closer and closer to the chosen value as N becomes larger.

The restriction on the class of estimators is that N times their mean quadratic error matrix must converge uniformly in a neighbourhood of the true value θ^0 of θ to a limit $W(\theta)$, continuous at θ^0 . We assume that both $W(\theta^0)$ and $H(\theta^0)$ are nonsingular. Furthermore we shall require some mild smoothness conditions on $H(\theta)$ in a neighbourhood of θ^0 : it must be continuous at θ^0 with bounded partial derivatives with respect to the parameter in a neighbourhood of θ^0 . Note that imposing regularity conditions on H is natural since it corresponds to supposing that the θ_i smoothly parametrize the allowed density matrices.

Suppose that as $N \rightarrow \infty$

$$NV^N(\theta) \rightarrow W(\theta)$$

uniformly in θ in a neighbourhood of θ^0 , with W continuous at θ^0 ; write $W^0 = W(\theta^0)$. Now in (71) let us make the following choices for the matrix functions C and D :

$$C(\theta) = W^{0^{-1}} H^{-1}(\theta) W^{0^{-1}},$$

$$D(\theta) = W^{0^{-1}} H^{-1}(\theta).$$

Then (71) (multiplied throughout by N) and (72) become

$$\begin{aligned} \mathbb{E}_\lambda \text{tr } W^{0^{-1}} H^{-1}(\Theta) W^{0^{-1}} NV^N(\Theta) &\geq \frac{(\mathbb{E}_\lambda \text{tr } W^{0^{-1}} H^{-1}(\Theta))^2}{\frac{1}{N} \mathbb{E}_\lambda \text{tr } H^{-1} I^N(M, \Theta) + \frac{1}{N} \tilde{\mathcal{I}}(\lambda)} \\ &\geq \frac{(\mathbb{E}_\lambda \text{tr } W^{0^{-1}} H^{-1}(\Theta))^2}{(d-1) + \frac{1}{N} \tilde{\mathcal{I}}(\lambda)} \end{aligned} \quad (80)$$

and

$$\tilde{\mathcal{I}}(\lambda) = \int d\theta \frac{1}{\lambda(\theta)} \sum_{ijkl} H_{ij}(\theta) \partial_{\theta_k} \{H_{ik}^{-1}(\theta) \lambda(\theta)\} \partial_{\theta_l} \{H_{jl}^{-1}(\theta) \lambda(\theta)\}, \quad (81)$$

where we have used our central inequality (27) to pass to (80). Now suppose that the quantity (81) is finite (we will give conditions for that in a moment). By the assumed uniform convergence of NV^N to W , upon letting $N \rightarrow \infty$ (80) becomes

$$\mathbb{E}_\lambda \text{tr } W^{0^{-1}} H^{-1}(\Theta) W^{0^{-1}} W(\Theta) \geq \frac{(\mathbb{E}_\lambda \text{tr } W^{0^{-1}} H^{-1}(\Theta))^2}{(d-1)}. \quad (82)$$

Now suppose the density λ in this equation (the probability density of Θ) is replaced by an element λ^m in a sequence of densities, concentrating on smaller and smaller neighbourhoods of θ^0 as $m \rightarrow \infty$. Assume that $H(\theta)$ is continuous at θ^0 . Recall our earlier assumption that $W(\theta)$ is also continuous at θ^0 , with $W^0 = W(\theta^0)$. Then taking the limit as $m \rightarrow \infty$ of (82) yields

$$\text{tr } W^{-1}(\theta^0) H^{-1}(\theta^0) \geq (\text{tr } W^{-1}(\theta^0) H^{-1}(\theta^0))^2 / (d-1)$$

or the required limiting form of (27),

$$\text{tr } W^{-1}(\theta^0) H^{-1}(\theta^0) \leq (d-1).$$

It remains to discuss whether it was reasonable to assume that $\tilde{\mathcal{I}}(\lambda^m)$ is finite (for each m separately). Note that this quantity only depends on the prior density λ and on $H(\theta)$, where λ is one of a sequence of densities supported by smaller and smaller neighbourhoods of θ^0 . We already assumed that $H(\theta)$ was continuous at θ^0 . It is certainly possible to specify prior densities λ^m concentrating on the ball of radius $1/m$, say, satisfying the smoothness assumptions in [21] and with, for each m , finite Fisher information matrix

$$\int d\theta \frac{1}{\lambda^m(\theta)} \partial_{\theta_k} \{\lambda^m(\theta)\} \partial_{\theta_l} \{\lambda^m(\theta)\}.$$

Consideration of (81) then shows that it suffices further just to assume that $\partial_{\theta_k} \{H_{ik}^{-1}(\theta)\}$ is, for each i, k , bounded in a neighbourhood of θ^0 .

