PhD Thesis

Efficient State Estimation for Quantum Systems

László Ruppert

Department of Mathematical Analysis,
Budapest University of Technology and Economics, Hungary

Supervisor: Prof. Dénes Petz

2012

Contents

1 Introduction

The measurement of a quantum mechanical system is of a probabilistic nature, so even determining a measurable quantity requires statistical methods. This is why state estimation is an important field in quantum information theory [36, 39, 42].

It is well-known that a measurement applied to a quantum system will change the state of it in an irreversible way depending on the measurement outcome. Additionally, D'Ariano and Yuen prove in [15] that it is generally not possible to determine the state of one single quantum system, whatever estimation scheme is used. We use therefore several copies of the same system being in the same state[1], perform a sequence of measurements, and estimate the unknown state from the outcome statistics. This process is usually called quantum state tomography [13] and can be traced back to the seventies [22, 24].

The problem of state estimation is quite old, however, the interest in a thorough mathematical analysis of quantum state estimation procedures has been flourishing recently [12, 30, 60]. For example, an adaptive observable selection strategy based on a Bloch vector parameterization in spherical coordinates and on a Bayesian estimation method of qubits in mixed states is reported in reference [17].

Quantum process tomography is a closely related field, an exhaustive description of possible tomography methods can be found in [34]. The difference between quantum state and quantum process tomography is that here we have a quantum channel and we are interested in its effect. For this purpose we send a prepared state (that is known) through the channel, and we make a tomography on the output state to estimate the channel effects. Many publications in this field deal with a very specific case, namely when we have a qubit and a channel with one parameter, e.g., the depolarization

[1]The preparation of identical states is in general rather difficult to implement in practice, however in certain physical circumstances, for example in quantum optics, it is natural to have several copies of the quantum system in the same state.

channel examined in [56]. However, there are some works that investigate the estimation of multi-parameter channels [7, 63, 37], and the multidimensional case also appears in Ref. [19, 35, 54].

Quantum control [6, 26, 51] is another problem similar to state estimation. Its purpose is the same as in classical control theory, i.e., to give an estimate of the unmeasured time-dependent state variable in order to be used in state feedback schemes such as noise reduction [10] or state purification [23]. While in state estimation problems we have a time-invariant problem, in quantum feedback control we have often a time-depending situation, hence quantum state filtering is essential [8, 9].

In quantum control it is important not to demolish the state of the system completely, therefore a so-called weak measurement is applied [25]. Thus, post-measurement state still contains information about the original. This property was used in [28, 29] to reverse the effect of some weak quantum measurements.[2] Weak measurements can be used for state estimation purposes, too. The majority of related articles take a continuous-time approach [27, 59], but there are some discrete-time models, too [20, 52, 53].

This thesis deals mainly with incorporating partial information into the state estimation process. In quantum tomography setups some a priori information about the state can be given in various ways. The most popular subject in this field is state discrimination: in that case we know that the system is in one of several given states, and we would like to determine which one it is [14]. We can have an a priori probability distribution of the true state, too, as another possibility. This idea was used in [16] to obtain the optimal phase estimation. Unlike in the state discrimination setup, the possible states do not form a discrete set, instead, all pure states are considered. In our setup we know that the state is in a given subset of the whole state space[3], i.e., some parameters of the state are known.

To set up a quantum state estimation two ingredients have to be given:

[2]Of course it can be done with probability less then 1.

[3]This can be the whole state space itself.

the measurement strategy used to obtain information, and the estimator mapping the measurement data to the state space. Most publications use maximum-likelihood (ML), Bayesian or some other simple method to obtain an estimator from measurement data. For the measurement part the spectrum of approaches is very wide. There are works [4, 21] which perform a single measurement on the compound system from identical copies and obtain optimality in an asymptotic sense. Other authors [43, 49] take measurements independently on states and deal with the properties of the estimate when only a finite number of measurements is available.

To obtain the optimality of a measurement setup we have to somehow measure the error of our estimation, which can be done many ways. In statistics the accuracy of the estimation is usually quantified by the covariance matrix. The matrices are typically not comparable by the positive semi-definiteness, hence if different estimation schemes are compared, the determinant of the covariance matrix can be used instead. This approach was introduced in references [43, 44]. Their result was that complementary von Neumann measurements are optimal in the qubit case. A more general context appears in [5], which is much closer to our approach. A similar result was obtained earlier by Wooters and Fields [62], but instead of using the covariance matrix of the estimator, they maximized the average information gain to obtain the optimality of complementarity. Another simple error function is the Hilbert-Schmidt distance, e.g., Scott used this to prove the optimality of SIC-POVMs [57].

In my thesis I examine different state estimation scenarios and give the best estimation schemes. I consider the case when we have multiple von Neumann measurements as well as that of a single POVM measurement. I analyze the problem of partial a priori information for qubits and multilevel systems as well. As a special case, I also consider the setup when no information is known beforehand. I introduce a new generalization of SIC-POVMs and examine its properties both analytically and numerically.

2 Mathematical background

2.1 Notations

$\mathbb{R}$	the set of real numbers
$\mathbb{R}^n$	the n-dimensional (real) Euclidean space
$\mathbb{C}$	the set of complex numbers
$x \in X$	x is an element of set X
$A^\intercal$	transpose of matrix A
A^*	conjugate transpose of matrix A
A^{-1}	inverse of matrix A
$[A]_{i,j}$	$(i,j)^{\text{th}}$ entry of matrix A
A_k	k-th element in a series of matrices $A_1, A_2, \ldots$
$\operatorname{Tr} A$	trace of matrix A
$\det A$	determinant of matrix A
I	unit matrix
v_i	i-th element of vector v
$v^\intercal$	transpose of vector v
$M_n(\mathbb{C})$	the set of $n \times n$ matrices with complex elements
$\langle X \rangle$	average value of X
$\mathbb{E}(X)$	expected value of random variable X
$\operatorname{Var}(X)$	variance/covariance matrix of random variable X
$\langle A, B \rangle$ or $\langle A \vert B \rangle$	scalar product of A and B
$\vert v \rangle$	vector from a Hilbert space (ket)
$\langle v \vert$	linear functional on kets (bra)
$\delta_{i,j}$	Kronecker delta function (equals to 1 if $i = j$, 0 otherwise)
$\operatorname{Prob}(X)$	the probability of outcome X
p_i	the probability of the i-th outcome
$A \perp B$	A is orthogonal to B
$A \otimes B$	tensor product of A and B
$A \oplus B$	direct sum (Cartesian product) of A and B

2.2 State of a quantum system

The states of finite quantum systems are represented by $n \times n$ density matrices ($\rho \in M_n(\mathbb{C})$), obeying the following two properties:

$$\mathrm{Tr}\,(\rho) = 1 \tag{1}$$

$$\rho \geq 0, \tag{2}$$

that is, the eigenvalues are non-negative numbers.[4]

The 2-dimensional case

The qubit (short for quantum bit) is the simplest of such states, and will bear special importance in our further investigations. In this case ρ is a 2×2 matrix, and the so-called Bloch parameterization, as we will see, gives a geometrically clear viewpoint of the state space:

$$\rho(\theta) = \frac{1}{2}\left(I + \theta_1\sigma_1 + \theta_2\sigma_2 + \theta_3\sigma_3\right), \tag{3}$$

with σ_i-s denoting the Pauli matrices:

$$\sigma_0 = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

$$\sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

that is

$$\rho(\theta) = \frac{1}{2} \begin{bmatrix} 1+\theta_3 & \theta_1 - i\cdot\theta_2 \\ \theta_1 + i\cdot\theta_2 & 1-\theta_3 \end{bmatrix}.$$

Thus the Bloch vector $\theta = (\theta_1, \theta_2, \theta_3)$ is equivalent to the matrix representation of qubit states. Using the notation $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ we can write instead of (3) the short version:

$$\rho(\theta) = \frac{1}{2}\left(I + \theta \cdot \sigma\right).$$

[4]Consequently, ρ is also self-adjoint.

It is easy to see that (1) holds; the second property of density matrices (2) transforms to

$$\theta_1^2 + \theta_2^2 + \theta_3^2 \leq 1,$$

in the Bloch parameterization. The state space is thus the unit ball in $\mathbb{R}^3$. States on the surface are called pure, others are reffered to as mixed states.[5]

The n-dimensional case

In the qubit case we used the Pauli matrices as an orthogonal basis to the self-adjoint matrices in $M_2(\mathbb{C})$ for parameterization with respect to the Hilbert-Schmidt inner product:

$$\langle A, B \rangle = \operatorname{Tr}(A^* B)$$

In the n-dimensional case we will use generalized Pauli matrices for the same purpose. We set an orthonormal basis on self-adjoint matrices

$$\forall i: \ \sigma_i = \sigma_i^*, \quad \operatorname{Tr} \sigma_i = 0 \ \text{ and } \quad \forall i, j: \ \operatorname{Tr} \sigma_i \sigma_j = \delta_{i,j},$$

where $1 \leq i, j \leq n^2 - 1$ and $\sigma_0 = \frac{1}{\sqrt{n}} I$. Note that the Pauli matrices in the previous section were not normalized because of tradition.

Using this basis for parameterization we have

$$\rho(\theta) = \sum_{i=0}^{n^2-1} \theta_i \sigma_i, \tag{4}$$

and we call $\theta = (\theta_1, \theta_2, \ldots, \theta_{n^2-1})$ the generalized Bloch vector.

Substituting representation (4) back into property (1), we arrived at

$$\operatorname{Tr} \rho(\theta) = \operatorname{Tr} \sum_{i=0}^{n^2-1} \theta_i \sigma_i = \operatorname{Tr} \theta_0 \sigma_0 = \operatorname{Tr} \theta_0 \frac{1}{\sqrt{n}} I = \theta_0 \sqrt{n} = 1,$$

so $\theta_0 = \frac{1}{\sqrt{n}}$. Eq. (2), however, cannot be reformulated as a closed expression of the elements of the Bloch vectors. In the following sections, the quantum states will either be represented as a density matrix or as a Bloch vector.

[5] Generally pure states have eigenvalues 0 or 1, i.e., $\rho = \rho^2$ holds.

Example 2.1 *For $n = 3$ the generalized Pauli matrices are proportional to the Gell-Mann matrices:*

$$\sigma_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \sigma_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \sigma_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\sigma_4 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \sigma_5 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{bmatrix}, \quad \sigma_6 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

$$\sigma_7 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{bmatrix}, \quad \sigma_8 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix},$$

Example 2.2 *For $n = 4$ we can obtain the generalized Pauli matrices as the tensor products of 2-dimensional Pauli-matrices (σ_i):*

$$\{\frac{1}{2} \cdot \sigma_i \otimes \sigma_j\}, \quad \textit{with} \quad i, j \in \{0, 1, 2, 3\}$$

2.3 Measurements

Measurements are of a probabilistic nature in quantum mechanics. Observable quantities correspond to $n \times n$ self-adjoint matrices. Let the spectral decomposition of an observable A be the following:

$$A = \sum_{i=1}^{k} \lambda_i P_i. \tag{5}$$

Here, λ_i are the different eigenvalues of A, and the P_i-s are the corresponding eigenprojections ($P_i^* = P_i = P_i^2$, $\sum_i P_i = I$, $P_i P_j = 0$ for $i \neq j$).

The possible outcomes of the measurement are the different λ_i eigenvalues and the corresponding probability is

$$\mathrm{Prob}\,(\lambda_i) = \mathrm{Tr}\,(\rho P_i). \tag{6}$$

A key element of quantum measurements is that it will change the actual state of the quantum system to

$$\rho_i = \frac{P_i \rho P_i}{\operatorname{Tr} P_i \rho P_i}, \tag{7}$$

if the outcome of measurement is λ_i. Thus each sample can only be used once for measuring a certain quantity.

Example 2.3 *Measurement of the Pauli operator σ_1:*

If one considers in the qubit case the measurement of the observable σ_1, then we have the spectral decomposition:

$$\sigma_1 = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} =: E_{+1} - E_{-1}.$$

The possible outcomes are the different eigenvalues of the observable, i.e., ± 1. The probabilities of these outcomes are

$$Prob(+1) = \operatorname{Tr} \rho E_{+1} = \tfrac{1+\theta_1}{2}$$

$$Prob(-1) = \operatorname{Tr} \rho E_{-1} = \tfrac{1-\theta_1}{2}$$

respectively. The state after measurement is either

$$\theta^{+1} = \begin{bmatrix} +1, & 0, & 0 \end{bmatrix}^\intercal \quad or \quad \theta^{-1} = \begin{bmatrix} -1, & 0, & 0 \end{bmatrix}^\intercal$$

depending on the outcome.

From the state estimation point of view, only the distribution of the outcomes ($\operatorname{Tr}(\rho P_i)$) bears important. The eigenvalues λ_i and the state after the measurement are irrelevant. A measurement can thus be given as a set of projections $(P_1, P_2, \ldots, P_k)$ on $M_n(\mathbb{C})$, and we will call it a von Neumann measurement.

We will put a special emphasis in the next sections on the case of von Neumann measurements with two elements: (P_1, P_2). We have then $P_2 = I - P_1$, and if P_1 is a projection, then P_2 is always a projection, since

$$P_2^* = (I - P_1)^* = I - P_1^* = I - P_1 = P_2$$

and

$$P_2^2 = (I - P_1)(I - P_1) = I - P_1 - P_1 + P_1^2 = I - 2P_1 + P_1 = I - P_1 = P_2.$$

So every von Neumann measurement with two elements can be characterized with a projection P by using POVM $(P, I - P)$. In Example 2.3 we have $P = E_{+1} = \frac{1}{2}(I + \sigma_1)$.

Positive operator valued measurements

We can easily generalize this scenario to the so-called positive operator valued measurements (POVMs).

A set of self-adjoint operators $(E_1, E_2, \ldots, E_k)$ forms a POVM if and only if

$$\forall i : E_i \geq 0 \quad \text{and} \quad \sum_i E_i = I. \tag{8}$$

The probability of observing an outcome related to E_i is

$$p_i = \mathrm{Tr}\,(\rho E_i).$$

Note that unlike von Neumann measurements, POVMs do not necessary consist solely of projections, only positivity is required.

2.4 The standard state estimation method

This section introduces the important state estimation concepts through the standard method.

The standard method for qubits

First we perform three different kinds of von Neumann measurements with two elements $(P_i, I - P_i)$. Let us have for the observables the three Pauli matrices $\sigma_1, \sigma_2, \sigma_3$, i.e., we have the projections: $P_i = \frac{1}{2}(I + \sigma_i),\ (i = 1, 2, 3)$.

Then the probability of outcome $+1$ is

$$p_i = Tr(\rho P_i) = \frac{1 + \theta_i}{2}. \tag{9}$$

similarly to Example 2.3. From the physical point of view, σ_1, σ_2 és σ_3 correspond to measuring the spin in directions x, y and z, respectively.

Because of the high symmetry and the independence of the components we can easily construct an estimation scheme for the state θ. Suppose that m measurements are performed in each direction. Then the relative frequency ν_i of the outcomes $+1$ is a sufficient statistic:

$$\nu_i := \frac{m_i}{m}, \quad i = 1, 2, 3$$

with denoting m_i the number of $(+1)$-s in direction i.

The least squares estimator is a well-known and widespread method that minimizes the squared error. If the relative frequencies resulting from the measurements are ν_i, then the deviation from the real value of the state can be written in the form (because of the independence of the measurements, there are no cross terms):

$$\varphi_{LS}(\nu, \theta) = \sum_{i=1,2,3} \left(\nu_i - \frac{1 + \theta_i}{2} \right)^2 \to \min.$$

We are going to minimize this expression. It is trivially minimal when the expressions in brackets are zero. This way, knowing the relative frequencies ν_i, an estimate can be given for the Bloch vector θ:

$$\hat{\theta}_m(\nu) = \begin{bmatrix} 2\nu_1 - 1 \\ 2\nu_2 - 1 \\ 2\nu_3 - 1 \end{bmatrix} = \begin{bmatrix} 2\frac{m_1}{m} - 1 \\ 2\frac{m_2}{m} - 1 \\ 2\frac{m_3}{m} - 1 \end{bmatrix}. \tag{10}$$

The notation $\hat{\theta}_m(\nu)$ expresses that the above expression is an estimator of θ, takes m, the number of measurements, as a parameter and also that its value is a function of the outcome statistics ν.

