Quantum Chi-Squared and Goodness of Fit Testing

Kristan Temme\textsuperscript{1} and Frank Verstraete\textsuperscript{1}

\textsuperscript{1}Fakultät für Physik, Universität Wien, Boltzmannsasse 5, 1090 Wien, Austria

The density matrix in quantum mechanics parameterizes the statistical properties of the system under observation, just like a classical probability distribution does for classical systems. The expectation value of observables cannot be measured directly, it can only be approximated by applying classical statistical methods to the frequencies by which certain measurement outcomes (clicks) are obtained. In this paper, we make a detailed study of the statistical fluctuations obtained during an experiment in which a hypothesis is tested, i.e. the hypothesis that a certain setup produces a given quantum state. Although the classical and quantum problem are very much related to each other, the quantum problem is much richer due to the additional optimization over the measurement basis. Just as in the case of classical hypothesis testing, the confidence in quantum hypothesis testing scales exponentially in the number of copies. In this paper, we will argue 1) that the physically relevant data of quantum experiments is only contained in the frequencies of the measurement outcomes, and that the statistical fluctuations of the experiment are essential, so that the correct formulation of the conclusions of a quantum experiment should be given in terms of hypothesis tests, 2) that the (classical) $\chi^2$ test for distinguishing two quantum states gives rise to the quantum $\chi^2$ divergence when optimized over the measurement basis, 3) present a max-min characterization for the optimal measurement basis for quantum goodness of fit testing, find the quantum measurement which leads both to the maximal Pitman and Bahadur efficiency, and determine the associated divergence rates.

PACS numbers:

The problem of quantum measurement has received a wide-ranging surge of interest because of ground-breaking experiments in quantum information processing \cite{1,2}. A fundamental feature of quantum measurements is the peculiar interplay between the quantum and classical world: a quantum measurement gives rise to “classical clicks”, i.e. individual samples, and the only information that can be obtained when observing a quantum system is contained in the frequencies of the possible measurement outcomes. Let us consider a quantum experiment in which we receive a large but finite amount of identical copies of the state $\sigma$. As the number of measurements that can be done is obviously bounded, there is no way by which two quantum systems whose density matrices are very close to each other can be distinguished exactly. In other words, it is fundamentally impossible to certify that a given system is in a particular quantum state $\sigma$: the only thing we can aim for is to certify that all the data collected in the experiment is compatible with the hypothesis that we sampled from the state $\sigma$.

Exactly the same problem is present in classical statistics \cite{7}: it is impossible to certify that one is sampling from a given distribution, but one can only gain confidence that the samples are compatible or not with the fact that they are taken from a given distribution. Formally, the only thing achievable in a classical statistical experiment is to accept or reject a hypothesis. In the given setting, we take as the null hypothesis the fact that the distribution that we are sampling from has certain features, and we want to check whether the obtained data are compatible with this hypothesis. In practice, this means that a confidence interval has to be defined in which the hypothesis is accepted or rejected; for example, in the case of a zero mean normally distributed random variable with standard deviation 1, the confidence interval corresponding to 95% confidence would be $[-2, 2]$. The hypothesis is rejected when the experiment yields an outcome that was outside of this confidence interval, and accepted otherwise. Note that acceptance of the hypothesis does not imply that the hypothesis is true, it only indicated that the observed data are compatible with the hypothesis.

Such a framework for hypothesis testing was developed one century ago by Pearson and Fisher \cite{7,8}, and forms the backbone for many more advanced techniques. One of the most successful tests is the so-called $\chi^2$ test. Its success has to do with the fact that it is universal \cite{9}: the confidence intervals that can be defined are independent of the details of the distribution corresponding to the null hypothesis, as only the number of degrees of freedom plays a role. Also, the $\chi^2$ test is in practice already applicable when relatively few samples are taken. The $\chi^2$ test essentially measures the fluctuations around the expected frequencies of the possible outcomes: if those fluctuations are too small or too large, the hypothesis is rejected.

Fluctuations obviously also play a central role in quantum measurements. The expectation value of an observable is not something that can be measured, it can only be sampled, and we get an increasingly better precision the more measurements are being done. This actually means that the expectation value of an observable is not physical: only the individual samples (clicks) are physical. Expectation values can only be approximated using the frequencies of the different outcomes.
As a consequence, quantum mechanics should be reformulated in terms of observable quantities, i.e. clicks, and expectation values of a quantum observable are certainly not observable. For example, the Heisenberg uncertainty relation is formulated in terms of an expectation value and therefore not physical; it has to be reformulated in terms of clicks such as to get an operational meaning.

The topic of this paper is to make a detailed analysis of how the $\chi^2$ hypothesis test, when applied to the frequencies obtained from quantum measurements, reveals information about the underlying quantum states. A particular complication in the quantum setting that makes the problem much richer is the fact that we have the additional choice of the basis in which the measurements are done. The specific questions that we will address are:

1. How to set up the $\chi^2$ test in the quantum setting; how many degrees of freedom does the test have?

2. Suppose that we want to gain confidence that we prepared a certain quantum state $\sigma$ in the lab; what is the optimal POVM measurement such that, for all states for which $||\rho - \sigma|| > \epsilon$, we would reject the hypothesis with the least amount of measurements if the state were $\rho$ instead of $\sigma$?

3. What is the associated divergence rate for rejecting a false hypothesis?

4. What is the relationship between the classical $\chi^2$ distance defined on measured frequencies versus the quantum $\chi^2$ distance?

This paper fits into a long series of papers that were concerned with quantum parameter estimation and quantum hypothesis testing. A wealth of results has been reported in the seminal books of Helstrom [10] and Holevo [11], in a series of papers of Wootters [12] and other pioneers of the field of quantum information theory [13, 14]. The more recent developments are covered in the books of Hayashi [15] and Petz [16]. Very recently, breakthroughs were obtained in defining confidence intervals in the context of quantum tomography and testing of fidelity [17, 20]. The present paper develops similar ideas in the context of hypothesis testing.

$\chi^2$ hypothesis testing is fundamentally different than the Neyman-Pearson test as usually discussed in quantum hypothesis testing [10]. As opposed to Neyman-Pearson tests, the $\chi^2$ test is perfectly well defined without a need of formulating an alternative hypothesis. Such a situation arises precisely when we want to test whether a certain quantum state has been created in the lab. We will also focus on separable measurements, i.e. individual measurements on individual samples, as opposed to entangled measurements such as typically considered in Neyman-Pearson tests [21]. Therefore, the analysis presented here can immediately be used in current experiments.

I. THE $\chi^2$ TEST FOR QUANTUM MEASUREMENTS

Let us assume that we have an experimental quantum apparatus that supposedly spits out quantum states characterized by the density matrix $\sigma$. We would like to gain confidence that this hypothesis is true by performing measurements on it. The most general measurement strategy would correspond to the case where different positive operator valued measurements (POVM) $E_{\alpha,i}$ are chosen, with $E_{\alpha,i} \geq 0$, $\sum_{i=1}^{r_i} E_{\alpha,i} = I$, and where the POVM $\{E_i\}$ is measured a predetermined $k_i$ times (the fact that $k_i$ is not a random number is important for the determination of the number of degrees of freedom in the $\chi^2$ test). Such a typical setup for qubits would correspond to choosing $n_i = n/3$ with $n$ the total number of measurement done, and von-Neumann measurements in the bases $\sigma_x = E_{1,1} - E_{2,1}$, $\sigma_y = E_{1,2} - E_{2,2}$, and $\sigma_z = E_{1,3} - E_{2,3}$ respectively. Alternatively, one could choose $k_1 = n$ and do $n$ measurements with an informational complete POVM. We will henceforth consider the situation where only 1 POVM is used to do the measurements; we will discuss how to modify the results in the case that different POVM’s are chosen deterministically, but the results essentially remain the same.

