On the distinguishability of random quantum states

Ashley Montanaro[*]

Department of Computer Science, University of Bristol,
Woodland Road, Bristol, BS8 1UB, UK.

July 11, 2006

Abstract

We develop two analytic lower bounds on the probability of success p of identifying a state picked from a known ensemble of pure states: a bound based on the pairwise inner products of the states, and a bound based on the eigenvalues of their Gram matrix. We use the latter to lower bound the asymptotic distinguishability of ensembles of n random quantum states in d dimensions, where n/d approaches a constant. In particular, for almost all ensembles of n states in n dimensions, $p > 0.72$. An application to distinguishing Boolean functions (the "oracle identification problem") in quantum computation is given.

1   Introduction

A fundamental property of quantum mechanics is that non-orthogonal pure quantum states may not be distinguished perfectly. This leads to the following *quantum detection problem*: given an unknown quantum state $|\psi_?\rangle$, picked from a known set $\mathcal{E}$ with known a priori probabilities, find the "optimal" measurement M^{opt} to determine $|\psi_?\rangle$. Several different criteria for optimality may been considered [12, 5, 6]; here we only concern ourselves with optimising the probability of success P^{opt}, and in particular the related *state distinguishability problem* of finding P^{opt} without necessarily finding M^{opt}. Efficient optimisation techniques can be used to estimate P^{opt} numerically [7]; however, the problem of finding an analytic expression for P^{opt} seems intractable. We are therefore led to attempting to produce bounds on P^{opt}.

This note derives two lower bounds on P^{opt}; one based on the pairwise distinguishability of the states in $\mathcal{E}$, and one based on the eigenvalues of their Gram matrix. We use the latter, and a powerful result from random matrix theory (the Marčenko-Pastur law [18]), to bound the probability of distinguishing a set of random quantum states, for a quite general notion of randomness. This has an application to quantum computation in the so-called oracle identification problem introduced by Ambainis et al [1], where we are given an n-bit

---

[*]montanar@cs.bris.ac.uk

Boolean function f picked from a known set of N functions, and must identify f with the minimum number of queries to f. We show that, for all but an exponentially small fraction of sets with $N = 2^n$, a quantum computer can perform this task successfully in a constant number of queries (with arbitrarily high probability), whereas classical computation requires n queries for all such sets.

As showing that a set of quantum states are quite distinguishable forms an essential part of proofs in many areas of quantum information theory, we hope that these results will find application elsewhere.

The organisation of the paper is as follows. Section 2 introduces notation and our main tool, the so-called "pretty good measurement", before moving on to give the lower bounds on P^{opt}. An extension of the lower bounds to mixed states is considered. Section 3 applies the bounds to a specific family of ensembles (those where all the states have constant inner product). Section 4 describes the random matrix theory we will be using, and applies it to the distinguishability of random quantum states. Section 5 gives the application to the oracle identification problem, and the paper closes with some discussion in section 6.

2   Bounds on the distinguishability of quantum states

We consider an ensemble $\mathcal{E}$ containing n d-dimensional pure states $|\psi_i\rangle$ with their a priori probabilities p_i. We will use $\{|\psi_i'\rangle\}$ to denote the set containing the same states, renormalised to reflect their probabilities (i.e. $|\psi_i'\rangle = \sqrt{p_i}|\psi_i\rangle$). Given an unknown state $|\psi_?\rangle$, picked in accordance with these probabilities, the quantity we are interested in is the average probability of success for a given generalised measurement to distinguish which state we were given. For a measurement M (given by a set of positive operators $\{M_i\}$ summing to the identity), let this probability be denoted by $P^M(\mathcal{E})$. Then we have

$$P^M(\mathcal{E}) = \sum_i \langle\psi_i'|M_i|\psi_i'\rangle = \sum_i p_i\langle\psi_i|M_i|\psi_i\rangle \tag{1}$$

$M^{opt}(\mathcal{E})$ will denote the measurement with the optimal probability of success, and in an abuse of notation $P^{opt}(\mathcal{E})$ will denote this optimal probability. We call this the optimal probability of distinguishing the states in $\mathcal{E}$.