Note that this estimation can provide a result that is physically meaningless (the state will be outside of the Bloch ball), but the probability of

having unphysical states vanishes exponentially as the number of measurements increases due to the large deviation theorem. So we will ignore this problem in the future examinations.

$\hat{\theta}_m(\nu)$ is an unbiased estimator since

$$\mathbb{E}(\hat{\theta}_m(\nu)) = \begin{bmatrix} 2\frac{\mathbb{E}(m_1)}{m} - 1 \\ \\ 2\frac{\mathbb{E}(m_2)}{m} - 1 \\ \\ 2\frac{\mathbb{E}(m_3)}{m} - 1 \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \\ \theta_2 \\ \\ \theta_3 \end{bmatrix} = \underline{\theta}, \tag{11}$$

since m_i is binomially distributed with parameters m and p_i, and so from (9), we know its expected value and variance

$$\mathbb{E}(m_i) = mp_i = m\frac{1+\theta_i}{2} \quad \text{and} \quad \mathrm{Var}\,(m_i) = mp_i(1-p_i) = m\frac{1-\theta_i^2}{4} \tag{12}$$

If we have a random variable $X = (X_1, X_2, \ldots, X_k)$ with expectation $\mathbb{E}(X) = (\mu_1, \mu_2, \ldots, \mu_k)$, we can also define the covariance matrix:

$$[\mathrm{Var}\,(X)]_{i,j} = \mathbb{E}\left((X_i - \mu_i)(X_j - \mu_j)\right)$$

From (12) we can easily calculate the covariance matrix of estimator (10), too:

$$\mathrm{Var}\,(\hat{\theta}_m(\nu)) = \frac{1}{m}\begin{bmatrix} 1-\theta_1^2 & 0 & 0 \\ 0 & 1-\theta_2^2 & 0 \\ 0 & 0 & 1-\theta_3^2 \end{bmatrix}, \tag{13}$$

since m_i and m_j are independent ($i \neq j$), so off-diagonal elements are zeros, while

$$\mathrm{Var}\,\left(2\frac{m_i}{m} - 1\right) = 4\frac{\mathrm{Var}\,(m_i)}{m^2} = \frac{1-\theta_i^2}{m}.$$

So we get that the variance converges to zero in order of $\frac{1}{m}$, where m is the number of measurements. Together with the unbiasedness (11) of this estimator we get that (10) is a consistent estimator, i.e., if m converges to infinity, then $\hat{\theta}_m(\nu) \to \theta$ with probability 1.

Generally, we can suppose that the estimator is a linear function of ν

$$\hat{\theta}_m(\nu) = a \cdot \nu + b$$

and also that we have identical, independent measurements, so we get:

$$\mathrm{Var}\left(\hat{\theta}_m(\nu_i)\right) = a^2 \mathrm{Var}\left(\nu_i\right) = a^2 \frac{\mathrm{Var}\left(m_i\right)}{m^2} = a^2 \frac{\mathrm{Var}\left(m_i^*\right) \cdot m}{m^2}, \tag{14}$$

where m_i^* is the number of $(+1)$-s in direction i, after a single measurement. Note that m_i is the sum of m independent and identically distributed variables, all of which are distributed according to m_i^*. Consequently, $Var(m_i) = m \cdot Var(m_i^*)$. So from (14) we obtain

$$\mathrm{Var}\left(\hat{\theta}_m(\nu_i)\right) = \frac{\mathrm{Var}\left(\hat{\theta}_1(\nu_i)\right)}{m},$$

that is, we have the same optimization problem for $m = 1$ as for an arbitrary m up to a normalizing factor. Therefore in further calculations we will always use the $m = 1$ case for simplicity, but we will assume that the number of measurements is sufficient to be close to the asymptotic properties, where we do not have problems with unphysical estimations.

Finally let us introduce the Fisher information[6] for an observable having probability distribution $(q_1, q_2, \ldots, q_l)$:

$$[I(\theta)]_{i,j} = \sum_\alpha \frac{1}{q_\alpha} \frac{\partial q_\alpha}{\partial \theta_i} \frac{\partial q_\alpha}{\partial \theta_j}$$

If we have one measurement in the i-th direction (that is $m = 1$), then we have Bernoulli distribution with probability distribution $(\frac{1+\theta_i}{2}, \frac{1-\theta_i}{2})$, which has Fisher information:

$$[I_1(\theta)]_{i,i} = \frac{2}{1+\theta_i} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{2}{1-\theta_i} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{1-\theta_i^2}.$$

Since m_i has binomial distribution, which is a sum of independent Bernoulli distributions, we obtain

$$I_m(\theta) = m I_1(\theta) = m \begin{bmatrix} \frac{1}{1-\theta_1^2} & 0 & 0 \\ 0 & \frac{1}{1-\theta_2^2} & 0 \\ 0 & 0 & \frac{1}{1-\theta_3^2} \end{bmatrix}.$$

[6]With respect to the state parameters $\theta = (\theta_1, \theta_2, \ldots \theta_{n^2-1})$.

The Cramér-Rao inequality provides us a lower bound for the variance of parameter estimators. Let us suppose that we have an unbiased estimator $\Psi(\theta)$ of θ, then we have

$$\mathrm{Var}\left(\Psi(\theta)\right) \geq I_m(\theta)^{-1}. \tag{15}$$

Since $\hat{\theta}_m(\nu)$ is an unbiased estimator of θ and it fulfills (15) with equality, so the estimator (10) is an efficient estimator.

The POVM case

Instead of using 3 different von Neumann measurements with two elements: $(P_i, I - P_i) = \left(\frac{1}{2}(I + \sigma_i), \frac{1}{2}(I - \sigma_i)\right)$ we can use a single POVM with six elements:

$$E_1 = \frac{1}{6}(I + \sigma_1), \quad E_2 = \frac{1}{6}(I - \sigma_1),$$
$$E_3 = \frac{1}{6}(I + \sigma_2), \quad E_4 = \frac{1}{6}(I - \sigma_2),$$
$$E_5 = \frac{1}{6}(I + \sigma_3), \quad E_6 = \frac{1}{6}(I - \sigma_3).$$

Then the probabilities of different outcomes are $p_i = \mathrm{Tr}\left(\rho E_i\right)$, so we get:

$$p_1 = \frac{1 + \theta_1}{6}, \quad p_2 = \frac{1 - \theta_1}{6},$$
$$p_3 = \frac{1 + \theta_2}{6}, \quad p_4 = \frac{1 - \theta_2}{6},$$
$$p_5 = \frac{1 + \theta_3}{6}, \quad p_6 = \frac{1 - \theta_3}{6}.$$

We can similarly construct an estimator as in the von Neumann case, we need to repeat the single measurement $\{E_1, E_2, E_3, E_4, E_5, E_6\}$. Let n_i denote the number of outcomes related to E_i, and $n = \sum n_i$ the total number of measurements. Then using the relative frequencies $\nu_i = \frac{n_i}{n}$ we can construct the estimator of θ:

$$\hat{\theta}_n(\nu) = \begin{bmatrix} \frac{n_1 - n_2}{n_1 + n_2} \\ \\ \frac{n_3 - n_4}{n_3 + n_4} \\ \\ \frac{n_5 - n_6}{n_5 + n_6} \end{bmatrix} = \begin{bmatrix} 2\tilde{\nu}_1 - 1 \\ \\ 2\tilde{\nu}_2 - 1 \\ \\ 2\tilde{\nu}_3 - 1 \end{bmatrix}, \tag{16}$$

where $\tilde{\nu}_1 = \frac{n_1}{n_1+n_2}$, $\tilde{\nu}_2 = \frac{n_3}{n_3+n_4}$, $\tilde{\nu}_3 = \frac{n_5}{n_5+n_6}$.

So this estimator is very similar to (10), the only difference is that here we measure $n_1 + n_2$ times in σ_1 direction, $n_3 + n_4$ times in σ_2 direction and $n_5 + n_6$ times in σ_3 direction, instead of m times in each. We can easily verify that (16) is unbiased, and asymptotically has the same variance for each state parameter as (10) with $m = n/3$: the law of large number ensures that $n_1 + n_2 \sim \frac{n}{3}$, $n_3 + n_4 \sim \frac{n}{3}$ and $n_5 + n_6 \sim \frac{n}{3}$.

In the literature (16) is often referred to as the standard estimator of a qubit (e.g., [49]). As we have seen, there is not much difference, so we will prefer the simpler one (10).

Generalization to n-level quantum systems

A straightforward modification of the measurement setup can lead us to the multidimensional case [43, 44].

Consider an n-level quantum system with density matrix $\rho \in M_n(\mathbb{C})$. Since the density matrices are self-adjoint matrices with unit trace, they can be characterized by $d = n^2 - 1$ real parameters.

Let E_{ij} denote the $n \times n$ matrix units[7] and set

$$\begin{aligned} Z_{ii} &:= E_{ii} && (1 \le i < n), \\ X_{ij} &:= E_{ij} + E_{ji} && (i < j), \\ Y_{ij} &:= -\mathrm{i}E_{ij} + \mathrm{i}E_{ji} && (i < j). \end{aligned}$$

The spectrum of Z_{ii} is $\{0, 1\}$ and the spectrum of X_{ij} and Y_{ij} is $\{-1, 0, 1\}$. These observables can be used to estimate the $n^2 - 1$ real parameters of the $n \times n$ density matrix selectively. If each observable is measured r times, then $m = r(n^2 - 1)$ copies of the quantum system are used.

2.5 Complementary and symmetric measurements

In this subsection we expound on two basic properties, that will take important role in the efficient state estimation.

[7] $[E_{ij}]_{k,l} = \delta_{i,k} \cdot \delta_{j,l}$

Complementarity

The heuristic concept of complementarity was born together with quantum theory. A mathematical definition is due to Accardi [1] and Kraus [31]. Ref. [40] provides an overview about complementarity; details are given in [41, 48].

Let $\mathcal{H}$ be an n-dimensional Hilbert space. Then the basis $e_1, e_2, \ldots, e_n$ is complementary to the basis $f_1, f_2, \ldots, f_n$ if

$$|\langle e_i, f_j \rangle|^2 = \frac{1}{n} \qquad (1 \leq i, j \leq n). \tag{17}$$

If this condition holds then the two bases are also called mutually unbiased.

If two observables A and B have eigenvectors $e_1, e_2, \ldots, e_n$ and $f_1, f_2, \ldots, f_n$ which are orthonormal bases, then they are complementary if the bases $e_1, e_2, \ldots, e_n$ and $f_1, f_2, \ldots, f_n$ are complementary.

The generalization can be made to POVMs. The POVMs $\{E_1, E_2, \ldots, E_k\}$ and $\{F_1, F_2, \ldots, F_m\}$ are complementary if

$$\operatorname{Tr} E_i F_j = \frac{1}{n} \operatorname{Tr} E_i \operatorname{Tr} F_j \qquad (1 \leq i \leq k, \quad 1 \leq j \leq m).$$

or equivalently, the traceless parts of the matrices are orthogonal

$$E_i - \frac{\operatorname{Tr} E_i}{n} I \quad \perp \quad F_j - \frac{\operatorname{Tr} F_j}{n} I, \qquad (1 \leq i \leq k, \quad 1 \leq j \leq m).$$

The latter property can be extended to subspaces $\mathcal{A}_1, \mathcal{A}_2 \subset M_n(\mathbb{C})$. $\mathcal{A}_1$ and $\mathcal{A}_2$ are complementary if the traceless part of matrices $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$ are orthogonal, so we have the orthogonality of the following subspaces:

$$\mathcal{A}_1 \ominus \mathbb{C}I \perp \mathcal{A}_2 \ominus \mathbb{C}I,$$

where $\mathcal{A}_1 \ominus \mathbb{C}I$ is the traceless subspace of $\mathcal{A}_1$. In the following sections we are referring to this property as quasi-orthogonality.

Symmetric measurements

A density matrix $\rho \in M_n(\mathbb{C})$ has $n^2 - 1$ real parameters. To cover all parameters the POVM has to contain at least n^2 elements, since the probability

distribution has one less degree of freedom than the number of POVM elements. We can take projections P_i, $1 \leq i \leq n^2$, such that

$$E_i = \frac{1}{n} P_i, \qquad \sum_{i=1}^{n^2} P_i = nI, \qquad \mathrm{Tr}\, P_i P_j = \frac{1}{n+1} \quad (i \neq j) \tag{18}$$

and this is called symmetric informationally complete POVM (SIC-POVM) by Zauner [64] and is currently a rather popular concept [2, 3, 11, 32, 50, 55, 65]. Zauner shoved the existence for $n \leq 5$, there has been some analytic and numerical progress [18, 58], but the existence of a SIC-POVM is still unknown for a general dimension n.

From the state estimation point of view, the SIC-POVM $\{E_i : 1 \leq i \leq n^2\}$ of an n-level system is optimal among the POVMs having n^2 elements, it minimizes the average squared Hilbert-Schmidt distance of the estimation and the true density [57].

The case when we consider less than n^2 projections with similar properties is more interesting. If $n^2 - k$ ($k < n^2$) parameters are known, then we want to estimate only the $k - 1$ unknown parameters, therefore a POVM $\{E_i : 1 \leq i \leq k\}$ is sufficient for full state reconstruction.

In this case we expect the same kind of symmetry as in the unconditional version (18). Let us take a set of projections $P_i,\ 1 \leq i \leq k$ such that

$$E_i = \frac{1}{\lambda} P_i, \quad \sum_{i=1}^{k} P_i = \lambda I \quad \text{and} \quad \mathrm{Tr}\, P_i P_j = \mu \quad (i \neq j). \tag{19}$$

The SIC-POVM case corresponds to $k = n^2$, $\lambda = n$ and $\mu = 1/(n+1)$, but there are other parameter sets for which there exists a POVM fulfilling the conditions of Eq. (19).

Naturally, the optimal POVM depends on the known parameters, therefore we use the term of conditional SIC-POVM. This is a new subject of quantum information theory, the existence of such conditional SIC-POVMs can be a fundamental question in different quantum tomography problems.

2.6 Literature review

In this section we will summarize the results from the literature which are closely related to our approach.

Optimal state determination by mutually unbiased measurements [62] The basic ideas of our work can be traced back to the fundamental work of Wooters and Fields in 1989. They maximized the information gain of the estimation

$$\mathcal{I}_\theta(\hat{\theta}) = h(\theta) - h(\hat{\theta}|\theta), \tag{20}$$

where $h(\theta) = -\int f_0(x) \log f_0(x) dx$ is the entropy of the prior distribution, while $h(\hat{\theta}|\theta) = -\int f_0(x) h(\hat{\theta}|\theta = x) dx$ is the conditional entropy of $\hat{\theta}$ given θ. A key step is that before maximization they take the average of (20) over the possible true states with respect to the prior distribution, and they obtain from that the optimality of complementary measurements.

Point estimation of states of finite quantum systems [43] This article of Petz, Hangos and Magyar from 2007 is mostly dealing with the case of n-dimensional state estimators, but mentions as a side result the optimality of the complementary measurement for qubits. This is a much weaker result than the previous one, but they optimize a different quantity: instead of maximizing the information gain, they minimize the determinant of the average covariance matrix

$$\det \langle \mathrm{Var}\,(\hat{\theta}) \rangle \to \min\,. \tag{21}$$

Complementarity and state estimation [5] The direct premise of our work is this article by Baier and Petz from 2010. They used (21) for proving optimality of complementary measurements in quite general settings, so improved the result of Wooters and Fields. Their approach, however, had its own limitations: they do not consider either the known parameter case, or the single POVM case. But they results suggested to us that the determi-

nant of the average covariance matrix might be a useful quantity for more complex state estimation scenarios, too.