We denote our hypothesis $H$ by the fact that the $n$ samples we have obtained originate from doing quantum measurements on identical copies of the quantum state $\sigma$. The measurement can be described by a POVM with $r$ elements $\{E_i\}_{i=1,r}$ which obey $\sum_{i=1}^{r} E_i = 1$ and where all the individual elements of the POVM are positive semidefinite $E_i \geq 0$. We say that the measurement has $r$ possible outcomes labeled by $i$ and associate a probability $p_i$ to each outcome which is given due to Born’s rule by $p_i = \text{Tr}[E_i \sigma]$. If we record the number of times $n_i$ that we have obtained some outcome $i$, then we can construct the empirical distribution $f_i = n_i/n$ for the total number $n$ samples. By the law of large numbers [9], we expect that as $n \to \infty$ the empirical distribution converges to $f_i \to p_i$. However, in any realistic scenario, we can only draw a finite number of samples. Due to the inherent randomness of the quantum measurements, there will be fluctuations.
A. The $\chi^2$ distribution

We are now confronted with the problem of accepting or rejecting the hypothesis $H$ in light of only finitely many samples. Since we cannot be certain that the hypothesis $H$ is false, we seek to give bounds on the error probability of rejecting the hypothesis. This is exactly the scenario of the classical $\chi^2$-test (see e.g. \cite{9}). The collection frequencies $\{n_i\}$ are distributed according to the multinomial distribution

$$P(n_1, \ldots, n_r) = \frac{n!}{n_1! \cdots n_r!} p_1^{n_1} \cdots p_r^{n_r}, \quad (1)$$

where $n = n_1 + n_2 + \ldots + n_r$. The distributions of the individual $n_i$ can be computed as the marginals and are distributed according to the binomial

$$P(n_i) = \binom{n}{n_i} (1 - p_i)^{n-n_i} p_i^{n_i}. \quad (2)$$

In the asymptotic limit, i.e. for large values of $n$, the multinomial distribution converges to the normal distribution

$$P(n_1, \ldots, n_r) \sim \exp \left( -\frac{1}{2} \sum_i \frac{(n_i - np_i)^2}{np_i} \right). \quad (3)$$

This suggests that the random variable

$$\chi^2 = \sum_{i=1}^r \frac{(n_i - np_i)^2}{np_i} \quad (4)$$

is a good measure for testing whether we are sampling from $\{p_i\}$, as it measures the deviation of the empirical distribution $f_i$ from the ideal distribution $p_i$. This random variable $\chi^2$ indeed forms the basis for the celebrated $\chi^2$-test, originally introduced by Pearson \cite{8}. $\chi^2$ is obviously a positive random variable. A crucial property of this random variable is the fact that its expectation value is independent of $n$ and is equal to $r - 1$ if the samples are indeed drawn from the distribution $\{p_i\}$. This follows directly from the fact that the individual random variables $n_i$ are distributed according to the binomial distribution and hence $\mathcal{E}(n_i - np_i)^2 = np_i(1 - p_i)$; it then follows that

$$\sum_{i=1}^r \frac{\mathcal{E}(n_i - np_i)^2}{p_in} = \sum_{i=1}^r (1 - p_i) = r - 1. \quad (5)$$

Similarly, the variance of the random variable $\chi^2$ is given by

$$\mathcal{E}(\chi^2 - (r - 1))^2 = 2(r - 1) + \frac{1}{n} \left( \sum_{i=1}^r \frac{1}{p_i} - r^2 - 2r + 2 \right)$$

which also converges fast to a finite value. In practice, when $\forall i : np_i \gtrsim 5$, statisticians use the following asymptotic form of the distribution for the $\chi^2$ variable:

$$P_{r-1}(x) = \frac{1}{2^{\frac{r-1}{2}} \Gamma\left(\frac{r-1}{2}\right)} x^{\frac{r-3}{2}} \exp\left(-\frac{x}{2}\right). \quad (5)$$
For obvious reasons, this distribution is called the $\chi^2$-distribution, and is also the distribution which is obtained by summing up $r - 1$ squares of random variables distributed following the normal distribution with expectation value 0 and variance 1. Note that this distribution does neither depend on the original distribution $\{p_i\}$, nor on the total number of measurements, but only on the number of possible independent measurement outcomes $r - 1$. This total number of degrees of freedom is equal to the number of independent $n_i$ that have to be specified. For example, in the case of a POVM with 4 elements, $r = 4$, but there is the constraint that $\sum_i n_i = n$, and we hence have 3 degrees of freedom. In the case of the independent $\sigma_x, \sigma_y, \sigma_z$ measurements, there are 6 frequencies $n_{i\alpha}$, but only 3 of them are independent, and hence we again have only 3 degrees of freedom.

For our purposes, the most important feature of the $\chi^2$ distribution is that the tails of this distribution decay exponentially fast. The area under the right and left tail are given by the upper and lower incomplete gamma functions. It is interesting to note that the weight under the tail of this distribution from 0 to $x$ is proportional to $x^{(r-1)/2}$ for small $x$. This means that, for large enough degrees of freedom, it is possible to reject a hypothesis because there are not enough fluctuations around the expected values: frequencies that match the expected frequencies too well are highly unlikely, and an experiment reporting such values should be discredited!

B. The $\chi^2$ divergence

Let us now study what will happen when the samples are not drawn from the quantum state $\sigma$ but from the state $\rho$. Then the measurement outcomes will not be distributed according to $p_i = \text{Tr}[E_i\sigma]$ but according to the distribution $q_i = \text{Tr}[E_i\rho]$. The expectation value of $\chi^2$ becomes

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{The Figure shows the plot of the $\chi^2$-distribution, i.e. Eqn. (5), for $r = 4$ (red solid line) and $r = 9$ (grey dashed line) respectively. The error probability $\alpha$, as given in the $\chi^2$ test protocol in Section III, can be computed as the integral $\alpha = \int_{\chi^2_{r-1}}^{\infty} P_{r-1}(x) \, dx$ and is indicated by the shaded area from $\chi^2_{r-1}$ to $\infty$ underneath the right tail for the $r = 9$ distribution. Note that the area underneath the tail decays exponentially fast.}
\end{figure}
\[
\mathcal{E}[\chi^2] = \mathcal{E} \left[ \sum_i \frac{(n_i - np_i)^2}{np_i} \right] = \sum_i \frac{\mathcal{E}[n_i^2]}{np_i} - n
\]

\[
= \sum_i \frac{n^2 q_i^2 + n q_i(1 - q_i)}{np_i} - n
\]

\[
= (n - 1) \left( \sum_i \frac{(q_i - p_i)^2}{p_i} \right) + \left( \sum_i \frac{q_i}{p_i} \right) - 1
\]

The expectation value of \(\chi^2\) grows linearly with the number of samples, and the multiplicative factor to this linear divergence is defined as the \(\chi^2\)-divergence

\[
\chi^2(p, q) = \sum_i \frac{(q_i - p_i)^2}{p_i}.
\]

Note that the \(\chi^2\) divergence also follows naturally from a different divergence, the Kullback-Leibler divergence, in the limit where the 2 distributions are close to each other:

\[
\sum_i p_i \log \frac{p_i}{q_i} = \frac{1}{2} \sum_i \frac{(p_i - q_i)^2}{q_i} + O(||p - q||^3)
\]

C. The quantum \(\chi^2\) divergence

In the paper [22], a family of quantum versions of the \(\chi^2\)-divergence was introduced to study the convergence and relaxation rates [23] of completely positive maps and general dissipative quantum systems. All members of this class of quantum \(\chi^2\)-divergences reduce to the classical \(\chi^2\) divergence when \(\rho\) and \(\sigma\) commute. The formulation of the quantum versions of the \(\chi^2\)-divergence follows from the framework of monotone Riemannian metrics [24–29] and can be seen as a special case of this family of metrics. It follows from the analysis of monotone Riemannian metrics that the family of \(\chi^2\)-divergences has a partial order with a smallest and largest element. A special role was played by the Bures \(\chi^2\) divergence [30, 31], as it is always the smallest one of those quantum divergencies. It is defined as

\[
\chi^2_B(\sigma, \rho) = \text{Tr} \left[ (\rho - \sigma) \Omega_\sigma (\rho - \sigma) \right]
\]

with \(\Omega_\sigma\) the superoperator whose inverse if given by

\[
\Omega_\sigma^{-1}(X) = \frac{\sigma X + X \sigma}{2}
\]

Let us now show that an operational meaning can be given to this quantity by comparing it to the classical \(\chi^2\) divergence maximized over all possible quantum measurements.