In conclusion we have shown that under mild smoothness conditions on $H(\theta)$, the limiting mean quadratic error matrix W of a sufficiently regular but otherwise arbitrary sequence of estimators must satisfy the asymptotic version of our central inequality $\text{tr } H^{-1} W^{-1} \leq d-1$.

We shall now show that the bounds (27), (29), (31) are sharp in the case of pure states of spin 1/2 systems and of separable measurements in the case of mixed states of spin 1/2 systems. In particular, in the limit of a large number of copies N any target scaled m.q.e. matrix W that satisfies $\text{tr } H^{-1}W^{-1} \leq 1$ can be attained (provided W is non singular). We shall show this by explicitly constructing a measurement strategy that attains the bound. In section VI we have already shown that if $\text{tr } H^{-1}W^{-1} > 1$, then it cannot be attained.

A Attaining the bound at a fixed point θ^0

The first step in the proof is to consider the case of one copy of the unknown state ($N = 1$) and fix a particular point θ^0 . Then we show that for any target information matrix $G(\theta^0)$ that satisfies $\text{tr } H^{-1}(\theta^0)G(\theta^0) \leq 1$, we can build a measurement $M = M^{\theta^0}$, *in general depending on θ^0* , such that $I(M^{\theta^0}, \theta^0) = G(\theta^0)$. In the next sections we shall show how to use this intermediate result to build a measurement and estimation strategy whose asymptotic m.q.e. is equal to $W(\theta) = G(\theta)^{-1}$ for *all* θ .

Let us first consider the case of pure states. At θ^0 , the state is $|\psi^0\rangle$. We introduce a parameterization θ_1, θ_2 such that in the vicinity of $|\psi^0\rangle$, the unknown state is

$$|\psi(\theta)\rangle = |\psi^0\rangle + (\theta_1 + i\theta_2)|\psi^1\rangle. \quad (83)$$

Thus the original point θ^0 corresponds to the new $\theta_1 = \theta_2 = 0$. In this parameterization, H is proportional to the identity at $\theta_1 = \theta_2 = 0$: $H_{\theta_1\theta_1}(0) = H_{\theta_2\theta_2}(0) = 1$, $H_{\theta_1\theta_2}(0) = 0$.

We now diagonalize the matrix G . Thus there exist new parameters $\theta'_1 = \cos \lambda \theta_1 + \sin \lambda \theta_2$, $\theta'_2 = -\sin \lambda \theta_1 + \cos \lambda \theta_2$ such that $G_{\theta'_1\theta'_1}(0) = g_1 \geq 0$, $G_{\theta'_2\theta'_2}(0) = g_2 \geq 0$, $G_{\theta'_1\theta'_2}(0) = 0$.

In terms of the parameters θ'_1, θ'_2 , the unknown state is written

$$|\psi^0\rangle = |\psi^0\rangle + (\theta'_1 + i\theta'_2)|\psi^{1'}\rangle \quad (84)$$

where $|\psi^{1'}\rangle = e^{i\lambda}|\psi^1\rangle$.

The POVM M^{θ^0} consists of measuring the observable $|\psi^0\rangle\langle\psi^{1'}| + |\psi^{1'}\rangle\langle\psi^0|$ with probability g_1 , of measuring the observable $i(|\psi^0\rangle\langle\psi^{1'}| - |\psi^{1'}\rangle\langle\psi^0|)$ with probability g_2 , and of measuring nothing (or measuring the identity) with probability $1 - g_1 - g_2$. It is straightforward to verify that the Fisher information at θ^0 in a measurement of the POVM M^{θ^0} is equal to $G(\theta^0)$.

Let us now turn to the case of mixed states. We suppose that there are three unknown parameters. We use a parameterization in which $\rho(\theta) = (1/2)(I + \theta \cdot \sigma)$, with $\|\theta\| < 1$. Without loss of generality we can suppose that $\theta^0 = (0, 0, n)$, so that $\rho(\theta^0) = (1/2 + n/2)|1\rangle\langle 1| + (1/2 - n/2)|2\rangle\langle 2| = \frac{1}{2}(I + n\sigma_z)$. The tangent space at ρ is spanned by the Pauli matrices $\rho_x = \sigma_x (= \frac{\rho_{12} + \rho_{21}}{2})$, $\rho_y = \sigma_y (= \frac{\rho_{12} - \rho_{21}}{2})$, $\rho_z = \sigma_z (= \rho_{11} - \rho_{22})$ where in parenthesis we have given the relation to the basis used in section VC. In this coordinate system $H(\theta^0)$ is diagonal with eigenvalues 1, 1, $1/(1 - n^2)$.