Minimal qubit tomography [49] Rehacek, Englert and Kaszlikowski argued in 2004 that although there are six measurement directions in the standard method, in the qubit case four of them are enough to obtain all possible states. So they did not measure in the direction of the axis of the Bloch ball, but the elements of POVM were related to the vertices of a regular tetrahedron,

$$E_i = \frac{1}{4}(I + a_i \cdot \sigma), \quad i = 1, 2, 3, 4, \tag{22}$$

with Bloch vectors

$$a_1 = \frac{1}{\sqrt{3}}(1, 1, 1), \quad a_2 = \frac{1}{\sqrt{3}}(1, -1, -1),$$

$$a_3 = \frac{1}{\sqrt{3}}(-1, 1, -1), \quad a_4 = \frac{1}{\sqrt{3}}(-1, -1, 1).$$

Which is a well-known example of the SIC-POVM in the qubit case.

They applied the maximum likelihood principle to get an estimation for θ. If $\theta \neq 0$, the variance of their estimator is higher than that of the standard method [8], however, since their POVM does not contain more elements than necessary, it is easier to generalize their method to higher dimensions.

Tight informationally complete quantum measurements [57] Scott described in 2006 a method, that minimizes the expectation of the squared Hilbert-Schmidt distance

$$\mathbb{E}\|\rho - \hat{\rho}\|^2 \to \min,$$

to obtain the optimal linear state estimation. His result was that if we have $\rho \in M_n(\mathbb{C})$ and a POVM with $k \geq n^2$, then the optimal POVM is a tight[9]

[8] They also suggested an adaptive measurement procedure, which provides better results for pure states than the standard method.

[9] Tightness actually means some strong kind of symmetry, for the exact formulation see Definition 11 in [57].

rank-one informationally complete POVM. The case when $k = n^2$, i.e., we have exactly the necessary number of POVM elements, the optimal POVM is the n-dimensional SIC-POVM. In our work we are interested in the $k < n^2$ case and we generalize Scott's method to obtain a result analogous to Lemma 17 in [57]. This enables us to calculate some basic properties of conditional SIC-POVMs.

3 Efficient state estimation

In this section I explain my results in detail. First, the case of von Neumann measurements is investigated. Detailed examples are presented to give motivation for the method choice. Results concerning the general, n-dimensional case are also given.

Afterwards, positive operator valued measurements are examined, both for the qubit and the n-level setup. For the n-dimensional system, results are only given in the symmetrical case. The asymmetrical variant is much more difficult, however, it is also very interesting: the optimal measurements give us a generalization of SIC-POVMs.

The last part investigates this asymmetrical case in more detail: first we show a numerical method that solves this optimization problem efficiently. Finally, a different quantity is introduced which settles the question in the general, n-level case; showing that conditional SIC-POVMs are indeed optimal.

3.1 Von Neumann measurements [45]

In this section we will use von Neumann measurements with two elements for state estimation. When multiple parameters are unknown we have to use multiple von Neumann measurements for complete state reconstruction. In order to have a unique estimator we use exactly the sufficient number of projections.

3.1.1 The qubit case

1. Assume first that θ_1 and θ_2 are known and we want to estimate θ_3. The assumption means that the reduced state is known on the subspace generated by σ_1 and σ_2. It is a complementary to the subspace generated by σ_3. A projection

$$E = \frac{1}{2}\left(I + \lambda \cdot \sigma\right)$$

is used to estimate from the result of several measurements; $\lambda_1^2 + \lambda_2^2 + \lambda_3^2 = 1$.

The probability of getting an outcome related to E is

$$p = \operatorname{Tr} \rho E = \operatorname{Tr}\left[\frac{1}{2}\left(I + \sum_{i=1}^{3} \theta_i \sigma_i\right) \cdot \frac{1}{2}\left(I + \sum_{j=1}^{3} \lambda_j \sigma_j\right)\right] =$$

$$= \frac{1}{4}\operatorname{Tr} I + \frac{1}{4}\left(\sum_{i,j=1}^{3} \theta_i \lambda_j \operatorname{Tr} \sigma_i \sigma_j\right) = \frac{1}{2}\left(1 + \sum_{i=1}^{3} \theta_i \lambda_i\right) = \frac{1}{2}\left(1 + \langle \theta, \lambda \rangle\right),$$

since $\operatorname{Tr} \sigma_i \sigma_j = 2 \cdot \delta_{i,j}$, so we have

$$\theta_3 \lambda_3 = 2p - 1 - \theta_1 \lambda_1 - \theta_2 \lambda_2 \,. \tag{23}$$

In (23) $\lambda_1, \lambda_2, \lambda_3$ are the parameters of the measurements, θ_1 and θ_2 are known constants and θ_3 is the parameter to estimate. Our aim is to find the optimal measurement setup λ that minimizes the estimation error of θ_3. We proceed similarly as we did in Section 2.4. Denote by ν the random outcome of the measurement of E. The random variable ν is Bernoulli distributed with expectation p. The natural unbiased estimate of θ_3 is

$$\hat{\theta_3} = \frac{1}{\lambda_3}\left(2\nu - 1 - \theta_1 \lambda_1 - \theta_2 \lambda_2\right). \tag{24}$$

So the variance of the estimator related to measurement E is

$$\mathrm{Var}\,(\hat{\theta}_3) = 4\frac{\mathrm{Var}\,\nu}{\lambda_3^2} = \frac{1-\langle\lambda,\theta\rangle^2}{\lambda_3^2}\,. \tag{25}$$

Note that if r measurements are performed, then instead of a Bernoulli distribution we will have binomial, so in (24) ν is replaced by the relative frequency of the outcome 1 and the variance (25) should be divided by r. The variance changes by a constant factor, hence the minimization problem is the same.

We want to argue that the optimal estimate corresponds to the projection $\lambda_1 = \lambda_2 = 0,\ \lambda_3 = \pm 1$, i.e., we measure the coefficient of σ_3 the most efficiently if we measure in the direction corresponding to σ_3. This estimator has a variance of $1-\theta_3^2$, however, the inequality

$$\mathrm{Var}\,(\hat{\theta}_3) = \frac{1-\langle\lambda,\theta\rangle^2}{\lambda_3^2} \geq 1-\theta_3^2 \tag{26}$$

is not true in general. For example $\lambda = \theta$ is possible for pure states and then left-hand side is 0. So it is not true that σ_3 is generally the best observable, but as we will show that it is optimal if we take the average of the variance on the unitarily invariant states.

Since $\theta_1,\ \theta_2$ are given, the only unitarily invariant vector to $(\theta_1,\theta_2,\theta_3)$ is $(\theta_1,\theta_2,-\theta_3)$, we should average $\mathrm{Var}\,(\hat{\theta}_3)$ on them:

$$\Big\langle \mathrm{Var}\,(\hat{\theta}_3)\Big\rangle = \frac{1}{2}\cdot\frac{1-(\lambda_1\theta_1+\lambda_2\theta_2+\lambda_3\theta_3)^2}{\lambda_3^2} + \frac{1}{2}\cdot\frac{1-(\lambda_1\theta_1+\lambda_2\theta_2-\lambda_3\theta_3)^2}{\lambda_3^2} =$$

$$= \frac{1-(\lambda_1\theta_1+\lambda_2\theta_2)^2-\lambda_3^2\theta_3^2}{\lambda_3^2}$$

We claim that although (26) is not true in every case, it will be true in the average sense:

$$\Big\langle \mathrm{Var}\,(\hat{\theta}_3)\Big\rangle \geq 1-\theta_3^2. \tag{27}$$

Note that the average of the right-hand side of (26) will remain the same, since $1-(-\theta_3)^2 = 1-\theta_3^2$.

Eq. (27) is equivalent

$$(\lambda_1\theta_1+\lambda_2\theta_2)^2 \leq 1-\lambda_3^2,$$

and this can be easily proved using the Schwartz inequality:

$$(\lambda_1\theta_1 + \lambda_2\theta_2)^2 \leq (\lambda_1^2 + \lambda_2^2)(\theta_1^2 + \theta_2^2) \leq 1 - \lambda_3^2$$

so the inequality (27) holds in every case and the equality is possible only in case $\lambda_3 = \pm 1$.

Theorem 3.1 *Assume that θ_1 and θ_2 are known parameters and $\theta_1^2+\theta_2^2 < 1$. In this setting the average covariance will be minimal if $\lambda_1 = \lambda_2 = 0$ and $\lambda_3 = \pm 1$, i.e., the projection $\frac{I\pm\sigma_3}{2}$ is measured.*

So we have that the measurement $\left(\frac{I+\sigma_3}{2}, \frac{I-\sigma_3}{2}\right)$ is really optimal but only in the average sense.

2. Assume next that θ_1, θ_2 and θ_3 are not known. Let the observables

$$E_x = \frac{1}{2}(I + x \cdot \sigma) \qquad (x = a, b, c)$$

be measured in the true state ρ, where a, b, c are unit vectors in $\mathbb{R}^3$ ($\|a\|_2 = \|b\|_2 = \|c\|_2 = 1$) and $\theta = (\theta_1, \theta_2, \theta_3)$ are the unknown parameters of the state. The probabilities are

$$p_x := \frac{1 + \langle x, \theta\rangle}{2}, \quad p := (p_a, p_b, p_c)\,.$$

If the measurements are performed r times, then p_x is estimated by the relative frequency ν_x of the outcome 1. Just like earlier, the value of r does not contribute to the optimization problem, so in what follows, we will assume that $r = 1$. The estimate $\hat{\theta}$ can be found by solving

$$\nu_x = \frac{1 + \langle x, \hat{\theta}\rangle}{2} \qquad (x = a, b, c).$$

Since it can be written in the form

$$2\begin{bmatrix}\nu_a\\ \nu_b\\ \nu_c\end{bmatrix} = \begin{bmatrix}1\\ 1\\ 1\end{bmatrix} + T\begin{bmatrix}\hat{\theta}_1\\ \hat{\theta}_2\\ \hat{\theta}_3\end{bmatrix}$$

with the matrix

$$T := \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix},$$

the solution is

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{bmatrix} = T^{-1} \begin{bmatrix} 2\nu_a - 1 \\ 2\nu_b - 1 \\ 2\nu_c - 1 \end{bmatrix}. \tag{28}$$

This is once again an unbiased estimator, the relative frequencies (ν_x) are independent and have a Bernoulli distribution, so the covariance matrix of this estimator is

$$V(\theta) = T^{-1} \begin{bmatrix} 1 - \langle a, \theta \rangle^2 & 0 & 0 \\ 0 & 1 - \langle b, \theta \rangle^2 & 0 \\ 0 & 0 & 1 - \langle c, \theta \rangle^2 \end{bmatrix} (T^{-1})^*. \tag{29}$$

We can get the average covariance matrix with the integral

$$\int_H V(\theta)\, d\mu(\theta) =$$

$$= T^{-1} \begin{bmatrix} \int_H 1 - \langle a, \theta \rangle^2 \, d\mu(\theta) & 0 & 0 \\ 0 & \int_H 1 - \langle b, \theta \rangle^2 \, d\mu(\theta) & 0 \\ 0 & 0 & \int_H 1 - \langle c, \theta \rangle^2 \, d\mu(\theta) \end{bmatrix} (T^{-1})^*;$$

where we take the integral on the unitarily invariant states $H = \{(\theta_1, \theta_2, \theta_3) : \theta_1^2 + \theta_2^2 + \theta_3^2 = R^2\}$ with respect to the normalized Lebesgue measure. Due to the symmetry we can obtain that

$$\int_H 1 - \langle a, \theta \rangle^2 d\mu(\theta) = \int_H 1 - \langle b, \theta \rangle^2 d\mu(\theta) = \int_H 1 - \langle c, \theta \rangle^2 d\mu(\theta) = C$$

where $C > 0$ is a constant depending on the radius of the sphere ($R \leq 1$). So the average is

$$\int V(\theta)\, d\mu(\theta) = C\, (T^* T)^{-1}.$$

The determinant is minimal if $\det(T^*T) = (\det T)^2$ is maximal. $\det T$ is the volume of the parallelepiped determined by vectors a, b and c, and it is maximal when they are orthogonal, i.e., E_a, E_b and E_c are quasi-orthogonal.

Theorem 3.2 *Assume that $\theta_1, \theta_2, \theta_3$ are unknown parameters. Then the average quadratic error will be minimal if the three measurements are complementary.*

A possibility is the measurements originated from the three Pauli matrices: $\frac{I+\sigma_i}{2}$, as we can see in Section 2.4.

3. In the next essential example, θ_1, θ_2 are not known, but θ_3 is given. Assume that the observables

$$E_x = \frac{1}{2}(I + x \cdot \sigma) \qquad (x = a, b) \tag{30}$$

are measured in the true state ρ, where a, b are unit vectors in $\mathbb{R}^3$.

In this case the equation to solve has the form

$$\begin{bmatrix} \nu_a \\ \nu_b \end{bmatrix} = \frac{1}{2}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} + U \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} + \begin{bmatrix} a_3\theta_3 \\ b_3\theta_3 \end{bmatrix} \right),$$

where

$$U = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}.$$

The latter equation yields the estimator

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = U^{-1}\left(2\begin{bmatrix} \nu_a \\ \nu_b \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \theta_3 \begin{bmatrix} a_3 \\ b_3 \end{bmatrix} \right).$$

The mean quadratic error matrix is

$$U^{-1} \begin{bmatrix} 1 - \langle a, \theta \rangle^2 & 0 \\ 0 & 1 - \langle b, \theta \rangle^2 \end{bmatrix} (U^{-1})^*. \tag{31}$$

We integrate with respect to the unknown θ_1 and θ_2 only. The average variance is the question on the circle $M = \{(\theta_1, \theta_2) : \theta_1^2 + \theta_2^2 = R^2\}$:

$$\frac{1}{2R\pi} \int_M 1 - \langle \lambda, \theta \rangle^2 \, d\theta = 1 - (\lambda_1^2 + \lambda_2^2)R^2/2 - \lambda_3^2\theta_3^2.$$

According to (31) we have the average covariance matrix

$$U^{-1} \begin{bmatrix} 1 - (a_1^2 + a_2^2)R^2/2 - a_3^2\theta_3^2 & 0 \\ 0 & 1 - (b_1^2 + b_2^2)R^2/2 - b_3^2\theta_3^2 \end{bmatrix} (U^{-1})^*.$$

Minimization of a matrix cannot be performed, so our idea is to minimize the determinant

$$\frac{(1-(a_1^2+a_2^2)R^2/2-a_3^2\theta_3^2)\,(1-(b_1^2+b_2^2)R^2/2-b_3^2\theta_3^2)}{\det U^2} \tag{32}$$

under the conditions $a_1^2+a_2^2+a_3^2=1$ and $b_1^2+b_2^2+b_3^2=1$.

Let us introduce the notations $c=a_1^2+a_2^2$ and $d=b_1^2+b_2^2$. Then we have to minimize

$$\left(1-cR^2/2-(1-c)\theta_3^2\right)\left(1-dR^2/2-(1-d)\theta_3^2\right)\frac{1}{\det U^2}. \tag{33}$$

$\det U^2 \leq cd$, so if c and d are given we should choose a and b such that $(a_1,a_2)\perp(b_1,b_2)$. Then $\det U^2=cd$ and (33) takes the form:

$$\left(\frac{1-\theta_3^2}{c}-(R^2/2-\theta_3^2)\right)\left(\frac{1-\theta_3^2}{d}-(R^2/2-\theta_3^2)\right).$$

On the other hand, we know that $c\leq 1$, $d\leq 1$, and $R^2\leq 1-\theta_3^2$, so both factors are always positive, on the other hand they are monotonous in c and d, hence the optimal case is achieved when $c=d=1$. In the parameterization (30) this means a and b are orthogonal and $a_3=b_3=0$.

Theorem 3.3 *Assume that θ_3 is the unknown parameter and $|\theta_3|\neq 1$. Then the determinant of the average covariance matrix is minimal if the two von Neumann measurements are complementary to each other and to σ_3.*

3.1.2 The n-dimensional case

In this section we generalize Theorem 3.3 to systems which have more than two levels. Assume that we have some known parameters and some parameters to estimate. Let us use the decomposition

$$M_n(\mathbb{C})=\mathbb{C}I\oplus\mathcal{A}\oplus\mathcal{B},$$

where $\mathcal{A}$ and $\mathcal{B}$ are linear subspaces and orthogonality is defined with respect to the Hilbert-Schmidt inner product $\langle A,B\rangle=\operatorname{Tr}A^*B$.