**Lemma 1** For two states \(\sigma\) and \(\rho\) we denote the probability distributions \(p_i = \text{Tr}[E_i \sigma]\) and \(q_i = \text{Tr}[E_i \rho]\) for some POVM \(\{E_i\}_{i=1,...,r}\). Then, the Bures \(\chi^2_B\)-divergence is equal to the maximum value of \(\chi^2(p, q)\) when optimized over all possible POVM measurements:

\[
\chi^2_B(\rho, \sigma) = \max_{\{E_i\}} \chi^2(p, q),
\]

Furthermore, the measurement maximizing this \(\chi^2\) divergence is a projective von-Neumann measurement in the eigenbasis of \(\Omega_\sigma(\rho) = \sum_i \lambda_i |\psi_i\rangle \langle \psi_i|\).
After completion of this work, it was brought to our attention that this theorem first appeared in a paper by Braunstein and Caves about the geometry of quantum states [13]. The proof presented here has no similarity to this original proof, and we present this new proof here because the tools used will turn out to be relevant for the later sections; a central role is played by Woodberry’s matrix identity [32].

**Proof:** Let us first prove that Bures $\chi^2$ divergence forms an upper bound to the $\chi^2$ divergence with respect to any POVM $\{E_i\}$. We write for any operator $A \in M_D(\mathbb{C})$ on the Hilbert space $\mathbb{C}^D$ its vectorization as $|A\rangle = A \otimes \mathbb{I} |I\rangle$, where $|I\rangle = \sum_{k=1}^D |kk\rangle$ denotes the unnormalized maximally entangled state. We use this notation because superoperators become simple matrices in this representation:

$$\hat{\Omega}_B^{\sigma} = \frac{2}{\sigma \otimes \mathbb{I} + \mathbb{I} \otimes \sigma}$$

(14)

It is easy to see that the $\chi^2$ divergence is given by

$$\chi^2(p, q) = \langle p | \sum_i |E_i\rangle \langle E_i| \sigma \rangle |p\rangle - 1$$

(15)

and the Bures divergence by

$$\chi^2_B(\sigma, \rho) = \langle \rho | \frac{2}{\sigma \otimes \mathbb{I} + \mathbb{I} \otimes \sigma} | \rho\rangle - 1$$

(16)

It is therefore enough to prove the semidefinite matrix inequality

$$\frac{2}{\sigma \otimes \mathbb{I} + \mathbb{I} \otimes \sigma} - \sum_i |E_i\rangle \langle E_i| \sigma \rangle \geq 0$$

(17)

for all possible POVM’s $\{E_i\}$. A matrix is positive if and only if its inverse is positive, and the inverse can easily be calculated by making use of Woodberry’s identity [32]

$$(A - UCU^\dagger)^{-1} = A^{-1} + A^{-1} U \left( C^{-1} - U^\dagger A^{-1} U \right)^{-1} U^\dagger A^{-1}.$$ \hspace{1cm} (18)

Equation (17) is exactly of that form by choosing an orthonormal basis $|i\rangle$ with a number of elements equal to the total number of POVM elements and

$$A = \frac{2}{\sigma \otimes \mathbb{I} + \mathbb{I} \otimes \sigma}$$

(19)

$$U = \sum_i |E_i\rangle \langle i|$$

(20)

$$C = \sum_i |i\rangle \langle i| \frac{1}{\text{Tr}[E_i \sigma]}$$

(21)

As the matrix $A$ is obviously positive, (17) will hold if

$$C^{-1} - U^\dagger A^{-1} U = \sum_i \text{Tr}[E_i \sigma] |i\rangle \langle i| - \sum_{ij} |i\rangle \langle j| \frac{\sigma \otimes \mathbb{I} + \mathbb{I} \otimes \sigma}{2} |E_j\rangle \geq 0$$

(22)

or equivalently if the matrix $L = \sum L_{ij} |i\rangle \langle j|$, with the entries

$$L_{ij} = \begin{cases} \text{Tr} [E_i (\mathbb{I} - E_i) \sigma] & i = j \\ -\text{Tr} [E_i E_j \sigma] & i \neq j \end{cases}$$
is positive semidefinite. This is indeed true, as \(-L\) is the generator of a Markovian semi-group as occurring in master equations (the elements in all columns sum up to zero, and all off-diagonal elements are positive), which is well known to have only negative eigenvalues.

Note that we can make \(L\) equal to zero by choosing all POVM elements orthogonal to each other, i.e. by choosing a von Neumann measurement. The null space of the matrix occurring in (17) can now easily be seen to be spanned by the vectors in \(A^{-1}U\). \(ρ\) will therefore be in the null space and saturate the inequality if there exist numbers \(\{λ_i\}\) for which

\[
|ρ⟩ = \sum_i λ_i \langle i | A^{-1}U | i ⟩ = \sum_i λ_i \frac{σ ⊗ II + II ⊗ σ}{2} |E_i⟩.
\]

By writing \(E_i = |ψ_i⟩ ⟨ψ_i|\), this equation is equivalent to

\[
\sum_i λ_i |ψ_i⟩ ⟨ψ_i| = Ω^B_σ(ρ)
\]

which shows that a von Neumann measurement in the eigenbasis of \(Ω^B_σ(ρ)\) will give equality.

This shows that the quantum \(χ^2\) divergences have indeed an operational meaning. It also illustrates the fact that the problem of quantum hypothesis testing is much richer than the classical one: we have the extra choice optimization over the measurement basis.

II. GOODNESS OF FIT FOR QUANTUM MEASUREMENTS

We now come to the central part the paper, which is concerned with the problem of testing whether the data acquired during an experiment is compatible with the fact that it is sampled from a given quantum state \(σ\). Obviously, if we would like to make the measurement which reveals the most information, it should be the one that would allow to reject the hypothesis as soon as possible if the hypothesis is false.

We therefore define an \(ε\)-ball around our hypothesis state \(σ\), and will optimize over all possible POVM measurements in such a way that we require that the (classical!) \(χ^2\) divergence with respect to all possible density matrices \(ρ\) outside of this ball \(∥ρ − σ∥ ≥ ε\) is as large as possible. Due to the quadratic nature of the \(χ^2\) divergence, the natural norm to use is the Frobenius norm (i.e. \(∥X∥ = \sqrt{Tr[X^†X]}\)); all bounds derived for the Frobenius norm can however be converted to any other norm such as the infinity or trace distance by using well known inequalities.

Clearly, the optimal POVM should be an informationally complete POVM, as otherwise there would always be directions in which the divergence is zero. The properties of the optimal POVM will be discussed in the following section III.