We use three matrix norms: the 1-norm $\|A\|_1 = \sum_{i,j} |A_{ij}|$, the Euclidean (Frobenius) norm $\|A\|_2 = \sqrt{\sum_{i,j} |A_{ij}|^2}$, and the trace norm $\|A\|_{tr} = \mathrm{tr}\sqrt{A^\dagger A} = \sum_i \sigma_i(A)$, where $\sigma_i(A)$ denotes the i'th singular value of A. We will often use the $d \times n$ state matrix $S = S(\mathcal{E}) = (|\psi_1'\rangle, ..., |\psi_n'\rangle)$ whose i'th column is the state $|\psi_i'\rangle$. Then $G = S^\dagger S$ gives the $n \times n$ Gram matrix [14] encoding all the inner products between the renormalised states in $\mathcal{E}$. If $n < d$, G will have $d - n$ zero eigenvalues. Note that every rectangular matrix M with $\|M\|_2 = 1$ is a state matrix. ρ will represent the density matrix of the ensemble:

$$\rho = \sum_{i=1}^n |\psi_i'\rangle\langle\psi_i'| \tag{2}$$

It is well-known [15] that G and ρ have the same non-zero eigenvalues.

2.1 Use of the "pretty good measurement"

We will use a specific measurement to provide bounds on $P^{opt}(\mathcal{E})$, which is "canonical" in the sense that it performs reasonably well for any ensemble $\mathcal{E}$. This is the so-called *pretty good measurement* (PGM), which was independently identified by several authors (e.g. [9], [10]) and has a number of useful properties. It is usually defined as a set of projectors $\{|\nu_i\rangle\langle\nu_i|\}$ onto "measurement vectors" $|\nu_i\rangle$, where $|\nu_i\rangle = \rho^{-1/2}|\psi_i'\rangle$ (the inverse only being taken on the support of ρ). However, it may also be defined implicitly, which brings out its "canonical" nature.

To this end, consider an arbitrary measurement M for $\mathcal{E}$ that consists of a set of n rank 1 projectors onto unnormalised measurement vectors $|\mu_i\rangle$, where each measurement vector corresponds to a state $|\psi_i'\rangle$ in the ensemble. (In fact, it turns out that the optimal measurement for an ensemble of pure states always falls into this category [7].) The probability of getting measurement outcome i and receiving state j is then $|\langle\mu_i|\psi_j'\rangle|^2$, and the overall probability of success of this measurement is $\sum_{i=1}^{n} |\langle\mu_i|\psi_i'\rangle|^2$. We may thus encode all the inner products (and hence the probabilities) in a matrix P, where $P_{ij} = \langle\mu_i|\psi_j'\rangle$; and rather than looking for an optimal measurement M, we can rephrase our task as looking for an optimal matrix P that corresponds to a valid measurement.

We have the following requirement on P, from the fact that M must be a valid POVM.

$$(P^\dagger P)_{ij} = \sum_{k=1}^{n}\langle\psi_i'|\mu_k\rangle\langle\mu_k|\psi_j'\rangle = \langle\psi_i'|\left(\sum_{k=1}^{n}|\mu_k\rangle\langle\mu_k|\right)|\psi_j'\rangle = G_{ij} = (S^\dagger S)_{ij} \qquad (3)$$

A natural way to produce a matrix P that satisfies this condition from any given S is to take $P = \sqrt{G}$, the positive semidefinite square root of G. The PGM turns out to be a measurement corresponding to this matrix P, for, if $P_{ij} = \langle\nu_i|\psi_j'\rangle$, then

$$(P^2)_{ij} = \sum_{k=1}^{n}\langle\psi_i'|\rho^{-1/2}|\psi_k'\rangle\langle\psi_k'|\rho^{-1/2}|\psi_j'\rangle = \langle\psi_i'|\left(\rho^{-1/2}\sum_{k=1}^{n}|\psi_k'\rangle\langle\psi_k'|\rho^{-1/2}\right)|\psi_j'\rangle = G_{ij} \quad (4)$$

The probability of success for the PGM is thus given by $P^{pgm}(\mathcal{E}) = \sum_{i=1}^{n}(\sqrt{G})_{ii}^2$. Barnum and Knill have proved [3] that the PGM has the further property that it is almost optimal in the following sense.