A state has the density matrix

$$\rho = I/n + \rho_{\mathcal{A}} + \rho_{\mathcal{B}}. \tag{34}$$

We assume that the component $\rho_{\mathcal{A}}$ is known and $\rho_{\mathcal{B}}$ should be estimated. Let the dimension of $\mathcal{B}$ be k. The positive contractions $E^1, \dots, E^k$ are used for independent measurements on several identical copies of the n-level system. A measurement corresponds to the POVM $\{E^i, I - E^i\}$. These operators have expansion

$$E^i = e_i I + E^i_{\mathcal{A}} + E^i_{\mathcal{B}} \qquad (1 \le i \le k).$$

The expectations are

$$p_i := \operatorname{Tr} \rho E^i = e_i + \operatorname{Tr} \rho_{\mathcal{A}} E^i_{\mathcal{A}} + \operatorname{Tr} \rho_{\mathcal{B}} E^i_{\mathcal{B}} \qquad (1 \le i \le k).$$

We fix an orthonormal basis $F_1, \dots, F_k$ in $\mathcal{B}$ and the unknown component has the expansion

$$\rho_{\mathcal{B}} = \theta_1 F_1 + \ldots + \theta_k F_k,$$

where $\theta = (\theta_1, \dots, \theta_k)$ are the parameters to be estimated.

Similarly we make the parameterization for the measurements:

$$E^i_{\mathcal{B}} = e_{i1} F_1 + e_{i2} F_2 + \ldots + e_{ik} F_k.$$

Then

$$p_i = e_i + \operatorname{Tr} \rho_{\mathcal{A}} E^i_{\mathcal{A}} + \sum_{j=1}^{k} e_{ij} \theta_j \qquad (1 \le i \le k).$$

From that we can make estimates on $\hat{\theta}_i$ by solving the equations

$$\varepsilon_i = e_i + \operatorname{Tr} \rho_{\mathcal{A}} E^i_{\mathcal{A}} + \sum_{j=1}^{k} e_{ij} \hat{\theta}_j \qquad (1 \le i \le k),$$

where ε_i is the random result of the ith measurement. In another form

$$\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_k \end{bmatrix} = \begin{bmatrix} e_1 \\ \vdots \\ e_k \end{bmatrix} + \begin{bmatrix} \operatorname{Tr} \rho_{\mathcal{A}} E^1_{\mathcal{A}} \\ \vdots \\ \operatorname{Tr} \rho_{\mathcal{A}} E^k_{\mathcal{A}} \end{bmatrix} + \begin{bmatrix} e_{11} & \cdots & e_{1k} \\ \vdots & \ddots & \vdots \\ e_{k1} & \cdots & e_{kk} \end{bmatrix} \hat{\theta}^{\intercal}$$

or in a different notation

$$\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_k \end{bmatrix} = \begin{bmatrix} e_1 + \operatorname{Tr} \rho_{\mathcal{A}} E^1_{\mathcal{A}} \\ \vdots \\ e_k + \operatorname{Tr} \rho_{\mathcal{A}} E^k_{\mathcal{A}} \end{bmatrix} + T\hat{\theta}^{\intercal},$$

where T is the $k \times k$ matrix from the previous formula. Therefore,

$$\hat{\theta}^{\intercal} = T^{-1} \left(\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_k \end{bmatrix} - \begin{bmatrix} e_1 + \operatorname{Tr} \rho_{\mathcal{A}} E^1_{\mathcal{A}} \\ \vdots \\ e_k + \operatorname{Tr} \rho_{\mathcal{A}} E^k_{\mathcal{A}} \end{bmatrix} \right).$$

Since ε_i is an unbiased estimation on p_i, $\hat{\theta}$ is an unbiased estimation on θ, hence the covariance matrix is

$$\operatorname{Var}(\hat{\theta}) = T^{-1} \operatorname{Var}(\varepsilon)(T^{-1})^{\intercal},$$

where we used the notation $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots \varepsilon_k)^{\intercal}$.

Due to the independence of the measurements the expected value $\operatorname{Var}(\varepsilon)$ will be diagonal. The random variables ε_i have a Bernoulli distribution and the variance is

$$(1 - \operatorname{Tr} \rho E^i) \operatorname{Tr} \rho E^i.$$

We want to take the average:

$$\int (1 - \operatorname{Tr} \rho E^i) \operatorname{Tr} \rho E^i \, d\mu(\rho),$$

where the integration is going on the unitarily invariant states, and μ is a the corresponding normalized Haar-measure. For the sake of simplicity assume that the operators E^i have the same spectrum. Then the integral is a constant (c), that does not depend on the actual E_i and so the average covariance matrix is

$$\langle \operatorname{Var}(\hat{\theta}) \rangle = T^{-1} \begin{bmatrix} c & 0 & \cdots & 0 \\ 0 & c & \cdots & 0 \\ \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & c \end{bmatrix} (T^{-1})^{\intercal} = cT^{-1}(T^{-1})^{\intercal}.$$

The determinant is minimal if the determinant of the matrix T is maximal. Geometrically, the determinant is the volume of the parallelepiped determined by the row vectors. To maximize the determinant, the row vectors should be long. This implies that $E^i_{\mathcal{A}} = 0$ in (54), since otherwise we could project E^i on to $\mathbb{C}I \oplus \mathcal{B}$, and still have the same elements in T. So we have

$$E^i = e + e_{i1}F_1 + \ldots + e_{ik}F_k \qquad (1 \leq i \leq k),$$

and then the determinant of the matrix

$$\begin{bmatrix} 1/n & 0 & \cdots & 0 \\ e & e_{11} & \cdots & e_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ e & e_{k1} & \cdots & e_{kk} \end{bmatrix}$$

is $\det T$. The angle of the first row and any other row is fixed. To have a large determinant the rows of T should be orthogonal. In this case, the operators $E^1, \ldots, E^k$ are quasi-orthogonal.

Theorem 3.4 *If the positive contractions $E^1, \ldots, E^k$ have the same spectrum, then the determinant of the average of the quadratic error matrix is minimal if the operators $E^1, \ldots, E^k$ are quasi-orthogonal to each other and to $\mathcal{A}$.*

Example 3.1 *If two qubits are given and the reduced states of both qubits are known, then the ideal state estimation is connected to the observables*

$$\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12}, \sigma_{23}, \sigma_{31}, \sigma_{13}, \sigma_{21}, \sigma_{32},$$

with $\sigma_{ij} = \sigma_i \otimes \sigma_j$. □

3.2 Estimation with a single POVM [45]

If we want to use one positive operator-valued measure instead of multiple von Neumann measurements, we can easily extend the a method detailed previously. In the upcoming part we will use POVMs which have exactly the sufficient number of components to make the state estimation. Of course we can use greater number of components but in that case is not trivial the construction of the estimator.

3.2.1 The qubit case

1. Assume that we have three unknown parameters. They are $\theta = (\theta_1, \theta_2, \theta_3)$ and the positive operator-valued measure $\{E_a, E_b, E_c, E_d\}$ is performed. We write

$$E_x = x_0(I + x \cdot \sigma), \quad (x = a, b, c, d),$$

where $a, b, c, d \in \mathbb{R}^3$ and $a_0, b_0, c_0, d_0 \in \mathbb{R}$. Then from $\sum E_x = I$, we have

$$\begin{aligned} E_d &= I - E_a - E_b - E_c = (1 - a_0 - b_0 - c_0)I - (a_0 a + b_0 b + c_0 c) \cdot \sigma \\ &= d_0(I + d \cdot \sigma). \end{aligned}$$

The positivity conditions for E_a are $a_0 \geq 0$ and $1 \geq \langle a, a \rangle$. Similar conditions should hold for the other operators E_b, E_c, E_d, therefore the conditions of positivity are:

$$a_0 \geq 0, \quad b_0 \geq 0, \quad c_0 \geq 0, \quad d_0 \geq 0, \tag{35}$$

$$1 \geq \langle a, a \rangle, \quad 1 \geq \langle b, b \rangle, \quad 1 \geq \langle c, c \rangle, \quad 1 \geq \langle d, d \rangle. \tag{36}$$

The probabilities of different outcomes are

$$p_x = \operatorname{Tr} E_x \rho = x_0 + x_0 \langle x, \theta \rangle \quad (x = a, b, c).$$

In matrix notation we have

$$\begin{bmatrix} p_a \\ p_b \\ p_c \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \\ c_0 \end{bmatrix} + T \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

with matrix T defined as

$$T := \begin{bmatrix} a_0a_1 & a_0a_2 & a_0a_3 \\ b_0b_1 & b_0b_2 & b_0b_3 \\ c_0c_1 & c_0c_2 & c_0c_3 \end{bmatrix}.$$

If ν_a, ν_b, ν_c are the relative frequencies of the outcomes (of different measurements on identical copies), then the solution of the equation

$$\begin{bmatrix} \nu_a \\ \nu_b \\ \nu_c \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \\ c_0 \end{bmatrix} + T \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{bmatrix}$$

yield the state estimate:

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{bmatrix} = T^{-1} \begin{bmatrix} \nu_a - a_0 \\ \nu_b - b_0 \\ \nu_c - c_0 \end{bmatrix}. \tag{37}$$

The mean quadratic error matrix is

$$V(\theta) = T^{-1}\, W\, (T^{-1})^*$$

where W is the covariance matrix of the random variables ν_a, ν_b, ν_c. Since they have multinomial distribution we have:

$$W = \frac{1}{r} \begin{bmatrix} p_a(1-p_a) & -p_ap_b & -p_ap_c \\ -p_ap_b & p_b(1-p_b) & -p_bp_c \\ -p_ap_c & -p_bp_c & p_c(1-p_c) \end{bmatrix},$$

where r is the number of measurements; we fix $r = 1$. To get the average mean quadratic error matrix we integrate this on the rotation invariant states with respect to the normalized Lebesgue measure μ. We have to calculate two types of integrals

$$\int_H -p_ap_b d\mu(\theta) = -\int_H \Big(a_0 + a_0\langle a, \theta\rangle\Big)\Big(b_0 + b_0\langle b, \theta\rangle\Big)\, d\mu(\theta)$$

$$= -\int_H a_0b_0 d\mu(\theta) - a_0b_0 \int_H \langle a, \theta\rangle d\mu(\theta) - \int_H a_0b_0\langle b, \theta\rangle d\mu(\theta) - a_0b_0 \int_H \langle a, \theta\rangle\langle b, \theta\rangle d\mu(\theta)$$

$$= -a_0 b_0 + 0 + 0 - a_0 b_0 \frac{R^2}{3} \langle a, b \rangle,$$

where $H = \{(\theta_1, \theta_2, \theta_3) : \theta_1^2 + \theta_2^2 + \theta_3^2 = R^2\}$. The middle integrals are zeros because of the symmetry: $\int_H \langle v, \theta \rangle d\mu(\theta) = 0, \ \forall v$. The last integral can be evaluated as

$$\int_H \langle a, \theta \rangle \langle b, \theta \rangle d\mu(\theta) = \int_H \left(\sum_{\ell=1}^{3} a_\ell \theta_\ell \right) \left(\sum_{m=1}^{3} b_m \theta_m \right) d\mu(\theta) = \sum_{\ell=1}^{3} \left(\int_H \theta_\ell^2 d\mu(\theta) \right) a_\ell b_\ell$$

The quantity $\int_H \theta_\ell^2 d\mu(\theta)$ does not depend on ℓ, it is a constant, and

$$\sum_{\ell=1}^{3} \int_H \theta_\ell^2 d\mu(\theta) = \int_H \sum_{\ell=1}^{3} \theta_\ell^2 d\mu(\theta) = \int_H R^2 d\mu(\theta) = R^2.$$

Similarly,

$$\int_H p_a(1-p_a) d\mu(\theta) = \int_H \Big(a_0 + a_0 \langle a, \theta \rangle \Big) \Big(1 - a_0 - a_0 \langle a, \theta \rangle \Big) d\mu(\theta) = a_0(1-a_0) - a_0^2 \, \frac{R^2}{3} \, \langle a, a \rangle.$$

Let us use the notation $\alpha = \frac{R^2}{3}$. The average mean quadratic error matrix is then

$$\int_H V(\theta) \, d\mu(\theta) = T^{-1} W_0 (T^\intercal)^{-1},$$

with

$$W_0 = \begin{bmatrix} a_0 & 0 & 0 \\ 0 & b_0 & 0 \\ 0 & 0 & c_0 \end{bmatrix} - \begin{bmatrix} a_0 \\ b_0 \\ c_0 \end{bmatrix} \begin{bmatrix} a_0 & b_0 & c_0 \end{bmatrix} - \alpha T T^\intercal.$$

The next step is to minimize the determinant

$$\det \left(T^{-1} W_0 (T^\intercal)^{-1} \right) = \frac{\det (W_0)}{\det{}^2 (T)} = \frac{A}{B},$$

where after some straightforward simplifications we arrive at

$$A = \det \begin{bmatrix} a_0^{-1} - (1 + \alpha \langle a, a \rangle) & -(1 + \alpha \langle a, b \rangle) & -(1 + \alpha \langle a, c \rangle) \\ -(1 + \alpha \langle a, b \rangle) & b_0^{-1} - (1 + \alpha \langle b, b \rangle) & -(1 + \alpha \langle b, c \rangle) \\ -(1 + \alpha \langle a, c \rangle) & -(1 + \alpha \langle b, c \rangle) & c_0^{-1} - (1 + \alpha \langle c, c \rangle) \end{bmatrix}$$

and

$$B = \det{}^2 \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} = \det \begin{bmatrix} \langle a, a \rangle & \langle a, b \rangle & \langle a, c \rangle \\ \langle a, b \rangle & \langle b, b \rangle & \langle b, c \rangle \\ \langle a, c \rangle & \langle b, c \rangle & \langle c, c \rangle \end{bmatrix}.$$

The minimizer $(a_{opt}, b_{opt}, c_{opt})$ should be symmetric, hence $a_0 = b_0 = c_0 = d_0 = 1/4$, $\langle a, a\rangle = \langle b, b\rangle = \langle c, c\rangle = z$ and $\langle a, b\rangle = \langle a, c\rangle = \langle b, c\rangle = y$. We suppose that the minimizer does not depend on the radius of the sphere, so we can suppose $R = \alpha = 0$. (For arbitrary α see the n dimensional case.) Then A is a constant and

$$B = (z - y)^2(z + 2y)$$

The positivity conditions from (35) changes to

$$z \leq 1, \quad 3(z + 2y) \leq 1$$

The maximum of B is taken at $z = 1$ and $y = -1/3$. In this situation $2E_i$ is a projection P_i and $\operatorname{Tr} P_i P_j = 1/3$.

Theorem 3.5 *The optimal POVM is described by projections P_i $(1 \leq i \leq 4)$ such that*

$$E_i = P_i/2, \quad \sum_i P_i = 2I, \quad \operatorname{Tr} P_i P_j = 1/3 \quad (i \neq j).$$

This optimal POVM is termed **minimal qubit tomography** [49].

