Limitations on quantum dimensionality reduction

Citation: Harrow, Aram W., Ashley Montanaro, and Anthony J. Short. "Limitations on Quantum Dimensionality Reduction." Int. J. Quantum Inform. 13, no. 04 (June 2015): 1440001.

As Published: http://dx.doi.org/10.1142/S0219749914400012

Publisher: World Scientific

Persistent URL: http://hdl.handle.net/1721.1/98056

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

International Journal of Quantum Information
Vol. 13, No. 4 (2015) 1440001 (19 pages)

DOI: 10.1142/S0219749914400012

Limitations on quantum dimensionality reduction*

Aram W. Harrow

Center for Theoretical Physics,
Massachusetts Institute of Technology,
Cambridge, MA, USA
aram@mit.edu

Ashley Montanaro

Department of Computer Science,
University of Bristol,
Bristol, UK
ashley@cs.bris.ac.uk

Anthony J. Short

School of Physics, University of Bristol,
Bristol, UK
tony.short@bristol.ac.uk

Received 10 October 2013
Accepted 11 May 2014
Published 3 June 2014

The Johnson–Lindenstrauss Lemma is a classic result which implies that any set of n real vectors can be compressed to $O(\log n)$ dimensions while only distorting pairwise Euclidean distances by a constant factor. Here we consider potential extensions of this result to the compression of quantum states. We show that, by contrast with the classical case, there does not exist any distribution over quantum channels that significantly reduces the dimension of quantum states while preserving the 2-norm distance with high probability. We discuss two tasks for which the 2-norm distance is indeed the correct figure of merit. In the case of the trace norm, we show that the dimension of low-rank mixed states can be reduced by up to a square root, but that essentially no dimensionality reduction is possible for highly mixed states.

Keywords: Dimensionality reduction; quantum channels.

*An extended abstract of this work previously appeared in the proceedings of ICALP'11.

1. Introduction

The Johnson–Lindenstrauss (JL) Lemma[1] is a dimensionality reduction result which has found a vast array of applications in computer science and elsewhere (see e.g. Refs. 2–4). It can be stated as follows.

Theorem 1 (Johnson–Lindenstrauss Lemma[1]). *For all dimensions d, e, there is a distribution $\mathcal{D}$ over linear maps $\mathcal{E} : \mathbb{R}^d \to \mathbb{R}^e$ such that, for all real vectors v, w,*

$$\Pr_{\mathcal{E}\sim\mathcal{D}}[(1-\epsilon)\|v-w\|_2 \leq \|\mathcal{E}(v)-\mathcal{E}(w)\|_2 \leq \|v-w\|_2] \geq 1-\exp(-\Omega(\epsilon^2 e)),$$

where $\|\cdot\|_2$ is the Euclidean (ℓ_2) distance.

The lemma is usually applied via the following corollary, which follows by taking a union bound.

Corollary 2. *Given a set S of n d-dimensional real vectors, there is a linear map $\mathcal{E} : \mathbb{R}^d \to \mathbb{R}^{O(\log n/\epsilon^2)}$ that preserves all Euclidean distances in S, up to a multiple of $1-\epsilon$. Further, there is an efficient randomized algorithm to find and implement $\mathcal{E}$.*

There are several remarkable aspects of this result. First, the target dimension does not depend on the source dimension d at all. Second, the randomized algorithm can be simply stated as: choose a random e-dimensional subspace with $e = O(\log n/\epsilon^2)$, project each vector in S onto this subspace, and rescale the result by a constant that does not depend on S. Third, this algorithm is *oblivious*: in other words, $\mathcal{E}$ does not depend on the vectors whose dimensionality is to be reduced.

More generally, let ℓ_p^d be the vector space $\mathbb{R}^d$ equipped with the ℓ_p-norm $\|\cdot\|_p$. A *randomized embedding* from ℓ_p^d to ℓ_p^e with distortion[a] $1/(1-\epsilon)$ and failure probability δ is a distribution $\mathcal{D}$ over maps $\mathcal{E} : \mathbb{R}^d \to \mathbb{R}^e$ such that, for all $v, w \in \mathbb{R}^d$,

$$\Pr_{\mathcal{E}\sim\mathcal{D}}[(1-\epsilon)\|v-w\|_p \leq \|\mathcal{E}(v)-\mathcal{E}(w)\|_p \leq \|v-w\|_p] \geq 1-\delta.$$

This definition does not allow the distance between vectors to increase; such embeddings are called contractive. The JL Lemma states that there exists a randomized embedding from ℓ_2^d to ℓ_2^e with distortion $1/(1-\epsilon)$ and failure probability $\exp(-\Omega(\epsilon^2 e))$. Another natural norm to consider in this context is ℓ_1. In this case the situation is less favorable: it has been shown by Charikar and Sahai[5] that there exist $O(d)$ points in ℓ_1^d such that any linear embedding into ℓ_1^e must incur distortion $\Omega(\sqrt{d/e})$. Brinkman and Charikar later gave a set of n points for which any (even nonlinear) embedding achieving distortion D requires $n^{\Omega(1/D^2)}$ dimensions.[6]

1.1. *The JL Lemma in quantum information theory*

The JL Lemma immediately gives rise to a protocol for *quantum fingerprinting*,[7] or in other words efficient equality testing. Imagine that Alice and Bob each have an n-bit

[a] We use this somewhat clumsy definition of distortion for consistency with prior work.

string, and are required to send quantum states of the shortest possible length to a referee, who has to use these states to determine if their bit strings are equal (this is the so-called SMP, or simultaneous message passing, model of communication complexity[8]). Associate each bit string with an orthonormal basis vector of $\mathbb{R}^{2^n}$. Then the JL Lemma guarantees that there exists a map from $\mathbb{R}^{2^n}$ into $\mathbb{R}^{O(n)}$ such that the inner products between all of these 2^n vectors are preserved, up to a small constant. So Alice and Bob each simply apply this map to their vectors, renormalize the output (which makes very little difference to the inner products), and send the $O(\log n)$ qubit states corresponding to the resulting $O(n)$-dimensional vectors to the referee, who applies the swap test to the states.[7] Given two states $|\psi\rangle$, $|\phi\rangle$, this test accepts with probability $\frac{1}{2} + \frac{1}{2}|\langle\psi|\phi\rangle|^2$. As the inner products are approximately preserved by the map into $\mathbb{R}^{O(n)}$, the referee can distinguish between the two cases of the states he receives being equal or distinct, with constant probability.

More generally, Alice and Bob can use a similar SMP protocol to solve the following task: given quantum states $|\psi_A\rangle$, $|\psi_B\rangle$, each picked from a set of k states, determine $\langle\psi_A|\psi_B\rangle$ up to a constant. Whatever the initial dimension of the states, the JL Lemma (strictly speaking, an easy extension of the JL Lemma to complex vectors) guarantees that they can be compressed to $O(\log k)$ dimensions with at most constant distortion, implying that the referee can estimate $\langle\psi_A|\psi_B\rangle$ up to a constant using only $O(\log\log k)$ qubits of communication.

However, there is a problem with this protocol. While it is oblivious in the sense that it does not depend on the k states which are given as input, it is not oblivious in the following quantum sense: Alice and Bob each need to know what their states are in order to apply the embedding.[b] One would expect the right quantum analogue of a randomized embedding to map quantum states to quantum states in an oblivious fashion. Such an algorithm can be expressed as a distribution over quantum channels (completely positive, trace preserving (CPTP) maps[9,10]), which are the class of physically implementable operations in quantum theory.