Theorem 2.1. (Barnum, Knill) [3] $P^{pgm}(\mathcal{E}) \geq P^{opt}(\mathcal{E})^2$.

So there is the overall relationship $P^{opt}(\mathcal{E})^2 \leq P^{pgm}(\mathcal{E}) \leq P^{opt}(\mathcal{E})$. For completeness, we include (in Appendix A) a simplified proof of Barnum and Knill's result in the case of pure states.

2.2 Bounds from the pairwise inner products

A set of states that are pairwise almost orthogonal are pairwise almost distinguishable. It thus seems intuitively clear that, given such a set, the probability of success in distinguishing

one state from *all* the others must also be high. However, this intuition is wrong. This was noted by Jozsa and Schlienz [15], who showed that the inner products of an ensemble of states may all be reduced, while simultaneously reducing the von Neumann entropy of the ensemble (which gives a measure of overall distinguishability). This effect also manifests itself in quantum fingerprinting [4]. Here, d-dimensional states are "compressed" to $\log d$-dimensional "fingerprint" states that can be distinguished pairwise. However, given such a fingerprint the corresponding original state may not be identified, as this would violate Holevo's theorem [13].

Nevertheless, for certain ensembles the pairwise inner products can give a good lower bound on the overall distinguishability, as noted by several authors [9, 3]. In this section, we derive such a bound. Our approach is based on that of Hausladen et al. [9], who found a parabola forming a lower bound on the square root function, which is useful because of the following lemma.

Lemma 2.2. *If the function $\sqrt{x}$ is bounded below by $f(x) = ax + bx^2$ for $x \geq 0$, then $(\sqrt{G})_{ii} \geq aG_{ii} + b\sum_{j=1}^{n}|G_{ij}|^2$.*

Proof. G is a positive semidefinite matrix and thus may be diagonalised: $G = UDU^{\dagger}$, where $D = diag(\{\lambda_i\})$ and $U = (u_{ij})$ is unitary. Working out the matrix algebra shows that $(\sqrt{G})_{ii} = \sum_{k=1}^{n}\sqrt{\lambda_k}|u_{ik}|^2$, so $(\sqrt{G})_{ii} \geq \sum_{k=1}^{n}f(\lambda_k)|u_{ik}|^2 = f(G)_{ii}$. But $f(G)_{ii} = (aG + bG^2)_{ii} = aG_{ii} + b\sum_{j=1}^{n}G_{ij}G_{ji} = aG_{ii} + b\sum_{j=1}^{n}|G_{ij}|^2$. □

Our goal will be to find a and b to parametrise f such that $aG_{ii} + b\sum_{j=1}^{n}|G_{ij}|^2$ is maximised. It is clear that, for this to be maximised, $f(r)$ must equal $\sqrt{r}$ for some r (or we could just increase a or b). So we will pick a and b such that $f(r) = \sqrt{r}$ and $f'(r) = \frac{1}{2\sqrt{r}}$ (i.e. the curves are tangent at this point). This leads to the simultaneous equations

$$ar + br^2 = \sqrt{r}, \; a + 2br = \frac{1}{2\sqrt{r}} \tag{5}$$

Solving for a and b gives the optimal values

$$a = \frac{3}{2\sqrt{r}}, \; b = -\frac{1}{2r^{3/2}} \tag{6}$$

To see that $f(x)$ actually is a lower bound for $\sqrt{x}$ for any positive value of r (with these values for a and b), note that the only solutions to the related equation $f(x)^2 = x$ are $x = 0$, $x = r$, or $x = 4r$. As $f(4r)$ is negative, we have that $f(x) = \sqrt{x}$ if and only if $x = 0$ or $x = r$. So the only remaining possibility is that $f(x) > \sqrt{x}$ for all $0 < x < r$. Plugging in a suitable value of x (e.g. $r/2$) shows that this is not the case. The expression $aG_{ii} + b\sum_{j=1}^{n}|G_{ij}|^2$ may now be expressed solely in terms of r. Optimising this for r gives that the maximum is found at the point

