Quantum State Certification

Costin Bădescu
Computer Science Department,
Carnegie Mellon University
USA
cbadescu@cs.cmu.edu

Ryan O'Donnell
Computer Science Department,
Carnegie Mellon University
USA
odonnell@cs.cmu.edu

John Wright
Center for Theoretical Physics,
Massachusetts Institute of Technology
USA
jswright@mit.edu

ABSTRACT

We consider the problem of *quantum state certification*, where one is given n copies of an unknown d -dimensional quantum mixed state ρ , and one wants to test whether ρ is equal to some known mixed state σ or else is ϵ -far from σ . The goal is to use notably fewer copies than the $\Omega(d^2)$ needed for full tomography on ρ (i.e., density estimation). We give two robust state certification algorithms: one with respect to fidelity using $n = O(d/\epsilon)$ copies, and one with respect to trace distance using $n = O(d/\epsilon^2)$ copies. The latter algorithm also applies when σ is unknown as well. These copy complexities are optimal up to constant factors.

CCS CONCEPTS

• **Theory of computation** → **Quantum computation theory**;
Sample complexity and generalization bounds.

KEYWORDS

quantum property testing

ACM Reference Format:

Costin Bădescu, Ryan O'Donnell, and John Wright. 2019. Quantum State Certification. In *Proceedings of the 51st Annual ACM SIGACT Symposium on the Theory of Computing (STOC '19)*, June 23–26, 2019, Phoenix, AZ, USA. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3313276.3316344>

1 INTRODUCTION

A key step in building quantum devices is verifying that they work as intended. Typically, a quantum device is designed with the intent of outputting some known d -dimensional (mixed) state $\sigma \in \mathbb{C}^{d \times d}$, but the possibility of imperfections in the device's construction and noise in the device's operation mean that its actual output state $\rho \in \mathbb{C}^{d \times d}$ is unknown. *Quantum state certification* refers to the problem of testing whether ρ equals σ or is far from σ , given the ability to produce $\rho^{\otimes n}$ (i.e., n copies of ρ). This is the quantum (noncommutative) generalization of the classical statistical problem of testing identity of probability distributions [Can15].

A standard approach for quantum state certification is to first estimate ρ from $\rho^{\otimes n}$ using a *quantum state tomography (estimation)* procedure, then to check that the estimate is close to σ . Given

that ρ has $d^2 - 1$ real parameters, it is natural that the number of copies needed to estimate it should scale roughly as d^2 . This was confirmed in a trio of recent papers [HHJ⁺16, OW16, OW17]; among other things, those works show that $n = \tilde{\Theta}(d^2/\epsilon)$ copies of ρ are necessary and sufficient to produce an estimate $\hat{\rho}$ satisfying the fidelity bound $F(\rho, \hat{\rho}) \geq 1 - \epsilon$. (See Section 2.1 for more on prior work, and Section 3.1 for a review of distance measures such as fidelity, trace distance, χ^2 -divergence, etc.)

Unfortunately, even small scale quantum systems can have large dimension; for example, a system of q qubits has $d = 2^q$ dimensions. For such systems, the quadratic scaling in d required by full tomography (density estimation) can be prohibitively expensive. For example, a 2005 experiment [HHR⁺05] designed to produce the entangled 8-particle W -state ($d = 256$) used $n = 656100$ copies to estimate the actually-produced state. (The fidelity to the target state ended up being estimated as .85.)

However for the quantum state certification problem, the goal is not to learn the unknown state $\rho \in \mathbb{C}^{d \times d}$ but merely to test whether it is close to a target σ , or far from it. Learning the entire density matrix might be wasting copies of ρ to gain irrelevant information. As such, it is natural to ask: can we outperform tomography?

1.1 Our Results

In this work, we give a unified framework for analyzing the number of copies of ρ needed to estimate symmetric polynomials of ρ 's spectrum and hence perform various quantum state certification tasks. One of our main results is the following:

THEOREM 1.1. *For a fixed mixed state $\sigma \in \mathbb{C}^{d \times d}$ and $\epsilon > 0$, there is an algorithm that, given $n = O(d/\epsilon)$ copies of ρ , distinguishes (with high probability¹) between the cases $\rho = \sigma$ and $F(\rho, \sigma) < 1 - \epsilon$.*

In fact, the above theorem is a simplified (strictly weaker) form of the following *robust* state certification result. Here “robustness” refers to the fact that the testing algorithm accepts not just if $\rho = \sigma$ but also if ρ is sufficiently close to σ .

THEOREM 1.2. *Let $\sigma \in \mathbb{C}^{d \times d}$ be a fixed mixed state, and let $\epsilon > 0$. There is an algorithm that, given $n = O(d/\epsilon)$ copies of ρ , performs a measurement and then reports either “close” or “far”. The algorithm has the following guarantee (whp): If it reports “close” then we have the fidelity bound $F(\rho, \sigma) \geq 1 - \epsilon$. If it reports “far” then we have the Bures χ^2 -divergence² bound $D_{\chi^2}(\rho \parallel \sigma) > .49\epsilon$.*

To put it another way, if $D_{\chi^2}(\rho \parallel \sigma) \leq .49\epsilon$ (in particular, if $\rho = \sigma$) then the algorithm reports “close” and if $F(\rho, \sigma) < 1 - \epsilon$ then the

¹Henceforth abbreviated “whp”. We may take this to mean probability at least, say, $2/3$; however, by standard means this probability can be boosted to $1 - \delta$ at the expense of multiplying n by $O(\log(1/\delta))$.

²The Bures χ^2 -divergence is reviewed in Section 3.1.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
STOC '19, June 23–26, 2019, Phoenix, AZ, USA
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6705-9/19/06...\$15.00
<https://doi.org/10.1145/3313276.3316344>

algorithm reports “far” (whp). It should be noted that these notions of “close” and “far” in [Theorem 1.2](#) are nearly complementary, since it’s known that every pair of states ρ, σ satisfies either $F(\rho, \sigma) \geq 1 - \epsilon$ or $D_{\chi^2}(\rho \parallel \sigma) > .5\epsilon$.

[Theorem 1.2](#) is an easy consequence (see [Section 6.3](#)) of the following even stronger result, a certification procedure for “well-conditioned” states that is robust with respect to Bures χ^2 -divergence:

THEOREM 1.3. *Let $c > 0$ be any small constant. Fix a d -dimensional mixed state σ with smallest eigenvalue at least ce^2/d . Then there is an algorithm that, given $n = O(d/\epsilon^2)$ copies of ρ , (whp) outputs “close” if $D_{\chi^2}(\rho \parallel \sigma) \leq .99\epsilon^2$ and outputs “far” if $D_{\chi^2}(\rho \parallel \sigma) > \epsilon^2$.*

We also obtain a new sample-efficient certification algorithm in the case of *two unknown* states. Here one is given n copies each of mixed states ρ, σ and one wants to distinguish whether $\rho = \sigma$ or ρ is far from σ . Our main result here uses trace distance as the figure of merit, and it again improves on the copy-complexity of full tomography (namely, $\Theta(d^2/\epsilon^2)$ [[HHJ⁺16](#), [OW16](#)]) by a factor of d :

THEOREM 1.4. *There is an algorithm that, given $n = O(d/\epsilon^2)$ copies each of unknown mixed states $\rho, \sigma \in \mathbb{C}^{d \times d}$, (whp) distinguishes between the cases $\rho = \sigma$ and $D_{\text{tr}}(\rho, \sigma) > \epsilon$.*

We stated this theorem as above for simplicity, but with slightly more care (see [Section 5.3](#)) one also derive the following strictly stronger and much more precise result for trace-distance certification. It has both “robustness”, and significantly improved sample complexity when one of the states is close to having low rank:

COROLLARY 1.5. *Assume that one of the two unknown states — say, σ — is close to having rank at most k , in the sense that the sum of its largest k eigenvalues is at least $1 - \delta$. Then there is an algorithm that, given $n = O(k/\epsilon^2)$ copies each of $\rho, \sigma \in \mathbb{C}^{d \times d}$, (whp) distinguishes between the cases $D_{\text{HS}}(\rho, \sigma) \leq .58\epsilon$ and $D_{\text{tr}}(\rho, \sigma) > \delta + \epsilon$. (The constant .58 can be anything smaller than $2 - \sqrt{2}$.)*

Note that the above result has *no* sample complexity dependence on d . It is also already interesting even the case when one state σ is a known state (in which case the algorithm can simply prepare n copies of σ by itself).

We prove our [Theorem 1.4](#) (and [Corollary 1.5](#)) as a direct consequence — via Cauchy–Schwarz — of the following two-unknown-state certification algorithm, robust with respect to the Hilbert–Schmidt distance:

THEOREM 1.6. *There is an algorithm that, given $n = O(1/\epsilon^2)$ copies each of unknown mixed states $\rho, \sigma \in \mathbb{C}^{d \times d}$, (whp) outputs “close” if $D_{\text{HS}}(\rho, \sigma) \leq .99\epsilon$ and outputs “far” if $D_{\text{HS}}(\rho, \sigma) > \epsilon$.*

A Few Words Regarding the Hilbert–Schmidt Distance. We briefly emphasize here that [Theorem 1.6](#), our theorem with Hilbert–Schmidt distance as the figure of merit, is only a “means to an end”. Namely, it immediately implies our [Theorem 1.4](#) via the Cauchy–Schwarz bound $D_{\text{tr}}(\rho, \sigma) \leq \sqrt{d} \cdot D_{\text{HS}}(\rho, \sigma)$. We do not particularly advocate [Theorem 1.6](#) in and of itself, since Hilbert–Schmidt distance is not especially operationally meaningful for mixed states. We would also like to emphasize the following somewhat subtle point. There is a very obvious

way to estimate $D_{\text{HS}}(\rho, \sigma)$ for two unknown states, by repeatedly performing the SWAP Test. It’s very easy to calculate that $O(1/\epsilon^2)$ copies suffice to (whp) estimate $D_{\text{HS}}^2(\rho, \sigma)$ to within ϵ , but note this is for the *squared* Hilbert–Schmidt distance. For the (unsquared) Hilbert–Schmidt distance itself, the SWAP Test therefore requires $O(1/\epsilon^4)$ copies. This should be compared with our [Theorem 1.6](#), which uses only $O(1/\epsilon^2)$ copies. At first one might not be overly concerned with the distinction between quadratic and quartic dependence on the error parameter ϵ , but in fact it makes all the difference! As mentioned, if we use Cauchy–Schwarz to convert Hilbert–Schmidt distance to the operationally meaningful trace distance, then a $O(1/\epsilon^4)$ -copy Hilbert–Schmidt algorithm leads to a $O(d^2/\epsilon^4)$ -copy trace distance algorithm, which is *worse* than the trivial technique of using tomography. It is only by obtaining the tight $O(1/\epsilon^2)$ -copy Hilbert–Schmidt algorithm that one can obtain the tight $O(d/\epsilon^2)$ -copy trace distance algorithm.

Finally, we observe that even the simplest versions of our results — [Theorem 1.1](#) and [Theorem 1.4](#) — have optimal sample complexity (up to a constant). Indeed, this is the case even when σ is promised to be the maximally mixed state $1/d$. This is a consequence of the following lower bound from [[OW15](#)]:

THEOREM 1.7 ([[OW15](#)]). *Given even d and $0 \leq \epsilon \leq 1/2$, let $\sigma = 1/d$ and let C_ϵ denote the class of states with eigenvalues $\frac{1+2\epsilon}{d}, \frac{1-2\epsilon}{d}, \frac{1+2\epsilon}{d}, \frac{1-2\epsilon}{d}, \dots, \frac{1+2\epsilon}{d}, \frac{1-2\epsilon}{d}$. For any $\rho \in C_\epsilon$, one has*

$$D_{\text{tr}}(\rho, \sigma) = \epsilon, \quad F(\rho, \sigma) = 1 - \frac{1}{2}\epsilon^2 - O(\epsilon^4),$$

$$D_{\text{HS}}(\rho, \sigma) = 2\epsilon/\sqrt{d}, \quad D_{\chi^2}(\rho \parallel \sigma) = 4\epsilon^2 + O(\epsilon^4).$$

Then any measurement strategy that can distinguish (with probability advantage at least $1/3$) the case $\rho = \sigma$ from the case $\rho \in C_\epsilon$ using n samples from ρ must have $n > .15d/\epsilon^2$.