2. Assume that θ_1, θ_3 are not known, but θ_2 is given. This particular situation means that we are interested in real matrices. A POVM of three components $\{E_a, E_b, E_c\}$ is sufficient. We use the parameterization

$$E_x = x_0(I + x \cdot \sigma), \quad (x = a, b, c),$$

where a, b, c are vectors in $\mathbb{R}^3$, $a_0, b_0, c_0 \in \mathbb{R}$ and $\theta = (\theta_1, \theta_2, \theta_3)$ are the parameters of the state. From $\sum E_x = I$, we have $a_0 + b_0 + c_0 = 1$ and $a_0 a + b_0 b + c_0 c = 0$. The positivity conditions are:

$$0 \leq a_0, b_0, c_0 \qquad \|a\|, \|b\|, \|c\| \leq 1.$$

Similarly to the previous situation we have

$$\begin{bmatrix} p(a) \\ p(b) \end{bmatrix} = \theta_2 \begin{bmatrix} a_0(1 - a_2) \\ b_0(1 - b_2) \end{bmatrix} + \begin{bmatrix} a_0 a_1 & a_0 a_3 \\ b_0 b_1 & b_0 b_3 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_3 \end{bmatrix}$$

and define matrix T as

$$T := \begin{bmatrix} a_0a_1 & a_0a_3 \\ b_0b_1 & b_0b_3 \end{bmatrix} = \begin{bmatrix} a_0 & 0 \\ 0 & b_0 \end{bmatrix} \begin{bmatrix} a_1 & a_3 \\ b_1 & b_3 \end{bmatrix}$$

If $\nu(a), \nu(b)$ are the relative frequencies of the outcomes (of different measurements on identical copies), then the solution of equation

$$\begin{bmatrix} \nu(a) \\ \nu(b) \end{bmatrix} = \theta_2 \begin{bmatrix} a_0(1-a_2) \\ b_0(1-b_2) \end{bmatrix} + T \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_3 \end{bmatrix}$$

yields the estimator of $(\theta_1, \theta_3)^\intercal$

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_3 \end{bmatrix} = T^{-1} \begin{bmatrix} \nu(a) - \theta_2 a_0(1-a_2) \\ \nu(b) - \theta_2 b_0(1-b_2) \end{bmatrix}.$$

If we take the covariance matrix of this estimator, all terms are constants except $\hat{\rho}(a)$ and $\hat{\rho}(b)$ which are multinomially distributed, so we get

$$\mathrm{Var}\,(\hat{\theta}) = T^{-1} \begin{bmatrix} p(a)(1-p(a)) & -p(a)p(b) \\ -p(a)p(b) & p(b)(1-p(b)) \end{bmatrix} (T^{-1})^* = T^{-1} W (T^{-1})^*.$$

We integrate this on the domain $H = \{(\theta_1, \theta_3) : \theta_1^2 + \theta_3^2 = R^2 - \theta_2^2\}$, where R is a constant ($|\theta_2| \leq R \leq 1$). T does not depend on θ, so

$$\int_H \mathrm{Var}\,(\hat{\theta}) d\mu(\theta) = T^{-1} \left(\int_H W(\theta) d\mu(\theta) \right) (T^{-1})^* \tag{38}$$

and μ is the normalized Lebesgue measure on H. We calculate $\int_H W(\theta) d\mu(\theta)$ as follows:

$$\int_H -p(a)p(b)\, d\mu(\theta) = \int_H -a_0b_0(1+a_1\theta_1+a_2\theta_2+a_3\theta_3)(1+b_1\theta_1+b_2\theta_2+b_3\theta_3)\, d\mu(\theta)$$

$$= -a_0b_0 \int_H \left[(1+a_2\theta_2) + (a_1\theta_1 + a_3\theta_3)\right] \left[(1+b_2\theta_2) + (b_1\theta_1 + b_3\theta_3)\right]\, d\mu(\theta)$$

We use this separation, because $1 + a_2\theta_2$ is a constant, and $a_1\theta_1 + a_3\theta_3$ has expectation value 0, since by symmetry $\mathbb{E}(\theta_1) = \mathbb{E}(\theta_3) = 0$.

$$\int_H (1+a_2\theta_2)(1+b_2\theta_2) d\mu(\theta) = (1+a_2\theta_2)(1+b_2\theta_2)$$

$$\int_H (1+a_2\theta_2)(b_1\theta_1 + b_3\theta_3) d\mu(\theta) = 0$$

$$\int_H (a_1\theta_1 + a_3\theta_3)(1 + b_2\theta_2)d\mu(\theta) = 0$$

$$\int_H (a_1\theta_1 + a_3\theta_3)(b_1\theta_1 + b_3\theta_3)d\mu(\theta) = \int_H a_1b_1\theta_1^2 + a_3b_3\theta_3^2 d\mu(\theta),$$

because $\mathbb{E}(\theta_1\theta_3) = 0$, finally we can see that

$$\int_H \theta_1^2 d\mu(\theta) = \int_H \theta_3^2 d\mu(\theta) =: \beta \tag{39}$$

where β is a constant from interval $[0, 1 - \theta_2^2]$, so we obtain

$$\int_H -p(a)p(b)d\mu(\theta) = -a_0b_0\left[(1 + a_2\theta_2)(1 + b_2\theta_2) + \beta(a_1b_1 + a_3b_3)\right].$$

Similar calculations give

$$\int_H (1 - p(a))p(a)d\mu(\theta) = a_0(1 + a_2\theta_2) - a_0^2\left[(1 + a_2\theta_2)^2 + \beta(a_1^2 + a_3^2)\right]$$

and

$$\int_H (1 - p(b))p(b)d\mu(\theta) = b_0(1 + b_2\theta_2) - b_0^2\left[(1 + b_2\theta_2)^2 + \beta(b_1^2 + b_3^2)\right].$$

Using the notations

$$q_a = 1 + a_2\theta_2, \quad q_b = 1 + b_2\theta_2, \quad a^* = (a_1, a_3), \quad b^* = (b_1, b_3)$$

we have the integral

$$\int_H W(\theta)d\mu(\theta) = \begin{bmatrix} a_0q_a - a_0^2\left[q_a^2 + \beta\langle a^*, a^*\rangle\right] & -a_0b_0\left[q_aq_b + \beta\langle a^*, b^*\rangle\right] \\ -a_0b_0\left[q_aq_b + \beta\langle a^*, b^*\rangle\right] & b_0q_b - b_0^2\left[q_b^2 + \beta\langle b^*, b^*\rangle\right] \end{bmatrix}$$

$$= \begin{bmatrix} a_0q_a - a_0^2q_a^2 & -a_0b_0q_aq_b \\ -a_0b_0q_aq_b & b_0q_b - b_0^2q_b^2 \end{bmatrix} - \beta \begin{bmatrix} -a_0^2\langle a^*, a^*\rangle & -a_0b_0\langle a^*, b^*\rangle \\ -a_0b_0\langle a^*, b^*\rangle & -b_0^2\langle b^*, b^*\rangle \end{bmatrix}.$$

Dividing the first rows and columns with a_0 and the second ones with b_0, we have

$$\det \int_H W(\theta)d\mu(\theta) = a_0^2b_0^2 \det\left(\begin{bmatrix} q_aa_0^{-1} - q_a^2 & -q_aq_b \\ -q_aq_b & q_bb_0^{-1} - q_b^2 \end{bmatrix} - \beta \begin{bmatrix} \langle a^*, a^*\rangle & \langle a^*, b^*\rangle \\ \langle a^*, b^*\rangle & \langle b^*, b^*\rangle \end{bmatrix}\right).$$

On the other hand,

$$\det T^2 = a_0^2b_0^2 \det{}^2 \begin{bmatrix} a_1 & a_3 \\ b_1 & b_3 \end{bmatrix} = a_0^2b_0^2 \det \begin{bmatrix} \langle a^*, a^*\rangle & \langle a^*, b^*\rangle \\ \langle a^*, b^*\rangle & \langle b^*, b^*\rangle \end{bmatrix}.$$

So we should minimize a determinant of the form A/B, where

$$A = \det V, \quad B = \det \begin{bmatrix} \langle a^*, a^* \rangle & \langle a^*, b^* \rangle \\ \langle a^*, b^* \rangle & \langle b^*, b^* \rangle \end{bmatrix} =: \det C,$$

$$V = \begin{bmatrix} q_a a_0^{-1} - q_a^2 & -q_a q_b \\ -q_a q_b & q_b b_0^{-1} - q_b^2 \end{bmatrix} - \beta \begin{bmatrix} \langle a^*, a^* \rangle & \langle a^*, b^* \rangle \\ \langle a^*, b^* \rangle & \langle b^*, b^* \rangle \end{bmatrix} =: D - \beta \cdot C.$$

For the minimizer (a_{opt}, b_{opt})we can assume some symmetry conditions: $a_0 = b_0 = c_0 = \frac{1}{3}$, and we will minimize A and maximize B independently to obtain the optimum:

$$A = (d_{11} - \beta\ c_{11})(d_{22} - \beta\ c_{22}) - (d_{12} - \beta\ c_{12})^2 \to \min, \tag{40}$$

$$B = c_{11}c_{22} - (c_{12})^2 \to \max. \tag{41}$$

Let us suppose first that a_2 and b_2 are given (then the elements of D are constants) and we want to optimize the other variables.

We know that

$$\langle a + b, a + b \rangle = c_{11} + c_{22} + 2c_{12} + (a_2 + b_2)^2 \leq 1. \tag{42}$$

If $c_{12} \geq 0$, then from (42) we have $c_{11} + c_{22} \leq 1$, hence $B \leq 1/4$. If $c_{12} < 0$, then B is maximal if c_{12} is maximal, from (42) we have an upper bound:

$$c_{12} \leq \frac{1 - (a_2 + b_2)^2 - c_{11} - c_{22}}{2}. \tag{43}$$

Substituting this upper bound in (41) we have to maximize it in c_{11} and c_{22}. Using derivation we can conclude that it is maximal if c_{11} and c_{22} are maximal:

$$c_{11} = a_1^2 + a_3^2 \leq 1 - a_2^2 \quad \text{and} \quad c_{22} = b_1^2 + b_3^2 \leq 1 - b_2^2. \tag{44}$$

Substituting this upper bound in (41) we get

$$B = \frac{3}{4} - a_2^2 - a_2 b_2 - b_2^2\,. \tag{45}$$

which is optimal if $a_2 = b_2 = 0$. Then $B = 3/4$, so it is a global optimum.

Lemma 3.1 *The following inequality is always true:*

$$d_{12} - \beta\ c_{12} \leq 0$$

Proof. We have

$$d_{12} - \beta\ c_{12} = -(1 + \theta_2 a_2)(1 + \theta_2 b_2) - \beta(a_1 b_1 + a_3 b_3).$$

Since $c \leq 1 - \theta_2^2$ and

$$a_1 b_1 + a_3 b_3 = \langle a^*, b^* \rangle \geq -\|a^*\|\|b^*\| \geq -\sqrt{1 - a_2^2}\sqrt{1 - b_2^2},$$

so it is enough to show that

$$(1 + \theta_2 a_2)(1 + \theta_2 b_2) \geq (1 - \theta_2^2)\sqrt{1 - a_2^2}\sqrt{1 - b_2^2}.$$

The right-hand side does not depend on the signs, the left-hand side is minimal if $\theta_2 a_2 \leq 0,\ \theta_2 b_2 \leq 0$, so it suffices to prove for positive a_2, b_2, θ_2 that

$$(1 - \theta_2 a_2)(1 - \theta_2 b_2) \geq (1 - \theta_2^2)\sqrt{1 - a_2^2}\sqrt{1 - b_2^2}.$$

This is true since from the Cauchy-Schwarz inequality it follows that

$$a_2\theta_2 + \sqrt{1 - a_2^2}\sqrt{1 - \theta_2^2} \leq 1 \quad \Longrightarrow \quad 1 - a_2\theta_2 \geq \sqrt{1 - a_2^2}\sqrt{1 - \theta_2^2},$$

and a similar statement is true for b_2. □

Using this lemma, we get that A is minimal if c_{12} is maximal, and from there the solution is almost the same as in the previous case, we only have more complicated calculations. We substitute the upper bound (43) into (40) and we can obtain that it is minimal if c_{11} and c_{22} are maximal. Using the bounds (44) we get for A a function of a_2 and b_2; using differentiation we can obtain that $a_2 = b_2 = 0$ gives the optimal solution here, too.

So in both cases we have equality in (44), so $\langle a^*, a^* \rangle = \langle b^*, b^* \rangle = 1$ and equality in (43), so $\langle a^*, b^* \rangle = -1/2$. Since $\|a\| = \|b\| = 1$,

$$\frac{3}{2}E_a = \frac{1}{2}(I + a \cdot \sigma), \qquad \frac{3}{2}E_b = \frac{1}{2}(I + b \cdot \sigma)$$

are projections. We have $\|c\| = \|a + b\| = 1$ and this implies that $3E_c/2$ is a projection as well.

Theorem 3.6 *The optimal POVM for the unknown θ_1 and θ_3 case can be described by projections P_i, $1 \le i \le 3$:*

$$E_i = \frac{2}{3}P_i, \quad \sum_{i=1}^{3} P_i = \frac{3}{2}I, \quad \mathrm{Tr}\,\sigma_2 P_i = 0, \quad \mathrm{Tr}\,P_i P_j = \frac{1}{4} \quad \textit{for} \quad i \neq j.$$

The optimal POVM is quasi-orthogonal to the subalgebra generated by σ_2, and symmetrical in the other directions. This POVM can be called conditional symmetric informationally complete POVM, the concept is examined in detail in Section 3.3.

3. Assume that only θ_3 is not known, θ_1 and θ_2 are known constants. In this case the POVM has only two components $\{E_a, E_b\}$. Again we use the parameterization

$$E_a = a_0(I + a \cdot \sigma), \quad E_b = I - E_a.$$

The conditions for positivity are

$$0 \le a_0 \le 1, \quad \|a\| \le 1, \quad \|a\| \le \frac{1 - a_0}{a_0} \tag{46}$$

The probability of the first outcome is $p = a_0(1 + \langle a, \theta \rangle)$. From this the estimation for θ_3:

$$\hat{\theta}_3 = \frac{1}{a_0 a_3}(\nu - a_0 - a_0 a_1 \theta_1 - a_0 a_2 \theta_2).$$

The average of the variance for $(\theta_1, \theta_2, \theta_3)$ and $(\theta_1, \theta_2, -\theta_3)$ is

$$\frac{1}{2} \sum_{\{\theta_3, -\theta_3\}} \frac{1}{a_0^2 a_3^2} p(1-p) = \frac{L - L^2}{a_0^2 a_3^2} - \theta_3^2 \tag{47}$$

where $L = a_0(1 + a_1\theta_1 + a_2\theta_2)$. The first term is independent from θ_3, and it is minimal if $a_0 = 1/2$ and $(a_1, a_2, a_3) = (0, 0, 1)$. So we have obtained:

Theorem 3.7 *The average quadratic error is minimal if E_a and E_b are the spectral projections of σ_3.*

We can see that the optimal POVM is the von Neumann measurement discussed in Theorem 3.1, so we do not get any improvement by using POVMs. Note that the optimal von Neumann measurement is again a conditional symmetric informationally complete POVM.

3.2.2 The n-dimensional case

Now assume that we have the multidimensional system like in Section 3.1.2, with k unknown parameters.

Let us have the POVM with elements

$$E_i = e_i(I + f_i \cdot \sigma) \qquad (1 \le i \le k+1),$$

where $e_i \in \mathbb{R}$, $f_i \in \mathbb{R}^{n^2-1}$, $f_i \cdot \sigma = \sum_j f_{ij}\sigma_j$ and $\{\sigma_j : 1 \le j \le n^2-1\}$ are generalized Pauli matrices ($\operatorname{Tr}\sigma_i = 0$, $\operatorname{Tr}\sigma_i\sigma_j = \delta_{i,j}$). The positivity condition for E_i is not known, but we have a necessary condition:

Lemma 3.2 *If $n \times n$ matrix $I + g \cdot \sigma$ is positive, then $\sum_j g_j^2 \le n^2 - n$. If $\sum_j g_j^2 = n^2 - n$, then $I + g \cdot \sigma = nP$ with a projection P of rank 1.*

Proof. $A = g \cdot \sigma$ is self-adjoint, $\operatorname{Tr} A = 0$. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of A. Then

$$\sum_{j=1}^{n^2-1} g_j^2 = \operatorname{Tr} A^2 = \sum_{t=1}^{n} \lambda_t^2.$$

Since $\sum_{t=1}^n \lambda_t = 0$ from $\operatorname{Tr} A = 0$ and $\lambda_t \ge -1$ from $I + A \ge 0$, we have the upper bound. Namely, $\sum_{t=1}^n \lambda_t^2$ is maximal if $\lambda_1, \lambda_2, \ldots, \lambda_n$ is a permutation of the numbers $-1, -1, \ldots, -1, n-1$. In this case $I + A$ has eigenvalues $0, 0, \ldots, 0, n$, so it is a multiple of a projection. □

The probabilities of different outcomes are

$$p_i = \operatorname{Tr} E_i\rho = e_i + e_i\langle f_i, \theta\rangle, \quad (1 \le i \le k).$$

In matrix notation we have

$$\begin{bmatrix} p_1 \\ \vdots \\ p_k \end{bmatrix} = \begin{bmatrix} e_1 \\ \vdots \\ e_k \end{bmatrix} + T \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_k \end{bmatrix}$$

with matrix T defined as

$$T := \begin{bmatrix} e_1 \cdot (f_1)_1 & \ldots & e_1 \cdot (f_1)_k \\ \vdots & \ddots & \vdots \\ e_k \cdot (f_k)_1 & \ldots & e_k \cdot (f_k)_k \end{bmatrix},$$

where $(f_i)_j$ is the j-th component of f_i.