$$r = \frac{\sum_{j=1}^{n}|G_{ij}|^2}{G_{ii}} \tag{7}$$

Returning to the original inequality, we have

$$(\sqrt{G})_{ii} \geq \frac{3}{2\sqrt{r}}G_{ii} - \frac{1}{2r^{3/2}}\sum_{j=1}^{n}|G_{ij}|^2 \Rightarrow (\sqrt{G})_{ii}^2 \geq \frac{G_{ii}^3}{\sum_{j=1}^{n}|G_{ij}|^2} \tag{8}$$

We thus have the following bound on the probability of distinguishing the states in $\mathcal{E}$.

$$P^{pgm}(\mathcal{E}) \geq \sum_{i=1}^{n}\frac{\langle\psi_i'|\psi_i'\rangle^3}{\sum_{j=1}^{n}|\langle\psi_i'|\psi_j'\rangle|^2} = \sum_{i=1}^{n}\frac{p_i^2}{\sum_{j=1}^{n}p_j|\langle\psi_i|\psi_j\rangle|^2} \tag{9}$$

If all the states have equal a priori probabilities, the bound simplifies further to

$$P^{pgm}(\mathcal{E}) \geq \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sum_{j=1}^{n}|\langle\psi_i|\psi_j\rangle|^2} \tag{10}$$

Unlike previous bounds obtained by other authors for the probability of success of the PGM [9, 3], the bound (9) is always positive and greater than or equal to $\sum_{i=1}^{n}p_i^2$, thus showing that the PGM always does at least as well as the "non-measurement" of guessing which state was received in accordance with their a priori probabilities.

2.3 Bounds from eigenvalues

The eigenvalues of a Hermitian matrix are closely related to its diagonal elements; indeed, the former majorises the latter [14]. With this in mind, we look for a bound on the unknown diagonal elements of $\sqrt{G}$ in terms of the known eigenvalues $\{\lambda_i\}$ of G.

Lemma 2.3. $P^{pgm}(\mathcal{E}) \geq \frac{1}{n}\left(\sum_{i=1}^{n}\sqrt{\lambda_i}\right)^2 = \frac{1}{n}\|S\|_{tr}^2$.

Proof. By the fact that the trace of a matrix is the sum of its eigenvalues, we have

$$\sum_{i=1}^{n}(\sqrt{G})_{ii} = \sum_{i=1}^{n}\sqrt{\lambda_i} \tag{11}$$

$$\Rightarrow \left(\sum_{i=1}^{n}(\sqrt{G})_{ii}\right)^2 = \left(\sum_{i=1}^{n}\sqrt{\lambda_i}\right)^2 \tag{12}$$

$$\Rightarrow n\sum_{i=1}^{n}(\sqrt{G})_{ii}^2 \geq \left(\sum_{i=1}^{n}\sqrt{\lambda_i}\right)^2 \tag{13}$$

$$\Rightarrow P^{pgm}(\mathcal{E}) \geq \frac{1}{n}\left(\sum_{i=1}^{n}\sqrt{\lambda_i}\right)^2 \tag{14}$$

where in (13) we used a Cauchy-Schwarz inequality, showing that equality can only be attained in step (13) when all the $(\sqrt{G})_{ii}$ are equal. $\qquad\square$

Interestingly, this bound is the same as the fidelity of G with the maximally mixed state I/n, where the fidelity $F(\rho, \sigma)$ is defined as $\left(\text{tr} \sqrt{\rho^{1/2} \sigma \rho^{1/2}} \right)^2$ [19].