Our quantum certification algorithm from [Theorem 1.6](#) is not just copy-efficient, it can be carried out by polynomial-sized (i.e., $\text{poly}(n, d)$ -gate) quantum circuits. The details of the implementation appear in the full version of the paper.

1.2 Outline of the Remainder of the Paper

In [Section 2](#) we review prior work on quantum tomography and state discrimination. In [Section 3.1](#) we recall various measures of quantum state distance that will be important in this work. [Sections 3.2](#) and [3.3](#) are devoted to background on quantum probability and representation theory. In [Section 4](#), we develop a framework for finding the most efficient (lowest-variance) estimators for symmetric polynomial functions of unknown quantum states. These results are not strictly necessary for our proof of [Theorem 1.6](#) in [Section 5](#); however, they justify that the estimators used therein are optimal. [Section 6](#) contains our proof of [Theorems 1.2](#) and [1.3](#), as well as a diagonality tester for quantum states.

2 PRIOR WORK ON QUANTUM DENSITY TESTING/ESTIMATION

In this section we review some results on learning and testing unknown quantum states. As this area is extremely broad, we cannot completely review all known literature; we will simply give pointers to some of the best known and most relevant results.

2.1 Prior Quantum Density Estimation, Testing, and Certification

2.1.1 Tomography (Density Estimation). Before discussing state certification, we start by reviewing the best known results for the baseline problem of *tomography*; i.e., producing an estimate $\hat{\rho}$ of an unknown density matrix $\rho \in \mathbb{C}^{d \times d}$, given n copies $\rho^{\otimes n}$, up to error ϵ (whp) for some notion of “distance”. We will also let k denote the *rank* of ρ , which is 1 when ρ is a pure state, and in general is at most d . The best results achievable depend on the “figure of merit” — i.e., distance measure — chosen (see Section 3.1 for a review).

In [HHJ⁺16] it was shown that $n = O(kd/\epsilon) \cdot \log(d/\epsilon)$ copies suffice to obtain infidelity ϵ (i.e., $F(\rho, \hat{\rho}) \leq 1 - \epsilon$); this also implies that $n = O(kd/\epsilon^2) \cdot \log(d/\epsilon)$ copies suffice to obtain trace distance ϵ (i.e., $D_{\text{tr}}(\rho, \hat{\rho}) \leq \epsilon$). Those authors also showed that $n = \Omega(kd/\epsilon^2)/\log(d/k\epsilon)$ copies are necessary, with the log factor being removable in the case $k = d$. Independently, in [OW16] it was shown that $n = O(d/\epsilon^2)$ copies suffice to obtain Hilbert–Schmidt distance ϵ (i.e., $D_{\text{HS}}(\rho, \hat{\rho}) \leq \epsilon$); this also implies a copy complexity of $n = O(kd/\epsilon^2)$ for trace distance (slightly better than in [HHJ⁺16]). More generally, [OW16] showed a kind of “PCA” result: for ρ of any rank, $n = O(kd/\epsilon^2)$ copies suffice to produce an estimate $\hat{\rho}$ whose trace distance from ρ is at most ϵ more than that of the best rank- k approximator. Finally, a followup work [OW17] gave an alternate proof of the $n = O(kd/\epsilon) \cdot \log(d/\epsilon)$ bound for infidelity, showed also an $n = O(k^2d/\epsilon)$ bound, and extended these bounds to the PCA case.

2.1.2 Density Testing. Tomography results suffer from the inherent issue that $n = \Theta(d^2)$ copies are needed in the general case (except when the figure of merit is Hilbert–Schmidt distance, but this metric is not considered to be very meaningful, operationally). Thus as mentioned, it is natural to focus on restricted problems like state certification, distance estimation, and other *property testing* problems that can potentially be carried out with $n = O(d)$ or better. Montanaro and de Wolf [MW16] have given an excellent survey on property testing of quantum states; we review a few of the known results here.

A typical quantum property testing problem would involve two disjoint classes C_1, C_2 of d -dimensional quantum states; given n copies of an unknown ρ , promised to be in either C_1 or C_2 , the task is to distinguish which is the case (whp) using few copies of ρ . In particular, the *quantum state certification* problem for fixed state $\sigma \in \mathbb{C}^{d \times d}$ is the case when $C_1 = \{\sigma\}$ and $C_2 = \{\rho : D(\rho, \sigma) > \epsilon\}$ for some notion $D(\cdot, \cdot)$ of distance and some parameter ϵ .

When σ is a *pure* state, it is straightforward to show (see, e.g., [MW16]) that the associated quantum state certification task, with infidelity as the distance measure, can be done using $n = O(1/\epsilon)$ copies (and this implies $n = O(1/\epsilon^2)$ copies suffice for trace distance). Indeed, the same is possible when both ρ and σ are unknown pure states, and one is given n copies of each. For practical purposes, it may be useful to have a state certification algorithm for a known pure σ that only uses simple measurements; e.g., Pauli observables. For this problem, it has been shown [FL11, dSLP11, AGKE15] that for σ known and pure, one can solve the certification problem given $n = O(d/\epsilon^2)$ copies of an

unknown ρ with infidelity as the distance metric — indeed, with this many copies one can estimate the fidelity $F(\rho, \sigma)$ to $\pm\epsilon$.

For the state certification problem when σ is mixed (not pure), not much is known except in one case: when $\sigma = 1/d$, the “maximally mixed” state. For this problem, it was shown in [OW15] that $n = \Theta(d/\epsilon^2)$ copies are necessary and sufficient, when the distance measure is trace distance. In fact, for the $n = O(d/\epsilon^2)$ upper bound, [OW15] effectively show that one can estimate the *purity* $\text{tr}(\rho^2)$ of ρ sufficiently well so as to distinguish between purity $1/d$ (achieved by the maximally mixed state) and purity exceeding $1/d + \epsilon^2/d$. Note that the latter case is equivalent to ρ being $\epsilon/\sqrt{d}$ -far from $1/d$ in Hilbert–Schmidt distance and ϵ^2 -far from $1/d$ in Bures χ^2 -divergence. The lower bound was mentioned earlier as Theorem 1.7.

2.1.3 The Asymptotic Regime for State Discrimination. There is a related class of work that we refer to as the “asymptotic regime”. Consider the simplest quantum property testing problem, *state discrimination*, in which $C_1 = \{\sigma_1\}$ and $C_2 = \{\sigma_2\}$ for two known states $\sigma_1, \sigma_2 \in \mathbb{C}^{d \times d}$. The perspective we take in this paper involves determining the least number of copies n such that one can distinguish $\rho = \sigma_1$ from $\rho = \sigma_2$ with high probability — say, with both “type I” and “type II” errors having probability at most $\delta = 1/3$. One can reduce this δ to any small positive constant at the expense of making n a constant factor larger. We refer to this perspective as the *non-asymptotic regime*, because we do not consider any limiting error rate as $n \rightarrow \infty$; rather, we wish to find a concrete upper bound on the n that suffices, depending only on d , the distance between σ_1 and σ_2 , and nothing else.

On the other hand, there is substantial work on the *asymptotic regime*, sometimes going under the name *quantum hypothesis testing*, in which the focus is on how exponentially fast the error rate goes to 0 in the limit as $n \rightarrow \infty$. Here one might seek the best (smallest) constant R such that, given n copies, one can ensure type I and type II errors have probability at most $(R + o(1))^n$, where the $o(1)$ refers to $n \rightarrow \infty$. A downside of such results is that they do not a priori give any information about how large n needs to be before error bounds “kick in”; e.g., the $o(1)$ function might not be less than, say, .1 until n is larger than some uncontrolled function of d (e.g., 2^d) or of some other parameters (e.g., the smallest nonzero eigenvalue of σ_1 or σ_2).

A good survey of the results in the asymptotic regime appears in [ANSV08]; they review known quantum versions of Stein’s Lemma and Sanov’s Theorem, and prove quantum versions of Chernoff’s Bound and the Hoeffding–Blahut–Csiszár–Longo bound. For example, in the basic hypothesis testing problem described above, they prove that the best rate R is given by $Q_{\min}(\sigma_1 \parallel \sigma_2) = \min_{0 \leq s \leq 1} \text{tr}(\sigma_1^s \sigma_2^{1-s})$ (a quantity that is within a factor of 2 of the infidelity between σ_1 and σ_2).

We end this section by stating a version of Chebyshev’s inequality which will be useful in what follows.

LEMMA 2.1. *Let $X^{(n)}$ be a sequence of estimators for a number $\mu \geq 0$, meaning $E[X^{(n)}] = \mu$ for all n . Suppose we have a variance bound of the form*

$$\text{Var}[X^{(n)}] \leq O\left(\frac{b(\mu)}{n^2} + \frac{v(\mu)}{n}\right), \quad (1)$$

where

$$b(\mu), v(\mu), \frac{\mu^2}{b(\mu)}, \frac{\mu^2}{v(\mu)} \text{ are increasing functions of } \mu \geq 0. \quad (2)$$

(The $O(\cdot)$ should hide a universal constant.) Let $\theta > 0$ be a parameter. Then provided

$$n \geq C \max \left\{ \sqrt{\frac{b(\theta)}{\theta^2}}, \frac{v(\theta)}{\theta^2} \right\}, \quad (3)$$

one can use $X^{(n)}$ to distinguish (with high probability) whether $\mu \leq .99\theta$ or $\mu > \theta$. Here C is another universal constant. (More generally, to achieve $1 - \gamma$ in place of .99, one should take $\gamma^2\theta^2$ in place of θ^2 in the denominators in (3).)

PROOF. We report “ $\mu \leq .99\theta$ ” if $X^{(n)} \leq .995\theta$ and report “ $\mu > \theta$ ” if $X^{(n)} > .995\theta$. The analysis is straightforward by Chebyshev’s bound. $\square$

3 PRELIMINARIES

3.1 Classical and Quantum Distances and Divergences

3.1.1 Distances and Divergences for Quantum Mixed States. There are again many distances and divergences used for comparing two quantum states ρ and σ ; see, e.g., [GLN05], [BZ07, Chap. 13], [Aud12] for some surveys. All of the quantum distances we review will be *unitarily invariant*, meaning that $D(\rho, \sigma) = D(U\rho U^\dagger, U\sigma U^\dagger)$ for all unitaries U .

Many classical distances have a quantum analogue, and indeed some have *several* quantum analogues. Typically, a quantum distance between ρ and σ reduces to the analogous classical distance between p and q in the case that $\rho = \text{diag}(p)$ and $\sigma = \text{diag}(q)$ are diagonal.

In particular, for every classical f -divergence one can form either the “standard quantum f -divergence” (introduced by Petz) or the “measured quantum f -divergence” — see [HM17]. We will only consider the latter. Given a classical f -divergence $d_f(\cdot, \cdot)$, one obtains the corresponding measured quantum f -divergence $D_f(\cdot, \cdot)$ as follows:

$$D_f(\rho, \sigma) = \sup_{\text{POVMs } \{E_i\}_{i=1}^N} \{d_f(p_\rho, p_\sigma)\}, \quad (4)$$

where $p_\xi = (\text{tr}(\xi E_1), \dots, \text{tr}(\xi E_N))$.

In other words, the quantum divergence is defined as the maximum classical divergence that can be achieved when applying the same POVM to both states. In this section we will encounter the measured quantum f -divergence corresponding to total variation distance, squared Hellinger distance, and χ^2 -divergence.