Take any symmetric positive matrix G satisfying $\text{tr } GH^{-1}(\theta^0) \leq 1$. Define the matrix $F = H^{-\frac{1}{2}}GH^{-\frac{1}{2}} = \sum_i \gamma_i f_i \otimes f_i$, where γ_i and f_i are the eigenvalues and eigenvectors of F . The condition $\text{tr } GH^{-1}(\theta^0) \leq 1$ can then be rewritten $\sum_i \gamma_i \leq 1$. If we define $g_i = H^{\frac{1}{2}}f_i$, then we can write $G = \sum_i \gamma_i g_i \otimes g_i$. Denote $m_i = g_i/\|g_i\|$.

Consider the measurement of the spin along the direction m_i . This is the POVM consisting of the two projectors $P_{+m_i} = \frac{1}{2}(I + m_i \cdot \sigma)$ and $P_{-m_i} = \frac{1}{2}(I - m_i \cdot \sigma)$. The information matrix for this measurement is

$$I(P_{\pm m_i})_{kl} = \sum_{\pm} \frac{\text{tr}(P_{\pm m_i} \sigma_k) \text{tr}(P_{\pm m_i} \sigma_l)}{\text{tr}(P_{\pm m_i} \rho)} = \frac{(m_i)_k (m_i)_l}{(1 - n^2 (m_i)_z^2)} \quad (85)$$

where $(m_i)_k$ is component k of vector m_i . Therefore this information matrix is proportional to $g_i \otimes g_i$. One verifies that it obeys $\text{tr } H^{-1}I(P_{\pm m_i}) = 1$, as it must by our findings in section V since the measurement is exhaustive, $N = 1$, and $p = d^2 - 1$. Therefore the coefficient of proportionality is 1 and

$$I(P_{\pm m_i}) = g_i \otimes g_i. \quad (86)$$

We now combine such POVM's to obtain the POVM whose elements are

$$\gamma_1 P_{+m_1}, \gamma_1 P_{-m_1}, \gamma_2 P_{+m_2}, \gamma_2 P_{-m_2}, \gamma_3 P_{+m_3}, \gamma_3 P_{-m_3}, (1 - \gamma_1 - \gamma_2 - \gamma_3). \quad (87)$$

The information matrix for this measurement is just the sum $\gamma_1 I(P_{\pm m_1}) + \gamma_2 I(P_{\pm m_2}) + \gamma_3 I(P_{\pm m_3}) = \sum_i \gamma_i g_i \otimes g_i = G$. Thus the POVM (87) attains the target information G at the point θ^0 .

B Attaining the bound for every θ and arbitrary N by separable measurements

We now prove Theorem VII that states that in the case of spin half particles we can attain the bound (29) for every θ . Give yourself a continuous matrix $W(\theta)$, the target scaled m.q.e. matrix, satisfying (29) for every θ . Define $G(\theta) = W(\theta)^{-1}$, the target scaled information matrix, which satisfies therefore (27). We will show that there exists a separable measurement and an estimation strategy on N copies of the state $\rho(\theta)$ such that the m.q.e. matrix V^N of the estimator satisfies

$$V^N(\theta)_{ij} = E_\theta((\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)) = \frac{W_{ij}(\theta)}{N} + o\left(\frac{1}{N}\right) \quad (88)$$

for all θ . In fact this holds uniformly in θ in a sufficiently small neighbourhood of any given point. This is proven by constructing explicitly a measurement and estimation strategy that satisfies (88), following the lines of [15].

The measurement and estimation strategy we propose is the following: first take a fraction $N_0 = O(N^a)$ of the states, for some fixed $0 < a < 1$, and on one third of them measure σ_x , on one third σ_y and on one third σ_z . One obtains from each measurement of σ_x the outcome ± 1 with probabilities $\frac{1}{2}(1 \pm \theta_x)$, and similarly for σ_y, σ_z . Using this data we make a first estimate of θ , call it $\tilde{\theta}$, for instance by equating the observed relative frequencies of ± 1 in the three kinds of measurement to their theoretical values. If the state is pure this determines a first estimate of the direction of polarization. If the state is mixed it is possible that the initial estimate suggests that the Bloch vector lies outside the unit sphere. This only occurs with exponentially small probability (in N_0) and if this is the case the measurement is discarded. As discussed below this only affects the mean quadratic error by $o(1/N)$.