The aforementioned discussion leads us to define the following quantity

**Definition 2** The divergence rate \(ξ\) for the quantum \(χ^2\) goodness of fit test for the state \(σ\) is given by

\[
ξ(σ) = \frac{1}{ε^2} \max_{\{E_i\}} \min \left\{ \frac{1}{∥ρ−σ∥≥ε} \chi^2(p,q) \right\},
\]

where we have defined the classical \(χ^2\)-divergence

\[
\chi^2(p,q) = \left( \sum_i \frac{q_i^2}{p_i} - 1 \right),
\]

with respect to the induced probability distributions \(p_i = Tr[E_iσ]\) and \(q_i = Tr[E_iρ]\). The optimization is performed over all possible POVM \(\{E_i\}_{i=1...r}\) and states \(ρ\) for which \(∥ρ−σ∥ ≥ ε\) as measured by the Frobenius norm.
Note that, due to the quadratic nature of $\chi^2$, $\xi(\sigma)$ is independent of $\epsilon$. As will be proved in the next section, the divergence rate $\xi(\sigma)$ is guaranteed to lie in a small interval:

$$\frac{2}{3} \leq \xi(\sigma) \leq 1$$ (27)

This bound is actually very important: it shows that the prefactor of the linear term of the expectation value of $\chi^2$ is independent of the dimension of the Hilbert space, which is of course crucial for the quantum $\chi^2$ hypothesis testing to make sense and to be scalable. Furthermore, $\xi(\sigma)$ and the corresponding optimal POVM can be calculated exactly as the solution of a simple eigenvalue problem: see theorem 3. As discussed later, the optimal POVM turns out to be optimal both in the sense of Pitman [33] and Bahadur [34].

A goodness of fit test protocol for the state $\sigma$ is then given as follows:

1. Choose the POVM $r$ element $\{E^*_i\}$ that optimizes $\xi$ as given in definition 2.
2. Measure $\{E^*_i\}$ on $n$ independent samples of the state $\rho$ and record the frequencies $n_i$ of the $i$'th outcome.
3. Compute the test statistic $c^2 = \sum_{i=1}^{r} \frac{(n_i - p_i n)^2}{p_i n}$, where $p_i = \text{Tr}[E_i \sigma]$ corresponds to the hypothesis $H$.
4. Reject the Hypothesis with error probability $\alpha$ if $c^2 \geq \chi^2_{\alpha}$, where the constant $\chi^2_{\alpha}$ is determined via

$$\alpha = \int_{\chi^2_{\alpha}}^{\infty} P_{r-1}(x)dx$$ (28)

5. If the test statistic $c^2$ is smaller than $\chi^2_{\alpha}$, we state that the observed data is consistent with the hypothesis $H$ up to a statistical error $\alpha$.

Note that we assumed the large $n$ limit to compute the distribution function for the $\chi^2$ variable. This assumption is generally well satisfied if we take sufficiently many samples. A good estimate is that we have to take to make sure that $np_i > 5$ for all $i$. Furthermore, note that the test can be made slightly more advanced by also rejecting the hypothesis when the fluctuations are not large enough, i.e. if the value $c^2$ is too small!

If we now turn to the Definition 2 of the divergence rate, we can give it a meaningful interpretation in the light of the test protocol. The goal of the optimization is to construct a test, i.e. a quantum measurement, which rules out the hypothesis $H$ with as little samples as possible if it is not true. That is, we want that the test statistics $c^2$ grows as fast as possible with the number of samples $n$. In light of Eqn. 29, we see that the expectation value of the $\chi^2$ random variable grows linearly in the number of samples $n$ with the prefactor $\chi^2(p,q)$. In the case where $\rho = \sigma$ and thus $p = q$, i.e. the $H$ is true, the classical $\chi^2$ vanishes and we obtain the expectation value $r-1$ and a standard deviation of $\sqrt{2/(r-1)}$. When $\rho \neq \sigma$, the goal is to find the measurement that reaches the critical region indicated by $\chi^2_{\alpha}$ as fast as possible, in the worst case.

We therefore have a class of estimators, parameterized by the different possible POVM’s $E$, and we want to find the most efficient one. Associated to every POVM $E$, there is a worst case state $\rho_E$ with $\|\rho_E - \sigma\|_2 \geq \epsilon$ which gives rise to a divergence rate $\xi_E$. The expected number of samples $n$ needed to exceed the power $\alpha$ of the test statistic is given by the formula

$$(r-1) + (n-1)c^2\xi_E \simeq \chi^2_{\alpha}$$ (29)

or

$$n \simeq \frac{\chi^2_{\alpha} - (r-1)}{c^2\xi_E}$$ (30)

This is the number of expected samples which are necessary to reject the hypothesis if it is untrue.

Now there are several possible notions of efficiencies for asymptotic tests. For the so-called Pitman efficiency [33], we compare tests in such a way that $\alpha$ is fixed but for which $\epsilon \to 0$ gradually, and look at the scaling
of $n$ as a function of $\epsilon$. Obviously, the POVM that minimizes $n$ is the one for which $\xi_E$ is maximal, i.e. the POVM that corresponds to the optimal one with respect to the definition of $\xi(\sigma)$. Note that this POVM is also optimal according to Pitman for the maximum likelihood. Different tests can also be compared with respect to the Bahadur efficiency [34]. In the framework of Bahadur, $\epsilon$ is fixed, but the error $\alpha$ is made smaller and smaller (which corresponds to a larger and larger $\chi^2_n$), and the scaling of $n$ with respect to $\alpha$ is compared. The optimal POVM which maximizes $\xi$ is obviously also the one with maximal Bahadur efficiency. The optimal quantum measurement is therefore the one with maximal Pitman and Bahadur efficiency within the class of all quantum $\chi^2$ tests.

Note that the standard deviation of $\chi^2$ is $\sqrt{2(r-1)}$. Therefore, $\chi^2_n - (r-1)$ for a fixed $\alpha$ but varying dimension of the Hilbert space is proportional to the square root of the number of degrees of freedom, i.e. linear in the dimension of the Hilbert space.

III. DIVERGENCE RATE AND OPTIMAL POVM

Let us next get some insights into the structure of the optimal POVM measurement. If the state $\sigma$ is full rank, the POVM must be informationally complete, so the number of POVM elements has to be at least equal to the square of the dimension of the Hilbert space, i.e. $r \geq D^2$, as otherwise there are always perturbations $X$ around the state $\sigma$ for which $\text{Tr}[E_i X] = 0$. We will now prove that all the elements $E_i$ of the POVM must be pure, which is intuitively obvious. Then we will go on proving matching upper and lower bounds to the quantity $\xi(\sigma)$. The lower bound is constructive, and hence gives an explicit construction for the optimal measurement to perform that maximizes the discriminating power.

Lemma 3 If the POVM $\{E_i\}$ is optimal in the sense that it maximizes the divergence rate, then all its elements can be chosen to be pure: $E_i = p_i |\psi_i \rangle \langle \psi_i |$.

Proof: Assume that the first element of the POVM with $r$ elements $\{E_i\}$ has rank $k_1 > 1$, i.e. $E_1 = \sum_{i=2}^{k_1} p_i |\psi_i \rangle \langle \psi_i |$. We will show that we can construct another POVM with $r + 1$ elements which leads to a larger error rate, and for which the rank of $E_1$ is $k_1 - 1$ and the rank of $E_{r+1}$ is equal to 1. Then the proof follows by induction. Let us therefore define $\tilde{E}_1 = \sum_{i=2}^{k_1} p_i |\psi_i \rangle \langle \psi_i |$ and $E_{r+1} = p_1 |\psi_1 \rangle \langle \psi_1 |$. The it is enough to prove the semidefinite inequality

$$\left( |E_1 \rangle \langle E_1 | + |E_{r+1} \rangle \langle E_{r+1} | \right) \leq \left( |\tilde{E}_1 \rangle \langle \tilde{E}_1 | + |E_{r+1} \rangle \langle E_{r+1} | \right)$$

(31)

Indeed, if this inequality is true, then the optimization over $\rho$ will necessarily yield a larger value. Since we are working in an effective 2-dimensional subspace spanned by $\tilde{E}_1$ and $E_{r+1}$, we have to prove that the $2 \times 2$ matrix

$$M = q_1 |\tilde{E}_1 \rangle \langle \tilde{E}_1 | + q_2 |E_{r+1} \rangle \langle E_{r+1} | - r \left( |E_{r+1} \rangle \langle \tilde{E}_1 | + |\tilde{E}_1 \rangle \langle E_{r+1} | \right) \geq 0,$$