Definition 3.1. The *trace distance* between ρ and σ is

$$D_{\text{tr}}(\rho, \sigma) = \frac{1}{2} \|\rho - \sigma\|_1.$$

The trace distance is a metric and it has a maximum value of 1, occurring when ρ and σ have orthogonal support. Helstrom [Hel76] showed that trace distance is the measured version of classical total variation distance in the sense of Equation (4). It therefore equals the maximum probability with which the states ρ and σ

can be discriminated by some measurement. It also follows that it satisfies the “quantum data processing inequality”. In other words, it can never increase when the same quantum channel (completely positive trace-preserving map) is applied to both ρ and σ .

Definition 3.2. The *Hilbert–Schmidt distance* (or *Frobenius distance*) between ρ and σ is

$$D_{\text{HS}}(\rho, \sigma) = \|\rho - \sigma\|_{\text{HS}} = \left(\sum_{i,j=1}^d |\rho_{ij} - \sigma_{ij}|^2 \right)^{1/2} = \text{tr}((\rho - \sigma)^2)^{1/2}.$$

This metric can be seen as analogue of the classical ℓ_2 distance. It is not, however, a direct analogue in the sense of Equation (4); this is related to the fact that it does not satisfy the quantum data processing inequality. Nevertheless, it is useful by virtue of the fact that the squared Hilbert–Schmidt distance, $D_{\text{HS}}^2(\rho, \sigma) = \text{tr}((\rho - \sigma)^2)$, is extremely easy to compute, and that it can be related to the trace distance via Cauchy–Schwarz for matrices: $\frac{1}{2} D_{\text{HS}}(\rho, \sigma) \leq D_{\text{tr}}(\rho, \sigma) \leq \frac{1}{2} \sqrt{d} \cdot D_{\text{HS}}(\rho, \sigma)$.

Definition 3.3. The *Bures distance* between ρ and σ is

$$D_{\text{B}}(\rho, \sigma) = (2(1 - F(\rho, \sigma)))^{1/2},$$

where

$$F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1$$

is the *fidelity* between ρ and σ . (The quantity $1 - F(\rho, \sigma)$ is termed the *infidelity*).

The Bures distance is a metric and it has a maximum value of $\sqrt{2}$, occurring when ρ and σ have orthogonal support. The work of Fuchs and Caves [FC95] shows that the (squared) Bures distance is the measured version of classical (squared) Hellinger distance in the sense of Equation (4). It follows that $\frac{1}{2} D_{\text{B}}^2(p, q) \leq D_{\text{tr}}(p, q) \leq D_{\text{B}}(p, q)$. It also follows that the Bures distance satisfies the quantum data processing inequality.

We more often consider the square of the Bures distance, $D_{\text{B}}^2(\rho, \sigma)$, which is simply twice the infidelity. It is also quite common to consider the squared fidelity, $F^2(\rho, \sigma)$. The squared fidelity, as shown by Uhlmann [Uhl76], is the maximum overlap between purifications of ρ and σ , where the *overlap* of (mixed) quantum states ρ' and σ' is defined to be $\text{tr}(\rho'\sigma')$.

Note that when ρ and σ are “close”, with $F(\rho, \sigma) = 1 - \epsilon$, we have that $1 - F^2(\rho, \sigma) \approx 2\epsilon$. Thus there is not much difference if one defines infidelity as $1 - F(\rho, \sigma)$ or $1 - F^2(\rho, \sigma)$; these quantities are always within a factor 2 of each other, and also of the squared Bures distance. Also very closely related is the *quantum Hellinger affinity*, $Q_{1/2}(\rho, \sigma) = \text{tr}(\sqrt{\rho}\sqrt{\sigma})$. It satisfies $F^2(\rho, \sigma) \leq Q_{1/2}(\rho, \sigma) \leq F(\rho, \sigma)$ and has been used to define a “quantum Hellinger distance” by $D_{\text{H}}^2(\rho, \sigma) = 2(1 - Q_{1/2}(\rho, \sigma))$; see [ANSV08]. The same bound $F^2(\rho, \sigma) \leq Q_{\min}(\rho, \sigma) \leq F(\rho, \sigma)$ also holds [Aud12] for the quantity $Q_{\min}(\rho \parallel \sigma)$ arising in the quantum Chernoff bound mentioned in Section 2.1.3.

Definition 3.4. Assume σ has full rank. The *Bures χ^2 -divergence* of ρ from σ is

$$D_{\chi^2}(\rho \parallel \sigma) = \text{tr}((\rho - \sigma) \cdot \Omega_\sigma(\rho - \sigma)),$$

where Ω_σ is the linear operator whose inverse is defined by $\Omega_\sigma^{-1}(A) = \frac{1}{2}(\sigma A + A\sigma)$. (There is a simple generalization to the case

where σ does not have full rank, so long as ρ 's support is a subset of σ 's; we will not need it, however.) In case $\sigma = \text{diag}(\beta_1, \dots, \beta_d)$, we obtain the following more explicit formula:

$$D_{\chi^2}(\rho \parallel \sigma) = \sum_{i,j=1}^d \frac{2}{\beta_i + \beta_j} |\Delta_{ij}|^2, \quad \text{where } \Delta = \rho - \sigma.$$

The Bures χ^2 -divergence is the measured version of the classical χ^2 -divergence in the sense of Equation (4), as shown in [BC94, TV15]. As such, it satisfies the quantum data processing inequality, and we can infer from the classical case that $D_B^2(\rho, \sigma) \leq D_{\chi^2}(\rho \parallel \sigma)$. Indeed, it is known [TKR⁺10] that the *quantum relative entropy*, $S(\rho \parallel \sigma) = \text{tr}(\rho(\log \rho - \log \sigma))$ is sandwiched in between: $D_B^2(\rho, \sigma) \leq S(\rho \parallel \sigma) \leq D_{\chi^2}(\rho \parallel \sigma)$. As in the classical case, such bounds are what makes the Bures χ^2 -divergence useful, together with its having a relatively simple formula when σ is considered to be “fixed”.

We close this section by commenting that, although we focus on Bures χ^2 -divergence, there are many generalizations of χ^2 -divergence to the quantum case. For example, the “standard quantum f -divergence” version is $\text{tr}((\rho - \sigma)^2 \sigma^{-1})$. More generally, one may consider $\text{tr}((\rho - \sigma) \sigma^{-\alpha} (\rho - \sigma) \sigma^{\alpha-1})$ for any $\alpha \in [0, 1/2]$, and there are further possibilities. See, e.g., [Pet96, TKR⁺10], wherein it is explained that the Bures χ^2 -divergence takes on the smallest value among a wide family of generalizations.

3.2 Quantum Probability

Let V be a finite-dimensional vector space over $\mathbb{C}$ and let $\text{End}(V)$ denote the algebra of linear operators on V . An operator $X \in \text{End}(V)$ is self-adjoint or Hermitian if $X^\dagger = X$, where $X^\dagger$ denotes the conjugate-transpose of X ; X is positive if there exists an operator $Y \in \text{End}(V)$ such that $X = Y^\dagger Y$. For self-adjoint operators $X, Y \in \text{End}(V)$ we write $X \leq Y$ provided $Y - X$ is positive. The identity operator is denoted by 1 , with the dimension of the underlying vector space being inferred from the context.

Definition 3.5. A *quantum state* ϱ is defined to be a positive operator $\varrho \in \text{End}(V)$ with $\text{tr}(\varrho) = 1$.

Definition 3.6. A *positive-operator valued measurement* (POVM) $\mathcal{M}$ consists of a set of positive operators that sum to the identity operator 1 . When a measurement $\mathcal{M} = \{E_1, \dots, E_k\}$ is applied to a quantum state ϱ , the *outcome* is $i \in [k]$ with probability $p_i = \text{tr}(\varrho E_i)$.

Definition 3.7. An *observable* O is a self-adjoint operator $O \in \text{End}(V)$. It has a unique spectral decomposition $O = \lambda_1 \Pi_1 + \dots + \lambda_k \Pi_k$, where the λ_i 's are the distinct real eigenvalues of O , and the Π_i 's are the orthogonal projections onto the associated eigenspaces. The projections $\{\Pi_i : i \in [k]\}$ form a POVM.

Suppose we perform this POVM on a quantum state $\varrho \in \text{End}(V)$ and then report the eigenvalue λ_i upon receiving outcome i . Then we obtain a discrete real-valued random variable $\mathbf{x}$, which takes value λ_i with probability $\text{tr}(\varrho \Pi_i)$ for $i = 1, \dots, k$.

FACT 3.8. Given an observable O and associated real-valued random variable $\mathbf{x}$, it holds that $\mathbb{E}[\mathbf{x}] = \text{tr}(\varrho O)$. It also holds that the observable O^2 is associated to the random variable $\mathbf{x}^2$. Thus we can compute $\text{Var}[\mathbf{x}]$ as $\text{tr}(\varrho O^2) - \text{tr}(\varrho O)^2$.

In light of these facts, it is reasonable to define the notation $\mathbb{E}_\varrho[O]$ and $\text{Var}_\varrho[O]$. In fact, we will extend this notation to all operators, not just self-adjoint ones.

Definition 3.9. The *expectation* of operator $X \in \text{End}(V)$ with respect to state ϱ is defined by $\mathbb{E}_\varrho[X] = \text{tr}(\varrho X)$.

Since $\mathbb{E}_\varrho[1] = 1$, $\mathbb{E}_\varrho[X^\dagger] = \overline{\mathbb{E}_\varrho[X]}$, and $\mathbb{E}_\varrho[X^\dagger X] \geq 0$ for all $X \in \text{End}(V)$, the map $\mathbb{E}_\varrho[\cdot]$ defines a positive linear functional of norm 1 on $\text{End}(V)$. Moreover, $\mathbb{E}_{\varrho \otimes \varrho'}[\cdot]$ satisfies the following tensorization property: $\mathbb{E}_{\varrho \otimes \varrho'}[O \otimes O'] = \mathbb{E}_\varrho[O] \cdot \mathbb{E}_{\varrho'}[O']$ for all observables $O, O' \in \text{End}(V)$. The following straightforward fact says that $\mathbb{E}_\varrho[\cdot]$ is also monotone with respect to the Löwner partial order.

FACT 3.10. If $O_1, O_2 \in \text{End}(V)$ are observables, then $O_1 \leq O_2$ if and only if $\mathbb{E}_\varrho[O_1] \leq \mathbb{E}_\varrho[O_2]$ for all states $\varrho \in \text{End}(V)$.

Definition 3.11. The *covariance* of two operators $X_1, X_2 \in \text{End}(V)$ with respect to state ϱ is the sesquilinear form defined by

$$\text{Cov}_\varrho[X_1, X_2] = \mathbb{E}_\varrho[(X_1 - \mu_1 1)^\dagger (X_2 - \mu_2 1)] = \mathbb{E}_\varrho[X_1^\dagger X_2] - \mu_1^\dagger \mu_2, \\ \text{where } \mu_i = \mathbb{E}_\varrho[X_i].$$

Since $\text{Cov}_\varrho[1, \cdot] = \text{Cov}_\varrho[\cdot, 1] = 0$, it follows that $\text{Cov}_\varrho[\cdot, \cdot]$ is also translation-invariant in each argument; i.e., $\text{Cov}_\varrho[O_1 + a1, O_2 + b1] = \text{Cov}_\varrho[O_1, O_2]$ for all $a, b \in \mathbb{C}$. Furthermore, $\text{Cov}_{\varrho \otimes \varrho'}[\cdot, \cdot]$ satisfies the following tensorization property,

$$\text{Cov}_{\varrho \otimes \varrho'}[X_1 \otimes Y_1, X_2 \otimes Y_2] = \text{Cov}_\varrho[X_1, X_2] \cdot \text{Cov}_{\varrho'}[Y_1, Y_2],$$

for all operators $X_1, X_2, Y_1, Y_2 \in \text{End}(V)$. Hence,

$$\text{Cov}_{\varrho \otimes \varrho'}[X_1 \otimes 1, 1 \otimes X_2] = 0. \quad (5)$$

When X_1 and X_2 are observables, the equality above is a quantum analogue of the classical fact that the covariance of independent random variables is zero.