If $\nu_i,\ 1 \leq i \leq k$ are the relative frequencies of the i-th outcomes then we have the estimator

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_k \end{bmatrix} = T^{-1} \begin{bmatrix} \nu_1 - e_1 \\ \vdots \\ \nu_k - e_k \end{bmatrix}. \tag{48}$$

The covariance matrix of this estimator is

$$\mathrm{Var}\,(\hat{\theta}) = T^{-1}\, W\, (T^{-1})^*,$$

where W is the covariance matrix of the random variables ν_i:

$$W = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & \dots & -p_1p_k \\ -p_1p_2 & p_2(1-p_2) & \dots & -p_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -p_1p_k & -p_2p_k & \dots & p_k(1-p_k) \end{bmatrix},$$

The average mean quadratic error matrix is then

$$\langle \mathrm{Var}\,(\hat{\theta}))\rangle = \int_H \mathrm{Var}\,(\hat{\theta})\, d\mu(\theta) = T^{-1} \int_H W d\mu(\theta) (T^{\intercal})^{-1},$$

where we integrate on the unitarily invariant states fulfilling the conditions for known parameters with respect to the normalized Haar-measure μ.

The final step is to minimize the determinant

$$\det \langle \mathrm{Var}\,(\hat{\theta}))\rangle = \frac{\det\left(\int_H W d\mu(\theta)\right)}{\det{}^2(T)} = \frac{A}{B} \to \min. \tag{49}$$

From what we have seen in the qubit case, general calculations seem only feasible in case $\mathcal{A} = \{0\}$. In our further calculations we will restrict ourselves to this case ($k = n^2 - 1$).

Then by using similar arguments as in the qubit case (see the calculations before Theorem 3.5), we get

$$\int_H -p_ip_j d\mu(\theta) = -e_ie_j - e_ie_j\alpha\langle f_i, f_j\rangle,$$

$$\int_H p_i(1-p_i) d\mu(\theta) = e_i(1-e_i) - e_i^2\,\alpha\,\langle f_i, f_i\rangle,$$

where $\alpha > 0$ is the constant depending on the domain of the integration.

After simplifications we arrive at

$$A = \det \begin{bmatrix} e_1^{-1} - 1 - \alpha\langle f_1, f_1\rangle & -1 - \alpha\langle f_1, f_2\rangle & \cdots & -1 - \alpha\langle f_1, f_{n^2-1}\rangle \\ -1 - \alpha\langle f_1, f_2\rangle & e_2^{-1} - 1 - \alpha\langle f_2, f_2\rangle & \cdots & -1 - \alpha\langle f_2, f_{n^2-1}\rangle \\ \vdots & \vdots & & \vdots \\ -1 - \alpha\langle f_1, f_{n^2-1}\rangle & -1 - \alpha\langle f_2, f_{n^2-1}\rangle & \cdots & e_{n^2-1}^{-1} - 1 - \alpha\langle f_{n^2-1}, f_{n^2-1}\rangle \end{bmatrix}$$

and

$$B = \det{}^2 \begin{bmatrix} (f_1)_1 & \cdots & (f_1)_{n^2-1} \\ \vdots & \ddots & \vdots \\ (f_{n^2-1})_1 & \cdots & (f_{n^2-1})_{n^2-1} \end{bmatrix} = \det \begin{bmatrix} \langle f_1, f_1\rangle & \cdots & \langle f_1, f_{n^2-1}\rangle \\ \vdots & \ddots & \vdots \\ \langle f_1, f_{n^2-1}\rangle & \cdots & \langle f_{n^2-1}, f_{n^2-1}\rangle \end{bmatrix}.$$

The minimizer should be symmetric, hence $e_i = 1/n^2$, $\langle f_i, f_i\rangle = x$ and $\langle f_i, f_j\rangle = y$, if $i \neq j$, $i, j \leq n^2 - 1$. Thus, we obtain

$$A = (n^2 - \alpha(x - y))^{n^2-2}(1 - \alpha(x + (n^2 - 2)y))$$

and

$$B = (x - y)^{n^2-2} \cdot (x + (n^2 - 2)y)$$

Therefore, we minimize

$$\frac{A}{B} = \left(\frac{n^2}{x - y} - \alpha\right)^{n^2-2} \left(\frac{1}{x + (n^2 - 2)y} - \alpha\right). \tag{50}$$

We can also calculate the length of $f_{n^2} = -\sum_{i=1}^{n^2-1} f_i$:

$$\langle f_{n^2}, f_{n^2}\rangle = (n^2 - 1) \cdot (x + (n^2 - 2)y) \leq n^2 - n \tag{51}$$

where the latter inequality is the condition for the positivity, see Lemma 3.2. On the other hand, we have condition

$$x \leq n^2 - n \tag{52}$$

If both inequalities were sharp, then (50) would not be minimal, because we could increase both x and y with a sufficiently small ε and then the value

of (50) would be smaller. If the equality holds in (51) then the second term of (50) will be constant, so we want to have as big a difference of x and y as possible, hence x should be maximal (and then y is minimal). If the equality holds in (52), then from (51) we have $y \leq -(n^2-n)/(n^2-1)$, and on this domain (50) has a negative derivative. So the minimum is taken at

$$x = n^2 - n, \quad \text{and} \quad y = -\frac{n^2-n}{n^2-1}.$$

Lemma 3.2 gives

$$E_i = \frac{1}{n^2}(I + f_i \cdot \sigma) = \frac{1}{n} P_i$$

with some projections P_i and

$$\operatorname{Tr} P_i P_j = \frac{1}{n^2}\langle I + f_i\sigma, I + f_j\sigma\rangle = \frac{1}{n} + \frac{1}{n^2}y = \frac{1}{n+1}.$$

Note that we arrived at a system described by Eq. (18), a SIC-POVM. The following statement is obtained.

Theorem 3.8 *If a symmetric informationally complete system exists, then the optimal POVM is described by its projections P_i as $E_i = P_i/n$ ($1 \leq i \leq n^2$).*

For a qubit the existence of the symmetric informationally complete POVM is obvious, there are other known examples in low dimensions. The question of existence is for a general n, however, unknown.

3.3 Conditional SIC-POVMs

In the previous section we obtained a very interesting result in Theorem 3.6: the optimality of conditional SIC-POVMs fulfilling the combination of symmetricalness and complementarity. Interestingly, this was the most difficult case to prove, the total symmetricalness made the proof easier even in the general, n-dimensional case (Theorem 3.8). The simplicity of $M_2(\mathbb{C})$ made the calculations feasible, but the question is whether there are similar objects in higher dimensions, and if they exist, whether they are the optimal measurements in the conditional case.

So we propose in Section 3.3.1 a numerical method which solve the state estimation problem efficiently and lead us to analytic examples in higher dimensions.

Finally, in Section 3.3.2 we solve the conditional case generally by introducing a new quantity for state estimation efficiency.

3.3.1 The numerical approach [46]

In this section, we will show a method for solving the previously described optimization problem numerically, and then we give the optimal POVMs in some higher dimensional settings.

Problem statement

The problem is the same as detailed in Section 3.2.2:

- We have an n-dimensional system, with k unknown parameters.
- We have an estimator for the unknown parameters (48).
- We have a quantity to minimize: the determinant of the average covariance matrix (49), which is a function of POVM $E = (E_1, \ldots, E_{k+1})$:

$$DACM(E) = \det \langle \mathrm{Var}\,(\hat{\theta}))\rangle = \frac{\det\left(\int_H W d\mu(\theta)\right)}{\det{}^2(T)} \to \min.$$

It is the term

$$\int_H W d\mu(\theta) = \int \chi_H(U\rho U^*) W(U\rho U^*)\, d\lambda(U), \tag{53}$$

that is problematic to calculate, where λ is the Haar-measure on the unitaries and χ is the characteristic function $\chi_H(x) = 1$, if $x \in H$ and $\chi_H(x) = 0$, if $x \notin H$).

We can approximate the value of (53) by numerical integration, so in the following part we define an algorithm which solves this optimization problem effectively.

The algorithm

We introduce the method through an example: let us assume that we have a 3-dimensional system, i.e., a qutrit ($n = 3$), and we know the diagonal entries of the density matrix ρ, so $k = 6$.

We parameterize $M_3(\mathbb{C})$ using the Gell-Mann matrices, we use a dense enough grid on the parameter space $\mathbb{R}^8$ and check for each grid-point whether it is an element of H. Actually, the Bloch vector has only 6 parameters since the diagonal entries of the states are known. The actual calculation consists simply of checking for all grid points the positive definiteness of the matrix determined by the actual generalized Bloch-vector. Then we cluster the grid points of H according to their eigenvalues: we partition the interval $[0, 1]$; two states will belong to the same cluster if their eigenvalues are in the same cells. We choose one cluster, this means all the states with the "same" eigenvalues (i.e., achievable states using unitary transformations) and we take the sum of W in these points. Let us note that we do not use a normalized measure (we do not have to use it in (53), either), since it is not necessary: we get the same optimization problem up to a constant factor. Another remark is that if we choose a small cluster, the computation will be less precise than for a large one, but much faster.

The next problem is how to select new POVMs to get better and better estimations. We choose an arbitrary initial point in the interior of the state

space and in each step, we take a new random POVM by perturbing the parameters using a normal distribution with a given variance. This means that for each E_j, $(j = 1, ...6)$, we calculate $\hat{a}^{(j)} = a^{(j)} + \mathcal{N}_8(0, s(t))$, we repeat the random realization of normal vectors for E_1, while $\hat{a}^{(1)}$ will determine a positive matrix, then we continue the realization with E_2, and so on. The variance of the normal distribution $(s(t))$ is decreased in time: first we need a larger variance for faster convergence, but near the boundary of the state space of POVMs we will easily get negative eigenvalues if the disturbance is too high. If we have a new Bloch vector for all the 6 POVM elements, we take all the variation of $a^{(j)}$ and $\hat{a}^{(j)}$ $(\tilde{a}^{(j)} \in \{a^{(j)}, \hat{a}^{(j)}\})$, and we check for all the $2^6 = 64$ cases whether the correlated $E_7 = I - E_1 - E_2 - \ldots - E_6$ will be a physically possible state or not. Then we go through the valid POVMs and we use simulated annealing [33] for this series of POVMs. Let the current best POVM be E and the next in the line to check is $\tilde{E}$, then we change the best POVM to $\tilde{E}$ with probability:

$$\mathbb{P}(E \to \tilde{E}) = \frac{1}{1 + \exp\left(\frac{\log(DACM(\tilde{E})) - \log(DACM(E))}{T}\right)},$$

where T is the so-called temperature. For high temperatures, the probabilities are close to $1/2$ so the optimal POVM can roam freely, but for low temperatures, we change the current best POVM only if the new POVM is really better. This transition probability determines a special kind of Glauber dynamics, so there is a good chance that it will converge to the global optimum. The reason why we use simulated annealing instead of simply selecting the best POVM from the line is because otherwise the algorithm tends to set in one direction and it only converges to the boundary of the state space. The simulated annealing is useful here because it can change this path by overcoming potential barriers. Also we increase the temperature from time to time to help escape from local optima.

Results

The implementation of this algorithm was made with Mathematica [61], the typical running time is a few minutes and we get quite good convergence to the optimal POVM. This algorithm helps us find optimal POVMs in different scenarios, even analytically:

Proposition 3.1 *If $\rho \in M_3(\mathbb{C})$ and we do know the diagonal elements of ρ, then the optimal POVM is*

$$E_1 = \frac{1}{7}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, E_2 = \frac{1}{7}\begin{bmatrix} 1 & \varepsilon^6 & \varepsilon^2 \\ \varepsilon & 1 & \varepsilon^3 \\ \varepsilon^5 & \varepsilon^4 & 1 \end{bmatrix}, E_3 = \frac{1}{7}\begin{bmatrix} 1 & \varepsilon^2 & \varepsilon^3 \\ \varepsilon^5 & 1 & \varepsilon \\ \varepsilon^4 & \varepsilon^6 & 1 \end{bmatrix},$$

$$E_4 = \frac{1}{7}\begin{bmatrix} 1 & \varepsilon^4 & \varepsilon^6 \\ \varepsilon^3 & 1 & \varepsilon^2 \\ \varepsilon & \varepsilon^5 & 1 \end{bmatrix}, E_5 = \overline{E_2},\ E_6 = \overline{E_3},\ E_7 = \overline{E_4},$$

where $\varepsilon = \exp\left(\frac{2\pi i}{7}\right)$.

We can easily check that this POVM fulfills the following conditions in (19) with constants $k = 7$, $\lambda = 7/3$, $\mu = 2/9$.

Proposition 3.2 *If $\rho \in M_4(\mathbb{C})$ and we know the off-diagonal elements of ρ, then the optimal POVM is*

$$E_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},\ E_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$E_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},\ E_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

So the POVM contains the diagonal matrix units., which have the properties (19) too, with constants $k = 4$, $\lambda = 1$, $\mu = 0$.

Proposition 3.3 *If $\rho \in M_4(\mathbb{C})$ and we do not know the parameters related to $\{\sigma_1 \otimes I,\ \sigma_2 \otimes I,\ \sigma_3 \otimes I\}$, i.e., we want to estimate $M_2 \otimes I$, then the optimal POVM is*

$$E_i = F_i \otimes I, \quad i = 1, 2, 3, 4,$$

where F_i-s are the elements of the 2-dimensional SIC-POVM (22).

In this case P_i-s are projections of rank two and (19) holds with $k = 4$, $\lambda = 2$, $\mu = \frac{2}{3}$.

Proposition 3.4 *If $\rho \in M_4(\mathbb{C})$ and we do not know the parameters related to $\{\sigma_1 \otimes I,\ \sigma_2 \otimes I,\ \sigma_3 \otimes I,\ I \otimes \sigma_1,\ I \otimes \sigma_2,\ I \otimes \sigma_3\}$, then the optimal POVM has the following properties:*

- *E_1, E_2, E_3 are in $M_2 \otimes I$ and have the eigenvalues: $(\frac{2}{7}, \frac{2}{7}, 0, 0)$*
- *E_4, E_5, E_6 are in $I \otimes M_2$ and have the eigenvalues: $(\frac{2}{7}, \frac{2}{7}, 0, 0)$*
- *E_7 has eigenvalues $(\frac{2}{7}, \frac{1}{7}, \frac{1}{7}, 0)$*

This does not fulfill the conditions in (19), although one can observe some kind of symmetry.

Let us note that the speed of search for the optimal POVM mainly depends on the number of unknown parameters, so if we do not have many unknown parameters the problem is not much more difficult in 4 dimensions than in the detailed 3-dimensional case.

Another useful remark is that if we are interested in finding conditional SIC-POVMs of rank one, then we can create a much faster algorithm using the condition (19) and using the parameterization of pure states.

3.3.2 The n-dimensional case [47]

In the following we examine the case of $M_n(\mathbb{C})$ and use the decomposition to three orthogonal subspaces:

$$M_n(\mathbb{C}) = A \oplus B \oplus C, \tag{54}$$

where $A := \{\lambda I : \lambda \in \mathbb{C}\}$ is one dimensional. Denote the orthogonal projections to the subspaces A, B, C by $\mathbf{A}, \mathbf{B}, \mathbf{C}$.

Then the density matrix $\rho \in M_n(\mathbb{C})$ has the decomposition

$$\rho = \frac{I}{n} + \mathbf{B}\rho + \mathbf{C}\rho,$$

since $\mathbf{A}\rho = \frac{I}{n}$. Assume that $\mathbf{B}\rho$ is the known traceless part of ρ and $\mathbf{C}\rho$ is the unknown traceless part of ρ. If the dimension of B is m, then the dimension of C, i.e., the number of unknown parameters is $n^2 - m - 1$. For the state estimation we have to use a POVM with at least $N = n^2 - m$ elements. To get a unique solution we will use a POVM with exactly N elements: $\{E_1, E_2, \dots, E_N\}$.

