

## Estimating the spectrum of a density matrix with LOCC

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2006 J. Phys. A: Math. Gen. 39 1645

(<http://iopscience.iop.org/0305-4470/39/7/009>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.108

The article was downloaded on 03/06/2010 at 05:00

Please note that [terms and conditions apply](#).

# Estimating the spectrum of a density matrix with LOCC

**Manuel A Ballester**

Department of Mathematics, University of Utrecht, Box 80010, 3508 TA Utrecht,  
The Netherlands

E-mail: [ballester@math.uu.nl](mailto:ballester@math.uu.nl)

Received 18 July 2005, in final form 15 December 2005

Published 1 February 2006

Online at [stacks.iop.org/JPhysA/39/1645](http://stacks.iop.org/JPhysA/39/1645)

## Abstract

The problem of estimating the spectrum of a density matrix is considered. Other problems, such as bipartite pure state entanglement, can be reduced to spectrum estimation. A local operations and classical communication (LOCC) measurement strategy is shown which is asymptotically optimal. This means that, for a very large number of copies, it becomes unnecessary to perform collective measurements which should be more difficult to implement in practice.

PACS numbers: 03.65.Wj, 03.67.–a, 03.67.Mn

## 1. Introduction

Estimating a mixed state density matrix optimally, when one has  $N$  copies of it available, is a difficult problem. The problem has been solved for qubits by Vidal *et al* (1999), Bagan *et al* (2004) and Hayashi and Matsumoto (2004) and it is known that optimal collective measurements perform strictly better than any measurement which can be implemented with local operations and classical communication (LOCC). For mixed qudits, i.e., mixed states on a Hilbert space of dimension  $d$ , not much work on finding optimal collective measurements has been done. In the present work, a simpler case is studied, the estimation of the spectrum of a qudit density matrix. This problem has already been studied from the large deviation point of view by Keyl and Werner (2001) and for the qubit case by Bagan *et al* (2005).

In addition to being interesting in itself, spectrum estimation is useful because other problems can be reduced to it.

- Estimation of bipartite pure state entanglement. This problem has been studied for  $d = 2$  by Sancho and Huelga (2000) and Acín *et al* (2000).
- Estimation of the generalized Pauli channel. This problem has been studied by Fujiwara and Imai (2003) and the depolarizing channel (special case of the Pauli channel) by Sasaki *et al* (2002).

In the present paper, an LOCC asymptotically optimal<sup>1</sup> strategy will be described. The optimality of this LOCC strategy will be established by showing that it asymptotically satisfies the quantum Cramér-Rao bound (QCRB), stated by Helstrom (1976). The QCRB is a bound on the mean square error of ‘reasonable’ estimators.

This paper is organized as follows. In section 2, the necessary concepts are introduced and it is specified what is meant by optimality. In section 3, the estimation strategy is described and the main result is stated more precisely (equation (4)). In section 4, the conditional mean square error matrix (MSE) is calculated; this is needed for the next two sections. A heuristic argument supporting the main result is given in section 5 and a proof is given in section 6 (theorem 3).

## 2. Preliminaries

The density matrix  $\rho$  ( $\rho \geq 0$ ,  $\text{tr } \rho = 1$ ) will be parametrized in the following way:

$$\rho(p) = \sum_{k=1}^{d-1} p_k |k\rangle\langle k| + \left(1 - \sum_{l=1}^{d-1} p_l\right) |d\rangle\langle d|,$$

where  $p \in \Theta \subset \mathbb{R}^{d-1}$  is the parameter of interest,

$$\Theta = \left\{ (p_1, \dots, p_{d-1}) : 0 \leq p_k \leq 1, \sum_{k=1}^{d-1} p_k \leq 1 \right\}$$

is the set of possible values of the parameter and  $\{|1\rangle, \dots, |d\rangle\}$  is a basis of eigenvectors of  $\rho$ .

The quantum estimation problem that will be studied in this paper is that, given  $N$  copies of a completely unknown  $\rho$ , one is only interested in estimating its eigenvalues. Some of the needed concepts and results will be introduced next for the  $N = 1$  case.