Definition 3.12. The *variance* of operator $X \in \text{End}(V)$ with respect to state ϱ is defined by $\text{Var}_\varrho[X] = \text{Cov}_\varrho[X, X]$.

It holds that $\text{Var}_\varrho[X] \geq 0$ for all X , $\text{Var}_\varrho[cO] = |c|^2 \text{Var}_\varrho[O]$ for all $c \in \mathbb{C}$, and

$$\text{Var}_\varrho \left[\sum_{i=1}^k X_i \right] = \sum_{i=1}^k \text{Var}_\varrho[X_i] + \sum_{\substack{i,j=1 \\ i \neq j}}^k \text{Cov}_\varrho[X_i, X_j]$$

for all operators $X_1, \dots, X_k \in \text{End}(V)$.

We end this section with a definition and lemma that will assist us in finding observables with low variance. Let V_1 and V_2 denote finite-dimensional vector spaces over $\mathbb{C}$ and let $\Phi : \text{End}(V_1) \rightarrow \text{End}(V_2)$ be a linear map.

Definition 3.13. $\Phi : \text{End}(V_1) \rightarrow \text{End}(V_2)$ is *positive* if $\Phi(X) \geq 0$ for all $X \in \text{End}(V_1)$ with $X \geq 0$. And, Φ is *unital* if $\Phi(1) = 1$.

Suppose that $V_1 = V_2 = V$ and Φ is positive and unital. Then the following result holds:

LEMMA 3.14. If $\mathbb{E}_\varrho \circ \Phi = \mathbb{E}_\varrho$, then $\text{Var}_\varrho[\Phi(O)] \leq \text{Var}_\varrho[O]$ for all observables $O \in \text{End}(V)$.

PROOF. Let $O \in \text{End}(V)$ be an observable. Since $E_\varrho[\Phi(O)] = E_\varrho[O]$, it suffices to show that $E_\varrho[\Phi(O)^2] \leq E_\varrho[O^2]$. Since O is self-adjoint and Φ is positive and unital, $\Phi(O)^2 \leq \Phi(O^2)$, by the Kadison–Schwarz inequality [Kad52]. Hence, by Fact 3.10, $E_\varrho[\Phi(O)^2] \leq E_\varrho[\Phi(O^2)]$. Since $E_\varrho \circ \Phi = E_\varrho$, it follows that $E_\varrho[\Phi(O)^2] \leq E_\varrho[O^2]$, as needed. $\square$

Thus, the class of mean-preserving positive unital maps is variance-nonincreasing.

3.3 Representation Theory

Let $\mathfrak{S}_n$ denote the symmetric group on the alphabet $[n]$ and let $U(d)$ denote the group of $d \times d$ unitary matrices.

Definition 3.15. A *partition* λ is a nonincreasing sequence of non-negative integers of finite support. If $\lambda_1 + \lambda_2 + \dots = n$, then λ is said to be a partition of n , denoted by $\lambda \vdash n$. The size of the support of λ is called the *length* of the partition and is denoted by $\ell(\lambda)$. The *power sum* symmetric polynomial in d variables $p_\lambda(x_1, \dots, x_d)$ associated to a partition λ of length k is defined by $p_\lambda = p_{\lambda_1} p_{\lambda_2} \dots p_{\lambda_k}$, where $p_r(x_1, \dots, x_d) = x_1^r + \dots + x_d^r$ for all $r \geq 0$.

The cycle type of a permutation $\pi \in \mathfrak{S}_n$ is denoted by $\text{cyc}(\pi)$. Sorted in nonincreasing order, $\text{cyc}(\pi)$ is a partition of n . Thus, the partitions of n index the conjugacy classes of $\mathfrak{S}_n$.

Definition 3.16. Let $\mathcal{P}$ denote the unitary representation of $\mathfrak{S}_n$ on $(\mathbb{C}^d)^{\otimes n}$ defined by $\mathcal{P}(\pi) |x_1\rangle \otimes \dots \otimes |x_n\rangle = |x_{\pi^{-1}(1)}\rangle \otimes \dots \otimes |x_{\pi^{-1}(n)}\rangle$, for all $|x_1\rangle, \dots, |x_n\rangle \in \mathbb{C}^d$ and $\pi \in \mathfrak{S}_n$. Furthermore, let Ad_U be the linear map on observables defined by $\text{Ad}_U(X) = (U^{\otimes n} X (U^{\otimes n})^\dagger)$ for all $U \in U(d)$.

Definition 3.17. The *symmetric group algebra* $\mathbb{C}\mathfrak{S}_n$ is the algebra of functions $f : \mathfrak{S}_n \rightarrow \mathbb{C}$. The functions $1_\pi : \mathfrak{S}_n \rightarrow \mathbb{C}$ with $\pi \in \mathfrak{S}_n$ form a basis of $\mathbb{C}\mathfrak{S}_n$, where 1_π is defined by $1_\pi(\tau) = 1$ if $\pi = \tau$ and $1_\pi(\tau) = 0$ otherwise. With a slight abuse of notation, we use π to denote the function 1_π and think of elements of $\mathbb{C}\mathfrak{S}_n$ as linear combinations of permutations $\pi \in \mathfrak{S}_n$. Thus, the product in $\mathbb{C}\mathfrak{S}_n$ is obtained by extending the product in $\mathfrak{S}_n$ to a bilinear map. $\mathbb{C}\mathfrak{S}_n$ also admits a conjugate-linear involution $X \mapsto X^\dagger$ defined by $\pi^\dagger = \pi^{-1}$ for all $\pi \in \mathfrak{S}_n$.

The representation $\mathcal{P}$ of $\mathfrak{S}_n$ extends to a $*$ -representation of the $*$ -algebra $\mathbb{C}\mathfrak{S}_n$ as follows:

$$X = \sum_{\pi \in \mathfrak{S}_n} a_\pi \pi \mapsto \sum_{\pi \in \mathfrak{S}_n} a_\pi \mathcal{P}(\pi) = \mathcal{P}(X).$$

Since the representation $\mathcal{P}$ is unitary, it follows that $\mathcal{P}(X^\dagger) = \mathcal{P}(X)^\dagger$ for all $X \in \mathbb{C}\mathfrak{S}_n$.

The center of $\mathbb{C}\mathfrak{S}_n$, denoted by $Z(\mathbb{C}\mathfrak{S}_n)$, is the set of elements $X \in \mathbb{C}\mathfrak{S}_n$ with the property that $XY = YX$ for all $Y \in \mathbb{C}\mathfrak{S}_n$. For all partitions $\kappa \vdash n$, let $O_\kappa \in \mathbb{C}\mathfrak{S}_n$ be defined by

$$O_\kappa = \frac{\text{avg}}{\pi \in \mathfrak{S}_n, \text{cyc}(\pi)=\kappa} \{\pi\}.$$

In other words, O_κ is the normalized indicator function of the conjugacy class of permutations of cycle type κ . The following elementary result relates the elements O_κ to the center of $\mathbb{C}\mathfrak{S}_n$.

PROPOSITION 3.18. $\{O_\kappa \mid \kappa \vdash n\}$ is a linear basis for $Z(\mathbb{C}\mathfrak{S}_n)$.

For a proof, see [GW09, Proposition 4.3.7]. Since $O_\kappa^\dagger = O_\kappa$ for all $\kappa \vdash n$, it follows that $\{O_\kappa \mid \kappa \vdash n\}$ is also a basis for the real vector space of self-adjoint elements of $Z(\mathbb{C}\mathfrak{S}_n)$.

4 EFFICIENT QUANTUM ESTIMATORS

The connection between observables and random variables presented in Section 3.2 allows us to import notions from classical statistics into the quantum setting. In this section, this connection is used to define quantum estimators and introduce the notion of statistical efficiency of a quantum estimator. These notions are used to formulate a structure theorem for efficient quantum estimators in situations where the statistic of interest is unitarily invariant.

As before, let V be a finite-dimensional vector space over $\mathbb{C}$. Let S denote a set of quantum states on V and let $f : S \rightarrow \mathbb{R}$ be a statistic on S . The set S serves to restrict an estimation problem to a particular class of quantum states. S will be gradually restricted, as needed, from an arbitrary set of quantum states to a set of multipartite quantum states of the form $\rho^{\otimes n}$ or $\rho^{\otimes m} \otimes \sigma^{\otimes n}$, where ρ and σ are quantum states on $\mathbb{C}^d$.

Definition 4.1. An *estimator* for f is an observable $O \in \text{End}(V)$ such that $E_\varrho[O] = f(\varrho)$ for all $\varrho \in S$. An estimator O is *efficient* if $\text{Var}_\varrho[O] \leq \text{Var}_\varrho[O']$ for all estimators $O' \in \text{End}(V)$ for f .

Henceforth, fix $V = (\mathbb{C}^d)^{\otimes n}$ and let S denote the set of states of the form $\rho_1 \otimes \dots \otimes \rho_n$, where $\rho_1, \dots, \rho_n$ are quantum states on $\mathbb{C}^d$.

Definition 4.2. A statistic $f : S \rightarrow \mathbb{R}$ is *unitarily invariant* if $f \circ \text{Ad}_U = f$ for all $U \in U(d)$. An observable $O \in \text{End}(V)$ is *unitarily invariant* if $\text{Ad}_U(O) = O$ for all $U \in U(d)$.

Let Φ be the map on observables $O \in \text{End}(V)$ defined by

$$\Phi(O) = \int_{U(d)} \text{Ad}_U(O) dU,$$

where dU denotes Haar measure. Note that Φ preserves self-adjointness and, hence, maps observables to observables.

PROPOSITION 4.3. *If O is an estimator for a unitarily invariant statistic f , then $\Phi(O)$ is also an estimator for f , and $\text{Var}_\varrho[\Phi(O)] \leq \text{Var}_\varrho[O]$ for all $\varrho \in S$.*

PROOF. The map Φ is positive and unital. Since f is unitarily invariant,

$$\begin{aligned} E_\varrho[\Phi(O)] &= \int_{U(d)} \text{tr}(\text{Ad}_{U^\dagger}(\varrho) O) dU = \int_{U(d)} f(\text{Ad}_{U^\dagger}(\varrho)) dU \\ &= \int_{U(d)} f(\varrho) dU = E_\varrho[O]. \end{aligned}$$

Hence, by Lemma 3.14, $\text{Var}_\varrho[\Phi(O)] \leq \text{Var}_\varrho[O]$. $\square$

The following result relates the image of the map Φ to the symmetric group algebra $\mathbb{C}\mathfrak{S}_n$ and the representation $\mathcal{P}$. It uses the Schur–Weyl duality theorem. For a proof, see e.g. [CS06, Proposition 2.2].

PROPOSITION 4.4. *The map Φ is a projection into $\mathcal{P}(\mathbb{C}\mathfrak{S}_n)$.*

Thus, if O is an efficient estimator for a unitarily invariant statistic f , then $\Phi(O)$ is also an efficient estimator for f . Hence, the next corollary follows immediately from [Proposition 4.4](#).

COROLLARY 4.5. *To find an efficient estimator for a unitarily invariant statistic $f : S \rightarrow \mathbb{R}$, it suffices to consider estimators of the form $\mathcal{P}(X)$ with $X \in \mathbb{C}\mathfrak{S}_n$.*

In light of [Corollary 4.5](#), we introduce the following notation:

Notation 4.6. Let E_ϱ be extended to a map on elements $X \in \mathbb{C}\mathfrak{S}_n$ defined by $E_\varrho[X] = E_\varrho[\mathcal{P}(X)]$. Thus, E_ϱ , Cov_ϱ , and Var_ϱ are defined directly on elements of $\mathbb{C}\mathfrak{S}_n$ via the representation $\mathcal{P}$.