Let $\mathcal{B}(d)$ denote the set of d-dimensional Hermitian operators. The distance between quantum states ρ, $\sigma \in \mathcal{B}(d)$ can be measured using the Schatten p-norm $\|\rho - \sigma\|_p$, which is defined as $\|X\|_p = (\sum_i |\lambda_i(X)|^p)^{1/p}$, where $\lambda_i(X)$ is the ith eigenvalue of X. The case $p = 1$ is known as the trace norm, and $p = 2$ is sometimes known as the Hilbert–Schmidt norm. We have the following definition.

Definition 1. A *quantum embedding* from $S \subseteq \mathcal{B}(d)$ to $\mathcal{B}(e)$ in the Schatten p-norm, with distortion $1/(1-\epsilon)$ and failure probability δ, is a distribution $\mathcal{D}$ over quantum channels $\mathcal{E} : \mathcal{B}(d) \to \mathcal{B}(e)$ such that, for all ρ, $\sigma \in S$,

$$\Pr_{\mathcal{E}\sim\mathcal{D}}[(1-\epsilon)\|\rho-\sigma\|_p \leq \|\mathcal{E}(\rho) - \mathcal{E}(\sigma)\|_p \leq \|\rho-\sigma\|_p] \geq 1-\delta.$$

Rather than only considering embeddings that succeed for all states in $\mathcal{B}(d)$, we generalize the definition to subsets of states. An interesting such subset is the pure

[b] On the other hand, if the unphysical operation of postselection is allowed, the JL Lemma can be applied directly.

states, for which one might imagine stronger embeddings can be obtained. Indeed, a closely related notion has been studied before by Winter,[11] and more recently Hayden and Winter,[12] under the name of quantum identification for the identity channel. In this setting, the sender Alice has a pure state $|\psi\rangle \in \mathbb{C}^d$ and the receiver Bob is given the description of a pure state $|\phi\rangle \in \mathbb{C}^d$. Alice encodes her state $|\psi\rangle$ as a quantum message using a quantum channel $\mathcal{E} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^e)$ and sends it to Bob, who performs a measurement $(D_\phi, I - D_\phi)$ on the message. The goal is to obtain approximately the same measurement statistics as if Bob had performed the measurement $(|\phi\rangle\langle\phi|, I - |\phi\rangle\langle\phi|)$ on $|\psi\rangle$:

$$\forall\ |\psi\rangle, |\phi\rangle, \quad |\mathrm{tr}[D_\phi\, \mathcal{E}(|\psi\rangle\langle\psi|)] - |\langle\psi|\phi\rangle|^2| \leq \epsilon.$$

Winter showed in Ref. 11 that, for constant ϵ, this can be achieved with $e = O(\sqrt{d})$; note that the resulting states $\mathcal{E}(|\psi\rangle\langle\psi|)$ are highly mixed. Winter's result allows the development of a one-way protocol for testing equality of n-bit strings using $\frac{1}{2}\log_2 n + O(1)$ qubits of communication from Alice to Bob, which is still the best known separation between one-way quantum and classical communication complexity for total functions.[13] In our terminology, the result of Ref. 11 shows that there exists a quantum embedding from $\mathcal{B}(d)$ to $\mathcal{B}(O(\sqrt{d}))$ that approximately preserves the trace distance between (initially) pure states. But note that one aspect of Winter's result is stronger than we need: he showed the existence of a channel such that the distance is approximately preserved between *all* pairs of states. Here, we are interested in finding distributions $\mathcal{D}$ over channels $\mathcal{E}$ such that, for an arbitrary pair of states, the distance is approximately preserved with high probability; this is potentially a weaker notion. In particular, it is not necessarily true that the individual channel obtained by averaging over $\mathcal{D}$ will preserve the distance between an arbitrary pair of states.

We pause to mention that the JL Lemma has found some other uses in quantum information theory. Cleve *et al.*[14] used it to give an upper bound on the amount of shared entanglement required to win a particular class of non-local games. Gavinsky, Kempe and de Wolf[15] used it to give a simulation of arbitrary quantum communication protocols by quantum SMP protocols (with exponential overhead). Embeddings between norms have also been used. Aubrun, Szarek and Werner[16,17] have used a version of Dvoretzky's theorem on "almost-Euclidean" subspaces of matrices under Schatten norms to give counterexamples to the additivity conjectures of quantum information theory. And, more recently, Fawzi, Hayden and Sen[18] have used ideas from the theory of low-distortion embeddings of the "$\ell_1(\ell_2)$"-norm to prove the existence of strong entropic uncertainty relations.

1.2. *Our results*

In this paper, we show that the dimensionality reduction that can be achieved by quantum embeddings is very limited. We begin, in Sec. 2, by considering the Schatten 2-norm (which is just the vector 2-norm on matrices). We show that, in stark contrast to the JL Lemma, any quantum embedding which preserves the 2-norm distance

between (say) orthogonal pure states with constant distortion and constant failure probability can only achieve at most a constant reduction in dimension.

One potential criticism of this result is that the 2-norm is not usually seen as a physically meaningful distance measure, as compared with the trace norm. However, we argue in Sec. 3 that for certain problems the 2-norm is indeed the correct distance measure. We discuss two problems — equality testing without a reference frame and state discrimination with a random measurement — where the 2-norm appears naturally as the figure of merit.

In Sec. 4 we turn to the trace norm, for which we have upper and lower bounds. On the upper bound side, we extend the result of Winter[11] to show that low-rank mixed states are also amenable to dimensionality reduction; roughly speaking, d-dimensional mixed states of rank r can be embedded into $O(\sqrt{rd})$ dimensions with constant distortion. On the other hand, we show using the 2-norm lower bound that highly mixed states cannot be embedded into low dimension: there is a lower bound of $\Omega(\sqrt{d}\frac{\|\rho-\sigma\|_1}{\|\rho-\sigma\|_2})$ on the target dimension of any constant distortion trace norm embedding that succeeds with constant probability for the pairs $U\rho U^\dagger$, $U\sigma U^\dagger$ for all unitary operators U. In particular, this implies an $\Omega(\sqrt{d})$ lower bound for any embedding which succeeds for a unitarily invariant set of states. In the case that $|\rho-\sigma|$ is proportional to a projector (i.e. all non-zero eigenvalues of $\rho-\sigma$ are equal in absolute value), our upper and lower bounds coincide.

Finally, some notes on miscellaneous notation. F_d will denote the unitary operator which swaps (or flips) two d-dimensional quantum systems (i.e. $F_d = \sum_{i,j=1}^{d} |i\rangle\langle j| \otimes |j\rangle\langle i|$), and I_d will denote the d-dimensional identity matrix. Whenever we say that $U \in U(d)$ is a random unitary operator, we mean that U is picked uniformly at random according to Haar measure on the unitary group $U(d)$.

2. Dimensionality Reduction in the 2-Norm

We now show that quantum dimensionality reduction in the 2-norm is very limited.