Let  $M$  be a measurement with outcomes in a finite set  $\Omega$ , i.e., a collection of matrices  $\{M_\xi : \xi \in \Omega\}$  satisfying  $M_\xi \geq 0$  and  $\sum_{\xi \in \Omega} M_\xi = \mathbb{1}$ , and let  $\hat{p} = (\hat{p}_1, \dots, \hat{p}_{d-1})$  be an estimator of  $p$ , i.e., a map from  $\Omega$  to  $\Theta$ . The performance of such a measurement–estimator pair will be quantified by the MSE

$$\text{MSE}(\hat{p}, p, M)_{kl} = \mathbb{E}[(\hat{p}_k - p_k)(\hat{p}_l - p_l)] = \sum_{\xi \in \Omega} \text{tr}[\rho(p)M_\xi](\hat{p}_{\xi k} - p_k)(\hat{p}_{\xi l} - p_l),$$

where  $\mathbb{E}f$  means expectation of  $f$ .

The QCRB states that any unbiased<sup>2</sup> measurement–estimator pair  $(\hat{p}, M)$  of  $p$  satisfies

$$\text{MSE}(\hat{p}, p, M) \geq H(p)^{-1},$$

where  $H$  is the quantum Fisher information (QFI) (see for example Helstrom (1976) or Holevo (1982)). The QFI can be defined as the matrix with elements

$$H(p)_{kl} = \text{Re } \text{tr}[\rho(p)\lambda_k(p)\lambda_l(p)],$$

<sup>1</sup> That is, it performs asymptotically as well as any other measurement strategy.

<sup>2</sup> Unbiased means that

$$\mathbb{E}\hat{p}_k = \sum_{\xi \in \Omega} \text{tr}[\rho(p)M_\xi]\hat{p}_{\xi k} = p_k.$$

where  $\{\lambda_1(p), \dots, \lambda_{d-1}(p)\}$  are the symmetric logarithmic derivatives (SLD). The SLD are defined as self-adjoint solutions to the equation

$$\partial_k \rho(p) = \frac{\rho(p)\lambda_k(p) + \lambda_k(p)\rho(p)}{2}, \tag{1}$$

where  $\partial_k$  means partial derivative with respect to  $p_k$ .

The SLD for the model studied in this paper are easy to calculate, indeed, writing (1) on the basis of eigenvectors we get

$$\langle i|[|k\rangle\langle k| - |d\rangle\langle d|]|j\rangle = \frac{p_i + p_j}{2} \langle i|\lambda_k(p)|j\rangle,$$

or

$$\lambda_k(p) = \frac{|k\rangle\langle k|}{p_k} - \frac{|d\rangle\langle d|}{p_d}.$$

From the SLD one can then calculate the QFI to get

$$H(p)_{kl} = \frac{\delta_{kl}}{p_k} + \frac{1}{p_d}, \quad k, l \in \{1, \dots, d-1\},$$

where  $p_d = 1 - \sum_{l=1}^{d-1} p_l$ ; the inverse of  $H$  is

$$H(p)_{kl}^{-1} = p_k \delta_{kl} - p_k p_l, \quad k, l \in \{1, \dots, d-1\}. \tag{2}$$

When one has  $N$  copies of  $\rho$ , i.e., the model is of the form  $\rho(p)^{\otimes N}$ , the QCRB becomes

$$\text{MSE}(\hat{p}, p, M)^{(N)} \geq \frac{H(p)^{-1}}{N},$$

and this bound is valid for *any* measurement  $M$  (i.e. LOCC or not), as long as the measurement-estimator pair  $(\hat{p}, M)$  is unbiased.

The class of unbiased estimators, however, is too restrictive since in most practical situations one deals with biased ones. Gill and Levit (1995) used a multivariate extension of an inequality due to van Trees (1968) to prove a more general bound. From their result and an inequality due to Braunstein and Caves (1994), it can be shown that, under some regularity conditions, if  $\sqrt{N}(\hat{p} - p) \xrightarrow{D} Z(p)$  then

$$\text{Var } Z(p) \geq H(p)^{-1}, \tag{3}$$

where ' $\xrightarrow{D}$ ' means convergence in distribution. This means that the variance of the limiting distribution of any regular estimator satisfies the QCRB.

### 3. Estimation strategy

Suppose now that one knows the basis of eigenvectors, and let us consider the measurement with elements  $M_k = |k\rangle\langle k|$ . For this measurement, the probability of outcome  $k$  is

$$\text{tr}[\rho(p)M_k] = p_k.$$

Now suppose that this measurement is performed on  $N$  copies of  $\rho$ ; let  $N_k$  be the number of times that outcome  $k$  was observed. Then  $\{N_1, \dots, N_{d-1}\}$  have a multinomial distribution, i.e.,

$$\text{Pr}(N_1 = n_1, \dots, N_{d-1} = n_{d-1}) = \frac{N!}{\prod_{k=1}^{d-1} n_k!} \prod_{k=1}^{d-1} p_k^{n_k},$$

where  $n_d = N - \sum_{k=1}^{d-1} n_k$ . The estimator

$$\hat{p}_k = \frac{N_k}{N}$$

is unbiased and a simple calculation shows that its MSE equals the inverse of the QFI divided by  $N$ , which means that it saturates the QCRB and therefore it is optimal.