If $\gamma = (i_1 i_2 \dots i_\ell) \in \mathfrak{S}_n$, let tr_γ be defined by $\text{tr}_\gamma(\varrho) = \text{tr}(\rho_{i_1} \rho_{i_2} \dots \rho_{i_\ell})$. The following proposition establishes a formula for the expectation $E_\varrho[\pi]$ of a permutation $\pi \in \mathfrak{S}_n$ with respect to a state $\varrho \in S$. The proof is a straightforward calculation which appears in the full version of the paper.

PROPOSITION 4.7. *Let $\pi \in \mathfrak{S}_n$ be an arbitrary permutation. If $\pi = \gamma_1 \dots \gamma_k$ is a decomposition of π into disjoint cycles, including cycles of length 1, then*

$$E[\pi^{-1}] = \prod_{i=1}^k \text{tr}_{\gamma_i}(\varrho).$$

Remark 4.8. In describing the cycle type of a permutation $\pi \in \mathfrak{S}_n$, it is common to omit mentioning 1-cycles. Conveniently, this would have no effect in [Proposition 4.7](#), since $\text{tr}(\rho_i) = 1$ anyway for all i .

Definition 4.9. The group Γ of *permutation invariants* of the set of states S is defined by

$$\Gamma = \{\pi \in \mathfrak{S}_n \mid \forall \varrho \in S, \forall X \in \mathbb{C}\mathfrak{S}_n, E[\pi^{-1}X\pi] = E[X]\}.$$

Note that the definition of Γ depends on S . For all $X \in \mathbb{C}\mathfrak{S}_n$, let $X^\Gamma \in \mathbb{C}\mathfrak{S}_n$ be defined by

$$X^\Gamma = \frac{1}{|\Gamma|} \sum_{\pi \in \Gamma} \pi^{-1}X\pi.$$

Thus, $X^\Gamma \tau = \tau X^\Gamma$ for all $\tau \in \Gamma$ and $X \in \mathbb{C}\mathfrak{S}_n$.

PROPOSITION 4.10. *For all self-adjoint elements $O \in \mathbb{C}\mathfrak{S}_n$, $\text{Var}_\varrho[O^\Gamma] \leq \text{Var}_\varrho[O]$.*

PROOF. The map $O \mapsto O^\Gamma$ is positive and unital. Moreover, $E_\varrho[O^\Gamma] = E_\varrho[O]$ for all $O \in \mathbb{C}\mathfrak{S}_n$. Hence, by [Lemma 3.14](#), $\text{Var}_\varrho[O^\Gamma] \leq \text{Var}_\varrho[O]$. $\square$

COROLLARY 4.11. *To find an efficient estimator for a unitarily invariant statistic $f : S \rightarrow \mathbb{R}$, it suffices to consider estimators of the form $\mathcal{P}(X)$ with $X \in \mathbb{C}\mathfrak{S}_n$ and $X\tau = \tau X$ for all $\tau \in \Gamma$.*

The group Γ acts on $\mathfrak{S}_n$ by conjugation, viz. $\tau \in \Gamma$ acts on $\mathfrak{S}_n$ by $\pi \mapsto \tau^{-1}\pi\tau$. This action partitions the group $\mathfrak{S}_n$ into disjoint orbits: $\mathfrak{S}_n = O_1 \cup \dots \cup O_\ell$, where two permutations π_1 and π_2 belong to the same orbit O_i for $i \in [\ell]$ if and only if there exists $\tau \in \Gamma$ such that $\tau^{-1}\pi_1\tau = \pi_2$. It is easy to see that an element $X \in \mathbb{C}\mathfrak{S}_{m+n}$ commutes with all elements of Γ if and only if X is constant on the orbits $O_1, \dots, O_\ell$ defined by Γ . Let $\phi_i \in \mathbb{C}\mathfrak{S}_{m+n}$ denote the indicator function of the orbit O_i for $i \in [\ell]$. Thus, the set $\{\phi_1, \dots, \phi_\ell\}$ forms a basis for the elements $X \in \mathbb{C}\mathfrak{S}_n$ that are

constant on the orbits $O_1, \dots, O_\ell$. Therefore, by [Corollary 4.11](#), it holds that:

PROPOSITION 4.12. *To find an efficient estimator for a unitarily invariant statistic $f : S \rightarrow \mathbb{R}$, it suffices to consider estimators of the form $\mathcal{P}(X)$ with $X = a_1\phi_1 + \dots + a_\ell\phi_\ell$, where $a_1, \dots, a_\ell \in \mathbb{C}$.*

4.0.1 Case: $\varrho = \rho^{\otimes n}$. Let S denote the set of states of the form $\varrho = \rho^{\otimes n}$, where ρ is a quantum state on $\mathbb{C}^d$. Let $\alpha \in \mathbb{R}^d$ denote the spectrum of ρ (taken in some arbitrary order).

When ϱ is a state of the form $\rho^{\otimes n}$, the expectation $E_\varrho[\pi]$ of $\pi \in \mathfrak{S}_n$ has a particularly simple formula:

PROPOSITION 4.13. *For all $\pi \in \mathfrak{S}_n$ with $\text{cyc}(\pi) = \kappa$, $E_\varrho[\pi] = p_\kappa(\alpha)$.*

PROOF. Let ℓ denote the number of disjoint cycles in the decomposition of π . By [Proposition 4.7](#),

$$E[\pi] = \text{tr}(\rho^{\kappa_1}) \dots \text{tr}(\rho^{\kappa_\ell}) = p_{\kappa_1}(\alpha) \dots p_{\kappa_\ell}(\alpha) = p_\kappa(\alpha). \quad \square$$

Thus, $E_\varrho[\pi]$ depends only on the cycle type of π . Since the cycle types of $\pi_1\pi_2$ and $\pi_2\pi_1$ are equal for all $\pi_1, \pi_2 \in \mathfrak{S}_n$, the following result holds:

PROPOSITION 4.14. *For all $X, Y \in \mathbb{C}\mathfrak{S}_n$, $E_\varrho[XY] = E_\varrho[YX]$.*

PROOF. For all $\pi_1, \pi_2 \in \mathfrak{S}_n$, $\text{cyc}(\pi_1\pi_2) = \text{cyc}(\pi_2\pi_1)$. Hence, by [Proposition 4.13](#), $E_\varrho[\pi_1\pi_2] = E_\varrho[\pi_2\pi_1] = p_\kappa(\alpha)$, where $\kappa = \text{cyc}(\pi_1\pi_2)$. It follows by linearity that $E_\varrho[XY] = E_\varrho[YX]$ for all $X, Y \in \mathbb{C}\mathfrak{S}_n$. $\square$

Thus, we obtain the following strengthening of [Corollary 4.5](#):

PROPOSITION 4.15. *To find an efficient estimator for a unitarily invariant statistic $f : S \rightarrow \mathbb{R}$, it suffices to consider estimators of the form $\mathcal{P}(X)$ with $X \in Z(\mathbb{C}\mathfrak{S}_n)$.*

PROOF. By [Proposition 4.14](#), $\Gamma = \mathfrak{S}_n$. The statement follows immediately from [Corollary 4.11](#). $\square$

The expectation $E_\varrho[X]$ of an estimator $X \in Z(\mathbb{C}\mathfrak{S}_n)$ can be expressed as a linear combination of $p_\kappa(\alpha)$ with $\kappa \vdash n$ where, recall, α is the spectrum of ρ . By [Proposition 3.18](#), the elements $O_\kappa \in \mathbb{C}\mathfrak{S}_n$ with $\kappa \vdash n$ form a basis for the real vector space of self-adjoint elements of $Z(\mathbb{C}\mathfrak{S}_n)$. Hence, an estimator $X \in Z(\mathbb{C}\mathfrak{S}_n)$ can be expressed uniquely as a linear combination of the form

$$X = \sum_{\kappa \vdash n} a_\kappa O_\kappa,$$

where $a_\kappa \in \mathbb{R}$ for all $\kappa \vdash n$. Thus, by [Proposition 4.13](#),

$$E_\varrho[X] = \sum_{\kappa \vdash n} a_\kappa E_\varrho[O_\kappa] = \sum_{\kappa \vdash n} a_\kappa p_\kappa(\alpha).$$

Moreover, it is easy to see that an estimator $X \in Z(\mathbb{C}\mathfrak{S}_n)$ is unique:

PROPOSITION 4.16. *If $X_1, X_2 \in Z(\mathbb{C}\mathfrak{S}_n)$ are estimators for $f : S \rightarrow \mathbb{R}$, then $X_1 = X_2$.*

Therefore, all observables in the center of $\mathbb{C}\mathfrak{S}_n$ are efficient estimators:

COROLLARY 4.17. *If $X \in Z(\mathbb{C}\mathfrak{S}_n)$ is an estimator for $f : S \rightarrow \mathbb{R}$, then X is efficient.*

PROOF. The result follows from [Proposition 4.15](#) and [Proposition 4.16](#). $\square$

Example 4.18. By [Corollary 4.17](#), O_κ is an efficient estimator for $f(\varrho) = p_\kappa(\alpha)$. In particular, suppose $\kappa = (k, 1, 1, \dots)$, which we will denote simply as (k) (recalling [Remark 4.8](#)). Then $O_{(k)}$ is an efficient estimator of $f(\varrho) = \text{tr}(\rho^k)$.

4.0.2 Case: $\varrho = \rho^{\otimes m} \otimes \sigma^{\otimes n}$. Let S denote the set of states of the form $\varrho = \rho^{\otimes m} \otimes \sigma^{\otimes n}$, where ρ and σ are quantum states on $\mathbb{C}^d$. Let $\alpha \in \mathbb{R}^d$ and $\beta \in \mathbb{R}^d$ denote the spectra of ρ and σ , respectively. The group Γ of permutation invariants of S can be described as follows:

PROPOSITION 4.19. $\Gamma \cong \mathfrak{S}_m \times \mathfrak{S}_n$, where $(\pi_1, \pi_2) \in \Gamma$ embeds in $\mathfrak{S}_{m+n}$ in the natural way, viz. by applying π_1 to $\{1, \dots, m\}$ and applying π_2 to $\{m+1, \dots, m+n\}$.

To find an efficient estimator with respect to S , it is sufficient, by [Proposition 4.12](#), to consider functions $X \in \mathbb{C}\mathfrak{S}_{m+n}$ which are constant on the orbits defined by the action of Γ on $\mathfrak{S}_{m+n}$.

Notation 4.20. Since Γ acts on $\mathfrak{S}_{m+n}$ by conjugation, the orbits of Γ refine the conjugacy classes of $\mathfrak{S}_{m+n}$. An orbit of Γ is uniquely determined by a signature consisting of a cycle type and a map that associates each index in the cycle type with either ρ or σ . For instance, the signature $(\rho \sigma)$ identifies the orbit of Γ which consists of all transpositions that exchange an index in $\{1, \dots, m\}$ with an index in $\{m+1, \dots, m+n\}$. Note that $(\rho \sigma) = (\sigma \rho)$. Similarly, $(\rho \rho \sigma)$ denotes the set of 3-cycles with two indices in $\{1, \dots, m\}$ and one index in $\{m+1, \dots, m+n\}$.

If $\mathfrak{s}$ is the signature of an orbit of Γ , let $O_{\mathfrak{s}} \in \mathbb{C}\mathfrak{S}_{m+n}$ denote the average of all elements in the orbit. For example, $O_{(\rho \sigma)}$ denotes the average of all transpositions in the $(\rho \sigma)$ orbit described above.