On the remaining $N' = N - N_0$ states we carry out the measurement $M = M^{\tilde{\theta}}$ such that $I(M^{\tilde{\theta}}, \tilde{\theta}) = G(\tilde{\theta})$ which we have just shown how to construct. Note that $I(M^{\tilde{\theta}}, \theta) = G(\theta)$ is only guaranteed when θ is precisely equal to $\tilde{\theta}$. Write $I(M, \theta; \tilde{\theta})$ for the Fisher information about θ , based on the measurement $M^{\tilde{\theta}}$ optimal at $\tilde{\theta}$, while the true value of the parameter is actually θ . Given $\tilde{\theta}$, each of the N' second stage measurements represents one draw from the probability distribution $p(\xi|\theta; \tilde{\theta}) = \text{tr } M_\xi^{\tilde{\theta}} \rho(\theta)$. We use the classical m.l.e. based on this data only (with $\tilde{\theta}$ fixed at its observed value) to estimate what is the value of θ . Call this estimated value $\hat{\theta}$.

Let $\epsilon > 0$ be fixed, arbitrarily small. Let θ^0 denote the true value of θ . For given $\delta > 0$ let $B(\theta^0, \delta)$ denote the ball of radius δ about θ^0 . Fix a convenient matrix norm $\|\cdot\|$. We have the exponential bound

$$\Pr\{\tilde{\theta} \in B(\theta^0, \delta)\} \geq 1 - Ce^{-DN_0\delta^2} \quad (89)$$

for some positive numbers C and D (depending on δ). The reason we take N_0 proportional to N^a for some $0 < a < 1$ is that this ensures that $1 - Ce^{-DN_0} = o(1/N)$.

Modern results [23] on the m.l.e. $\hat{\theta}$ state that, under certain regularity conditions, the conditional m.q.e. matrix of $\hat{\theta}$ satisfies (at $\theta = \theta^0$, and conditional on $\tilde{\theta}$)

$$V^{N'}(\theta^0; \tilde{\theta}) = \frac{I(M, \theta^0; \tilde{\theta})^{-1}}{N'} + o\left(\frac{1}{N'}\right) \quad (90)$$

uniformly in θ^0 . We need however for the next step in our argument that this same result is true uniformly in $\tilde{\theta}$ for given θ^0 . This could be verified by careful reworking of the proof in [23]. Rather than doing that, we will explicitly calculate in subsection VII C and VIID the conditional m.q.e. matrix of our estimator and show that it satisfies (90) uniformly in $\tilde{\theta}$ in a small enough neighbourhood $B(\theta^0, \delta)$ of θ^0 . The 'little o ' in (90) refers to the chosen matrix norm.

We will also need that $I(M, \theta^0; \tilde{\theta})^{-1}$ is continuous in $\tilde{\theta}$ at $\tilde{\theta} = \theta^0$, at which point it equals by our construction the target scaled m.q.e. $W(\theta^0)$. This is also established in subsection VII C. Therefore, replacing if necessary δ by a smaller value, we can guarantee that $I(M, \theta^0; \tilde{\theta})^{-1}$ is within ϵ of $I(M, \theta^0; \theta^0)^{-1} = W(\theta^0)$ for all $\tilde{\theta} \in B(\theta^0, \delta)$.

If $\tilde{\theta}$ is outside the domain $B(\theta^0, \delta)$, then the norm of $V^{N'}(\theta^0; \tilde{\theta})$ is bounded by a constant A since θ belongs to a compact domain.

Putting everything together we find that

$$\begin{aligned} \|N'V^N(\theta^0) - W(\theta^0)\| &= \left\| \int \left(N'V^{N'}(\theta^0; \tilde{\theta}) - W(\theta^0) \right) dP(\tilde{\theta}) \right\| \\ &\leq \int_{B(\theta^0, \delta)} \|N'V^{N'}(\theta^0; \tilde{\theta}) - W(\theta^0)\| dP(\tilde{\theta}) + AN'C'e^{-DN_0} \\ &= \int_{B(\theta^0, \delta)} \|I(M, \theta^0; \tilde{\theta})^{-1} + o(1) - W(\theta^0)\| dP(\tilde{\theta}) + o(1) \\ &\leq \epsilon + o(1) + o(1). \end{aligned}$$

It follows since $N'/N \rightarrow 1$ as $N \rightarrow \infty$ that $\limsup_{N \rightarrow \infty} \|NV^N(\theta^0) - W(\theta^0)\| \leq \epsilon$. Since ϵ was arbitrary, we obtain (88).