This would be the whole story, except for the fact that we have assumed that the eigenbasis of  $\rho$  is known. If the eigenbasis is not known, one can try to use a two-step adaptive strategy such as the one considered by Gill and Massar (2000). The idea is to make an initial rough estimate of  $\rho$  on an asymptotically vanishing fraction of the copies, e.g.,  $N^\mu$  with  $0 < \mu < 1$ . Let  $\sigma$  be that initial estimate of  $\rho$  and  $|\psi_k\rangle$  be its (not necessarily unique) eigenbasis. On the rest of the copies ( $N - N^\mu$ ) of  $\rho$ , the measurement with elements  $M_k = |\psi_k\rangle\langle\psi_k|$  is performed.

In the rest of this paper, it will be shown that this method is asymptotically optimal, i.e., it asymptotically achieves the QCRB:

$$\lim_{N \rightarrow \infty} N \text{MSE}(\hat{p}, p, M)^{(N)} = H(p)^{-1}, \quad (4)$$

provided  $\mu$  is chosen strictly larger than  $1/2$ .

#### 4. The MSE in the adaptive scheme

Let  $N_i = N^\mu$  and  $N_f = N - N^\mu$  be the sample sizes for the first and second stages respectively. In the second stage, the probability of outcome  $k$ , given the initial estimate  $\sigma$ , is

$$q_k = \text{tr} M_k \rho(p) = \langle \psi_k | \rho(p) | \psi_k \rangle.$$

These probabilities are also a random variable.

Next the MSE of the second stage (i.e. assuming fixed  $q$ 's) will be calculated. A condition for obtaining (4) will be derived from it.

Just as before, let  $N_k$  be the number of times that outcome  $k$  is observed and let us estimate  $p_k$  as

$$\hat{p}_k = \frac{N_k}{N_f}.$$

The expectation of this estimator conditioned on  $\sigma$  is

$$\mathbb{E}[\hat{p}_k | \sigma] = q_k,$$

so that in general it is a biased estimator. A simple calculation shows that the MSE conditioned on the first rough estimate of  $\rho$  is

$$\mathbb{E}[(\hat{p}_k - p_k)(\hat{p}_l - p_l) | \sigma] = \frac{q_k \delta_{kl} - q_k q_l}{N_f} + (p_k - q_k)(p_l - q_l). \quad (5)$$

The second term is the square of the bias; the MSE itself is

$$\text{MSE}(\hat{p}, p, M)^{(N)} = \mathbb{E}[\mathbb{E}[(\hat{p}_k - p_k)(\hat{p}_l - p_l) | \sigma]].$$

Comparing (2) and (5) and using the fact that  $N/N_f \rightarrow 1$  as  $N \rightarrow \infty$ , it is easy to see that in order to get (4) it is sufficient that

$$\lim_{N \rightarrow \infty} \mathbb{E}[N(q_k - p_k)(q_l - p_l)] = 0. \tag{6}$$

Indeed, if this is true, then it also holds that  $\mathbb{E}[q_k] \rightarrow p_k$  and  $\mathbb{E}[q_k q_l] \rightarrow p_k p_l$ .

**5. Heuristic argument**

Suppose for simplicity that all eigenvalues of  $\rho$  are different; then one expects that after the first estimate, the eigenbasis of  $\rho$  and the eigenbasis of  $\sigma$  are related by a unitary matrix which is very close to the identity, i.e.,

$$|\psi_k\rangle = U|k\rangle,$$

with

$$U = \exp\left(i \sum_{\alpha=1}^{d^2-1} \eta_\alpha T_\alpha\right) = e^{i\eta \cdot T},$$

where  $\{T_1, \dots, T_{d^2-1}\}$  is a basis of  $\mathfrak{su}(d)$  satisfying  $\text{tr } T_\alpha T_\beta = \delta_{\alpha\beta}$ ,  $\eta \in \mathbb{R}^{d^2-1}$  and  $\|\eta\|$  is small. One can then expand  $U$  in Taylor series about  $\eta = 0$ ,

$$U = \mathbb{1} + i\eta \cdot T - \frac{1}{2} (\eta \cdot T)^2 + o(\|\eta\|^2).$$

For any decent initial estimation strategy,  $\eta$  is expected to go to 0 as  $N \rightarrow \infty$  at a rate of  $N_i^{-1/2} = N^{-\mu/2}$ .