Example 4.21. By [Proposition 4.7](#), $O_{(\rho \sigma)}$ is an estimator for $f(\varrho) = \text{tr}(\rho\sigma)$.

Moreover, $O_{(\rho \sigma)}$ satisfies the following uniqueness property:

PROPOSITION 4.22. If $X \in \mathbb{C}\mathfrak{S}_{m+n}$ is an estimator for the statistic $f : S \rightarrow \mathbb{R}$ defined by $f(\varrho) = \text{tr}(\rho\sigma)$ and X is of the form presented in [Proposition 4.12](#), then $X = O_{(\rho \sigma)}$.

A similar argument proves the following:

PROPOSITION 4.23. If $X \in \mathbb{C}\mathfrak{S}_{m+n}$ is an estimator for the statistic $f : S \rightarrow \mathbb{R}$ defined by $f(\varrho) = D_{\text{HS}}^2(\rho, \sigma)$ and X is of the form presented in [Proposition 4.12](#), then $X = O_{(\rho \rho)} + O_{(\sigma \sigma)} - 2O_{(\rho \sigma)}$.

Thus, the estimators obtained for $\text{tr}(\rho\sigma)$ and $D_{\text{HS}}^2(\rho, \sigma)$ are efficient:

COROLLARY 4.24. $O_{(\rho \sigma)}$ is an efficient estimator for $f(\varrho) = \text{tr}(\rho\sigma)$.

COROLLARY 4.25. $O_{(\rho \rho)} + O_{(\sigma \sigma)} - 2O_{(\rho \sigma)}$ is an efficient estimator for $f(\varrho) = D_{\text{HS}}^2(\rho, \sigma)$.

5 HILBERT-SCHMIDT DISTANCE AND RELATED ESTIMATION

5.1 Purity, and Testing Identity to the Maximally Mixed State

Let ρ be a quantum state on $\mathbb{C}^d$, let $\varrho = \rho^{\otimes n}$, and define $f(\varrho) = \text{tr}(\rho^2)$. The quantity $\text{tr}(\rho^2)$ is called the *purity* of ρ . One can also

easily compute that the purity is the same as the squared Hilbert-Schmidt distance to the maximally mixed state, up to an additive constant: $D_{\text{HS}}^2(\rho, 1/d) = \text{tr}(\rho^2) - 1/d$.

By [Example 4.18](#), the observable $O_{(2)}$ is an efficient estimator for the statistic f . The variance of $O_{(2)}$ can be explicitly computed; the details of this calculation appear in the full version of the paper.

LEMMA 5.1.

$$\text{Var}_{\varrho}[O_{(2)}] = \frac{1}{\binom{n}{2}}(1 - p_2(\alpha)^2) + \frac{2(n-2)}{\binom{n}{2}}(p_3(\alpha) - p_2(\alpha)^2).$$

At this point, we show how to prove our [Theorem 1.6](#) in the special case that σ is known to be the maximally mixed state. (This result was originally proven, in a slightly more opaque way, in [\[OW15, Theorem 4.1\]](#).)

PROPOSITION 5.2. (Special case of [Theorem 1.6](#).) There is an algorithm that, given $n = O(1/\epsilon^2)$ copies of the state $\rho \in \mathbb{C}^{d \times d}$, (whp) outputs “close” if $D_{\text{HS}}(\rho, 1/d) \leq .99\epsilon$ and outputs “far” if $D_{\text{HS}}(\rho, 1/d) > \epsilon$.

PROOF. Since $D_{\text{HS}}^2(\rho, 1/d) = \text{tr}(\rho^2) - 1/d$, the observable $O_{(2)} - 1/d$ is an unbiased estimator of $D_{\text{HS}}^2(\rho, 1/d)$. Let $\alpha \in \mathbb{R}^d$ denote the spectrum of ρ and let $\Delta_i = \alpha_i - 1/d$ for all $i \in [d]$. Thus,

$$p_3(\alpha) - p_2(\alpha)^2 = \frac{p_2(\Delta)}{d} + p_3(\Delta) - p_2(\Delta)^2 \leq p_2(\Delta) = D_{\text{HS}}^2(\rho, 1/d).$$

Hence, by [Lemma 5.1](#),

$$\text{Var}_{\varrho}\left[O_{(2)} - \frac{1}{d}\right] = \text{Var}_{\varrho}[O_{(2)}] \leq O\left(\frac{1}{n^2} + \frac{p_2(\Delta)}{n}\right).$$

The result now follows from [Lemma 2.1](#). $\square$

5.2 Squared Hilbert-Schmidt Distance

Let ρ and σ be quantum states on $\mathbb{C}^d$, let $\varrho = \rho^{\otimes m} \otimes \sigma^{\otimes n}$, and define $f(\varrho) = D_{\text{HS}}^2(\rho, \sigma) = \text{tr}(\rho^2) + \text{tr}(\sigma^2) - 2\text{tr}(\rho\sigma)$. By [Corollary 4.25](#), $O_{(\rho \rho)} + O_{(\sigma \sigma)} - 2O_{(\rho \sigma)}$ is an efficient estimator for the statistic f . The following result, proved in the full version of the paper, bounds the variance of this estimator:

PROPOSITION 5.3. When $m = n$,

$$\text{Var}_{\varrho}[O_{(\rho \rho)} + O_{(\sigma \sigma)} - 2O_{(\rho \sigma)}] = O\left(\frac{1}{n^2} + \frac{D_{\text{HS}}^2(\rho, \sigma)}{n}\right).$$

5.3 Consequences for Testing

[Theorem 1.6](#), which uses $O(1/\epsilon^2)$ -copies of unknown ρ, σ to distinguish $D_{\text{HS}}(\rho, \sigma) \leq .99\epsilon$ from $D_{\text{HS}}(\rho, \sigma) > \epsilon$, is now an immediate consequence of [Lemma 2.1](#) and [Proposition 5.3](#).

In the remainder of this section we give the proof of [Corollary 1.5](#):

PROOF. The testing algorithm does not need to know δ , nor which of ρ or σ is δ -close to rank k : it simply applies the robust Hilbert-Schmidt tester [Theorem 1.6](#) with error parameter $c\epsilon/\sqrt{k}$, where $c = \frac{1}{1+1/\sqrt{2}}$. All we need to show is an elementary fact of pure matrix analysis: assuming $D_{\text{HS}}(\rho, \sigma) \leq c\epsilon/\sqrt{k}$, it holds that $D_{\text{tr}}(\rho, \sigma) \leq \delta + \epsilon$. Since the Hilbert-Schmidt and trace distances are symmetric we may assume that it is σ that is close to rank k ;

and, since these distances are unitarily invariant, we may assume that $\sigma = \text{diag}(\beta_1, \dots, \beta_d)$, where $\beta_1 + \dots + \beta_k \geq 1 - \delta$.

Write ρ_A for the top-left $k \times k$ block of ρ , write ρ_B for its bottom-right $(d-k) \times (d-k)$ block, and write ρ_{off} for the “off-diagonal” $d \times d$ matrix given by zeroing out those two blocks. Similarly define σ_A , σ_B , and σ_{off} , so $\sigma_A = \text{diag}(\beta_1, \dots, \beta_k)$, $\sigma_B = \text{diag}(\beta_{k+1}, \dots, \beta_d)$, and $\sigma_{\text{off}} = 0$. Now

$$2D_{\text{tr}}(\rho, \sigma) = \|\rho - \sigma\|_1 \leq \|\rho_A - \sigma_A\|_1 + \|\rho_{\text{off}} - \sigma_{\text{off}}\|_1 + \|\rho_B - \sigma_B\|_1, \quad (6)$$

by the triangle inequality. The matrix $\rho_A - \sigma_A$ of course has rank at most k , and the matrix $\rho_{\text{off}} - \sigma_{\text{off}}$ has rank at most $2k$ (being the sum of a $k \times (d-k)$ matrix and a $(d-k) \times k$ matrix). Thus we use Cauchy–Schwarz to bound the first two terms on the right of (6) by

$$\begin{aligned} & \sqrt{k}\|\rho_A - \sigma_A\|_{\text{HS}} + \sqrt{2k}\|\rho_{\text{off}} - \sigma_{\text{off}}\|_{\text{HS}} \\ & \leq \sqrt{k}D_{\text{HS}}(\rho, \sigma) + \sqrt{2k}D_{\text{HS}}(\rho, \sigma) \leq (1 + \sqrt{2})c\epsilon. \end{aligned}$$

Now if we can show

$$\|\rho_B - \sigma_B\|_1 \leq 2\delta + c\epsilon, \quad (7)$$

we will have bounded $2D_{\text{tr}}(\rho, \sigma)$ by $2\delta + (2 + \sqrt{2})c\epsilon = 2\delta + 2\epsilon$, as needed.

To show (7), we begin with the triangle inequality:

$$\begin{aligned} \|\rho_B - \sigma_B\|_1 & \leq \|\rho_B\|_1 + \|\sigma_B\|_1 \\ & = \text{tr}(\rho_B) + \text{tr}(\sigma_B) = (1 - \text{tr}(\rho_A)) + (1 - \text{tr}(\sigma_A)), \end{aligned}$$

where the first equality used that ρ_B and σ_B are positive, and the second used that ρ and σ have trace 1. Continuing,

$$\begin{aligned} & (1 - \text{tr}(\rho_A)) + (1 - \text{tr}(\sigma_A)) \\ & = 2 - 2\text{tr}(\sigma_A) + \text{tr}(\sigma_A - \rho_A) \\ & \leq 2\delta + \|\sigma_A - \rho_A\|_1 \leq 2\delta + \sqrt{k}\|\sigma_A - \rho_A\|_{\text{HS}}, \end{aligned}$$

where we used $1 - \text{tr}(\sigma_A) = 1 - (\beta_1 + \dots + \beta_k) \leq \delta$, and also Cauchy–Schwarz again. Now (7) follows since $\|\sigma_A - \rho_A\|_{\text{HS}} \leq D_{\text{HS}}(\rho, \sigma) \leq c\epsilon/\sqrt{k}$. $\square$

6 QUANTUM CHI-SQUARED ESTIMATION

6.1 A Chi-Squared Observable

In this section, σ will denote a fixed full-rank d -dimensional density matrix, and we will develop a natural unbiased estimator for the Bures χ^2 -divergence $\text{tr}((\rho - \sigma) \cdot \Omega_\sigma(\rho - \sigma))$. This formula suggests a natural bilinear form:

Definition 6.1. For matrices $S, T \in \mathbb{C}^{d \times d}$, define the bilinear form

$$\omega_\sigma^{(2)}(S, T) = \text{tr}(S \cdot \Omega_\sigma T).$$

This bilinear form has the following “contraction” property:

PROPOSITION 6.2. For any $S \in \mathbb{C}^{d \times d}$ it holds that $\omega_\sigma^{(2)}(S, \sigma) = \text{tr}(S) = \omega_\sigma^{(2)}(\sigma, S)$.

PROOF. Both identities are direct from the definition of the Ω_σ : the first uses $\Omega_\sigma \sigma = \mathbf{1}$; the second uses $S = \frac{1}{2} \sigma \cdot \Omega_\sigma S + \frac{1}{2} \Omega_\sigma S \cdot \sigma$. $\square$

It follows that

$$\begin{aligned} D_{\chi^2}(\rho \parallel \sigma) & = \omega_\sigma^{(2)}(\rho - \sigma, \rho - \sigma) \\ & = \omega_\sigma^{(2)}(\rho, \rho) - \omega_\sigma^{(2)}(\sigma, \rho) - \omega_\sigma^{(2)}(\rho, \sigma) + \omega_\sigma^{(2)}(\sigma, \sigma) \\ & = \omega_\sigma^{(2)}(\rho, \rho) - \text{tr}(\rho) - \text{tr}(\rho) + \text{tr}(\sigma), \end{aligned}$$

and from this we arrive at another standard formula for the Bures χ^2 -divergence:

PROPOSITION 6.3. If ρ is a d -dimensional density matrix, then

$$D_{\chi^2}(\rho \parallel \sigma) = \omega_\sigma^{(2)}(\rho, \rho) - 1 = \text{tr}(\rho \cdot \Omega_\sigma \rho) - 1.$$

If $\sigma = \text{diag}(\beta_1, \dots, \beta_d)$, then Ω_σ acts by multiplying the ij -th entry by $\frac{2}{\beta_i + \beta_j} = \text{avg}\{\beta_i, \beta_j\}^{-1}$; thus in this case,

$$D_{\chi^2}(\rho \parallel \sigma) = \left(\sum_{i,j=1}^d \frac{|\rho_{ij}|^2}{\text{avg}\{\beta_i, \beta_j\}} \right) - 1.$$

In light of the above, it is natural to define the following observable.