It is worth noting that no upper bound on the success probability in terms of the eigenvalues alone can be found, for the following reason. Any set of eigenvalues $\{\lambda_i\}$ summing to 1 can give rise to a Gram matrix G where $G_{ii} = \lambda_i$, and $G_{ij} = 0$ (for $i \neq j$). Such matrices correspond to an ensemble $\mathcal{E}$ of perfectly distinguishable states where $P^{pgm}(\mathcal{E}) = 1$. As future work, it would be interesting to determine whether an upper bound (or an improved lower bound) could be produced by considering the diagonal entries of G as well as its eigenvalues.

2.4 Distinguishing mixed states

It is natural to ask to what extent these lower bounds hold for the generalised problem of distinguishing an ensemble $\mathcal{E}$ consisting of mixed states $\{\rho_i\}$. The following lemma allows the problem to be related to that of distinguishing pure states.

Lemma 2.4. *Let $\mathcal{E}$ be an ensemble of n d-dimensional mixed states $\{\rho_i\}$ with a priori probabilities $\{p_i\}$, and having spectral decompositions $\rho_i = \sum_{k=1}^{d} \lambda_{ik} |v_{ik}\rangle \langle v_{ik}|$. Let $\mathcal{F}$ be an ensemble of the nd pure states given by the eigenvectors $\{|v_{ik}\rangle\}$ with a priori probabilities $\{p_i \lambda_{ik}\}$. Then $P^{pgm}(\mathcal{E}) \geq P^{pgm}(\mathcal{F})$.*

Proof. For mixed states, the PGM is defined by the following measurement operators $\{M_i\}$:

$$M_i = \rho^{-1/2} \rho_i' \rho^{-1/2}, \text{ where } \rho_i' = p_i \rho_i \text{ and } \rho = \sum_{i=1}^{n} \rho_i' \tag{15}$$

So the probability of success can be bounded as follows, where we use the renormalised eigenvectors $|v_{ik}'\rangle = \sqrt{p_i} \sqrt{\lambda_{ik}} |v_{ik}\rangle$.

$$P^{pgm}(\mathcal{E}) = \sum_{i=1}^{n} \text{tr} \left(\rho^{-1/2} \rho_i' \rho^{-1/2} \rho_i' \right) \tag{16}$$

$$= \sum_{i=1}^{n} \text{tr} \left(\rho^{-1/2} \left(\sum_{k=1}^{d} |v_{ik}'\rangle \langle v_{ik}'| \right) \rho^{-1/2} \left(\sum_{l=1}^{d} |v_{il}'\rangle \langle v_{il}'| \right) \right) \tag{17}$$

$$= \sum_{i=1}^{n} \sum_{k,l=1}^{d} \text{tr} \left(\rho^{-1/2} |v_{ik}'\rangle \langle v_{ik}'| \rho^{-1/2} |v_{il}'\rangle \langle v_{il}'| \right) \tag{18}$$

$$= \sum_{i=1}^{n} \sum_{k,l=1}^{d} |\langle v_{ik}'| \rho^{-1/2} |v_{il}'\rangle|^2 \geq \sum_{i=1}^{n} \sum_{k=1}^{d} |\langle v_{ik}'| \rho^{-1/2} |v_{ik}'\rangle|^2 = P^{pgm}(\mathcal{F}) \tag{19}$$

$\square$

Therefore, if the eigenvalues and eigenvectors of the states $\{\rho_i\}$ are known, the lower bounds given previously may be applied. If not, a weaker lower bound based only on

6

the pairwise fidelities of the states may be given (where, as before, we set $F(\rho, \sigma) = \left(\text{tr} \sqrt{\rho^{1/2} \sigma \rho^{1/2}} \right)^2$).