The expression for  $q_k$  is

$$q_k = \sum_l p_l |\langle l|U|k\rangle|^2,$$

and

$$|\langle l|U|k\rangle|^2 = \delta_{kl} + \langle l|\eta \cdot T|k\rangle \langle k|\eta \cdot T|l\rangle - \delta_{kl} \langle k|(\eta \cdot T)^2|k\rangle + o(\|\eta\|^2);$$

therefore

$$q_k - p_k = \langle k|(\eta \cdot T)\rho(\eta \cdot T)|k\rangle - p_k \langle k|(\eta \cdot T)^2|k\rangle + o(\|\eta\|^2).$$

From the previous expression and the fact that  $\eta$  goes to zero at the rate  $N^{-\mu/2}$ , one can expect that

$$\mathbb{E}(q_k - p_k)^2 = \frac{c}{N^{2\mu}} + o(N^{-2\mu}),$$

where  $c$  is a constant possibly depending on  $p$ . From the previous equation it follows that

$$\lim_{n \rightarrow \infty} N \mathbb{E}(q_k - p_k)^2 = 0, \tag{7}$$

if and only if  $\mu > 1/2$ . Now, using (7) together with the Cauchy–Schwartz inequality

$$(\mathbb{E}[N(q_k - p_k)(q_l - p_l)])^2 \leq \mathbb{E}[N(q_k - p_k)^2] \mathbb{E}[N(q_l - p_l)^2],$$

(6) follows. As pointed out before, the desired result (4) is a consequence of (6).

## 6. Rigorous argument

### 6.1. Some intermediate results

If  $\rho = \mathbb{1}/d$ , then any basis chosen for the second stage will give  $(q_k - p_k) = 0$ , so in what follows it is assumed that  $\rho \neq \mathbb{1}/d$ , i.e.,  $\rho$  has at least two different eigenvalues.

The following intermediate result will be needed. Basically it states that if  $\rho$  and  $\sigma$  are close to each other, then so will be their eigenvalues and eigenspaces.

**Lemma 1.** *Let*

$$\rho = \sum_{a=1}^n p_a \Pi_a, \quad \sigma = \sum_{k=1}^d s_k |\psi_k\rangle\langle\psi_k|,$$

where  $p_a \neq p_b$  for  $a \neq b$ ,  $2 \leq n \leq d$  is the number of different eigenvalues and  $\Pi_a$  is a projector onto the eigenspace corresponding to eigenvalue  $p_a$ , and let  $d_a = \text{tr} \Pi_a$  be the degeneracy of  $p_a$ ; also let

$$\Delta = \min_a \min_{b \neq a} |p_a - p_b| > 0.$$

If

$$d_{\text{HS}}(\rho, \sigma) = \sqrt{\text{tr}(\rho - \sigma)^2} \leq \delta < \frac{\Delta}{1 + \sqrt{d}},$$

then

(i)  $\forall a, k$

$$|p_a - s_k| \sqrt{\langle\psi_k|\Pi_a|\psi_k\rangle} \leq \delta,$$

i.e., either  $p_a$  is close to  $s_k$  or  $|\psi_k\rangle$  is almost orthogonal to the eigenspace corresponding to  $p_a$ .

(ii)  $\forall a \exists k$  such that  $|p_a - s_k| \leq \delta$  and  $\forall k \exists a$  such that  $|p_a - s_k| \leq \delta$ , i.e., every eigenvalue of  $\sigma$  is close to an eigenvalue of  $\rho$  and vice versa. Let  $M_a = \{k : |p_a - s_k| \leq \delta\}$  and  $m_a = |M_a| > 0$ . Note that  $M_a \cap M_b = \emptyset$  for  $a \neq b$ .

(iii) Let  $a \neq b$ ; then if  $k \in M_b$ , then  $|p_a - s_k| \geq \Delta - \delta$  and

$$\sqrt{\langle\psi_k|\Pi_a|\psi_k\rangle} \leq \frac{\delta}{\Delta - \delta},$$

i.e. if  $s_k$  is within a distance  $\delta$  of  $p_b \neq p_a$ , then  $|\psi_k\rangle$  is almost orthogonal to the eigenspace corresponding to  $p_a$ .

(iv)  $m_a = d_a$ , i.e., for  $\delta$  small enough, the number of eigenvalues of  $\sigma$  within a distance  $\delta$  from  $p_a$  is equal to the degeneracy of  $p_a$ .