Let us suppose that σ_i is an orthonormal basis of self-adjoint matrices, such that $\sigma_0 \in A$ (then $\sigma_0 = \frac{1}{\sqrt{n}} I$) and for $1 \le i \le n^2 - 1$, σ_i are either in B or in C. We parameterize the quantum state accordingly:

$$\rho = \sum_{i=0}^{n^2-1} \theta_i \sigma_i.$$

Then from $\operatorname{Tr} \rho = 1$ we have $\theta_0 = \frac{1}{\sqrt{n}}$. The positivity condition can not be expressed in general, but a necessary condition for the coefficients can be obtained:

$$\sum_{i=0}^{n^2-1} \theta_i^2 = \operatorname{Tr} \rho^2 \le 1. \tag{55}$$

We use the notation $\rho_* = \rho - \mathbf{B}\rho$. The aim of the state estimation is to cover ρ_*.

If $\{Q_i : 1 \le i \le N\}$ are self-adjoint matrices satisfying the following equation

$$\rho_* = \frac{1}{n} I + \sum_{\sigma_i \in C} \theta_i \sigma_i = \sum_{i=1}^{N} p_i Q_i, \qquad p_i = \operatorname{Tr} \rho E_i,$$

then $\{Q_i : 1 \le i \le N\}$ is a dual frame of $\{E_i : 1 \le i \le N\}$. In this case the state estimation formula can be written as

$$\hat{\rho_*} = \sum_{i=1}^{N} \hat{p}_i Q_i.$$

We define the distance as

$$\|\rho_* - \hat{\rho}_*\|_2^2 = \operatorname{Tr}(\rho_* - \hat{\rho}_*)^2 = \sum_{i,j=1}^{N} (p_i - \hat{p}_i)(p_j - \hat{p}_j)\langle Q_i, Q_j\rangle$$

and its expectation is

$$\begin{aligned}
\mathbb{E}\|\rho_* - \hat{\rho}_*\|_2^2 &= \sum_{i,j=1}^{N} (p_i\delta(i,j) - p_i p_j)\langle Q_i, Q_j\rangle \\
&= \sum_{i=1}^{N} p_i\langle Q_i, Q_i\rangle - \langle \sum_{i=1}^{N} p_i Q_i, \sum_{j=1}^{N} p_j Q_j\rangle \\
&= \sum_{i=1}^{N} p_i\langle Q_i, Q_i\rangle - \operatorname{Tr}(\rho_*)^2.
\end{aligned}$$

We concentrate on the first term which is

$$\sum_{i=1}^{N} (\operatorname{Tr} E_i\rho)\langle Q_i, Q_i\rangle \tag{56}$$

and we take the integral with respect to the Haar measure on the unitaries $\mathrm{U}(n)$.

Note first that for any projection Π of rank 1

$$\int_{\mathrm{U}(n)} U\Pi U^*\, d\mu(U) = c$$

with some constant c. If $\sum_{i=1}^{n} \Pi_i = I$, then

$$nc = \sum_{i=1}^{n} \int_{\mathrm{U}(n)} U\Pi_i U^*\, d\mu(U) = I$$

and we have $c = I/n$. Therefore for $A = \sum_{i=1}^{n} \lambda_i \Pi_i$ we have

$$\int_{\mathrm{U}(n)} UAU^*\, d\mu(U) = \sum_{i=1}^{n} \lambda_i c = \frac{I}{n}\operatorname{Tr} A$$

and application to the integral of (56) gives

$$\int \operatorname{Tr} E_i(U\rho U^*)\, d\mu(U) = \frac{1}{n}\operatorname{Tr} E_i.$$

So we get the following quantity for the error of the state estimation:

$$T := \int \mathbb{E}\left(\|U\rho_* U^* - U\hat{\rho}_* U^*\|_2^2\right) d\mu(U) = \frac{1}{n}\sum_{i=1}^{N}(\operatorname{Tr} E_i)\langle Q_i, Q_i\rangle - \operatorname{Tr}(\rho^*)^2$$

This is the quantity to be minimized. Since the second part is constant, our task is to minimize the first part:

$$\sum_{i=1}^{N}(\operatorname{Tr} E_i)\langle Q_i, Q_i\rangle \to \min. \tag{57}$$

We define the superoperator:

$$\mathbf{F} : M_n \to A \oplus C, \quad \mathbf{F} = \sum_{i=1}^{N} |E_i\rangle\langle E_i|(\operatorname{Tr} E_i)^{-1}.$$

It will have rank N, so if $N < n^2$ the inverse of $\mathbf{F}$ does not exists, but we can use its pseudo-inverse $\mathbf{F}^-$, so that $\mathbf{F}^-|\sigma_i\rangle = 0$, if $\sigma_i \in B$.

R_i is the canonical dual frame of F_i, if

$$|R_i\rangle = \mathbf{F}^-|P_i\rangle, \tag{58}$$

where

$$P_i = (\operatorname{Tr} E_i)^{-1} E_i \tag{59}$$

.

Lemma 3.3 *For a fixed E_i, (57) is minimal if $Q_i = R_i$, i.e., if we use the canonical dual frame.*

Proof. Let us use the notation $W_i = Q_i - R_i$. Then

$$\begin{aligned}
\sum_{i=1}^{N} \operatorname{Tr} E_i|R_i\rangle\langle W_i| &= \sum_{i=1}^{N} \operatorname{Tr} E_i|R_i\rangle\langle Q_i| - \sum_{i=1}^{N} \operatorname{Tr} E_i|R_i\rangle\langle R_i| \\
&= \sum_{i=1}^{N} \operatorname{Tr} E_i\mathbf{F}^-|P_i\rangle\langle Q_i| - \sum_{i=1}^{N} \operatorname{Tr} E_i\mathbf{F}^-|P_i\rangle\langle P_i|\mathbf{F}^- \\
&= \mathbf{F}^- \sum_{i=1}^{N} \operatorname{Tr} E_i|P_i\rangle\langle Q_i| - \mathbf{F}^-\Big(\sum_{i=1}^{N} \operatorname{Tr} E_i|P_i\rangle\langle P_i|\Big)\mathbf{F}^-
\end{aligned}$$

$$= \quad \mathbf{F}^{-}\mathbf{\Pi} - \mathbf{F}^{-}\mathbf{F}\mathbf{F}^{-} = \mathbf{F}^{-}\mathbf{\Pi} - \mathbf{F}^{-}\mathbf{\Pi} = 0, \tag{60}$$

where $\mathbf{\Pi} = \mathbf{A} + \mathbf{C}$, and we use that from

$$|\rho_*\rangle = \sum_{i=1}^{N} \langle E_i||\rho\rangle |Q_i\rangle$$

it follows that

$$\mathbf{\Pi} = \sum_{i=1}^{N} |Q_i\rangle\langle E_i|.$$

So we have

$$\begin{aligned}
\sum_{i=1}^{N} \mathrm{Tr}\, E_i\langle Q_i, Q_i\rangle &= \sum_{i=1}^{N} \mathrm{Tr}\, E_i\langle W_i, W_i\rangle + \sum_{i=1}^{N} \mathrm{Tr}\, E_i\langle W_i, R_i\rangle \\
&\quad + \sum_{i=1}^{N} \mathrm{Tr}\, E_i\langle R_i, W_i\rangle + \sum_{i=1}^{N} \mathrm{Tr}\, E_i\langle R_i, R_i\rangle \\
&= \sum_{i=1}^{N} \mathrm{Tr}\, E_i\langle W_i, W_i\rangle + \sum_{i=1}^{N} \mathrm{Tr}\, E_i\langle R_i, R_i\rangle \\
&\geq \sum_{i=1}^{N} \mathrm{Tr}\, E_i\langle R_i, R_i\rangle.
\end{aligned}$$

□

From this lemma we know the optimal dual frame for a fixed POVM E_i, and the following lemma provides a property for the optimal POVM:

Lemma 3.4 *The quantity in (57) is minimal if*

$$\mathbf{F} = \mathbf{A} + \frac{n-1}{N-1}\mathbf{C}.$$

Proof. From (60) we have

$$\sum_{i=1}^{N} (\mathrm{Tr}\, E_i)|R_i\rangle\langle R_i| = \mathbf{F}^{-}\mathbf{\Pi} = \mathbf{F}^{-},$$

so we have the equation:

$$\sum_{i=1}^{N} (\mathrm{Tr}\, E_i)\langle R_i, R_i\rangle = \mathrm{Tr}\,(\mathbf{F}^{-}).$$

Let $\nu_1, \nu_2, \ldots, \nu_{n^2}$ be the eigenvalues of $\mathbf{F}$. Since the rank of $\mathbf{F}$ is N, we have $\nu_i = 0$ for $i > N$. We want to minimize

$$\mathrm{Tr}\,(\mathbf{F}^{-}) = \sum_{i=1}^{N} \frac{1}{\nu_i}. \tag{61}$$

It is easy to check that $\mathbf{A}$ is an eigenfunction of $\mathbf{F}$ with $\nu_1 = 1$ eigenvalue:

$$\mathbf{F}|I\rangle = \sum_{i=1}^{N} (\mathrm{Tr}\, E_i)|P_i\rangle\langle P_i, I\rangle = \sum_{i=1}^{N} (\mathrm{Tr}\, E_i)|P_i\rangle = \sum_{i=1}^{N} |E_i\rangle = |I\rangle$$

and we have the following condition:

$$\sum_{i=1}^{N} \nu_i = \mathrm{Tr}\,\mathbf{F} = \sum_{i=1}^{N} \langle P_i, P_i\rangle \mathrm{Tr}\, F_i \leq \sum_{i=1}^{N} \mathrm{Tr}\, F_i = \mathrm{Tr}\, I = n. \tag{62}$$

Combining these conditions we get that the measurement is optimal if $\nu_2 = \nu_3 = \ldots = \nu_N = \frac{n-1}{N-1}$. □

An important observation is, that to achieve the minimum in Lemma 3.4, the inequality (62) must hold as an equality. That means that in the optimal case we have

$$\langle P_i, P_i\rangle = 1. \tag{63}$$

From the definition of P_i (59) we know that $P_i \geq 0$ and $\mathrm{Tr}\, P_i = 1$. Adding (63) to this condition we can conclude that in the optimal case P_i is a rank-one projection ($\forall i : 1 \leq i \leq N$). From (61) we can see that the minimal value of (57) is $1 + \frac{(N-1)^2}{n-1}$. The only question that remains unanswered is whether we can achieve this lower bound and when.

Let us use the notation $\lambda_i = \mathrm{Tr}\, F_i$, then from Lemma 3.4 we obtain the following equation for the optimal POVM:

$$\mathbf{F} = \sum_{i=1}^{N} \lambda_i |P_i\rangle\langle P_i| = \mathbf{A} + \frac{n-1}{N-1}\mathbf{C}.$$

From that we can obtain

$$\sum_{i=1}^{N} \lambda_i \langle Q|P_i\rangle\langle P_i|Q\rangle = \langle Q|\mathbf{A} + \frac{n-1}{N-1}\mathbf{C}|Q\rangle, \tag{64}$$

with $Q := P_k - d \cdot I$. Since $\langle P_i|Q\rangle = \operatorname{Tr} P_i P_k - d$, the left-hand side of (64) becomes

$$\sum_{i=1}^{N} \lambda_i \langle Q|P_i\rangle\langle P_i|Q\rangle = \lambda_k (1-d)^2 + \sum_{i\neq k} \lambda_i (\operatorname{Tr} P_i P_k - d)^2.$$

We can compute the right-hand side as well:

$$\mathbf{A}(P_k - dI) = \mathbf{A}P_k - dI = \mathbf{A}(P_k - I/n) + I/n - dI = I(1/n - d),$$

$$\langle Q|\mathbf{A}|Q\rangle = (1/n - d)\operatorname{Tr}(P_k - dI) = n(1/n - d)^2$$

When $P_k = \sum_{i=0}^{N} c_i \sigma_i$, then

$$\mathbf{C}|Q\rangle = \sum_{\sigma_i \in C} c_i \sigma_i, \qquad \langle Q|\mathbf{C}|Q\rangle = \sum_{\sigma_i \in C} c_i^2.$$

So (64) becomes

$$\lambda_k (1-d)^2 + \sum_{i\neq k} \lambda_i (\operatorname{Tr} P_i P_k - d)^2 = n(1/n - d)^2 + \frac{n-1}{N-1} \sum_{\sigma_i \in C} c_i^2. \tag{65}$$

From (55) we have

$$\sum_{\sigma_i \in C} c_i^2 \leq 1 - c_0^2 = 1 - 1/n. \tag{66}$$

This implies

$$\lambda_k (1-d)^2 \leq n(1/n - d)^2 + \frac{n-1}{N-1}(1 - 1/n),$$

which is true for every value of d, so

$$\lambda_k \leq \min_d \frac{n(1/n - d)^2 + \frac{n-1}{N-1}(1 - 1/n)}{(1-d)^2}$$

By differentiating we can obtain that the right hand side is minimal if

$$d = \frac{N-n}{n(N-1)}$$

and then we get

$$\lambda_k \leq \frac{n}{N}.$$

Since $\sum_{i=k}^{N} \lambda_k = n$, we have $\lambda_1 = \lambda_2 = \ldots = \lambda_N = n/N$.

From that follows that there is an equality in (66), too. So we have

$$\sum_{\sigma_i \in C} c_i^2 = 1 - c_0^2 \quad \Rightarrow \quad c_i = 0, \text{ if } \sigma_i \in B \quad \Rightarrow \quad \operatorname{Tr} \sigma_i P_k = 0, \text{ if } \sigma_i \in B.$$

On the other hand from (65) we have

$$\sum_{i \neq k} \frac{n}{N} \Big(\operatorname{Tr} P_i P_k - \frac{N-n}{n(N-1)} \Big)^2 = 0.$$

Thus,

$$\operatorname{Tr} P_i P_k = \frac{N-n}{n(N-1)} \quad \text{if} \quad i \neq k.$$

So we can obtain the following result:

Theorem 3.9 *In the conditional case, the elements of optimal POVM can be described as multiples of rank-one projections satisfying the following properties* $(1 \leq i, j \leq N)$*:*

$$E_i = \frac{n}{N} P_i, \quad \operatorname{Tr} P_i P_j = \frac{N-n}{n(N-1)} \quad (i \neq j), \quad \operatorname{Tr} \sigma_l P_i = 0 \quad (\forall l : \sigma_l \in B).$$

So we proved that the conditional SIC-POVMs are the optimal (if they exist), and using the notations of (19), we get the constants: $k = N$, $\lambda = \frac{N}{n}$, $\mu = \frac{N-n}{n(N-1)}$. Now we present some examples related to the previous theorem with different N values.

Corollary 3.1 *If we do not have any a priori information about the state* $(m = 0, N = n^2)$*, then*

$$\operatorname{Tr} P_i P_j = \frac{1}{n+1} \quad (i \neq j)$$

so the optimal POVM is the well-known SIC-POVM (if it exists).

Corollary 3.2 *If we know the off-diagonal elements of the state and we want to estimate the diagonal entries* $(m = n^2 - n, N = n)$*, then from Theorem 3.9 it follows that the optimal POVM has the properties*

$$\operatorname{Tr} P_i P_j = 0 \quad (i \neq j), \quad \sum_{i=1}^{n} P_i = I, \quad \textit{and} \quad P_i \textit{ is diagonal.}$$

So the diagonal matrix units form an optimal POVM.