Theorem 2.5. *Let $\mathcal{E}$ be an ensemble of n d-dimensional mixed states $\{\rho_i\}$ with a priori probabilities $\{p_i\}$. Then*

$$P^{pgm}(\mathcal{E}) \geq \sum_{i=1}^{n} \frac{p_i^2 \, \text{tr}(\rho_i^2)}{\sum_{j=1}^{n} p_j F(\rho_i, \rho_j)} \tag{20}$$

Proof. From the bound (9) and Lemma 2.4, we have

$$P^{pgm}(\mathcal{E}) \; \geq \; \sum_{i=1}^{n} \sum_{k=1}^{d} \frac{p_i^2 \lambda_{ik}^2}{\sum_{j=1}^{n} \sum_{l=1}^{d} p_j \lambda_{jl} |\langle v_{ik} | v_{jl} \rangle|^2} \tag{21}$$

$$= \; \sum_{i=1}^{n} \sum_{k=1}^{d} \frac{p_i^2 \lambda_{ik}^2}{\sum_{j=1}^{n} p_j \langle v_{ik} | \left(\sum_{l=1}^{d} \lambda_{jl} |v_{jl}\rangle \langle v_{jl}| \right) |v_{ik}\rangle} \tag{22}$$

$$= \; \sum_{i=1}^{n} \sum_{k=1}^{d} \frac{p_i^2 \lambda_{ik}^2}{\sum_{j=1}^{n} p_j \langle v_{ik} | \rho_j | v_{ik} \rangle} \tag{23}$$

$$\geq \; \sum_{i=1}^{n} \sum_{k=1}^{d} \frac{p_i^2 \lambda_{ik}^2}{\sum_{j=1}^{n} p_j F(\rho_i, \rho_j)} = \sum_{i=1}^{n} \frac{p_i^2 \, \text{tr}(\rho_i^2)}{\sum_{j=1}^{n} p_j F(\rho_i, \rho_j)} \tag{24}$$

$\square$

This bound gets progressively worse as the states in $\mathcal{E}$ get more mixed. One might expect the following lower bound to hold for mixed states, as it is the obvious extension of the bound (9) for pure states, but interestingly it does not.

$$P^{pgm}(\mathcal{E}) \not\geq \sum_{i=1}^{n} \frac{p_i^2}{\sum_{j=1}^{n} p_j F(\rho_i, \rho_j)} \tag{25}$$

A simple counterexample is given by the equiprobable ensemble consisting of the following two three-dimensional states.

$$\rho_1 = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; \rho_2 = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} \tag{26}$$

3 The distinguishability of states with constant inner product

An illustrative case to apply these bounds to is that of equiprobable states where the pairwise inner products are all equal, so the states are all equally distinguishable from each other. Consider an ensemble $\mathcal{E}$ with Gram matrix G, where $G_{ii} = 1/n$ and $G_{ij} = p/n$ for

$i \neq j$ (and p is a positive real constant). In this case, the inner product bound of section 2.2 gives the bound

$$P^{pgm}(\mathcal{E}) \geq \frac{1}{1 + p^2(n - 1)} = O(1/n) \tag{27}$$

The eigenvalue bound, however, gives much better results. The symmetry of G shows immediately that it has an eigenvector $(1, 1, ..., 1)$; the corresponding eigenvalue is $\lambda_1 = p + (1 - p)/n$. The set of eigenvectors may be completed by taking any $n - 1$ vectors orthogonal to $(1, 1, ..., 1)$, which will be eigenvectors with eigenvalues $\lambda_{2...n} = (1 - p)/n$. We therefore have

$$P^{pgm}(\mathcal{E}) \;\geq\; \frac{1}{n}\left(\sqrt{p + \frac{1 - p}{n}} + (n - 1)\sqrt{\frac{1 - p}{n}}\right)^2 \tag{28}$$

$$\geq\; \frac{1}{n}\left((n - 1)^2 \frac{(1 - p)}{n}\right) \geq (1 - p) - \frac{2(1 - p)}{n} \tag{29}$$

so the probability of distinguishing these states approaches a constant as $n \to \infty$. In fact, one can show that inequality (28) is actually an equality giving the precise probability of success $P^{pgm}(\mathcal{E})$ (this follows from showing that the diagonal entries of $\sqrt{G}$ are all equal).