(v)  $\forall k \in M_a$ ,

$$|p_a - \langle\psi_k|\rho|\psi_k\rangle| \leq c(\rho)\delta^2,$$

where

$$c(\rho) = \frac{4(d-1)}{\Delta}.$$

The proof of this lemma is given in appendix A.

Now the way in which the first rough estimation is done will be specified. For this part, it is convenient to represent  $\rho$  and  $\sigma$  in the following way:

$$\rho = \frac{\mathbb{1}}{d} + \theta \cdot T, \quad \sigma = \frac{\mathbb{1}}{d} + \hat{\theta} \cdot T.$$

The initial measurement strategy (which will be called *plain tomography*) is to divide the initial number of copies  $N_i$  into  $d^2 - 1$  groups of size  $N_0 = N_i/(d^2 - 1)$ , and in group  $\alpha \in \{1, \dots, d^2 - 1\}$  perform the measurement

$$M_{\pm}^{(\alpha)} = \frac{\mathbb{1} \pm T_{\alpha}}{2}.$$

The probabilities are

$$p_{\pm}^{(\alpha)} = \frac{1 \pm \theta_{\alpha}}{2}.$$

Let  $w_{\alpha+}$  be the number of times that outcome + was obtained; it is binomially distributed  $w_{\alpha+} \sim \text{Bin}(N_0, (1 + \theta_{\alpha})/2)$ . The estimator for  $\theta_{\alpha}$  is taken to be

$$\hat{\theta}_{\alpha} = 2 \frac{w_{\alpha+}}{N_0} - 1.$$

The following result holds.

**Lemma 2.** *If  $\mu > 1/2$ , then  $\forall \epsilon > 0$  and  $\forall h \geq 0$*

$$\lim_{N \rightarrow \infty} (N^h \Pr[\sqrt{N}|q_k - p_k| \geq \epsilon]) = 0. \tag{8}$$

The proof of this lemma is given in appendix B.

### 6.2. Proof of the main result

**Theorem 3.** *If  $\mu > 1/2$ , then (4) holds.*

**Proof.** Let  $X_k^{(N)} = \sqrt{N}(q_k - p_k)$ ; clearly  $(X_k^{(N)})^2 \leq N$ . All that needs to be proven is that

$$\lim_{N \rightarrow \infty} \mathbb{E}[X_k^{(N)} X_l^{(N)}] = 0.$$

We have that

$$|\mathbb{E}[X_k^{(N)} X_l^{(N)}]| \leq \mathbb{E}[|X_k^{(N)} X_l^{(N)}|] \leq \sqrt{\mathbb{E}[(X_k^{(N)})^2] \mathbb{E}[(X_l^{(N)})^2]}, \tag{9}$$

where in the second inequality the Cauchy–Schwartz inequality has been used. Now choose any  $\epsilon > 0$ ,

$$\begin{aligned} \mathbb{E}[(X_k^{(N)})^2] &= \sum_{x \geq 0} x \Pr[(X_k^{(N)})^2 = x] \\ &= \sum_{0 \leq x < \epsilon^2} x \Pr[(X_k^{(N)})^2 = x] + \sum_{x > \epsilon^2} x \Pr[(X_k^{(N)})^2 = x] \\ &\leq \epsilon^2 \Pr[(X_k^{(N)})^2 < \epsilon^2] + N \Pr[(X_k^{(N)})^2 \geq \epsilon^2] \\ &\leq \epsilon^2 + N \Pr[|X_k^{(N)}| \geq \epsilon]. \end{aligned}$$



Now using (8) one gets that  $\forall \epsilon > 0$

$$\lim_{N \rightarrow \infty} \mathbb{E}[(X_k^{(N)})^2] \leq \epsilon^2,$$

which implies that it must be zero; this fact and (9) imply (6) and therefore the desired result (4).  $\square$

We have proven something about the limit of the MSE, but (3) is a bound to the variance of the limiting distribution. However, since the limit of the MSE cannot be smaller than the variance of the limiting distribution (which in this case can easily be proven to be Gaussian), it follows that our estimator achieves the bound (3).

## 7. Estimation of bipartite pure state entanglement

A bipartite entangled pure state  $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$  can be written as (Schmidt's decomposition)

$$|\psi_{AB}\rangle = \sum_{k=1}^d \sqrt{p_k} |k\rangle \otimes |e_k\rangle,$$

where  $\{|k\rangle\}$  and  $\{|e_k\rangle\}$  are orthonormal bases of  $\mathcal{H}_A$  and  $\mathcal{H}_B$  which are both of dimension  $d$ .