Theorem 3. *Let $\mathcal{D}$ be a distribution over quantum channels (CPTP maps) $\mathcal{E} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^e)$ such that, for fixed quantum states $\rho \neq \sigma$ and for all unitary operators $U \in U(d)$,*

$$\Pr_{\mathcal{E}\sim\mathcal{D}}[\|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2 \geq (1-\epsilon)\|U\rho U^\dagger - U\sigma U^\dagger\|_2] \geq 1-\delta$$

for some $0 \leq \epsilon, \delta \leq 1$. Then $e \geq (1-\delta)(1-\epsilon)^2 d$.

Note that the above lower bound on target dimension holds for any embedding of a unitarily invariant set of states. For example, taking ρ and σ to be orthogonal pure states and inserting $\epsilon = \delta = 0$ recovers the (unsurprising) result that any embedding that exactly preserves distances between all orthogonal pure states with certainty must satisfy $e \geq d$. More generally, if we have an embedding which succeeds with constant probability and has constant distortion, the target dimension can be no

smaller than $\Omega(d)$. In order to prove the theorem, we will need the following two technical lemmas, which are proved in Appendix A.

Lemma 4. *Let $\mathcal{E} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^e)$ be a quantum channel (CPTP map). Then*

$$\mathrm{tr}[F_e \, \mathcal{E}^{\otimes 2}(F_d)] \leq de.$$

Lemma 5. *Let ρ and σ be d-dimensional quantum states. Then*

$$\int U^{\otimes 2}(\rho - \sigma)^{\otimes 2}(U^\dagger)^{\otimes 2} dU = \frac{\|\rho - \sigma\|_2^2}{d^2 - 1}\left(F_d - \frac{I_{d^2}}{d}\right).$$

The following lemma is the key to most of the results in this paper.

Lemma 6. *Let ρ and σ be quantum states and let $\mathcal{E} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^e)$ be a quantum channel. Then*

$$\int \|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2^2 \, dU \leq \frac{d(e^2 - 1)}{e(d^2 - 1)} \|\rho - \sigma\|_2^2.$$

Proof. We have

$$\begin{aligned}
\int \|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2^2 \, dU &= \int \|\mathcal{E}(U(\rho - \sigma)U^\dagger)\|_2^2 \, dU \\
&= \int \mathrm{tr}[F_e \, \mathcal{E}(U(\rho - \sigma)U^\dagger)^{\otimes 2}] \, dU \\
&= \mathrm{tr}\left[F_e \, \mathcal{E}^{\otimes 2}\left(\int U^{\otimes 2}(\rho - \sigma)^{\otimes 2}(U^\dagger)^{\otimes 2} \, dU\right)\right] \\
&= \frac{\|\rho - \sigma\|_2^2}{d^2 - 1} \mathrm{tr}\left[F_e \, \mathcal{E}^{\otimes 2}\left(F_d - \frac{I_{d^2}}{d}\right)\right] \\
&\leq \frac{\|\rho - \sigma\|_2^2}{d^2 - 1}(de - d\mathrm{tr}[\mathcal{E}(I_d/d)^2]) \\
&\leq \frac{d(e^2 - 1)}{e(d^2 - 1)} \|\rho - \sigma\|_2^2.
\end{aligned}$$

We use linearity of $\mathcal{E}$ in the first equality, and the second equality is the tensor product trick $\mathrm{tr}[X^2] = \mathrm{tr}[F_e X^{\otimes 2}]$ for e-dimensional operators X. The fourth equality is Lemma 5, the first inequality is Lemma 4, and the second inequality is simply $\mathrm{tr}\rho^2 \geq 1/e$ for all e-dimensional states ρ. □

We are finally ready to prove Theorem 3.

Proof of Theorem 3. We will prove something slightly stronger: that for a random U, the 2-norm is not approximately preserved under a map $\mathcal{E}$ picked from $\mathcal{D}$, unless e is almost as large as d. So assume

$$\Pr_{\mathcal{E} \sim \mathcal{D}, U \in U(d)}[\|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2 \geq (1 - \epsilon)\|U\rho U^\dagger - U\sigma U^\dagger\|_2] \geq 1 - \delta,$$

or equivalently

$$\Pr_{\mathcal{E}\sim\mathcal{D}, U\in U(d)}[\|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2^2 \geq (1-\epsilon)^2\|\rho-\sigma\|_2^2] \geq 1-\delta,$$

where we use the unitary invariance of the 2-norm. By Markov's inequality, this implies that

$$\int_{\mathcal{E}\sim\mathcal{D}} \int \|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2^2 \, dU \geq (1-\delta)(1-\epsilon)^2\|\rho-\sigma\|_2^2,$$

implying in turn that there must exist some $\mathcal{E}$ such that

$$\int \|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2^2 \, dU \geq (1-\delta)(1-\epsilon)^2\|\rho-\sigma\|_2^2.$$

So let $\mathcal{E} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^e)$ be a quantum channel that does satisfy this inequality. Then we have

$$(1-\delta)(1-\epsilon)^2\|\rho-\sigma\|_2^2 \leq \int \|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2^2 \, dU \leq \left(\frac{e}{d}\right)\|\rho-\sigma\|_2^2,$$

where the second inequality follows from Lemma 6, assuming that $e \leq d$. We have shown that $e \geq (1-\delta)(1-\epsilon)^2 d$, completing the proof of the theorem. □

3. Operational Meaning of the 2-Norm

In this section, we discuss the meaning of the 2-norm distance between quantum states. It is usually assumed that the trace norm is the "right" measure of distance between states, and proofs going via the 2-norm usually do so only for calculational simplicity. However, here we argue that the 2-norm is of interest in its own right, by giving two operational interpretations of this distance measure.

3.1. *Equality testing without a reference frame*

Consider the following equality-testing game. We are given a description of two different states ρ and σ. An adversary prepares two systems in one of the states $\rho \otimes \rho$, $\sigma \otimes \sigma$, $\rho \otimes \sigma$ or $\sigma \otimes \rho$, with equal probability of each. He then applies an unknown unitary U to each system (i.e. he applies $U \otimes U$ to the joint state). Our task is to determine whether the two systems have the same state or different states. This models equality testing in a two-party scenario in which the preparer and tester do not share a reference frame.[19] One protocol for solving this task is simply to apply the swap test[7] to the two states we are given, output "same" if the test accepts, and "different" otherwise. When applied to two states ρ, σ this test accepts with probability $\frac{1}{2} + \frac{1}{2}\mathrm{tr}\rho\sigma$, so for any U the overall probability of success is

$$\frac{1}{4}\left(\frac{1}{2} + \frac{1}{2}\mathrm{tr}[\rho^2]\right) + \frac{1}{4}\left(\frac{1}{2} + \frac{1}{2}\mathrm{tr}[\sigma^2]\right) + \frac{1}{2}\left(\frac{1}{2} - \frac{1}{2}\mathrm{tr}[\rho\sigma]\right) = \frac{1}{2} + \frac{1}{8}\|\rho-\sigma\|_2^2.$$

Using our previous result, we now show that this is optimal.

Theorem 7. *The maximal probability of success of the above game is* $\frac{1}{2} + \frac{1}{8}\|\rho - \sigma\|_2^2$.