Corollary 3.3 *If we know the diagonal elements of the state and we want to estimate the off-diagonal entries ($m = n - 1, N = n^2 - n + 1$), then from Theorem 3.9 it follows that the optimal POVM has the properties*

$$\operatorname{Tr} P_i P_j = \frac{n-1}{n^2} \quad (i \neq j), \quad \sum_{i=1}^{n} P_i = \frac{n^2 - n + 1}{n} I$$

and P_i has a constant diagonal. A 3-dimensional example for this case is shown in Proposition 3.1

4 Conclusions and discussion

A density matrix $\rho \in M_n(\mathbb{C})$ has $n^2 - 1$ real parameters. If the parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ are not known, but the others are known, we can use this information to estimate $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$. We also need an error function $f(\theta, \hat{\theta})$ to minimize. In our work, we mostly concentrate on minimizing the determinant of the average covariance matrix

$$f(\theta, \hat{\theta}) = \det \langle \mathrm{Var}\,(\hat{\theta}) \rangle \to \min\,. \tag{67}$$

In Section 3.1 the results about multiple von Neumann measurements are discussed. The thorough examination of the different 2-dimensional settings shows the origin and motivation of the usage of (67) for state estimation problems. We suppose that we have exactly the sufficient number of POVMs with two elements: $\{F_1, I - F_1\}$, $\{F_2, I - F_2\}$, $\ldots$, $\{F_k, I - F_k\}$. Theorem 3.4 summarizes the results: the optimal measurements are von Neumann measurements, they are complementary to each other and to the subspace generated by the known parameters.

This is not a surprising result, since Wooters and Fields showed that it is the observable set consisting of orthogonal observables that results in the most information [62], so it is a common anticipation that the complementary (quasi-orthogonal) von Neumann measurements are optimal. The main novelty here is the case when some parameters are known. Using this information less measurements are needed, and we also obtain the quasi-orthogonality of measurement for the known parameters. After we published our work online, similar idea appeared also in [60]. They minimize a different quantity, but get the same phenomenon, proving it is independent from the actual scenario.

In Section 3.2 we cover the case of a single positive operator valued measurement $\{E_1, E_2, \ldots, E_k\}$. We optimize the quantity introduced in the previous section, but here we get calculations that are much more difficult than in the qubit case. This suggests that the main result is not as general as

in the von Neumann case. In n dimensions, analytic results could only be achieved for the case when neither element of the density is assumed to be known. In that setting the symmetrical informationally complete POVMs (SIC-POVMs) are optimal (Theorem 3.8). That is exactly the result that we expected, but it proves the generality of the method used (67), since it yields strong results in very different settings.

We also have a very interesting result in Theorem 3.6: if some parameters are known and some parameters unknown the optimal POVM is symmetrical and complementary to the subspace of the known parameters. This theorem is a combination of Theorems 3.4 and 3.8. We can think of the optimal measurement as a generalization of SIC-POVMs, hence we named them conditional SIC-POVMs.

Sections 3.3 investigates the properties of this new concept. There are basically two different approaches, which provide different kinds of results. First we present a numerical algorithm to optimize (67), which gives us the optimal POVM in different scenarios. Proposition 3.2 provides a trivial example of conditional SIC-POVMs, namely the diagonal matrix units. This result can be extended to any dimension (see Corollary 3.2), providing a simple example for existence. Proposition 3.3 is an example for conditional SIC-POVMs that contain projections of rank 2. So it is not necessary to restrict our investigation to rank-one projections. On the one hand this is a good result, since the set of rank-one projection is very small; on the other hand, it is inconvenient, since rank-one projections are easier to handle. There is a conjecture that a SIC-POVM exists in every dimension. However, in the conditional case the situation is more complex, a conditional SIC-POVM does not necessarily exist (Proposition 3.4). Proposition 3.1 provides the first non-trivial example of a conditional SIC-POVM. Thus the determinant of the average covariance matrix (67), is not only suited to provide analytical results in many cases, but we can also use it efficiently in more difficult non-symmetric settings for numerical optimization.

In the second part we obtain the optimality of conditional SIC-POVMs

analytically. But instead of using (67) we minimize the square of the Hilbert-Schmidt distance. This quantity was used in [57] to obtain the optimality of SIC-POVMs, we generalized their method to the conditional case. Theorem 3.9 states that the optimal POVM

- contains rank-one projections,
- fulfills the symmetry conditions (19) with constants depending only on the dimensionality of the state and on the number of unknown parameters,
- is complementary to the subspace of known parameters.

This result is quite strong, but has some limitations, too. For example, it does not ensure the existence of such a POVM. From Corollary 3.1 we can see that the SIC-POVM is a special case of conditional SIC-POVMs. The existence of SIC-POVMs is not known in general, so the general existence of conditional SIC-POVMs is a problem at least as difficult. Moreover, Proposition 3.3 and 3.4 show examples when Theorem 3.9 is not applicable. But it also proves the optimality of existing conditional SIC-POVMs, see Corollary 3.2 and 3.3. The latter scenario is the generalization of Proposition 3.1. We have found a way to construct such POVMs for infinitely many dimensions [47], i.e., a whole class of non-trivial conditional SIC-POVMs. Let us also note that minimizing the average squared Hilbert-Schmidt distance (57) gives the same result as the determinant of the average covariance matrix (67) in all POVM cases examined. So it seems that in the conditional case the conditional SIC-POVM is the optimal one (if it exists) independently from the particular minimization problem, so it might play an important role in many quantum estimation problems.

5 References

[1] L. Accardi, *Some trends and problems in quantum probability*, in Quantum probability and applications to the quantum theory of irreversible processes, eds. L. Accardi, A. Frigerio and V. Gorini, Lecture Notes in Math. **1055**, 1–19. Springer, 1984.

[2] D. M. Appleby, *SIC-POVMs and MUBs: Geometrical relationships in prime dimension*, Foundations of probability and physics **5**, 223–232, AIP Conf. Proc., 1101, Amer. Inst. Phys., New York, 2009.

[3] D. M. Appleby, A. Ericsson and C. A. Fuchs, *Properties of qbist state spaces*, Foundations of Physics **41**, 564-579, 2011.

[4] E. Bagan, M.A. Ballester, R.D. Gill, A. Monras and R. Munoz-Tapia, *Optimal full estimation of qubit mixed states*, Phys. Rev. A, 73, 032301, 2006.

[5] T. Baier and D. Petz, *Complementarity and state estimation*, Rep. Math. Phys., **65**, 203–214, 2010.

[6] V.P. Belavkin, *Towards the theory of control in observable quantum systems*, Automatica and Remote Control, **44**, 178–188, 1983.

[7] A. Bendersky, F. Patawski, and J.P. Paz, *Selective and efficient estimation of parameters for quantum process tomography*, Phys. Rev. Lett., **100**, 190403, 2008.

[8] L. Bouten, R. van Handel and M. R. James, *An Introduction to Quantum Filtering* SIAM J. Control Optim., **46**, 2199-2241, 2007.

[9] L. Bouten, R. van Handel and M. R. James, *A Discrete Invitation to Quantum Filtering and Feedback Control* SIAM Review, **51**, pp. 239-316, 2009.

[10] A. M. Branczyk, P. E. M. F. Mendonca, A. Gilchrist, A. C. Doherty, and S. D. Bartlett, *Quantum Control of a Single Qubit*, Phys. Rev. A, **75**, 012329, 2007.

[11] S. Colin, J. Corbett, T. Durt and D. Gross, *About SIC POVMs and discrete Wigner Distributions*, J. Opt. B: Quantum Semiclass. Opt. **7** S778, 2005.

[12] M. Cramer, M. B. Plenio, S. T. Flammia, D. Gross, S. D. Bartlett, R. Somma, O. Landon-Cardinal, Y.-K. Liu, D. Poulin, *Efficient quantum state tomography*, Nat. Commun., **1 (9)** 149, 2010.

[13] G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi, *Quantum tomography*, Advances in Imaging and Electron Physics, **128**, 205-308, 2003.

[14] G. M. D'Ariano, M. F. Sacchi, J. Kahn, *Minimax quantum state discrimination*, Phys. Rev. A, **72**, 032310, 2005.

[15] G.M. D'Ariano and H.P. Yuen, *Impossibility of measuring the wave function of a single quantum system*, Physical Review Letters, **76**, 2832–2835, 1996.

[16] R. Demkowicz-Dobrzanski, *Beyond quantum Fisher information: optimal phase estimation with arbitrary a priori knowledge*, Phys. Rev. A, **83**, 061802, 2011.

[17] D.G. Fisher and M. Freyberger, *Estimating mixed quantum states*, Physics Letters A, **273**, 293-302, 2000.

[18] S. T. Flammia, *On SIC-POVMs in prime dimensions*, J. Phys. A: Math. Gen. **39** 13483, 2006.

[19] A. Fujiwara and H. Imai, *Quantum parameter estimation of a generalized Pauli channel*, Journal of Physics A: Mathematical and General, **36** 8093–8103, 2003.

[20] K.M. Hangos, L. Ruppert: *State estimation methods using indirect measurements*, Quantum Probability and Related Topics, World Scientific, p. 163-180, 2011

[21] M. Hayashi and K. Matsumoto, *Asymptotic performance of optimal state estimation in quantum two level system*, J. Math. Phys. **49**, 102101, 2008.

[22] C. W. Helstrøm, *Quantum decision and estimation theory*, Academic Press, New York, 1976.

[23] C. Hill and J. Ralph., *Weak measurement and rapid state reduction in bipartite quantum systems*, New J. Phys., **9**, 151, 2007.

[24] A. S. Holevo, *Probabilistic and statistical aspects of quantum theory*, North-Holland, 1982.

[25] K. Jacobs and D. A. Steck, *A straightforward introduction to continuous quantum measurement*, Contemporary Physics **47**, 279, 2006.

[26] K. Jacobs, *Feedback Control Using Only Quantum Back-Action*, New J. Phys. **12**, 043005, 2010.

[27] A.N. Jordan, B. Trauzettel, and G. Burkard, *Weak measurement of quantum dot spin qubits*, Physical Review B, **76**, 155324, 2007.

[28] Y.-S. Kim, Y.-W. Cho, Y.-S. Ra and Y.-H. Kim, *Reversing the weak quantum measurement for a photonic qubit*, Opt. Express, **17**, 11978, 2009.

[29] A.N. Korotkov and A.N. Jordan. *Undoing a weak quantum measurement of a solid-state qubit*, Physical Review Letters, **97**, 166805, 2006.

[30] R. L. Kosut, I. Walmsley, H. Rabitz, *Optimal experiment design for quantum state and process tomography and Hamiltonian parameter estimation*, `arXiv, quant-ph/0411093`, 2004.

[31] K. Kraus, *Complementary observables and uncertainty relations*, Phys. Rev. D (3) **35**, 3070–3075, 1987.

[32] Z. E. D. Medendorp, F. A. Torres-Ruiz, L. K. Shalm, G. N. M. Tabia, C. A. Fuchs, A. M. Steinberg, *Experimental characterization of qutrits using SIC-POVMs*, Phys. Rev. A **83**, 051801, 2011.

[33] N. Metropolis, A. W. Rosenbluth, M.N. Rosenbluth, A. H. Teller and E. Teller, *Equation of State Calculations by Fast Computing Machines*, The Journal of Chemical Physics **21** (6), 1087, 1953.

[34] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, *Quantum process tomography: Resource analysis of different strategies*, Physical Review A, **77**, 032322, 2008.

[35] M. Nathanson and M. B. Ruskai, *Pauli diagonal channels constant on axes*, Journal of Physics A Mathematical General, **40**, 8171–8204, 2007.

[36] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2000.

[37] J. Nunn, B. J. Smith, G. Puentes, I. A. Walmsley, and J. S. Lundeen. *Optimal experiment design for quantum state tomography: Fair, precise, and minimal tomography*, Physical Review A, **81**, 042109, 2010.

[38] H. Ohno, D. Petz and A. Szántó, *Quasi-orthogonal subalgebras of* 4×4 *matrices*, Linear Alg. Appl. **425**, 109–118, 2007.

[39] M. Paris and J. Rehacek, *Quantum State Estimation*, Springer-Verlag, Berlin, 2004.

[40] D. Petz, *Algebraic complementarity in quantum theory*, J. Math. Phys. **51**, 015215, 2010.

[41] D. Petz, *Complementarity in quantum systems*, Rep. Math. Phys. **59** 209–224, 2007.

[42] D. Petz, *Quantum Information Theory and Quantum Statistics*, Springer-Verlag, Heidelberg, 2008.

[43] D. Petz, K.M. Hangos, A. Magyar, *Point estimation of states of finite quantum systems*, J. Phys. A: Math. Theor., **40**, 7955-7969, 2007.

[44] D. Petz, K.M. Hangos and L. Ruppert, *Quantum state tomography with finite sample size*, in Quantum Bio-Informatics, eds. L. Accardi, W. Freudenberg, M. Ohya, World Scientific, pp. 247–257, 2008.

[45] D. Petz and L. Ruppert, *Efficient quantum tomography needs complementary and symmetric measurements*, accepted for publication in Rep. Math. Phys., `http://arxiv.org/abs/1011.5210`

[46] D. Petz and L. Ruppert, *Optimal quantum state tomography with known parameters*, Journal of Physics A: Math. Theor. **45**, 085306, 2012.

[47] D. Petz, L. Ruppert and A. Szántó, *Conditional SIC-POVMs*, to be published, `http://arxiv.org/abs/1202.5741`

[48] D. Petz, A. Szántó and M. Weiner, *Complementarity and the algebraic structure of 4-level quantum systems*, J. Infin. Dim. Analysis Quantum Prob. **12**, 99–116, 2009.

[49] J. Rehacek, B-G. Englert and D. Kaszlikowski, *Minimal qubit tomography*, Physical Review A, **70**, 052321, 2004.

[50] M. Renes, R. Blume-Kohout, A. J. Scott and C. M. Caves, *Symmetric informationally complete quantum measurements*, J. Math. Phys. **45**, 2171, 2004.

[51] R. Romano and D. D'Alessandro, *Incoherent control and entanglement for two-dimensional coupled systems*, Phys. Rev. A **73**, 022323, 2006.

[52] L. Ruppert, A. Magyar, K.M. Hangos, *Compromising non-demolition and information gaining for qubit state estimation*, Quantum Probability and Related Topics, World Scientific, p. 212-224, 2008.

[53] L. Ruppert, K.M. Hangos: *Martingale approach in quantum state estimation using indirect measurements*, Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems, p. 2049-2054, 2010.

[54] L. Ruppert, D. Virosztek, K.M. Hangos *Optimal parameter estimation of Pauli channels*, accepted for publication in Journal of Physics A: Math. Theor.

[55] M. B. Ruskai, *Some connections between frames, mutually unbiased bases, and POVM's in quantum information theory*, Acta Appl. Math. **108**, 709–719, 2009.

[56] M. Sasaki, M. Ban, and S. M. Barnett, *Optimal parameter estimation of a depolarizing channel*, Phys. Rev. A, **66(2)**, 022308, 2002.

[57] A. J. Scott, *Tight informationally complete quantum measurements*, J. Phys. A **39**, 13507, 2006.

[58] A. J. Scott and M. Grassl, *SIC-POVMs: A new computer study*, J. Math. Phys. **51**, 042203, 2010.

[59] G.A. Smith, A. Silberfarb, I.H. Deutsch, and P.S. Jessen, *Efficient quantum state estimation by continuous weak measurement and dynamical control*, Phys. Rev. Lett., **97**, 180403, 2006.

[60] Y. S. Teo, H. Zhu, B.-G. Englert, J. Rehacek, Z. Hradil, *Knowledge and ignorance in incomplete quantum state tomography*, Phys. Rev. Lett. **107**, 020404, 2011.

[61] Inc. Wolfram Research. *Mathematica Edition: Version 8.0.* Wolfram Research, Inc., Champaign, Illinois, 2010.

[62] W.K. Wooters and B.D. Fields, *Optimal state determination by mutually unbiased measurements*, Annals of Physics, **191**, 363–381, 1989.

[63] K. C. Young, M. Sarovar, R. Kosut, and K. B. Whaley, *Optimal quantum multiparameter estimation and application to dipole- and exchange-coupled qubits*, Phys. Rev. A, **79(6)**, 062301, 2009.

[64] G. Zauner, *Quantendesigns - Grundzüge einer nichtkommutativen Designtheorie*, PhD thesis (University of Vienna), 1999.

[65] H. Zhu, *SIC POVMs and Clifford groups in prime dimensions*, J. Phys. A: Math. Theor. **43** 305305, 2010.