The entanglement of  $|\psi_{AB}\rangle$  can be calculated as the entropy of one of the reduced states

$$E(|\psi_{AB}\rangle) = -\text{tr}(\rho_A \log_2 \rho_A) = -\sum_{k=1}^d p_k \log_2 p_k,$$

where  $\rho_A = \text{tr}_B |\psi_{AB}\rangle \langle \psi_{AB}|$ , i.e., entanglement is a function of the eigenvalues of the reduced density matrix. This means that entanglement can be estimated by performing measurements on  $\rho_A$  only, in order to estimate its spectrum. The question is whether this procedure is optimal. A quick calculation of the QFI for the parameters  $p_k$  in the model given by  $|\psi_{AB}\rangle$  shows that indeed the entanglement of  $|\psi_{AB}\rangle$  can be optimally estimated by estimating the spectrum of  $\rho_A$  using the procedure described above in this paper.

The same result<sup>3</sup> was obtained by Acín *et al* (2000) for  $d = 2$  using other tools.

## 8. Conclusions

The estimation of the spectrum of a finite-dimensional density matrix has been analysed. The following LOCC procedure has been studied.

- (i) Perform the so-called plain tomography on  $N^\mu$  copies, where  $\mu > 1/2$  and  $N$  is the total number of copies. From this, one gets an initial estimate of the whole density matrix, denoted by  $\sigma$ . Let  $|\psi_1\rangle, \dots, |\psi_d\rangle$  be a set of eigenvectors of  $\sigma$ .
- (ii) Perform the measurement with elements  $M_k = |\psi_k\rangle \langle \psi_k|$  on the remaining  $N - N^\mu$  copies and estimate  $p_k$  as the number of times the outcome  $k$  was obtained divided by  $N$ .

It has been shown that the above procedure performs asymptotically as well as any measurement (including collective ones). This means that in the asymptotic regime there is no need to perform the more complicated collective measurements for the estimation of the spectrum of a density matrix (or pure bipartite entanglement).

<sup>3</sup> That entanglement can be optimally estimated by estimating the spectrum of the reduced density matrix.

**Acknowledgments**

I would like to thank Richard Gill, Madalin Guță and Igor Grubišić for their very useful comments. This research was funded by the Netherlands Organization for Scientific Research (NWO); support from the RESQ (IST-2001-37559) project of the IST-FET programme of the European Union is also acknowledged.

**Appendix A. Proof of lemma 1**

(i) The square of the distance between  $\rho$  and  $\sigma$  can be written as

$$\begin{aligned} d_{\text{HS}}(\rho, \sigma)^2 &= \sum_{k=1}^d \sum_{a=1}^n \langle \psi_k | (\rho - \sigma) \Pi_a | (\rho - \sigma) | \psi_k \rangle \\ &= \sum_{k=1}^d \sum_{a=1}^n (p_a - s_k)^2 \langle \psi_k | \Pi_a | \psi_k \rangle \leq \delta^2. \end{aligned}$$

Since all terms are non-negative, this implies that all of them are less than or equal to  $\delta$  and this implies point (i).

(ii) For point (ii), only the first statement will be proven; the proof of the second is almost identical. Suppose that the opposite is true, i.e., that  $\exists a$  such that  $\forall k |p_a - s_k| > \delta$ ; then

$$\begin{aligned} d_{\text{HS}}(\rho, \sigma)^2 &= \sum_{k=1}^d \sum_{b=1}^n (p_b - s_k)^2 \langle \psi_k | \Pi_b | \psi_k \rangle \\ &\geq \sum_{k=1}^d (p_a - s_k)^2 \langle \psi_k | \Pi_a | \psi_k \rangle \\ &> \delta^2 \text{tr } \Pi_a \geq \delta^2, \end{aligned}$$

i.e.,  $d_{\text{HS}}(\rho, \sigma) > \delta$ , which is a contradiction.

(iii)  $|p_a - s_k| = |(p_a - p_b) + (p_b - s_k)| \geq |p_a - p_b| - |p_b - s_k| \geq \Delta - \delta$ ; the second statement follows from the previous inequality and point (i).