Definition 6.4. Assume henceforth that $\sigma = \text{diag}(\beta_1, \dots, \beta_d)$ is diagonal. We define the associated χ^2 observable, operating on $(\mathbb{C}^d)^{\otimes 2}$, as follows:

$$\mathcal{X}_\sigma = \sum_{i,j=1}^d \frac{|ji\rangle\langle ij|}{\text{avg}\{\beta_i, \beta_j\}}.$$

Evidently, $E_{\rho \otimes \rho}[\mathcal{X}_\sigma] = D_{\chi^2}(\rho \parallel \sigma) + 1$.

Definition 6.5. Given distinct $s, t \in [n]$, we write $\mathcal{X}_\sigma^{(s,t)}$ for the operator which acts on $(\mathbb{C}^d)^{\otimes n}$ by applying $\mathcal{X}_\sigma$ to the s -th and the t -th tensor copies of $\mathbb{C}^d$ and acting as the identity on the remaining copies. (The dependence on n in the notation is implicit.)

Observation 6.6. Observe that $\mathcal{X}_\sigma^{(s,t)}$ is rather similar to the observable $\mathcal{P}((s\ t))$; however, when it swaps letters i and j , it picks up a scalar factor of $\frac{2}{\beta_i + \beta_j}$. Thus in comparison with

$$\mathcal{P}((1\ 2)) \cdot \mathcal{P}((2\ 3)) = \mathcal{P}((1\ 2\ 3)) = \sum_{i,j,k=1}^d |ijk\rangle\langle jki|$$

we have

$$\mathcal{X}_\sigma^{(1,2)} \cdot \mathcal{X}_\sigma^{(2,3)} = \sum_{i,j,k=1}^d \frac{|ijk\rangle\langle jki|}{\text{avg}\{\beta_i, \beta_j\} \cdot \text{avg}\{\beta_i, \beta_k\}},$$

the scalar factors in the denominator arising because letters i and k are swapped, and then letters i and j are swapped. As a consequence, rather than the matrix trilinear form mapping (R, S, T) to

$$\begin{aligned} \text{tr}(\mathcal{P}((1\ 2)) \cdot \mathcal{P}((2\ 3)) \cdot R \otimes S \otimes T) & = \text{tr}(\mathcal{P}((1\ 2\ 3)) \cdot R \otimes S \otimes T) \\ & = \text{tr}(TSR) \end{aligned}$$

as in [Proposition 4.7](#), we obtain the trilinear form given in the subsequent definition.

Definition 6.7. For matrices $R, S, T \in \mathbb{C}^{d \times d}$, define the trilinear form

$$\begin{aligned}\omega_\sigma^{(3)}(R, S, T) &= \text{tr}(\mathcal{X}_\sigma^{(1,2)} \cdot \mathcal{X}_\sigma^{(2,3)} \cdot R \otimes S \otimes T) \\ &= \sum_{i,j,k=1}^d \frac{T_{ij}S_{jk}R_{ki}}{\text{avg}\{\beta_i, \beta_j\} \cdot \text{avg}\{\beta_i, \beta_k\}}.\end{aligned}$$

We again get a certain “contraction” property:

PROPOSITION 6.8. For any $S, T \in \mathbb{C}^{d \times d}$ it holds that $\omega_\sigma^{(3)}(S, T, \sigma) = \omega_\sigma^{(2)}(S, T) = \omega_\sigma^{(3)}(\sigma, S, T)$.

PROOF. We prove the second identity, the first being similar. When we substitute $R = \sigma$ into **Definition 6.7** we obtain

$$\omega_\sigma^{(3)}(\sigma, S, T) = \sum_{i,j,k=1}^d \frac{T_{ij}S_{jk}\sigma_{ki}}{\text{avg}\{\beta_i, \beta_j\} \cdot \text{avg}\{\beta_i, \beta_k\}}$$

Since σ is diagonal, the summands with $i \neq k$ vanish. When $i = k$ we have $\sigma_{kk} = \beta_k$, which cancels the factor of $\text{avg}\{\beta_i, \beta_k\}$. We are left with

$$\omega_\sigma^{(3)}(\sigma, S, T) = \sum_{j,k=1}^d \frac{T_{ij}S_{ji}}{\text{avg}\{\beta_i, \beta_j\}},$$

which is indeed $\omega_\sigma^{(2)}(S, T)$. $\square$

Finally:

Definition 6.9. For a given $n \geq 2$, we define the *averaged χ^2 observable* on $(\mathbb{C}^d)^{\otimes n}$ to be $\mathcal{O}_{\chi^2} = \text{avg}_{s \neq t} \{\mathcal{X}_\sigma^{(s,t)}\} - 1$, where the average is over all distinct ordered pairs $s, t \in [n]$.

Evidently:

PROPOSITION 6.10. $\mathbb{E}_{\rho^{\otimes n}}[\mathcal{O}_{\chi^2}] = \mathbb{D}_{\chi^2}(\rho \parallel \sigma)$ and $\text{Var}_{\rho^{\otimes n}}[\mathcal{O}_{\chi^2}] = \text{Var}_{\rho^{\otimes n}}[\text{avg}_{s \neq t} \{\mathcal{X}_\sigma^{(s,t)}\}]$.

6.2 Analyzing the Variance of the Averaged Chi-Squared Observable

The calculation of the variance of the averaged χ^2 observable, $\text{Var}_{\rho^{\otimes n}}[\text{avg}_{s \neq t} \{\mathcal{X}_\sigma^{(s,t)}\}]$, proceeds exactly as does the calculation of the variance of the purity observable in **Lemma 5.1**. We obtain:

PROPOSITION 6.11. The averaged χ^2 -observable has variance $\frac{1}{\binom{n}{2}} \left(\text{tr}(\mathcal{X}_\sigma^2 \rho^{\otimes 2}) - \omega_\sigma^{(2)}(\rho, \rho)^2 \right) + \frac{2(n-2)}{\binom{n}{2}} \left(\omega_\sigma^{(3)}(\rho, \rho, \rho) - \omega_\sigma^{(2)}(\rho, \rho)^2 \right)$.

Introducing the shorthand $\Delta = \rho - \sigma$, we analyze the terms in **Proposition 6.11**.

PROPOSITION 6.12.

$$\begin{aligned}\omega_\sigma^{(3)}(\rho, \rho, \rho) - \omega_\sigma^{(2)}(\rho, \rho)^2 \\ = \omega_\sigma^{(3)}(\Delta, \Delta, \Delta) + \omega_\sigma^{(3)}(\Delta, \sigma, \Delta) - \mathbb{D}_{\chi^2}(\rho \parallel \sigma)^2.\end{aligned}$$

PROOF. This is immediate from writing $\rho = \Delta + \sigma$ and using: multilinearity of $\omega_\sigma^{(3)}(\cdot, \cdot, \cdot)$; the contraction properties **Propositions 6.2** and **6.8**; $\text{tr}(\rho) = \text{tr}(\sigma) = 1$; and, $\mathbb{D}_{\chi^2}(\rho \parallel \sigma) = \omega_\sigma^{(2)}(\Delta, \Delta)$. $\square$

We will ignore the subtracted $\mathbb{D}_{\chi^2}(\rho \parallel \sigma)^2$ and use the following simple bound for $\omega_\sigma^{(3)}(\Delta, \sigma, \Delta)$:

PROPOSITION 6.13. $\omega_\sigma^{(3)}(\Delta, \sigma, \Delta) \leq 2\mathbb{D}_{\chi^2}(\rho \parallel \sigma)$.

PROOF. Recalling **Definition 6.7** and using $\sigma = \text{diag}(\beta_1, \dots, \beta_d)$ we get

$$\begin{aligned}\omega_\sigma^{(3)}(\Delta, \sigma, \Delta) &= \sum_{i,j=1}^d \frac{\Delta_{ij}\beta_j\Delta_{ji}}{\text{avg}\{\beta_i, \beta_j\}^2} \\ &\leq 2 \sum_{i,j=1}^d \frac{|\Delta_{ij}|^2}{\text{avg}\{\beta_i, \beta_j\}} = 2\mathbb{D}_{\chi^2}(\rho \parallel \sigma),\end{aligned}$$

where the inequality used $\frac{\beta_j}{\text{avg}\{\beta_i, \beta_j\}} \leq 2$. $\square$

We now come to the main term in **Proposition 6.12**:

PROPOSITION 6.14. Assume the smallest eigenvalue of σ is at least δ . Then

$$\omega_\sigma^{(3)}(\Delta, \Delta, \Delta) \leq \sqrt{2d/\delta} \cdot \mathbb{D}_{\chi^2}(\rho \parallel \sigma)^{3/2}.$$

PROOF. Applying Cauchy–Schwarz to the formula in **Definition 6.7** gives that $\omega_\sigma^{(3)}(\Delta, \Delta, \Delta)$ is at most

$$\sqrt{\sum_{i,j,k=1}^d \frac{|\Delta_{ij}|^2 |\Delta_{ki}|^2}{\text{avg}\{\beta_i, \beta_j\} \cdot \text{avg}\{\beta_i, \beta_k\}}} \cdot \sqrt{\sum_{i,j,k=1}^d \frac{|\Delta_{jk}|^2}{\text{avg}\{\beta_i, \beta_j\} \cdot \text{avg}\{\beta_i, \beta_k\}}}.$$

The sum inside the first square-root above is

$$\sum_{i=1}^d \left(\sum_{j=1}^d \frac{|\Delta_{ij}|^2}{\text{avg}\{\beta_i, \beta_j\}} \right)^2 \leq \left(\sum_{i,j=1}^d \frac{|\Delta_{ij}|^2}{\text{avg}\{\beta_i, \beta_j\}} \right)^2 = \mathbb{D}_{\chi^2}(\rho \parallel \sigma)^2.$$

For the sum inside the second square-root above, we use the elementary fact that

$$\text{avg}\{\beta_i, \beta_j\} \cdot \text{avg}\{\beta_i, \beta_k\} \geq (\delta/2) \cdot \text{avg}\{\beta_j, \beta_k\}$$

when $\delta \leq \beta_i, \beta_j, \beta_k \leq 1$. Thus this second sum is at most

$$d \cdot (2/\delta) \cdot \sum_{j,k} \frac{|\Delta_{jk}|^2}{\text{avg}\{\beta_j, \beta_k\}} = (2d/\delta) \cdot \mathbb{D}_{\chi^2}(\rho \parallel \sigma).$$

Combining the two bounds above completes the proof. $\square$

We now analyze the first term in **Proposition 6.11**, ignoring the subtracted $\omega_\sigma^{(2)}(\rho, \rho)^2$:

PROPOSITION 6.15. Assume the smallest eigenvalue of σ is at least δ . Then

$$\text{tr}(\mathcal{X}_\sigma^2 \rho^{\otimes 2}) \leq 2d^2 + (2d/\delta) \cdot \mathbb{D}_{\chi^2}(\rho \parallel \sigma).$$