Proof. Let $(M, I - M)$ be an arbitrary POVM where the operator M corresponds to the answer "same". Then the probability of success achieved by this POVM for a given U is $\frac{1}{2} + \frac{1}{2}B$, where B is the *bias*, which is equal to

$$\frac{1}{2}\mathrm{tr}[M((U\rho U^\dagger \otimes U\rho U^\dagger + U\sigma U^\dagger \otimes U\sigma U^\dagger) - (U\rho U^\dagger \otimes U\sigma U^\dagger + U\sigma U^\dagger \otimes U\rho U^\dagger))].$$

If the adversary adopts the strategy of picking U uniformly at random, the average bias obtained is

$$\frac{1}{2}\mathrm{tr}\left[M\int U^{\otimes 2}(\rho\otimes\rho + \sigma\otimes\sigma - \rho\otimes\sigma - \sigma\otimes\rho)(U^\dagger)^{\otimes 2}dU\right]$$
$$= \frac{1}{2}\mathrm{tr}\left[M\int U^{\otimes 2}(\rho-\sigma)^{\otimes 2}(U^\dagger)^{\otimes 2}\right],$$

which by Lemma 5 is equal to

$$\frac{\|\rho-\sigma\|_2^2}{2(d^2-1)}\mathrm{tr}\left[M\left(F_d - \frac{I_{d^2}}{d}\right)\right].$$

This expression is maximized by setting M equal to a projector onto the subspace spanned by the eigenvectors of $F_d - \frac{I_{d^2}}{d}$ with positive eigenvalues. As F_d has $d(d+1)/2$ eigenvalues equal to 1, and $d(d-1)/2$ eigenvalues equal to -1, we obtain $\mathrm{tr}[M(F_d - \frac{I_{d^2}}{d})] = (d^2-1)/2$. This implies that the average bias is at most $\frac{1}{4}\|\rho-\sigma\|_2^2$. As the worst-case bias can only be lower, this implies the claimed result. □

3.2. *Performing a random measurement*

The second game we will discuss is state discrimination with a fixed or random measurement. Imagine we are given a state which is promised to be either ρ or σ, with equal probability of each, and we wish to determine which is the case. It is well known that the largest bias achievable by choosing an appropriate measurement is $\frac{1}{2}\|\rho - \sigma\|_1$ (recall from the previous section that the bias B and the success probability p have the relationship $p = \frac{1}{2} + \frac{B}{2}$). But how well can we do if the measurement we apply does not in fact depend on ρ and σ?

We will see that $\|\rho - \sigma\|_2$ is closely related to the optimal bias achievable by performing one of the following two measurements, and deciding whether the state is ρ or σ based on the outcome.

- The uniform (isotropic) POVM whose measurement elements consist of normalized projectors onto all states $|\psi\rangle$.
- A projective measurement in a random basis (i.e. applying a random unitary operator and measuring in the computational basis).

In general, the largest bias achievable by measuring a POVM M which consists of measurement operators M_i can be written as

$$\frac{1}{2}\sum_i |\mathrm{tr}[M_i(\rho-\sigma)]|.$$

Each measurement operator of the uniform POVM is given by the projector onto some state $|\psi\rangle$, normalized by a factor of d (to check that this is right, note that

$$d\int d\psi |\psi\rangle\langle\psi| = d\left(\frac{I_d}{d}\right) = I_d$$

as expected). So the bias induced by the uniform POVM is

$$\frac{d}{2}\int d\psi |\langle\psi|(\rho-\sigma)|\psi\rangle|.$$

In the case of a measurement in a random basis $U \in U(d)$, we can calculate the *expected* bias as follows:

$$\begin{aligned}\frac{1}{2}\mathbb{E}_U\sum_{i=1}^{d}|\langle i|U^\dagger(\rho-\sigma)U|i\rangle| &= \frac{1}{2}\sum_{i=1}^{d}\mathbb{E}_U|\langle i|U^\dagger(\rho-\sigma)U|i\rangle| \\ &= \frac{1}{2}\sum_{i=1}^{d}\mathbb{E}_U|\langle 1|U^\dagger(\rho-\sigma)U|1\rangle| \\ &= \frac{d}{2}\int d\psi|\langle\psi|(\rho-\sigma)|\psi\rangle|;\end{aligned}$$

so these quantities are the same. They are also closely related to the 2-norm distance, as we will now see.

Theorem 8. *Let ρ, σ be d-dimensional quantum states. Then*

$$\frac{1}{3}\|\rho-\sigma\|_2 \le d\int d\psi|\langle\psi|(\rho-\sigma)|\psi\rangle| \le \|\rho-\sigma\|_2.$$

The lower bound in Theorem 8 was shown by Ambainis and Emerson[20] (see also the proof of Matthews, Wehner and Winter[21]), and the upper bound is not hard. However, as this result does not appear to be widely known, we include a proof (which is essentially the same as that of Ref. 21) in Appendix B.

In fact, the corresponding upper and lower bounds on the bias hold for any fixed POVM whose measurement vectors form a 4-design,[20] and the upper bound even holds for any fixed POVM whose vectors form a 2-design. This result can be useful in cases where one wishes to perform state discrimination without necessarily being able to construct the optimal measurement efficiently.[22] See Ref. 21 for much more detail on the bias achievable in state discrimination with fixed measurements.

4. Dimensionality Reduction in the Trace Norm

In this section we consider embeddings that reduce dimension while preserving the trace norm distance between states. As no quantum channel can increase this distance, we first observe that any such embedding will automatically be contractive.

4.1. *Upper bound*

It was previously shown by Winter[11] that, in our language, d-dimensional pure states can be embedded into $\mathcal{B}(O(\sqrt{d}))$ with constant distortion. We now extend this result to general mixed states, by showing that rank r mixed states can be embedded into dimension $O(\sqrt{rd})$ with constant distortion.

The embedding is conceptually very simple: apply a random unitary and trace out a subsystem. However, when the target dimension e does not divide d, we are forced to consider random isometries $V : \mathbb{C}^d \to \mathbb{C}^e \otimes \mathbb{C}^{\lceil d/e \rceil}$ instead of unitaries, where $\lceil x \rceil$ is the smallest integer y such that $y \geq x$. Recall that an isometry is a norm-preserving linear map, i.e. a map taking an orthonormal basis of one space to an orthonormal set of vectors in another (potentially larger) space. A random isometry is defined as a fixed isometry followed by a random unitary.

Formally, our embedding is a distribution over the following quantum channels $\mathcal{E}_V$.

Definition 2. Let d and e be positive integers such that $e \leq d$. For any isometry $V : \mathbb{C}^d \to \mathbb{C}^e \otimes \mathbb{C}^{\lceil d/e \rceil}$, let $\mathcal{E}_V : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^e)$ be the quantum channel that consists of performing V, then tracing out (discarding) the second subsystem.

We now analyze the performance of the embedding obtained by picking a random V and applying this channel.

Theorem 9. *Let d be a positive integer, and let ρ and σ be arbitrary d-dimensional mixed states such that ρ has rank r. Fix ϵ such that $0 < \epsilon < 1$. For any e such that $2\sqrt{rd/\epsilon} \leq e \leq d$, let $\mathcal{D}$ be the distribution on channels $\mathcal{E}_V : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^e)$ that is uniform on isometries $V : \mathbb{C}^d \to \mathbb{C}^e \otimes \mathbb{C}^{\lceil d/e \rceil}$. Then*

$$\Pr_{\mathcal{E}_V \sim \mathcal{D}}[\|\mathcal{E}_V(\rho) - \mathcal{E}_V(\sigma)\|_1 \geq (1-\epsilon)\|\rho - \sigma\|_1] \geq 1 - d\exp(-K\epsilon d),$$

for a universal constant K which may be taken to be $(1 - \ln 2)/(2 \ln 2) \approx 0.22$.