(iv)

$$\begin{aligned} m_a &= \sum_{k \in M_a} \langle \psi_k | \psi_k \rangle \geq \sum_{k \in M_a} \langle \psi_k | \Pi_a | \psi_k \rangle \\ &= \text{tr } \Pi_a - \sum_{k \notin M_a} \langle \psi_k | \Pi_a | \psi_k \rangle \\ &\geq \text{tr } \Pi_a - \sum_{k \notin M_a} \left( \frac{\delta}{\Delta - \delta} \right)^2 \\ &\geq \text{tr } \Pi_a - d \left( \frac{\delta}{\Delta - \delta} \right)^2, \end{aligned}$$

where point (iii) has been used. Now, since  $d_a = \text{tr } \Pi_a$ , we get

$$m_a \geq d_a - d \left( \frac{\delta}{\Delta - \delta} \right)^2.$$

Since  $\delta < \Delta / (1 + \sqrt{d})$ ,

$$d \left( \frac{\delta}{\Delta - \delta} \right)^2 < 1,$$

and since  $m_a$  is an integer, we have that  $m_a \geq d_a$ . Using the fact that  $\sum_a m_a = \sum_a d_a = d$ , we get that  $m_a = d_a$ .

(v) Let  $a \neq b$ , and  $k \in M_a$ ,

$$\begin{aligned} |p_a - p_b| \sqrt{\langle \psi_k | \Pi_b | \psi_k \rangle} &= |(p_a - s_k) + (s_k - p_b)| \sqrt{\langle \psi_k | \Pi_b | \psi_k \rangle} \\ &\leq [|p_a - s_k| + |s_k - p_b|] \sqrt{\langle \psi_k | \Pi_b | \psi_k \rangle} \\ &\leq [\delta \sqrt{\langle \psi_k | \Pi_b | \psi_k \rangle} + |s_k - p_b| \sqrt{\langle \psi_k | \Pi_b | \psi_k \rangle}] \\ &\leq [\delta \sqrt{\langle \psi_k | \Pi_b | \psi_k \rangle} + \delta] \leq 2\delta, \end{aligned}$$

where points (i) and (ii) have been used. Thus, we have that

$$\langle \psi_k | \Pi_b | \psi_k \rangle \leq \frac{4\delta^2}{(p_a - p_b)^2}.$$

Now we turn to the quantity of interest,

$$\begin{aligned} |p_a - \langle \psi_k | \rho | \psi_k \rangle| &= \left| p_a - \sum_b p_b \langle \psi_k | \Pi_b | \psi_k \rangle \right| \\ &= \left| \sum_b (p_a - p_b) \langle \psi_k | \Pi_b | \psi_k \rangle \right| \\ &\leq \sum_b |p_a - p_b| \langle \psi_k | \Pi_b | \psi_k \rangle \\ &= \sum_{b \neq a} |p_a - p_b| \langle \psi_k | \Pi_b | \psi_k \rangle \\ &\leq 4 \sum_{b \neq a} \frac{1}{|p_a - p_b|} \delta^2 \\ &\leq \frac{4(d-1)}{\Delta} \delta^2 = c(\rho) \delta^2. \end{aligned}$$

**Appendix B. Proof of lemma 2**

Now we enumerate the eigenvalues of  $\rho$  from 1 to  $d$  again, with some of them possibly equal. Points (ii) and (iv) of lemma 1 ensure that for every eigenvalue of  $\rho$ , the right number of eigenvalues of  $\sigma$  will satisfy point (v). From point (v) of lemma 1, we get that  $|q_k - p_k| \geq c(\rho) \delta^2$  implies  $d(\rho, \sigma)^2 \geq \delta^2$ ; we have

$$\begin{aligned} \Pr[|q_k - p_k| \geq c(\rho) \delta^2] &\leq \Pr[d(\rho, \sigma)^2 \geq \delta^2] \\ &= \Pr \left[ \sum_{\alpha=1}^{d^2-1} (\theta_\alpha - \hat{\theta}_\alpha)^2 \geq \delta^2 \right]. \end{aligned}$$

Since

$$\sum_{\alpha=1}^{d^2-1} (\theta_\alpha - \hat{\theta}_\alpha)^2 \geq \delta^2$$