C Analysis of the conditional mean quadratic error

We first consider the case of impure states, with the parameterization

$$\rho = \frac{1}{2}(I + \theta \cdot \sigma), \quad \text{with } \sum (\theta_i)^2 < 1. \quad (91)$$

where we have imposed that the state is never pure. This case turns out to allow the most explicit and straightforward analysis because the relation between the frequency of the outcomes and the parameters θ is linear. For other cases the analysis is more delicate and is discussed in the next subsection. In general, smoothness assumptions will have to be made on the parameterization $\rho = \rho(\theta)$.

We suppose that $W(\theta)$ is non-singular and continuous in θ . Consequently the γ_i (defined in section VII A) depend continuously on θ and are all strictly positive at the true value θ^0 of θ .

Given the initial estimate, the second stage measurement can be implemented as follows: for each of the $N' = N - N_0$ observations, independently of one another, with probability γ_i measure the projectors $P_{\pm m_i}$, in other words, measure the spin observable $m_i \cdot \sigma$. With probability $1 - \sum \gamma_i$ do nothing.

We emphasize that the γ_i and m_i all depend on the initial estimate $\tilde{\theta}$ through $W(\tilde{\theta})$ and $H(\tilde{\theta})$. In the following, all probability calculations are conditional on a given value of $\tilde{\theta}$.

For simplicity we will modify the procedure in the following two ways: firstly, rather than taking a random number of each of the three types of measurement, we will take the fixed (expected) numbers $\lfloor \gamma_i N' \rfloor$ (and neglect the difference between $\lfloor \gamma_i N' \rfloor$ and $\gamma_i N'$). Secondly, we will ignore the constraint $\sum (\theta_i)^2 \leq 1$. These two modifications make the maximum likelihood estimator easier to analyze, but do not change its asymptotic m.q.e. Later we will sketch how to extend the calculations to the original *constrained* maximum likelihood estimator based on *random* numbers of measurements of each observable.

Now measuring $m_i \cdot \sigma$ produces the values ± 1 with probabilities $p_{\pm i} = \frac{1}{2}(1 \pm \theta \cdot m_i)$. Since our data consists of three binomially distributed counts and we have three parameters $\theta_1, \theta_2, \theta_3$ the maximum likelihood estimator can be described, using the invariance of maximum likelihood estimators under 1-1 reparameterization, as follows: set the theoretical values $p_{\pm i}$ equal to their empirical counterparts (relative frequencies of ± 1 in the $\gamma_i N'$ observations of the i 'th spin) and solve the resulting three equations in three unknowns.

To be explicit, define $\eta_i = 2p_{+i} - 1 = \theta \cdot m_i$ and let $\hat{\eta}_i$ be its empirical counterpart. Recall that $m_i = g_i / \|g_i\|$, $g_i = H^{1/2} f_i$, where the f_i are the orthonormal eigenvectors of $H^{-1/2} G H^{-1/2}$, and where H and G are $H(\tilde{\theta})$, $G(\tilde{\theta})$, and $\tilde{\theta}$ is the preliminary estimate of θ . Then we can rewrite

$$\eta_i = \theta \cdot m_i = \theta \cdot g_i / \|g_i\| = \theta \cdot H^{1/2} f_i / \|H^{1/2} f_i\| = (H^{1/2} \theta) \cdot f_i / \|H^{1/2} f_i\|$$

from which we obtain

$$(H^{1/2}\theta).f_i = \|H^{1/2}f_i\|\eta_i$$

and hence

$$\theta = H^{-1/2} \sum_i \|H^{1/2}f_i\|\eta_i f_i.$$

The same relation holds between $\hat{\theta}$ and $\hat{\eta}$ and yields the sought for expression for $\hat{\theta}$ in terms of the empirical relative frequencies.

Observing that the $\hat{\eta}_i$ are independent with variance $4p_{+m_i}p_{-m_i}/(\gamma_i N') = (1 - (\theta.m_i)^2)/(\gamma_i N')$, the m.q.e. matrix of $\hat{\theta}$, conditional on the preliminary estimate $\tilde{\theta}$, is

$$V^{N'}(\theta^0; \tilde{\theta}) = \frac{1}{N'} \sum_i \frac{1}{\gamma_i} \left(1 - \frac{(\theta^0.H^{1/2}f_i)^2}{\|H^{1/2}f_i\|^2} \right) \|H^{1/2}f_i\|^2 H^{-1/2}(f_i \otimes f_i) H^{-1/2}. \quad (92)$$

There is no $o(1/N')$ term here so we do not have to check uniform convergence: the limiting value is attained exactly. Actually we cheated by replacing $\lfloor \gamma_i N' \rfloor$ by $\gamma_i N'$. This does introduce a $o(1/N')$ error into (92) uniformly in a neighbourhood of θ^0 in which the γ_i , which depend on $\tilde{\theta}$, are bounded away from zero, and H and its inverse are bounded.