In order to prove this theorem, we will need the following technical lemma, which is proven in Appendix C.

Lemma 10. *Let $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ be a finite-dimensional Hilbert space decomposed into subsystems A and B. For any projector P onto a subspace of $\mathcal{H}$, let $P^\perp = I - P$ be the projector onto the orthogonal subspace, and let D be the projector onto the*

support of $\mathrm{tr}_B P$. *Then, for any* $|\psi\rangle \in \mathcal{H}$,

$$\mathrm{tr}[(D \otimes I)P^{\perp}|\psi\rangle\langle\psi|P^{\perp}] \leq \mathrm{tr}[(D \otimes I)|\psi\rangle\langle\psi|]\mathrm{tr}[P^{\perp}|\psi\rangle\langle\psi|].$$

We will also need the following useful result of Bennett *et al.*[23] (see also Ref. 11).

Lemma 11. *Let* $|\psi\rangle$ *be a* d*-dimensional pure state, let* P *be the projector onto a* t*-dimensional subspace of* $\mathbb{C}^d$*, and let* $U \in U(d)$ *be picked according to Haar measure. Then, for any* $\delta \geq 0$,

$$\Pr_U \left[\mathrm{tr}[UPU^{\dagger}|\psi\rangle\langle\psi|] \geq (1+\delta)\frac{t}{d} \right] \leq \exp(-t(\delta - \ln(1+\delta))/(\ln 2)).$$

Proof of Theorem 9. We will upper bound the probability of the embedding failing, i.e.

$$\Pr_V[\|\mathcal{E}_V(\rho - \sigma)\|_1 < (1-\epsilon)\|\rho - \sigma\|_1].$$

Let S^+ and S^- be the disjoint sets of indices of $(\rho - \sigma)$'s positive and negative eigenvalues, respectively. Set $s = |S^+|$, and note that $s \leq \mathrm{rank}(\rho) = r$.[24] For a fixed V, expand $V(\rho - \sigma)V^{\dagger}$ as follows:

$$V(\rho - \sigma)V^{\dagger} = \sum_{i \in S^+} \lambda_i |\psi_i\rangle\langle\psi_i| - \sum_{i \in S^-} \mu_i |\psi_i\rangle\langle\psi_i|$$

for some orthonormal vectors $|\psi_i\rangle \in \mathbb{C}^e \otimes \mathbb{C}^{\lceil d/e \rceil}$ and positive coefficients λ_i, μ_i. Note that

$$\sum_{i \in S^+} \lambda_i = \sum_{i \in S^-} \mu_i = \|\rho - \sigma\|_1/2.$$

For any states ρ' and σ', it holds that

$$\|\rho' - \sigma'\|_1 = 2 \sup_{0 \leq M \leq I} \mathrm{tr} M(\rho' - \sigma');$$

in a protocol for distinguishing ρ' and σ', M is a measurement operator corresponding to the outcome that the state was ρ'. Thus, in order for it to hold that $\|\mathcal{E}_V(\rho - \sigma)\|_1 \geq (1-\epsilon)\|\rho - \sigma\|_1$, it suffices to exhibit an operator M such that $0 \leq M \leq I$ and

$$\mathrm{tr}[M(\mathcal{E}_V(\rho - \sigma))] \geq (1-\epsilon)\|\rho - \sigma\|_1/2 = (1-\epsilon)\sum_{i \in S^+} \lambda_i.$$

To find such an operator, set

$$P_V := \sum_{i \in S^+} |\psi_i\rangle\langle\psi_i|.$$

Note that P_V is the projector onto a random s-dimensional subspace of $\mathbb{C}^e \otimes \mathbb{C}^{\lceil d/e \rceil}$. Now let D_V be the projector onto the support of $\mathrm{tr}_B P_V$. Then

$$\mathrm{tr}[D_V \mathcal{E}_V(\rho - \sigma)] = \sum_{i \in S^+} \lambda_i \mathrm{tr}[D_V \mathrm{tr}_B |\psi_i\rangle\langle\psi_i|] - \sum_{i \in S^-} \mu_i \mathrm{tr}[D_V \mathrm{tr}_B |\psi_i\rangle\langle\psi_i|]. \quad (1)$$

For all $i \in S^+$, $\mathrm{tr}[D_V \mathrm{tr}_B |\psi_i\rangle\langle\psi_i|] = 1$, and for all $i \in S^-$, it holds that $\mathrm{tr}[P_V|\psi_i\rangle\langle\psi_i|] = 0$. Aside from this constraint, each individual state $|\psi_i\rangle$, $i \in S^-$, is picked at random and can be expressed in terms of a general random state $|\eta\rangle \in \mathbb{C}^e \otimes \mathbb{C}^{\lceil d/e \rceil}$ as

$$|\psi_i\rangle = \frac{P_V^\perp |\eta\rangle}{\|P_V^\perp |\eta\rangle\|_2},$$

where $P_V^\perp = I - P_V$ and the denominator is non-zero with probability 1. Then

$$\mathrm{tr}[(D_V \otimes I)|\psi_i\rangle\langle\psi_i|] = \frac{\mathrm{tr}[(D_V \otimes I)(P_V^\perp|\eta\rangle\langle\eta|P_V^\perp)]}{\mathrm{tr}[P_V^\perp|\eta\rangle\langle\eta|]} \leq \mathrm{tr}[(D_V \otimes I)|\eta\rangle\langle\eta|],$$

where the inequality is Lemma 10. For any e such that $e \geq s\lceil d/e \rceil$, D_V has rank $s\lceil d/e \rceil$ with probability 1. So, for any such e, $D_V \otimes I$ has rank $s\lceil d/e \rceil^2$ with probability 1. Applying Lemma 11, for any $\delta \geq 0$,

$$\Pr_{|\eta\rangle}\left[\mathrm{tr}[(D_V \otimes I)|\eta\rangle\langle\eta|] \geq (1+\delta)\frac{s\lceil d/e \rceil^2}{e\lceil d/e \rceil}\right] \leq \exp(-s\lceil d/e \rceil^2(\delta - \ln(1+\delta))/(\ln 2))$$

and hence

$$\Pr_V\left[\mathrm{tr}[(D_V \otimes I)|\psi_i\rangle\langle\psi_i|] \geq (1+\delta)\frac{s\lceil d/e \rceil}{e}\right] \leq \exp(-s\lceil d/e \rceil^2(\delta - \ln(1+\delta))/(\ln 2)).$$

Using a union bound over S^- in Eq. (1), for any e satisfying $e \geq s\lceil d/e \rceil$ it holds that

$$\Pr_V\left[\mathrm{tr}[D_V \mathcal{E}_V(\rho - \sigma)] \leq \sum_{i \in S^+} \lambda_i - (1+\delta)\frac{s\lceil d/e \rceil}{e}\sum_{i \in S^-}\mu_i\right] \leq d\exp(-s\lceil d/e \rceil^2(\delta - \ln(1+\delta))/(\ln 2)).$$