PROOF. Using $\text{avg}\{\beta_i, \beta_j\} \geq \sqrt{\beta_i \beta_j}$, we may bound $\text{tr}(\mathcal{X}_\sigma^2 \rho^{\otimes 2})$ as

$$\begin{aligned}\sum_{i,j=1}^d \frac{\rho_{ii}\rho_{jj}}{\text{avg}\{\beta_i, \beta_j\}^2} &\leq \sum_{i,j=1}^d \frac{\rho_{ii}\rho_{jj}}{\beta_i\beta_j} = \left(\sum_{i=1}^d \frac{\rho_{ii}}{\beta_i} \right)^2 \\ &= \left(d + \sum_{i=1}^d \frac{\Delta_{ii}}{\beta_i} \right)^2 \leq 2d^2 + 2 \left(\sum_{i=1}^d \frac{|\Delta_{ii}|}{\beta_i} \right)^2.\end{aligned}$$

Now using $\sqrt{\beta_i} \geq \sqrt{\delta}$ and then Cauchy–Schwarz,

$$\begin{aligned} \left(\sum_{i=1}^d \frac{|\Delta_{ii}|}{\beta_i} \right)^2 &\leq (1/\delta) \cdot \left(\sum_{i=1}^d \frac{|\Delta_{ii}|}{\sqrt{\beta_i}} \right)^2 \leq (d/\delta) \cdot \sum_{i=1}^d \frac{|\Delta_{ii}|^2}{\beta_i} \\ &\leq (d/\delta) \cdot \sum_{i,j=1}^d \frac{|\Delta_{ij}|^2}{\text{avg}\{\beta_i, \beta_j\}} = (d/\delta) \cdot D_{\chi^2}(\rho \parallel \sigma). \quad \square \end{aligned}$$

Combining all propositions in this section, we have established the following:

THEOREM 6.16. *Assume the smallest eigenvalue of σ is at least δ . Then*

$$\begin{aligned} \text{Var}_{\rho^{\otimes n}}[O_{\chi^2}] &\leq \frac{1}{\binom{n}{2}} \cdot \left(2d^2 + (2d/\delta) \cdot D_{\chi^2}(\rho \parallel \sigma) \right) \\ &\quad + \frac{2(n-2)}{\binom{n}{2}} \cdot \left(\sqrt{2d/\delta} \cdot D_{\chi^2}(\rho \parallel \sigma)^{3/2} + 2D_{\chi^2}(\rho \parallel \sigma) \right). \end{aligned}$$

6.3 Consequences for Testing

Assume σ is a fixed known density matrix, and we wish to estimate $D_{\chi^2}(\rho \parallel \sigma)$ given copies of an unknown density matrix ρ . Since we may first conjugate each copy of ρ by a unitary that diagonalizes σ , we may assume without loss of generality that σ is diagonal. Now the average χ^2 observable is an unbiased estimator for $D_{\chi^2}(\rho \parallel \sigma)$, and [Theorem 6.16](#) bounds its variance provided σ 's eigenvalues are not too small. Then from [Lemma 2.1](#) we immediately obtain [Theorem 1.3](#).

As mentioned, a corollary of [Theorem 1.3](#) is our main [Theorem 1.2](#), a robust “far-in-fidelity vs. close in χ^2 -divergence” tester with *no* assumption about σ 's eigenvalues. For convenience we restate and prove this theorem in the contrapositive and in terms of the squared Bures distance (which, recall, is exactly half the infidelity and is upper-bounded by the χ^2 -divergence):

COROLLARY 6.17 (EQUIVALENT TO [THEOREM 1.2](#)). *Fix a d -dimensional mixed state σ . Then there is an algorithm that, given $n = O(d/\epsilon)$ copies of ρ , (whp) outputs “close” if $D_{\chi^2}(\rho \parallel \sigma) \leq .49\epsilon$ and outputs “far” if $D_B^2(\rho, \sigma) > .5\epsilon$.*

PROOF. Let Φ_η denote the depolarizing channel, which maps a state $v \in \mathbb{C}^{d \times d}$ to the state $\Phi_\eta(v) = (1 - \eta)v + \eta \mathbf{1}/d$. Define $\rho' = \Phi_{c\epsilon}(\rho)$ and $\sigma' = \Phi_{c\epsilon}(\sigma)$, where $c > 0$ is a small absolute constant to be chosen later.

If $D_{\chi^2}(\rho \parallel \sigma) \leq .49\epsilon$ then $D_{\chi^2}(\rho' \parallel \sigma') \leq .49\epsilon$ by the quantum data processing inequality. On the other hand, in case $D_B^2(\rho, \sigma) > .5\epsilon$,

$$\sqrt{.5\epsilon} < D_B(\rho, \sigma) \leq D_B(\rho, \rho') + D_B(\rho', \sigma') + D_B(\sigma', \sigma) \quad (8)$$

by the triangle inequality. We can bound the first of these terms by

$$D_B^2(\rho, \rho') \leq 2D_{\text{tr}}(\rho, \rho') = \|\rho - \rho'\|_1 = \|c\epsilon\rho + c\epsilon\mathbf{1}/d\|_1 \leq 2c\epsilon,$$

where at the end we used the triangle inequality and $\|\rho\|_1 = \|\mathbf{1}/d\|_1 = 1$. A similar argument shows that $D_B^2(\sigma, \sigma') \leq 2c\epsilon$; i.e., $D_B(\sigma, \sigma') \leq \sqrt{2c\epsilon}$. Now taking c sufficiently small, (8) implies $D_B(\rho', \sigma') > \sqrt{.495\epsilon}$ and hence $D_{\chi^2}(\rho' \parallel \sigma') \geq D_B^2(\rho', \sigma') \geq .495\epsilon$.

In summary, if $D_{\chi^2}(\rho \parallel \sigma) \leq .49\epsilon$ then $D_{\chi^2}(\rho' \parallel \sigma') \leq .49\epsilon$, if $D_B^2(\rho, \sigma) > .5\epsilon$ then $D_{\chi^2}(\rho' \parallel \sigma') > .495\epsilon$, and all the eigenvalues

of σ' are at least $c\epsilon/d$. Thus we can obtain the desired tester by first applying the depolarizing channel $\Phi_{c\epsilon}$ to the n copies of ρ , producing n copies of ρ' , and then using the tester from [Theorem 1.3](#) with σ' in place of σ and $.5\epsilon$ in place of ϵ^2 . $\square$

We can also use this corollary to test if an unknown state is diagonal:

THEOREM 6.18. *Given $n = O(d/\epsilon)$ copies of a d -dimensional mixed state ρ , one can distinguish (whp) the case that ρ is diagonal (in the standard basis) from the case that ρ has infidelity more than ϵ with every diagonal state.*

PROOF. Let $p = (\rho_{11}, \dots, \rho_{dd})$ denote the diagonal of ρ , a probability distribution. We can obtain a sample from p given a copy of ρ simply by measuring ρ in the standard basis. Moreover, $O(d/\epsilon)$ samples suffice to produce an estimate $\hat{p}$ of p that satisfies $d_{\chi^2}(p \parallel \hat{p}_{\text{diag}}) \leq .49\epsilon$ (whp). The tester now applies [Corollary 6.17](#) with $\sigma = \text{diag}(\hat{p})$, using another $O(d/\epsilon)$ samples. If ρ is diagonal, then $D_{\chi^2}(\rho \parallel \sigma) = d_{\chi^2}(p \parallel \hat{p}) \leq .49\epsilon$ and the tester outputs “close” (whp). If ρ has infidelity more than ϵ with every diagonal state, then in particular $1 - F(\rho, \sigma) > \epsilon$; i.e., $D_B^2(\rho, \sigma) > .5\epsilon$, and the tester outputs “far” (whp). $\square$

ACKNOWLEDGMENTS

Costin Bădescu and Ryan O'Donnell are supported by NSF grant CCF-1618679. John Wright is supported by NSF grant CCF-6931885.

REFERENCES

- [AGKE15] Leandro Aolita, Christian Gogolin, Martin Kliesch, and Jens Eisert. Reliable quantum certification of photonic state preparations. *Nature Communications*, 6(8498), 2015.
- [ANSV08] Koenraad Audenaert, Michael Nussbaum, Arleta Szkola, and Frank Verstraete. Asymptotic error rates in quantum hypothesis testing. *Communications in Mathematical Physics*, 279(1):251–283, 2008.
- [Aud12] Koenraad Audenaert. Comparisons between quantum state distinguishability measures. Technical Report 1207.1197, arXiv, 2012.
- [BC94] Samuel Braunstein and Carlton Caves. Statistical distance and the geometry of quantum states. *Physical Review Letters*, 72(22):3439, 1994.
- [BZ07] Ingemar Bengtsson and Karol Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2007.
- [Can15] Clément Canonne. A survey on distribution testing: Your data is big. But is it blue? Technical Report 63, Electronic Colloquium on Computational Complexity, 2015.
- [CS06] Benoit Collins and Piotr Śniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264(3):773–795, 2006.
- [dSLP11] Marcus da Silva, Olivier Landon-Cardinal, and David Poulin. Practical characterization of quantum devices without tomography. *Physical Review Letters*, 107(21):210404, 2011.
- [FC95] Christopher Fuchs and Carlton Caves. Mathematical techniques for quantum communication theory. *Open Systems & Information Dynamics*, 3(3):345–356, 1995.
- [FL11] Steven Flammia and Yi-Kai Liu. Direct fidelity estimation from few Pauli measurements. *Physical Review Letters*, 106(23):230501, 2011.
- [GLN05] Alexei Gilchrist, Nathan Langford, and Michael Nielsen. Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71:062310, Jun 2005.
- [GW09] Roe Goodman and Nolan Wallach. *Symmetry, representations, and invariants*. Springer, 2009.
- [Hel76] Carl Helstrom. *Quantum Detection and Estimation Theory*. Academic Press, 1976.
- [HHJ⁺16] Jeongwan Haah, Aram Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, pages 913–925, 2016.
- [HHR⁺05] Hartmut Häfner, Wolfgang Hänsel, Christian Roos, Jan Benhelm, Michael Chwalla, Timo Körber, Umakant Rapol, Mark Riebe, Piet Schmidt, Christoph

- Becher, Otfried Günhe, Wolfgang Dür, and Rainer Blatt. Scalable multiparticle entanglement of trapped ions. *Nature*, 438(7068):643–646, 2005.
- [HM17] Fumio Hiai and Milán Mosonyi. Different quantum f -divergences and the reversibility of quantum operations. Technical Report 1604.03089, arXiv, 2017.
- [Kad52] Richard Kadison. A generalized Schwarz inequality and algebraic invariants for operator algebras. *Annals of Mathematics. Second Series*, 56:494–503, 1952.
- [MW16] Ashley Montanaro and Ronald de Wolf. *A Survey of Quantum Property Testing*. Number 7 in Graduate Surveys. Theory of Computing Library, 2016.
- [OW15] Ryan O'Donnell and John Wright. Quantum spectrum testing. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 529–538, 2015.
- [OW16] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, pages 899–912, 2016.
- [OW17] Ryan O'Donnell and John Wright. Efficient quantum tomography II. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing*, pages 962–974, 2017.
- [Pet96] Dénes Petz. Monotone metrics on matrix spaces. *Linear algebra and its applications*, 244:81–96, 1996.
- [TKR⁺10] Kristan Temme, Michael Kastoryano, Mary Ruskai, Michael Wolf, and Frank Verstraete. The χ^2 -divergence and mixing times of quantum Markov processes. *Journal of Mathematical Physics*, 51(12):122201, 2010.
- [TV15] Kristan Temme and Frank Verstraete. Quantum chi-squared and goodness of fit testing. *Journal of Mathematical Physics*, 56(1):012202, 18, 2015.
- [Uhl76] Armin Uhlmann. The “transition probability” in the state space of a $*$ -algebra. *Reports on Mathematical Physics*, 9(2):273–279, 1976.