implies that for at least one  $\alpha$

$$(\theta_\alpha - \hat{\theta}_\alpha)^2 \geq \frac{\delta^2}{d^2 - 1},$$

it follows that

$$\begin{aligned} \Pr \left[ \sum_{\alpha=1}^{d^2-1} (\theta_\alpha - \hat{\theta}_\alpha)^2 \geq \delta^2 \right] &\leq 1 - \Pr \left[ \forall \alpha, (\theta_\alpha - \hat{\theta}_\alpha)^2 < \frac{\delta^2}{d^2-1} \right] \\ &= 1 - \prod_{\alpha=1}^{d^2-1} \Pr \left[ |\theta_\alpha - \hat{\theta}_\alpha| < \frac{\delta}{\sqrt{d^2-1}} \right] \\ &= 1 - \prod_{\alpha=1}^{d^2-1} \Pr \left[ \left| w_{\alpha+} - \frac{1+\theta_\alpha}{2} N_0 \right| < \frac{N_0}{2} \frac{\delta}{\sqrt{d^2-1}} \right] \\ &= 1 - \prod_{\alpha=1}^{d^2-1} \left( 1 - \Pr \left[ \left| w_{\alpha+} - \frac{1+\theta_\alpha}{2} N_0 \right| \geq \frac{N_0}{2} \frac{\delta}{\sqrt{d^2-1}} \right] \right) \\ &\leq 1 - \left( 1 - 2 \exp \left[ -\frac{\delta^2}{2(d^2-1)} N_0 \right] \right)^{d^2-1}. \end{aligned}$$

In the last inequality, we have used a form of the Chernoff bound<sup>4</sup>. Thus, we finally have that

$$\Pr[|q_k - p_k| \geq c(\rho)\delta^2] \leq 1 - \left( 1 - 2 \exp \left[ -\frac{\delta^2}{2(d^2-1)} N_0 \right] \right)^{d^2-1}.$$

Now let  $c(\rho)\delta^2 = \epsilon N^{-1/2}$  and substitute  $N_0$  by its value,  $N^\mu/(d^2-1)$ ; the result is

$$\begin{aligned} \Pr[\sqrt{N}|q_k - p_k| \geq \epsilon] &\leq 1 - \left( 1 - 2 \exp \left[ -\frac{\epsilon N^{\mu-1/2}}{2c(\rho)(d^2-1)^2} \right] \right)^{d^2-1} \\ &\leq 2(d^2-1) \exp \left[ -\frac{\epsilon N^{\mu-1/2}}{2c(\rho)(d^2-1)^2} \right]. \end{aligned}$$

Multiplying by  $N^h$ , taking  $\mu > 1/2$  and  $N \rightarrow \infty$ , we get the desired result (8).

### References

- Acín A, Tarrach R and Vidal G 2000 Optimal estimation of two-qubit pure-state entanglement *Phys. Rev. A* **61** 062307 (Preprint [quant-ph/9911008](#))
- Bagan E, Baig M, Muñoz-Tapia R and Rodríguez A 2004 Collective versus local measurements in a qubit mixed-state estimation *Phys. Rev. A* **69** 010304(R) (Preprint [quant-ph/0307199](#))
- Bagan E, Ballester M A, Muñoz-Tapia R and Romero-Isart O 2005 Purity estimation with fully separable measurements *Phys. Rev. Lett.* **95** 110504 (Preprint [quant-ph/0505083](#))
- Braunstein S L and Caves C M 1994 Statistical distance and the geometry of quantum states *Phys. Rev. Lett.* **72** 3439
- Fujiwara A and Imai H 2003 Quantum parameter estimation of a generalized Pauli channel *J. Phys. A: Math. Gen.* **36** 8093–103
- Gill R D and Levit B Y 1995 Applications of the van Trees inequality: a Bayesian Cramér-Rao bound *Bernoulli* **1** 59–79
- Gill R D and Massar S 2000 State estimation for large ensembles *Phys. Rev. A* **61** 042312 (Preprint [quant-ph/9902063](#))
- Hayashi M and Matsumoto K 2004 Asymptotic performance of optimal state estimation in quantum two level system (Preprint [quant-ph/0411073](#))
- Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)

<sup>4</sup> If  $X \sim \text{Bin}(n, p)$ , then  $\Pr[|X - np| \geq \lambda] \leq 2 \exp(-2\lambda^2/n)$ .

- Holevo A 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam, New York, Oxford: North-Holland)
- Keyl M and Werner R F 2001 Estimating the spectrum of a density operator *Phys. Rev. A* **64** 052311 (*Preprint quant-ph/0102027*)
- Sancho J M G and Huelga S F 2000 Measuring the entanglement of bipartite pure states *Phys. Rev. A* **61** 042303 (*Preprint quant-ph/9910041*)
- Sasaki M, Ban M and Barnett S M 2002 Optimal parameter estimation of a depolarizing channel *Phys. Rev. A* **66** 022308 (*Preprint quant-ph/0203113*)
- van Trees H L 1968 *Detection, Estimation and Modulation Theory, Part 1* (New York: Wiley)
- Vidal G, Latorre J I, Pascual P and Tarrach R 1999 Optimal minimal measurements of mixed states *Phys. Rev. A* **60** 126–35 (*Preprint quant-ph/9812068*)