We now set $\delta = \frac{\epsilon e}{s\lceil d/e \rceil} - 1$. This gives the following bound, valid when $\epsilon e \geq s\lceil d/e \rceil$:

$$\begin{aligned}
&\Pr_V[\mathrm{tr}[D_V \mathcal{E}_V(\rho - \sigma)] \leq (1-\epsilon)\|\rho - \sigma\|_1/2] \\
&\quad \leq d\exp\left(-s\lceil d/e \rceil^2\left(\frac{\epsilon e}{s\lceil d/e \rceil} - 1 - \ln\left(\frac{\epsilon e}{s\lceil d/e \rceil}\right)\right)\Big/(\ln 2)\right) \\
&\quad \leq d\exp\left(-s(d/e)\lceil d/e \rceil\left(\frac{\epsilon e}{s\lceil d/e \rceil} - 1 - \ln\left(\frac{\epsilon e}{s\lceil d/e \rceil}\right)\right)\Big/(\ln 2)\right) \\
&\quad = d\exp\left(-\epsilon d\left(1 - \frac{s\lceil d/e \rceil}{\epsilon e}\left(1 + \ln\left(\frac{\epsilon e}{s\lceil d/e \rceil}\right)\right)\right)\Big/(\ln 2)\right).
\end{aligned}$$

Now the function $f(x) = x(1 + \ln(1/x))$ increases with x in the range $0 < x \leq 1$, so for any e such that $\frac{s\lceil d/e \rceil}{\epsilon e} \leq 1/2$, we have

$$\begin{aligned}
\Pr_V[\mathrm{tr}[D_V \mathcal{E}_V(\rho - \sigma)] \leq (1-\epsilon)\|\rho - \sigma\|_1/2] &\leq d\exp(-\epsilon d(1 - f(1/2))/(\ln 2)) \\
&= d\exp(-\epsilon d(1 - \ln 2)/(2\ln 2)).
\end{aligned}$$

Thus this inequality holds for any e such that $\epsilon e \geq 2s\lceil d/e \rceil$. As $\lceil d/e \rceil \leq 2d/e$ for $e \leq d$, this will be satisfied for any $e \geq 2\sqrt{sd/\epsilon}$, and in particular any $e \geq 2\sqrt{rd/\epsilon}$, implying

for any such e

$$\Pr_{\mathcal{E}_V \sim \mathcal{D}}[\|\mathcal{E}_V(\rho) - \mathcal{E}_V(\sigma)\|_1 \leq (1-\epsilon)\|\rho - \sigma\|_1] \leq d \exp(-\epsilon d(1 - \ln 2)/(2 \ln 2))$$

as required. □

Although this result is expressed in terms of the rank of the input states, a similar result would apply to states which are very close (in trace norm) to having low rank, but for simplicity we do not discuss this here.

4.2. *Lower bound*

It turns out that Lemma 6 is also strong enough to give a bound on embeddings of the trace norm, via a similar proof to that of Theorem 3. Charikar and Sahai[5] showed that there exists a set of $O(d)\,d$-dimensional vectors whose dimension cannot be significantly reduced while preserving their ℓ_1 distances. One might expect the same to be true for the trace norm, as the trace norm on diagonal matrices is just the ℓ_1-norm of the diagonal entries. However, note that this does not follow immediately from Charikar and Sahai's work, as it is conceivable that an embedding mapping diagonal to non-diagonal matrices could do better. Nevertheless, we now show that dimensionality reduction is impossible for some sets of highly mixed states.

Theorem 12. *Let $\mathcal{D}$ be a distribution over quantum channels (CPTP maps) $\mathcal{E} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^e)$ such that, for fixed quantum states $\rho \neq \sigma$ and for all unitary U,*

$$\Pr_{\mathcal{E} \sim \mathcal{D}}[\|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_1 \geq (1-\epsilon)\|U\rho U^\dagger - U\sigma U^\dagger\|_1] \geq 1 - \delta$$

for some $0 \leq \epsilon, \delta \leq 1$. Then

$$e \geq (1-\delta)(1-\epsilon)\sqrt{d}\,\frac{\|\rho - \sigma\|_1}{\|\rho - \sigma\|_2}.$$

In particular, if ρ and σ are orthogonal pure states, then $e \geq (1-\delta)(1-\epsilon)\sqrt{2d}$, and if ρ and σ are proportional to projectors onto orthogonal $d/2$-dimensional subspaces, $e \geq (1-\delta)(1-\epsilon)d$.

So we see that achieving any significant dimensionality reduction for arbitrary highly mixed states is impossible, and even for pure states the dimension can only be reduced by a square root (which was already known[11]).

Proof of Theorem 12. For a randomly chosen U, we have

$$\Pr_{\mathcal{E} \sim \mathcal{D}, U \in U(d)}[\|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_1 \geq (1-\epsilon)\|U\rho U^\dagger - U\sigma U^\dagger\|_1]\, dU \geq 1 - \delta,$$

and use Markov's inequality and the unitary invariance of the trace norm to obtain

$$\int_{\mathcal{E} \sim \mathcal{D}} \int \|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_1\, dU \geq (1-\delta)(1-\epsilon)\|\rho - \sigma\|_1.$$

Thus there must exist some $\mathcal{E}$ such that

$$\int \|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_1 \, dU \geq (1-\delta)(1-\epsilon)\|\rho - \sigma\|_1.$$

Simply estimating the 1-norm by the 2-norm and using Jensen's inequality, we get the bounds

$$\begin{aligned}(1-\delta)(1-\epsilon)\|\rho-\sigma\|_1 &\leq \sqrt{e}\int \|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2 \, dU \\ &\leq \sqrt{e}\left(\int \|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2^2 \, dU\right)^{1/2} \\ &\leq \left(\frac{e}{\sqrt{d}}\right)\|\rho-\sigma\|_2,\end{aligned}$$

where the last inequality follows from Lemma 6, assuming that $e \leq d$. Rearranging gives the theorem. □

This implies that the protocol of Theorem 9 is optimal for certain families of states, up to constant factors. Consider the family of pairs $U\rho U^\dagger$, $U\sigma U^\dagger$ for all $U \in U(d)$, where ρ and σ are proportional to projectors onto orthogonal r-dimensional subspaces of $\mathbb{C}^d$. Then

$$\frac{\|\rho-\sigma\|_1}{\|\rho-\sigma\|_2} = \sqrt{\operatorname{rank}(\rho-\sigma)} = \sqrt{2r},$$

implying that embeddings of this family with constant distortion and failure probability have a lower bound on the target dimension of $\Omega(\sqrt{rd})$, which is achieved by the embedding of Theorem 9.

5. Conclusions

We have shown that in the 2-norm, any constant-distortion embedding of a unitarily invariant set of d-dimensional states must have target dimension $\Omega(d)$, in contrast to the classical situation where an exponential reduction can be achieved. In the trace norm, the situation is somewhat better: d-dimensional states of rank r can be embedded in $O(\sqrt{rd})$ dimensions with constant distortion, but there is a lower bound of $\Omega(\sqrt{d}\frac{\|\rho-\sigma\|_1}{\|\rho-\sigma\|_2})$ dimensions on any constant distortion embedding that succeeds for the pairs of states $U\rho U^\dagger$ and $U\sigma U^\dagger$, for all unitary U.

Although the trace distance is often the most physically relevant distance measure to consider, we also argued that for certain tasks, the 2-norm distance is in fact the relevant distance measure between states. This occurs when the basis in which the states were prepared is unknown or the measurement apparatus does not depend on the states to be distinguished.