One may verify that (92) reduces to $W(\theta^0)/N'$ at $\tilde{\theta} = \theta^0$ (indeed at $\theta^0 = \tilde{\theta}$, $(\tilde{\theta}.H^{1/2}f_i)^2 = \frac{n^2 f_{iz}^2}{1-n^2}$ and $\|H^{1/2}f_i\|^2 = \frac{1-n^2+n^2 f_{iz}^2}{1-n^2}$). But this computation is really superfluous since at this point, we are computing the m.q.e. of the maximum likelihood estimator based on a measurement with, by our construction, Fisher information equal to the inverse of $W(\theta^0)$. (The modifications to our procedure do not alter the Fisher information). The two quantities must be equal by the classical large sample results for the maximum likelihood estimator.

We finally need to show the continuity in $\tilde{\theta}$ at $\tilde{\theta} = \theta^0$ of N' times the quantity in (92). This is evident if the γ_i are all different at θ^0 . Both the eigenvalues and the eigenvectors of $H^{-\frac{1}{2}}GH^{-\frac{1}{2}}$ are then continuous functions of $\tilde{\theta}$ at θ^0 . There is a potential difficulty however if some γ_i are equal to one another at $\tilde{\theta} = \theta^0$. In this case, the eigenvectors f_i are not continuous functions of $\tilde{\theta}$ at this point, and not even uniquely defined there. We argue as follows that this does not destroy continuity of the mean quadratic error. Consider a sequence of points $\tilde{\theta}^n$ approaching θ^0 . This generates a sequence of eigenvectors f_i^n and eigenvalues γ_i^n . The eigenvalues converge to the γ_i but the eigenvectors need not converge at all. However by compactness of the set of unit vectors in R^3 , there is a subsequence along which the eigenvectors f_i^n converge; and they must converge to a possible choice of eigenvectors at θ^0 . Thus along this subsequence the mean quadratic error (92) does converge to a limit given by the same formula evaluated at the limiting f_i etc. But this limit is equal by construction to the inverse of the target information matrix $G(\theta)$. A standard argument now shows that the limiting mean quadratic error is continuous at $\tilde{\theta} = \theta^0$.

The m.q.e. of $\hat{\theta}$ given $\tilde{\theta}$ (times N') therefore converges uniformly in a sufficiently small neighbourhood of θ^0 to a limit continuous at that point and equal to $W(\theta^0)$ there.

In our derivation of (88) we required the parameter and its estimator to be bounded. By dropping the constraint on the length of θ we have inadvertently lost this property. Suppose we replace our modified estimator $\hat{\theta}$ by the actual maximum likelihood estimator respecting the constraint. The two only differ when the unconstrained estimator lies outside the unit sphere; but this event only occurs with an exponentially small probability, uniformly in $\tilde{\theta}$, provided the γ_i are uniformly bounded away from 0 in the given neighbourhood of θ^0 . From this it can be shown that the mean quadratic error is altered by an amount $o(1/N')$ uniformly in $\tilde{\theta}$.

If we had worked with random numbers of measurements of each spin variable, when computing the mean quadratic error we would first have copied the computation above conditional on the numbers of measurements, say X_i , of each spin m_i . These numbers are binomially distributed with parameter N' and γ_i . The conditional mean quadratic error would be the same as the expression above but with $1/(\gamma_i N')$ replaced by $1/X_i$ (and special provision taken for the possible outcome $X_i = 0$). So to complete the argument we must show that $E(1/X_i) = 1/(\gamma_i N') + o(1/N')$ uniformly in $\tilde{\theta}$. This can also be shown to be true, using the fact that X_i/N' only differs from its mean by more than a fixed amount with exponentially small probability as $N' \rightarrow \infty$ and we restrict attention to $\tilde{\theta}$ in a neighbourhood of θ^0 where the γ_i are bounded away from zero.

Inspection of our argument shows that the convergence of the mean quadratic error is uniform in θ^0 as long as we keep away from the boundary of the parameter space.