(32)

where

$$q_1 = \frac{\text{Tr}[\tilde{E}_1 \sigma]^2}{\text{Tr}[\tilde{E}_1 \sigma]\text{Tr}[E_{r+1} \sigma](\text{Tr}[\tilde{E}_1 \sigma] + \text{Tr}[E_{r+1} \sigma])}$$

(33)

$$q_2 = \frac{\text{Tr}[E_{r+1} \sigma]^2}{\text{Tr}[\tilde{E}_1 \sigma]\text{Tr}[E_{r+1} \sigma](\text{Tr}[\tilde{E}_1 \sigma] + \text{Tr}[E_{r+1} \sigma])}$$

(34)

$$r = \frac{\text{Tr}[\tilde{E}_1 \sigma]\text{Tr}[E_{r+1} \sigma]}{\text{Tr}[\tilde{E}_1 \sigma]\text{Tr}[E_{r+1} \sigma](\text{Tr}[\tilde{E}_1 \sigma] + \text{Tr}[E_{r+1} \sigma])}.$$ 

Since this is a $2 \times 2$ matrix, it suffices to compute the trace and the determinant to verify positivity:

$$\det(M) = (q_1 q_2 - r^2) \left( \|\tilde{E}_1\|^2 \|E_{r+1}\|^2 - \|\tilde{E}_1\|E_{r+1}\|^2 \right) = 0$$

(35)

$$\text{Tr}[M] = \frac{\|\text{Tr}[\tilde{E}_1 \sigma] |\tilde{E}_1 \rangle - \text{Tr}[E_{r+1} \sigma] |E_{r+1}\|^2}{\text{Tr}[\tilde{E}_1 \sigma]\text{Tr}[E_{r+1} \sigma](\text{Tr}[\tilde{E}_1 \sigma] + \text{Tr}[E_{r+1} \sigma])} \geq 0.$$ 

(36)
This is obviously a positive rank 1 operator, which finishes the proof.

We are now ready to prove matching lower and upper bounds to $\xi(\sigma)$.

A. Upper bound to the divergence rate

An equivalent characterization of the divergence rate $\xi(\sigma)$ can be obtained by introducing the traceless operator $X = (\rho - \sigma)/\epsilon$:

$$
\xi(\sigma) = \max_{\{E_i\}} \min_X \left( \sum_i |E^i_i\rangle \left( \frac{p_i}{\langle E^i | \sigma \rangle} \right) \langle E^i | \right) |X\rangle \tag{37}
$$

under the conditions

$$
\begin{align*}
E_i &= |\psi^i\rangle \langle \psi^i | \\
\langle \psi^i | \psi^i \rangle &= 1 \\
\sum_{\alpha} p_i E^i &= \mathbb{1} \\
\text{Tr}[XX^+] &= 1 \\
\text{Tr}[X] &= 0 \\
X &= X^+
\end{align*}
$$

The sum over $\alpha$ is unlimited, i.e. there is no limit on the number of POVM elements, and the dimension of $X$ is the dimension of the Hilbert space corresponding to $\sigma$, i.e. $D$-dimensional. Note that $\epsilon$ factored out due to the quadratic dependence on $\rho - \sigma = \epsilon X$. Without loss of generality, we will work in the basis in which $\sigma$ is diagonal:

$$
\sigma = \sum_{\alpha=1}^{D} \lambda_\alpha |\alpha\rangle \langle \alpha |
$$

with the eigenvalues $\lambda_\alpha = (s_\alpha)^2$ ordered in decreasing order. We will also assume that $\sigma$ is full rank; if this condition is not satisfied, then we can always perturb $\sigma$ infinitesimally, and take the limit at the end.

We will prove the following lemma:

**Lemma 4** An upper bound to $\xi(\sigma)$ defined in [37] is given by the smallest nonzero eigenvalue of the matrix

$$
P_s \left( \sum_{\alpha=1}^{D} \frac{1}{1 + \lambda_\alpha} |\alpha\rangle \langle \alpha | \right) P_s \tag{38}
$$

with $P_s$ the projector on the subspace orthogonal to the vector $\sum_{\alpha} \sqrt{1/1 + \lambda_\alpha} |\alpha\rangle$.

A simple upper bound to this upper bound is

$$
\xi(\sigma) \leq \frac{1}{1 + \lambda_2} \leq 1
$$

with $\lambda_2$ the second largest eigenvalue of $\sigma$.

Note that this upper bound lies between $2/3$ and $1$ for any density matrix $\sigma$.

**Proof:** The proof of the theorem is a bit involved. In this proof, we will assume that the elements of the POVM are given by $p_i E_i$ with $E_i = |\psi_i\rangle \langle \psi_i |$, $\langle \psi_i | \psi_i \rangle = 1$, $\sum_i p_i E_i = \mathbb{1}$ and $p_i \leq 0$, $\sum_i p_i = D$. 
As a first step, we observe that a consequence of the fact that $\sigma$ is diagonal, thus we can twirl the POVM elements:

$$\text{Tr}[E_i \sigma] = \text{Tr}[E_i D(-\theta) \sigma D(\theta)] = \frac{\int d\theta_1 d\theta_2 \cdots \text{Tr}[D(\theta)E_i D(-\theta)\sigma]}{\int d\theta_1 d\theta_2 \cdots}$$

Here $D(\theta)$ is a diagonal matrix with elements $D_{kk} = \exp(i\theta_k)$. Therefore, two POVM's related by $E_i = D(\theta) \tilde{E}_i D(-\theta)$ will give the same value in the optimization of (37) as we can just transform the related $X$ to $\tilde{X} = D(-\theta)XD(\theta)$. It is therefore clear that an upper bound to (37) is obtained by solving the problem

$$\begin{align*}
\max_{\{E_i\}} \min_X & \frac{\int d\theta_1 d\theta_2 \cdots \langle X | D(-\theta) \otimes D(\theta) \left( \sum_i |E_i\rangle \langle E_i| \right) \otimes D(-\theta)|X \rangle}{\int d\theta_1 d\theta_2 \cdots}
\end{align*}$$

as this forces to use the same $X$ for different realizations of all equivalent POVM's related by such a "gauge transformation". This is equivalent to saying that the minimum eigenvalue of a convex combination of operators with the same eigenvalues is always larger then the minimum of the individual eigenvalues. This twirling integration can be done exactly, and by using the cyclicity of the trace we get

$$\hat{X} = \frac{\int d\theta_1 d\theta_2 \cdots D(\theta) \otimes D(-\theta)|X \rangle \langle X | D(-\theta) \otimes D(\theta)}{\int d\theta_1 d\theta_2 \cdots} = \sum_{\alpha,\beta=1}^D X_{\alpha\alpha} X_{\beta\beta} |\alpha\rangle \langle \alpha | \otimes |\beta\rangle \langle \beta | + \sum_{\alpha \neq \beta} |X_{\alpha\beta}|^2 |\alpha\rangle \otimes |\beta\rangle \langle \beta |$$

Substituting this into (37), we get

$$\xi(\sigma) \leq \max_{E^i} \min_X \sum_i p_i \frac{\langle E^i | \hat{X} | E^i \rangle}{\langle E^i | \sigma \rangle}$$

As $E^i = |\psi^i\rangle \langle \psi^i |$ are pure POVM elements,

$$\langle E^i | \alpha\rangle \langle \alpha | \otimes |\beta\rangle \langle \beta | E^i \rangle = E_{\alpha\alpha}^i E_{\beta\beta}^i = \langle E^i | \alpha\rangle \langle \beta | \beta | \alpha\rangle \langle \beta | E^i \rangle.$$  

Let’s now define a new vector $|e^i\rangle$ with $D$ components that contains the diagonal elements of $E^i$: $e^i_\alpha = E^i_{\alpha\alpha}$, and also the vector $|s\rangle$ with $D$ elements given by $s_\alpha = \sqrt{\lambda_\alpha}$ and $\lambda_\alpha$ the eigenvalues of $\sigma$.