The alert reader will have noticed that, in the case where one is interested in embedding a unitarily invariant set of states, the embedding might as well start by

performing a random unitary. Furthermore, as any quantum channel can be represented as an isometry into a larger space followed by tracing out a subsystem, this makes any embedding seem somewhat similar to the embedding used in Theorem 9. But note that the latter embedding is subtly different, as it can be seen as performing a fixed isometry followed by a random unitary, rather than vice versa. Further analysis of this embedding might allow the gap between the upper and lower bounds in the trace norm to be closed.

Another open question is whether bounds could be obtained on the possible dimensionality reduction when multiple copies of the input state are available. For example, if a very large number of copies are allowed, tomography can be performed, the input state can be approximately determined, and the JL Lemma applied. Presumably, even for a lower number of copies, stronger dimensionality reduction is possible than in the single-copy case. One could also ask whether stronger dimensionality reduction can be achieved by allowing some additional classical information; for some results in this direction, see Ref. 18.

Appendix A. Lemmas Relating to 2-Norm Embeddings

We now prove the subsidiary lemmas required for the proof of Lemma 6.

Proof of Lemma 4. Assume that $\mathcal{E}$ has the Kraus (operator-sum) decomposition

$$\mathcal{E}(\rho) = \sum_i A_i \rho A_i^\dagger$$

for some $e \times d$ matrices A_i such that $\sum_i A_i^\dagger A_i = I_d$, and $\mathrm{tr}[A_i^\dagger A_j] = 0$ if $i \neq j$. (Note that such a representation does indeed exist, from the unitary freedom in the Kraus decomposition.[9]) Then write

$$\begin{aligned}\mathrm{tr}[F_e\, \mathcal{E}^{\otimes 2}(F_d)] &= \mathrm{tr} \sum_{i,j} F_e (A_i \otimes A_j) F_d (A_i^\dagger \otimes A_j^\dagger) = \sum_{i,j} \mathrm{tr}[(A_j \otimes A_i)(A_i^\dagger \otimes A_j^\dagger)] \\ &= \sum_{i,j} \mathrm{tr}[A_j A_i^\dagger] \mathrm{tr}[A_i A_j^\dagger] = \sum_i (\mathrm{tr}[A_i^\dagger A_i])^2 \\ &\leq \left(\sum_i \mathrm{tr}[A_i^\dagger A_i] \right) \max_j \mathrm{tr}[A_j^\dagger A_j] \leq de.\end{aligned}$$

The fourth equality uses the orthogonality of the A_i and cyclicity of the trace, and the final inequality uses the facts that $\sum_i A_i^\dagger A_i = I_d$ and $\mathrm{tr}[A_i^\dagger A_i] \leq \|A_i^\dagger A_i\|_\infty \, \mathrm{rank}(A_i^\dagger A_i) \leq e$. □

Proof of Lemma 5. For brevity, set $\tau := \int U^{\otimes 2} (\rho - \sigma)^{\otimes 2} (U^\dagger)^{\otimes 2} dU$. Because of the averaging ("twirling") over the unitary group, τ must be a linear combination of the identity and swap operators on the space of two d-dimensional systems.[25] To evaluate this, we write $\tau = \alpha I_{d^2} + \beta F_d$ and calculate

$$\mathrm{tr}[\tau] = 0, \ \mathrm{tr}[F_d \tau] = \mathrm{tr}[(\rho - \sigma)^2],$$

implying that

$$\alpha d^2 + \beta d = 0, \quad \alpha d + \beta d^2 = \mathrm{tr}[(\rho - \sigma)^2].$$

Solving for α and β gives the claimed result. □

Appendix B. Proof of Theorem 8

We follow the strategy of Matthews, Wehner and Winter[21] to prove Theorem 8. We will use two subsidiary results, which are formalized as separate lemmas.

Lemma B.1. *Let ρ, σ be d-dimensional quantum states. Then*

$$\int d\psi \langle\psi|(\rho - \sigma)|\psi\rangle^2 = \frac{\mathrm{tr}[(\rho - \sigma)^2]}{d(d+1)}.$$

Proof. We use the tensor product trick:

$$\begin{aligned}\int d\psi \langle\psi|(\rho - \sigma)|\psi\rangle^2 &= \int d\psi \mathrm{tr}[(\rho - \sigma)^{\otimes 2}|\psi\rangle\langle\psi|^{\otimes 2}] = \mathrm{tr}\left[(\rho - \sigma)^{\otimes 2}\frac{I_{d^2} + F_d}{d(d+1)}\right] \\ &= \frac{\mathrm{tr}[(\rho - \sigma)^2]}{d(d+1)},\end{aligned}$$

noting that $\rho - \sigma$ is traceless and that $\int d\psi(|\psi\rangle\langle\psi|^{\otimes 2})$ is proportional to the projector onto the symmetric subspace of two d-dimensional systems. □

Lemma B.2. *Let ρ, σ be d-dimensional quantum states. Then*

$$\int d\psi \langle\psi|(\rho - \sigma)|\psi\rangle^4 \leq \frac{9\mathrm{tr}[(\rho - \sigma)^2]^2}{d(d+1)(d+2)(d+3)}.$$

Proof. This is the same technique as the previous lemma, but is a little more involved. Writing

$$\int d\psi \langle\psi|(\rho - \sigma)|\psi\rangle^4 = \mathrm{tr}\left[(\rho - \sigma)^{\otimes 4}\int d\psi(|\psi\rangle\langle\psi|^{\otimes 4})\right],$$

we note that $\int d\psi(|\psi\rangle\langle\psi|^{\otimes 4})$ is proportional to the projector onto the symmetric subspace of four d-dimensional systems, which we write as

$$P_{sym} = \frac{1}{4!}\sum_{\sigma \in S_4} P_\sigma,$$

where S_4 is the symmetric group of order 4 and P_σ is the operator that permutes the four systems according to the permutation σ. Let $\mathrm{Cyc}(\sigma)$ denote the sequence of cycle lengths in σ (e.g. $\mathrm{Cyc}((12)(3)) = (2,1)$). Then, for any d-dimensional operator X, it holds that

$$\mathrm{tr}[X^{\otimes 4}P_\sigma] = \prod_{c \in \mathrm{Cyc}(\sigma)} \mathrm{tr}[X^c],$$

which can be shown diagrammatically or by explicitly writing out the P_σ matrix. In particular, $\mathrm{tr}P_\sigma = d^{|\mathrm{Cyc}(\sigma)|}$. Permutations of four elements break down into five conjugacy classes, as follows: there is one of the form (1)(2)(3)(4); six of the form (12)(3)(4); three of the form (12)(34); eight of the form (123)(4); and six of the form (1234).

Thus

$$\mathrm{tr}P_{\mathrm{sym}} = \frac{1}{4!}(d^4 + 6d^3 + 11d^2 + 6d) = \frac{d(d+1)(d+2)(d+3)}{4!},$$

implying that

$$\int d\psi(|\psi\rangle\langle\psi|^{\otimes 4}) = \frac{1}{d(d+1)(d+2)(d+3)} \sum_{\sigma \in S_4} P_\sigma.$$

We can now calculate

$$\begin{aligned}&\mathrm{tr}\left[(\rho-\sigma)^{\otimes 4}\int d\psi(|\psi\rangle\langle\psi|^{\otimes 4})\right] \\ &\quad = \frac{1}{d(d+1)(d+2)(d+3)}(3\mathrm{tr}[(\rho-\sigma)^2]^2 + 6\mathrm{tr}[(\rho-\sigma)^4]),\end{aligned}$$

where we use the fact that $\rho - \sigma$ is traceless to ignore all terms corresponding to permutations with fixed points. The upper bound claimed in the statement of the theorem follows by simply noting that $\mathrm{tr}[(\rho-\sigma)^4] \leq \mathrm{tr}[(\rho-\sigma)^2]^2$. □

We are finally ready to prove Theorem 8, which we restate for convenience.