By the convergence of the normalized binomial distribution to the normal distribution, the representation of the estimator we gave above also shows that it is asymptotically normally distributed with asymptotic covariance matrix

equal to the target covariance matrix W . Moreover, if X has the binomial(n, p) distribution, then $n^{\frac{1}{2}}(X/n - p)$ converges in distribution to the normal with mean zero and variance $p(1 - p)$, uniformly in p . Thus the convergence in distribution is also uniform in θ^0 as long as we keep away from the boundary of the parameter space.

D Conditional mean quadratic error for other models

The preceding subsection gave a complete analysis of the mean quadratic error, given the preliminary estimate $\tilde{\theta}$ for the 3 unknown parameters θ_j of the parameterization (91). We shall first analyze the mean quadratic error when the unknown parameters are functions $\phi_i(\theta_j)$ of the parameters θ_j . We shall then consider the important case when the state is pure and depends on two unknown parameters, and finally the case when the state is pure or mixed and depends on one unknown parameter, or is mixed and depends on two unknown parameters.

Our first result is that if the change of parameters $\phi_i(\theta_j)$ is locally C^1 , then the m.q.e. matrix of the ϕ_i is obtained from the m.q.e. of the θ_j by the Jacobian $\partial\phi_i/\partial\theta_j$ except eventually at isolated points. This follows from the fact that under a smooth (locally C^1) parameterization, the delta method (first order Taylor expansion) allows us to conclude uniform convergence of the probability distribution of $\sqrt{N}(\hat{\phi}^N - \phi)$ to a normal limit with the target mean quadratic error. If the ϕ_i and their derivatives $\partial\phi_i/\partial\theta_j$ are bounded then this proves our claim. If there are points where the ϕ_i or their derivatives $\partial\phi_i/\partial\theta_j$ are infinite, then convergence in distribution does not necessarily imply convergence of moments. However a truncation device allows one to modify the estimate $\hat{\phi}$, replacing it by 0 if any component is larger than cN^a for given c and a (use the method of [23], Lemma II.8.2 together with the exponential inequality (89) for the multinomial distribution). With this minor modification one can show (uniform in ϕ in a neighbourhood of ϕ^0) convergence of the moments of the corresponding $\sqrt{N}(\hat{\phi} - \phi)$ to the moments of its limiting distribution, hence achievement of the bound in the sense of Theorem IV. In particular if the parameter ϕ is bounded then the truncation is superfluous.

Now turn to the pure state analog of model (91). Obtain a preliminary estimate of the location of ρ on the surface of the Poincaré sphere using the same method as in the mixed case, but always projecting onto the surface of the sphere. Next, after rotation to transform the preliminary estimate into ‘spin up’, reparameterize to $\rho = \frac{1}{2}(1 + \phi \cdot \sigma)$ where the parameters to be estimated are $(\phi_1, \phi_2) = (\theta'_1, \theta'_2)$ of the parameterization (84) while $\phi_3 = \sqrt{(1 - \phi_1^2 - \phi_2^2)}$. The preliminary estimate is at $\phi_1 = \phi_2 = 0$. The optimal measurement at this point according to Section VII A consists of measurements of the spins σ_1 and σ_2 on specified proportions of the remaining copies. The resulting estimator of the parameter (ϕ_1, ϕ_2) is a linear function of binomial counts and hence its mean quadratic error can be studied exactly as in section VII C. Then we must transfer back to the originally specified parameterization, for instance polar coordinates. This is done as in the preceding paragraph. If the transformation is locally C^1 then uniform convergence in distribution to the normal law also transfers back; convergence of mean quadratic error too if the original parameter space is bounded. Otherwise a truncation might be necessary. In any case, we can exhibit a procedure optimal in the sense of Theorem IV.

It remains to consider one- and two-dimensional sub-models of the full mixed model, and one-dimensional sub-models of the full pure model. We suppose that the model specifies a smooth curve or surface in the interior of the Poincaré sphere, or a smooth curve on its surface; smoothly parameterized by a one- or two-dimensional parameter as appropriate. The first stage of the procedure is just as before, finishing in projection of an estimated density matrix into the model. Then we reparameterize locally, augmenting the dimension of the parameter to convert the model into a full mixed or pure model respectively. The target information for the extra parameters is zero. Compute as before the optimal measurement at this point. Because of the zero values in the target information matrix, there will be zero eigenvalues γ_i in the computation of section VII A. Thus the optimal measurement will involve specified fractions of measurement of spin in the same number of directions as the dimension of the model. Compute the maximum likelihood estimator of the original parameters based on this data. If the parameterization is smooth enough the estimator will yet again achieve the bound of Theorem IV.