Substituting all this into the previous expressions, we get

$$\xi(\sigma) \leq \max_{e^i} \min_X \sum_i p_i \frac{\sum_{\alpha,\beta} \langle e^i | \alpha\rangle \langle \beta | e^i \rangle \left( |X_{\alpha\beta}|^2 (1 - \delta_{\alpha\beta}) + X_{\alpha\alpha} X_{\beta\beta} \right)}{\langle e^i | s^2 \rangle}$$

Note that we have the constraints

$$\sum_i p_i \langle e^i | \alpha \rangle = 1$$
$$\sum_{\alpha} X_{\alpha\alpha} = 0$$
$$\sum_{\alpha,\beta} |X_{\alpha\beta}|^2 = 1$$

The biggest problem in doing the optimization of equation (37) is the presence of the denominator. Now is the time to get rid of it: we will choose $X$ such that

$$|X_{\alpha\beta}|^2 (1 - \delta_{\alpha\beta}) + X_{\alpha\alpha} X_{\beta\beta} = \langle \alpha | s^2 \rangle \langle \beta | s^2 \rangle + \langle \alpha | t^2 \rangle \langle \beta | t^2 \rangle = s^2_{\alpha\beta} t^2 + s^2_{\alpha\beta} t^2_{\alpha\beta}$$
with the vector $|t^2\rangle$ with elements $\langle \alpha | t^2 \rangle = |t_\alpha|^2$ still to be determined. Note that any choice of $X$ will give us an upper bound as long as the constraints above are satisfied. If it is possible to choose such a $|t\rangle$, then the upper bound becomes equal to

$$\xi(\sigma) \leq 2 \sum_i p_i \langle e_i | t^2 \rangle \frac{\langle s^2 | e_i \rangle}{\langle s^2 | e_i \rangle} = 2 \sum_\alpha |t_\alpha|^2 \quad \text{(39)}$$

This implies that such $X$ and corresponding $t$ completely eliminates the $E^i$ from the upper bound, which was what we were looking for. It is indeed possible to choose such a $X$:

$$X_{\alpha\alpha} = \sqrt{2} s_\alpha t_\alpha$$

$$|X_{\alpha\beta}|^2 = (s_\alpha t_\beta - s_\beta t_\alpha)^2$$

The constraints on $X$ can now be written in terms of the new variables $t_\alpha$:

$$0 = \sum_\alpha s_\alpha t_\alpha$$

$$1 = \sum_{\alpha\beta} |X_{\alpha\beta}|^2$$

$$= \sum_{\alpha \neq \beta} (s_\alpha t_\beta - s_\beta t_\alpha)^2 + 2 \sum_\alpha (s_\alpha t_\alpha)^2$$

$$= 2 \left( \sum_\alpha (1 - s_\alpha^2)^2 t_\alpha^2 - \sum_{\alpha \neq \beta} s_\alpha s_\beta t_\alpha t_\beta + \sum_\alpha (s_\alpha t_\alpha)^2 \right)$$

$$= 2 \left( \sum_\alpha t_\alpha^2 + \sum_\alpha s_\alpha^2 t_\alpha^2 - \left( \sum_\alpha s_\alpha t_\alpha \right)^2 \right)$$

$$= 2 \sum_\alpha (1 + s_\alpha^2) t_\alpha^2$$

Note that we made use of the normalization of $\sigma$ in the form of $\sum_\alpha s_\alpha^2 = 1$ and also of the constraint $\sum_\alpha s_\alpha t_\alpha = 0$. Rescaling $t^2$ by a factor of 2, we get the optimization problem:

minimize $\sum_{\alpha=1}^D t_\alpha^2$

under the condition $\sum_{\alpha=1}^D s_\alpha t_\alpha = 0$

and $\sum_{\alpha=1}^D (1 + s_\alpha^2) t_\alpha^2 = 1$

This optimization problem can actually be written as an eigenvalue problem: define $y_\alpha = \sqrt{1 + s_\alpha^2} t_\alpha$ and $P_s$ the projector on the space orthogonal to the vector with components $s_\alpha / \sqrt{1 + s_\alpha^2}$. Then the upper bound is given by the second smallest eigenvalue (the smallest being 0) of the matrix

$$P_s \sum_\alpha \frac{1}{1 + s_\alpha^2} |\alpha\rangle \langle \alpha| P_s \quad \text{(40)}$$

This is the upper bound that we set out to prove. A simple upper bound to this upper bound can be found. By making use of the interlacing properties of eigenvalues of submatrices, we therefore know that the eigenvalues of this matrix obey

$$\mu_1 = 0 \leq \frac{1}{1 + s_1^2} \leq \mu_2 \leq \frac{1}{1 + s_2^2} \leq \ldots$$
which proves that

\[
\xi(\sigma) \leq \frac{1}{1 + s^2} \leq 1.
\]

This concludes the proof. \( \square \)

B. Lower bound to the divergence rate

Let us next prove a lower bound to the divergence rate \( \xi(\sigma) \). For this, we will have to guess class of good POVM's. We will do the optimization over the class of POVM's parameterized by 1 parameter \( 0 \leq p \leq 1 \):

\[
1 \leq i \leq D : \quad E_i = (1 - p) |i \rangle \langle i | \quad (41)
\]

\[
\text{if } j > D : \quad E_j = c(p) |\chi_j \rangle \langle \chi_j | \quad (42)
\]

\[
|\chi_j \rangle = \frac{1}{\sqrt{D}} \sum_k e^{i\theta_j^k} |k \rangle, \quad (43)
\]

where the \(|i \rangle\) label the eigenstates of \( \sigma \). All \(|\chi_j \rangle\) are chosen such that they have the same overlap with \( \sigma \): \( \langle \chi_j | \sigma | \chi_j \rangle = 1 \). Those states \(|\chi_j \rangle\) are hence only susceptible to the off-diagonal elements of \( \sigma \). In the case of \( D \) a prime or a power of prime, a possible choice of such a basis is given by the mutually unbiased basis, but as we only require unbiasedness with the standard basis, such a basis can easily be constructed in any dimension, e.g. by choosing basis labeled by the angles \( \{\theta^k_j\} \). We will choose such a basis that is invariant under any similarity transformation with diagonal elements \( D_{kk} = \exp(i\theta_k) \) (which is always possible), such that we have \( \sum_{j > D} E_j = c(p) \sum_{j > D} |\chi_j \rangle \langle \chi_j | = p \mathbb{1} \). This defines \( c(p) \) which we do not have to determine explicitly. It follows that

\[
\sum_{j > D} |E_j\rangle \frac{1}{\text{Tr}[E_j \sigma]} \langle E_j | = \frac{c(p)}{1/D} \sum_{j > D} |\chi_j \rangle \langle \chi_j | \langle \chi_j | = p \left( \sum_{ij \neq j} |ij\rangle \langle ij| + \sum_{i,j} |ii\rangle \langle jj| \right). \quad (44)
\]

This follows that the operator in invariant under twirling, and also because

\[
p_D = \text{Tr} \sum_j E_j = c(p) \text{Tr} \sum_j |\chi_j \rangle \langle \chi_j | = c(p) \text{Tr} \sum_j |\chi_j \rangle \langle \chi_j | \otimes |\chi_j \rangle \langle \chi_j |. \quad (45)
\]