Proof of Theorem 8. The upper bound is straightforward:

$$d\int d\psi|\langle\psi|(\rho-\sigma)|\psi\rangle| \leq d\left(\int d\psi\langle\psi|(\rho-\sigma)|\psi\rangle^2\right)^{1/2} = d\left(\frac{\mathrm{tr}[(\rho-\sigma)^2]}{d(d+1)}\right)^{1/2} \leq \|\rho-\sigma\|_2,$$

where the first inequality is Jensen's inequality, and the equality is Lemma B.1. For the lower bound, we use the fourth moment method of Berger[26] (which is just Hölder's inequality in disguise). This states that, for any real-valued random variable X,

$$\mathbb{E}[|X|] \geq \frac{\mathbb{E}[X^2]^{3/2}}{\mathbb{E}[X^4]^{1/2}}.$$

Applying this inequality gives

$$\begin{aligned}d\int d\psi|\langle\psi|(\rho-\sigma)|\psi\rangle| &\geq d\frac{(\int d\psi\langle\psi|(\rho-\sigma)|\psi\rangle^2)^{3/2}}{(\int d\psi\langle\psi|(\rho-\sigma)|\psi\rangle^4)^{1/2}} \\ &\geq d\left(\frac{\mathrm{tr}[(\rho-\sigma)^2]}{d(d+1)}\right)^{3/2}\left(\frac{d(d+1)(d+2)(d+3)}{9\mathrm{tr}[(\rho-\sigma)^2]^2}\right)^{1/2}\end{aligned}$$

by Lemmas B.1 and B.2, which simplifies to

$$d \int d\psi |\langle\psi|(\rho - \sigma)|\psi\rangle| \geq \frac{(d+2)^{1/2}(d+3)^{1/2}}{3(d+1)} \|\rho - \sigma\|_2 \geq \frac{1}{3}\|\rho - \sigma\|_2$$

as claimed. □

Appendix C. Proof of Lemma 10

We now prove Lemma 10, which we restate for convenience.

Proof of Lemma 10. The inequality clearly holds if $\mathrm{tr}[P^\perp|\psi\rangle\langle\psi|] = 0$, so assuming this is not the case and dividing both sides by $\mathrm{tr}[P^\perp|\psi\rangle\langle\psi|]$, the left-hand side is equal to

$$\frac{\mathrm{tr}[(D \otimes I)(I - P)|\psi\rangle\langle\psi|(I - P)]}{1 - \mathrm{tr}[P|\psi\rangle\langle\psi|]}.$$

The key observation which will allow us to simplify this expression is that $(D \otimes I)P = P = P(D \otimes I)$. To see this, note that the support of P is contained within the subspace onto which $D \otimes I$ projects, implying that $D \otimes I$ acts as the identity with respect to P. The left-hand side thus simplifies to

$$\begin{aligned}\frac{\mathrm{tr}[(D \otimes I)|\psi\rangle\langle\psi|] - \mathrm{tr}[P|\psi\rangle\langle\psi|]}{1 - \mathrm{tr}[P|\psi\rangle\langle\psi|]} &\leq \frac{\mathrm{tr}[(D \otimes I)|\psi\rangle\langle\psi|](1 - \mathrm{tr}[P|\psi\rangle\langle\psi|])}{1 - \mathrm{tr}[P|\psi\rangle\langle\psi|]} \\ &= \mathrm{tr}[(D \otimes I)|\psi\rangle\langle\psi|]\end{aligned}$$

as claimed. □

Acknowledgments

AWH was supported by the EC grant QESSENCE and the DARPA-MTO QuEST program through a grant from AFOSR. AM was supported by an EPSRC Postdoctoral Research Fellowship. AJS was supported by the Royal Society. This work was largely carried out while AWH was at the University of Washington, and AM and AJS were at the University of Cambridge.

References

1. W. Johnson and J. Lindenstrauss, *Contemp. Math.* **26** (1984) 189.
2. P. Indyk, in *Proc. 42nd Annual Symp. Foundations of Computer Science* (2001), pp. 10–33.
3. P. Indyk and R. Motwani, in *Proc. 30th Annual ACM Symp. Theory of Computing* (1998), pp. 604–613.
4. E. Kushilevitz, R. Ostrovsky and Y. Rabani, in *Proc. 30th Annual ACM Symp. Theory of Computing* (1998), pp. 614–623.
5. M. Charikar and A. Sahai, in *Proc. 43rd Annual Symp. Foundations of Computer Science* (2002), pp. 551–560.

6. B. Brinkman and M. Charikar, *J. ACM* **52**(5) (2005) (766).
7. H. Buhrman, R. Cleve, J. Watrous and R. de Wolf, *Phys. Rev. Lett.* **87**(16) (2001) 167902.
8. E. Kushilevitz and N. Nisan, *Communication Complexity* (Cambridge University Press, 1997).
9. M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).
10. J. Watrous, Theory of quantum information lecture notes, 2008. http://www.cs.uwaterloo.ca/∼watrous/quant-info/.
11. A. Winter, quant-ph/0401060.
12. P. Hayden and A. Winter, arXiv:1003.4994.
13. S. Aaronson, *Theory Comput.* **1** (2004) 1.
14. R. Cleve, P. Høyer, B. Toner and J. Watrous, in *Proc. 19th Annual IEEE Conf. Computational Complexity* (2004), pp. 236–249.
15. D. Gavinsky, J. Kempe and R. de Wolf, in *Proc. 21st Annual IEEE Conf. Computational Complexity* (2006), pp. 288–298.
16. G. Aubrun, S. Szarek and E. Werner, *J. Math. Phys.* **51** (2010) 022102.
17. G. Aubrun, S. Szarek and E. Werner, *Comm. Math. Phys.* **305** (2011) 85.
18. O. Fawzi, P. Hayden and P. Sen, in *Proc. 43rd Annual ACM Symp. Theory of Computing* (2011), pp. 773–782.
19. S. Bartlett, T. Rudolph and R. Spekkens, *Phys. Rev. Lett.* **91**(2) (2003) 027901.
20. A. Ambainis and J. Emerson, in *Proc. 22nd Annual IEEE Conf. Computational Complexity* (2007), pp. 129–140.
21. W. Matthews, S. Wehner and A. Winter, *Comm. Math. Phys.* **291**(3) (2009) 813.
22. P. Sen, in *Proc. 21st Annual IEEE Conf. Computational Complexity* (2006), p. 287.
23. C. H. Bennett, P. Hayden, D. Leung, P. Shor and A. Winter, *IEEE Trans. Inform. Theory* **51**(1) (2005) 56.
24. R. Bhatia, *Matrix Analysis* (Springer-Verlag, 1997).
25. R. Goodman and N. R. Wallach, *Symmetry, Representations and Invariants* (Springer, 2009).
26. B. Berger, *SIAM J. Comput.* **24**(6) (1997) 1188.