VIII CONCLUSIONS AND OPEN QUESTIONS

In this paper we have solved some of the theoretical problems that arise when trying to estimate the state of a quantum system of which one possesses a large number of copies. This constitutes a preliminary step towards solving the question with which Helstrom concluded his book [9]: “(. . .) mathematical statisticians are often concerned with asymptotic properties of decision strategies and estimators. (. . .) When the parameters of a quantum density operator are estimated on the basis of many observations, how does the accuracy of the estimates depend on the number of observations as that number grows very large? Under what conditions have the estimates asymptotic normal

distributions? Problems such as these, and still others that doubtless will occur to physicists and mathematicians, remain to be solved within the framework of the quantum-mechanical theory.”

In the case of pure states of spin $1/2$ particles the problem has been solved. The key result is that in the limit of large N , the variance of the estimate is bounded by (28), and the bound can be attained by separate von Neumann measurements on each particle.

In the case of mixed states of spin $1/2$ particles the state estimation problem for large N has been solved if one restricts oneself to separable measurements. However if one considers non separable measurements, then one can improve the quality of the estimate, which shows that the Fisher information, which in classical statistics is additive, is no longer so for quantum state estimation.

For the case of mixed states of spin $1/2$ particles, or for higher spins we do not know what the “outer” boundary of the set of (rescaled) achievable Fisher information matrices based on arbitrary (non separable) measurements of N systems looks like. We have some indications about the shape of this set (see section V F) and we know that it is convex and compact.

ACKNOWLEDGMENTS

S.M. thanks Utrecht University, Netherlands, where part of this work was carried out. He is a *chercheur qualifié* of the Belgium National Research Fund (FNRS). R.D.G. thanks the generous hospitality of the Department of Mathematics and Statistics, University of Western Australia.

REFERENCES

- [1] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Phys. Rev. Lett. **70** (1993) 1244
- [2] T. J. Dunn, I. A. Walmsley, and S. Mukamel, Phys. Rev. Lett. **74** (1995) 884
- [3] D. Leibfried, D. Meekhof, B. E. King, C. Monroe, W. M. Itano, and D. J. Wineland, Phys. Rev. Lett. **77** (1996) 4281
- [4] C. Kurtsiefer, T. Pfau, and J. Mlynek, Nature **386** (1997) 150
- [5] K. Vogel and H. Risken, Phys. Rev. A. **40** (1989) 2847
- [6] U. Leonhardt, Phys. Rev. Lett. **74** (1995) 4101
- [7] S. Weigert, *Reconstruction of spin states and its conceptual implications*, quant-ph/9809065
- [8] Z. Hradil, J. Summhammer, H. Rauch, *Quantum tomography as normalization of incompatible observations*, quant-ph/9806014
- [9] C. W. Helstrom, *Quantum Detection and Estimation Theory*, Academic, New York, 1976.
- [10] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, North Holland, Amsterdam, 1982.
- [11] S. Massar and S. Popescu, Phys. Rev. Lett. **74** (1995) 1259.
- [12] G. Vidal, J.I. Latorre, P. Pascual and R. Tarrach, *Optimal minimal measurements of mixed states*, quant-ph/9812068
- [13] W. K. Wootters, Phys. Rev. D **23** (1981) 357.
- [14] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72** (1994) 3439.
- [15] O. E. Barndorff-Nielsen and R. D. Gill, *an example of non attainability of expected quantum information*, quant-ph/9808009.
- [16] H. P. H. Yuen and M. Lax, IEEE Trans. IT-19(6) (1973) 740.
- [17] A. Peres and W. K. Wootters, Phys. Rev. Lett. **66**, 1119 (1991)
- [18] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **59** (1999) 1070
- [19] H. Cramér, *Mathematical methods of Statistics*, Princeton University, Princeton, NJ, 1946
- [20] H. L. van Trees, *Detection, Estimation and Modulation Theory*, Part 1, New York: Wiley, 1968.
- [21] R. D. Gill and B. Y. Levit, Bernoulli **1(1/2)** (1995) 59.

- [22] A. W. van der Vaart, ch. 27 pp. 397–410 in *Festschrift for Lucien Le Cam*, ed. D. Pollard, E. Torgersen, G.L. Yang; Springer, 1997.
- [23] I.A. Ibragimov, R.Z Has'minskii, *Statistical estimation—asymptotic theory*, Springer, 1981. Theorem III.1.1 (p. 174) is the theorem on uniform convergence of m.l.e.'s.
- [24] A.P.Dempster, N.M. Laird and D.R. Rubin, J. Roy. Statist. Soc. (B) 39 (1977), 1-38.