With those choices, there is hence only 1 parameter left, i.e. the weight \( p \) that weights the diagonal versus the off-diagonal parts of the density matrix \( \sigma \). A lower bound on \( \xi(\sigma) \) can now be obtained by the following optimization:

\[
\max_p \min_X \langle X | (1 - p) \sum_{i=1}^D \frac{1}{\lambda_i} |ii\rangle \langle ii| + p \sum_{ij \neq j} |ij\rangle \langle ij| + \sum_{i,j} |ii\rangle \langle jj| \rangle |X \rangle
\]

with \( X = (\rho - \sigma)/\epsilon \) a traceless hermitean operator with norm \( \|X\|_2 = 1 \). We therefore want to make the smallest eigenvalue of the matrix \( Q \) as large as possible, as this eigenvalue provides a lower bound to \( \xi \). The matrix \( Q \) is a direct sum \( Q_1 \oplus Q_2 \) where \( Q_1 \) is \( p \) times the identity matrix on the subspace spanned by \(|ij\rangle, i \neq j \), and \( Q_2 \) the \( D \times D \) matrix

\[
Q_2 = (1 - p) \sum_{\alpha=1}^D \frac{1}{\lambda_{\alpha}} |\alpha\rangle \langle \alpha| + p \sum_{\alpha, \beta=1}^D |\alpha\rangle \langle \beta| \quad (46)
\]

where we identified \(|\alpha\rangle = |ii\rangle\). Actually, this is not entirely correct, as we still have to include the constraint that \( \text{Tr}[X] = 0 \). This can easily be incorporated by projecting \( Q_2 \) on the subspace orthogonal to \(|\Omega\rangle = 1/\sqrt{D} \sum_\alpha |\alpha\rangle \).
Given $P = \mathbb{1} - |\Omega\rangle \langle \Omega|$, we therefore define $\tilde{Q}_2 = PQ_2P$.

The smallest eigenvalue of $Q_1$ is obviously proportional to $p$, while the smallest eigenvalue of $\tilde{Q}_2$ is monotonically decreasing with $p$. Therefore, the optimal value of $p$ will be the one for which the smallest eigenvalues of $Q_1$ and $\tilde{Q}_2$ coincide. This is equivalent to determining the largest $p$ for which

\[
(1 - p) \sum_{\alpha=1}^{D} \frac{1}{\lambda_\alpha} P |\alpha\rangle \langle \alpha| P + p \sum_{\alpha,\beta=1}^{D} P |\alpha\rangle \langle \beta| P \geq p P
\]

which is in turn equivalent to maximizing $p$ such that

\[
\sum_{\alpha} \frac{1}{\lambda_\alpha} P |\alpha\rangle \langle \alpha| P \geq p \left( \sum_{\alpha} \frac{1}{\lambda_\alpha} P |\alpha\rangle \langle \alpha| P - \sum_{\alpha \neq \beta} P |\alpha\rangle \langle \beta| P \right).
\]

(48)

This optimal $p$, which is the lower bound we were looking for, is then given by

\[
p = \frac{1}{\mu(S)}
\]

(49)

with $\mu$ the largest eigenvalue of the matrix

\[
S = \mathbb{1} - \frac{1}{D} \sum_{\alpha,\beta} |\alpha\rangle \langle \alpha| - \sum_{\alpha,\beta} \lambda_\alpha \lambda_\beta |\alpha\rangle \langle \beta|
\]

(50)

which is equivalent to $1$ plus the largest eigenvalue of the pseudo-inverse of the matrix $P\sigma^{-1}P$:

\[
\tilde{S} = \sum_{\alpha} \lambda_\alpha |\alpha\rangle \langle \alpha| - \sum_{\alpha,\beta} \lambda_\alpha \lambda_\beta |\alpha\rangle \langle \beta|
\]

(51)

$\tilde{S}$ is again the generator of a semi-group, and hence all its eigenvalues are larger or equal to zero. It is equal to zero for pure states, and the maximal possible eigenvalue is equal to $1/2$ and is obtained for the case $\lambda_1 = \lambda_2 = 1/2$, $\lambda_{i > 2} = 0$. Those 2 cases correspond to $\xi = 1$ and $\xi = 2/3$ respectively. It can easily be shown that the pseudo-inverse of the matrix $S$ has the same eigenvalues as the matrix $[\Omega]$ [30]. This means that our lower bound coincides with the upper bound! We have therefore proven:

**Theorem 5** The divergence rate $\xi(\sigma)$ is equal to $\xi(\sigma) = 1/(1 + \mu(S))$ with $\mu(S)$ the largest eigenvalue of the matrix

\[
S = \sum_{\alpha} \lambda_\alpha |\alpha\rangle \langle \alpha| - \sum_{\alpha,\beta} \lambda_\alpha \lambda_\beta |\alpha\rangle \langle \beta|
\]

(52)

where $\lambda_\alpha$ are the eigenvalues of $\sigma$ and $|\alpha\rangle$ the corresponding eigenvectors. In particular, this implies that

\[
\frac{2}{3} \leq \xi(\sigma) \leq 1
\]

(53)

with the value of $2/3$ obtained in the case where $\lambda_1 = \lambda_2 = 1/2$, $\lambda_{i > 2} = 0$, and the value $1$ when $\sigma$ is a pure state.

A possible choice for a POVM that gives the optimal error rate is given as follows:

\[
1 \leq i \leq D : \quad E_i = (1 - \xi) |i\rangle \langle i|
\]

\[
j > D : \quad E_j = c(\xi) |\chi_j\rangle \langle \chi_j|
\]

\[
|\chi_j\rangle = \frac{1}{\sqrt{D}} \sum_k e^{i\phi_j} |k\rangle
\]

(54)

(55)

(56)
with \( c(\xi) \) and the angles \( \{ \theta_j^k \} \) chosen such that the POVM is informationally complete and that

\[
\sum_{j > D} E_j = \xi \mathbb{I}
\]

Note that the degrees of freedom in the \( \chi^2 \) distribution corresponding to this optimal POVM can easily be reduced by dividing the POVM up in several resolutions of the identity, and fixing the number of times those different measurements are done by a fraction corresponding to their weight given in the theorem. For example, let us assume that the \( |\chi_j\rangle \) can be divided up into \( D \) orthonormal basis (as e.g. in the case of mutually unbiased bases), and that we want to do a total of \( N \) measurements. Then the von Neumann measurement in the basis \( |i\rangle \) can then be done \((1 - \xi).N\) times and the other von-Neumann measurements \( \xi/D.N \) times. The total degrees of freedom for the corresponding \( \chi^2 \) distribution is then given by \((D + 1).D - (D + 1) = D^2 - 1\) which is indeed equal to the total number of degrees of freedom in the density matrix. It is clear that exactly the same arguments for the error exponent carry through in this case.

**C. Examples of divergence rates**

Let us next look at some specific examples. A special role is played by the second largest eigenvalue of \( \sigma \): \( \xi \) is minimized when \( \lambda_2 \) is maximal, and maximized when \( \lambda_2 \) is minimal. The maximal divergence rate is obviously obtained for pure states and is exactly given by 1:

\[
\xi(|\psi\rangle \langle \psi|) = 1
\]

Furthermore, the states for which it is most difficult to do hypothesis testing are the ones corresponding to projectors on a 2-dimensional subspace:

\[
\xi \left( \frac{1}{2} \sum_{\alpha=1}^{2} |\alpha\rangle \langle \alpha| \right) = \frac{2}{3}
\]

\( \sigma = \frac{P}{2} \). However, there is clearly not a big discrepancy between \( \frac{2}{3} \) and 1, so the test will perform well for any state \( \sigma \).

Another interesting class of states contains all maximally mixed states: here

\[
\xi(\mathbb{I}/D) = \frac{1}{1 + 1/D}
\]

Finally, \( \xi(\sigma) \) can be calculated analytically for any density matrix defined on a 2-level system:

\[
\xi(\sigma) = \frac{1}{1 + 2\lambda_1\lambda_2}
\]

Following the constructive proof of the lower bound, A POVM with 6 elements that saturates this is given by

\[
\left\{ (1 - \xi(\sigma)) |0\rangle \langle 0|, (1 - \xi(\sigma)) |1\rangle \langle 1|, \frac{\xi(\sigma)}{2} |+\rangle \langle +|, \frac{\xi(\sigma)}{2} |-\rangle \langle -|, \frac{\xi(\sigma)}{2} |i\rangle \langle i|, \frac{\xi(\sigma)}{2} |-i\rangle \langle -i| \right\}
\]

where we work in the basis where \( \sigma \) is diagonal, and with \( |\pm\rangle \) and \( |\pm i\rangle \) the eigenbasis of the Pauli matrices \( \sigma_x \) and \( \sigma_y \). An optimal \( \chi^2 \) test with 3 degrees of freedom on \( N \) samples is then obtained by doing \((1 - \xi).N\) measurements in the computational basis and \( \xi.N/2 \) in both the \( \sigma_x \) and \( \sigma_y \) basis.
IV. DISCUSSION

We have argued that hypothesis testing provides the natural framework for describing quantum experiments that aim at verifying that a certain density matrix is prepared in the lab. This evades the artificial problems of enforcing positivity etc. encountered in quantum tomography.

We have studied the problem of hypothesis testing and goodness of fit testing of density matrices, and have focused on the $\chi^2$ test; as shown, most of the results that we derived also directly apply to the loglikelihood test. This provides a clear, simple and flexible framework for testing whether a given density matrix is produced by a certain experimental setup, and allows to define confidence intervals that are independent of the particular system under consideration. We were also able to characterize divergence rates $\xi(\sigma)$ by doing an optimization over all possible POVM measurements maximizing the information, and proved that $2/3 \leq \xi \leq 1$. This allowed to prove that, if we were sampling from a different density matrix $\rho$ instead of $\sigma$, that this would be detected in a number of measurements proportional to $D/(\xi(\sigma)\|\rho - \sigma\|^2)$ with $D$ the dimension of the Hilbert space. Furthermore, we showed that this measurement is both optimal from the point of view of Pitman and Bahadur efficiency.

Up till now, we assumed that there were no measurement errors. If such errors occur and the error model is known, the expected probabilities can easily be adjusted, and exactly the same analysis carries through.

We have also not yet touched upon the problem of the estimation of the parameters describing the density matrix (in principle, the $\chi^2$ test can also be used to estimate all or a few parameters in the density matrix, and it turns out that this estimator is asymptotically optimal; this will be discussed in future work.). The problem of estimation is obviously complementary to the topic of hypothesis testing; but independent of which procedure used to estimate the density matrix, the procedure should always be complemented by doing hypothesis testing on an independent sample set. The big advantage of hypothesis testing versus parameter estimation is the exponential scaling of the confidence in the number of samples: if the measurement data is compatible with the expected ones, we accept the hypothesis, and otherwise we reject it. Note also that continuous distributions can be tested by the $\chi^2$ method; this can be achieved by binning the data.

Also, we did not consider the question of entangled measurements on different copies. Just as in the case of tomography, this is a bit problematic as the total number of degrees of freedom increases exponentially with the number of copies on which joint measurements are done. However, it is possible to circumvent this problem and consider POVM measurements with few elements that only reveal full information about the marginals; this will also be discussed in future work.

From a more philosophical point of view, the topic of hypothesis testing forces us to rethink what it means for a quantity to be physical and what not. For example, the expectation value of an observable is not observable, but can only be sampled. The resulting fluctuations are an entire part of doing an experiment, and if an experiment would report frequencies that are too close or too far from the expected ones, then such an experiment can be categorized as suspicious. The only thing that is physical are the frequencies by which certain measurement outcomes are obtained, and the only goal of quantum mechanics is the prediction of those frequencies.

Acknowledgments

This work was supported by the EU Strep project QUEVADIS, the ERC grant QUERG, and the FWF SFB grants FoQuS and ViCoM. An important part of this work was done at Stony Brook.

[1] A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett. 49, 1804 (1982);
[2] C. Monroe, D. M. Meekhof, B. E. King, W. M. Itano, and D. J. Wineland, Phys. Rev. Lett. 75, 4714 (1995);
[3] D. Bouwmeester, J.W. Pan, K. Mattle, M. Eibl, H. Weinfurter and A. Zeilinger, Nature 390, 575 (1997);
[4] A. Furusawa, J. L. Sorensen, S. L. Braunstein, C. A. Fuchs, H. J. Kimble and E. S. Polzik, Science 282 (5389), 706 (1998);
[5] P. W. H. Pinkse, T. Fischer, P. Maunz and G. Rempe, Nature 404, 365 (2000);
[6] M. Riebe, H. Häffner, C. F. Roos, W. Hänsel, J. Benhelm, G. P. T. Lancaster, T. W. Körber, C. Becher, F. Schmidt-Kaler, D. F. V. James and R. Blatt, Nature 429, 734 (2004)
[7] R. A. Fisher, Statistical Methods, Experimental Design and Scientific Inference (Oxford University Press, Oxford, 1990)
[8] K. Pearson, Philosophical Magazine Series 5, Vol. 50, Iss. 302, 1900
[9] H. Cramer, Mathematical Methods of Statistics, (Princeton University Press, 1999)
[10] C. W. Helstrom, Quantum Detection and Estimation Theory (Academic, New York, 1976)
[11] A.S. Holevo, Probabilistic and Statistical Aspects of Quantum Theory (Nauka, 1980)
[12] W.K. Wootters, Phys. Rev. D 23, 357 (1981); W.K. Wootters, Ann. of Phys. 191, 363 (1989)
[13] S.L. Braunstein and C.M. Caves, Phys. Rev. Lett. 72, 3439 (1994)
[14] F. Hiai and D. Petz, Comm. in Math. Phys. 143, 99 (1991)
[15] M. Hayashi, Quantum Information: An Introduction (Springer, 2006).
[16] D. Petz, Quantum Information Theory and Quantum Statistics (Springer, 2008)
[17] R. Blume-Kohout, J. O.S. Yin, S. J. van Enk, Phys. Rev. Lett. 105, 170501 (2010)
[18] M.P. da Silva, O. Landon-Cardinal, D. Poulin, Phys. Rev. Lett. 107, 210404 (2011)
[19] S.T. Flammia, Y.K. Liu, Phys. Rev. Lett. 106, 230501 (2011).
[20] M. Christandl, R. Renner, arXiv:1108.5329 (2011)
[21] K.M.R. Audenaert et al., Phys. Rev. Lett. 98, 160501 (2007); K.M.R. Audenaert, M. Nussbaum, A. Szkoła, F. Verstraete, Comm. Math. Phys. 279, 251-283 (2008).
[22] K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf, F. Verstraete, J. Math. Phys. 51, 122201 (2010);
[23] David Reeb, Michael J. Kastoryano, Michael M. Wolf, J. Math. Phys. 52, 082201 (2011)
[24] E.A. Morozova and N.N. Chentsov, Itogi Nauki i Tekhniki 36:69-102 (1990).
[25] D. Petz, Reports on Math. Phys. 23, 57 – 65 (1986).
[26] D. Petz, Linear Algebra. Appl. 244, 81 – 96 (1996).
[27] D. Petz and Cs. Sudar, J. Math. Phys. 37, 2662 – 2673 (1996).
[28] D. Petz, M.B. Ruskai, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 1(1998), 83 – 90
[29] A. Lesniewski, M.B. Ruskai, J. Math. Phys. 40 (1999), 5702 5724
[30] D. J. C. Bures, Trans. Am. Math. Soc. 135, 199 (1969);
[31] A. Uhlmann, Rep. Math. Phys. 9, 273 (1976).
[32] M. A. Woodbury, The Stability of Out-Input Matrices. Chicago, Ill., 1949. 5 pp
[33] G.E. Noether, On a theorem by Pitman., Ann. Math. Statist., 26, 64-68. (1955)
[34] R.R. Bahadur, An optimal property of the likelihood ratio statistic. Proc. of the 5th Berkeley Symp. on Math. Stat. and Prob. (1965/66) 1,13-26.