Such an ensemble therefore provides a kind of converse to the ensemble of states used in quantum fingerprinting [4]: in this case, no matter how many states there are in the ensemble, their joint distinguishability is of the same order as their pairwise distinguishability. We will see below that this behaviour is not typical; however, it is perhaps not surprising, because $\mathcal{E}$ can only be realised in n dimensions. To see this, note that G is non-singular, so the states in $\mathcal{E}$ must be linearly independent.

4   The distinguishability of random quantum states

We will use Lemma 2.3 and some results from the theory of random matrices to put a lower bound on the probability of distinguishing random quantum states. The expected value of this lower bound will be obtained for a quite general notion of "randomness", but in order to get measure concentration results we will specialise to states distributed uniformly at random (according to the Haar measure). The results hold in the asymptotic regime where the number of states n and the dimension d approach a constant ratio.

4.1   A little random matrix theory

In this section, we will calculate the expected value of the trace norm of a random matrix. The distribution of the trace norm (i.e. the sum of singular values) of a matrix M is clearly related to that of the eigenvalues of the matrix $MM^{\dagger}$, which is known to statisticians as a (complex) *Wishart matrix*. The distribution of the eigenvalues of a Wishart matrix is given by the Marčenko-Pastur law [18], which is stated in the form we need in [2].

Theorem 4.1. (Marčenko/Pastur law) [18]
Let R_r be a family of $d \times n$ matrices with $n \geq d$ and $d/n \to r \in (0,1]$ as $n, d \to \infty$, where the entries of R_r are i.i.d. complex random variables with mean 0 and variance 1. Then, as $n, d \to \infty$, the eigenvalues of the rescaled matrix $\frac{1}{n}R_r R_r^\dagger$ tend to a limiting distribution with density

$$p_r(x) = \frac{\sqrt{(x - A^2)(B^2 - x)}}{2\pi r x} \tag{30}$$

for $A^2 \leq x \leq B^2$ (where $A = 1 - \sqrt{r}$, $B = 1 + \sqrt{r}$), and density 0 elsewhere.

We will translate this to a similar statement about the singular values of R_r. The following lemma is straightforward.

Lemma 4.2. *Let R_r be a family of $d \times n$ matrices with $k/m \to r \in (0,1]$ as $n, d \to \infty$, where $k = \min(n, d)$ and $m = \max(n, d)$, and the entries of R_r are i.i.d. complex random variables with mean 0 and variance 1. Then, as $n, d \to \infty$, the singular values of $R_r/\sqrt{m}$ tend to a limiting distribution with density*

$$p_r(y) = \frac{\sqrt{(y^2 - A^2)(B^2 - y^2)}}{\pi r y} \tag{31}$$

for $A \leq y \leq B$ (where $A = 1 - \sqrt{r}$, $B = 1 + \sqrt{r}$), and density 0 elsewhere.

Proof. The lemma follows from Theorem 4.1 for $n \geq d$ by substituting $y = \sqrt{x}$. For $n \leq d$, note that the singular values of R are the same as those of R^T, so the roles of n and d need merely be interchanged. ☐

Lemma 4.3. *Let R_r be a family of $d \times n$ matrices with $k/m \to r \in (0,1]$ as $n, d \to \infty$, where $k = \min(n, d)$ and $m = \max(n, d)$, and the entries of R_r are i.i.d. complex random variables with mean 0 and variance 1. Then, as $n, d \to \infty$, the expected trace norm of R_r is*

$$\mathbb{E}(\|R_r\|_{tr}) = \frac{m^{3/2}}{\pi} \int_A^B \sqrt{(y^2 - A^2)(B^2 - y^2)}\, dy \tag{32}$$

where $A = 1 - \sqrt{r}$, $B = 1 + \sqrt{r}$.

Proof. With probability 1, R_r will have k non-zero singular values. Let $\sigma_i(R_r)$ denote the value of the i'th (unsorted) singular value of R_r, for arbitrary i between 1 and k. We have

$$\mathbb{E}(\|R_r\|_{tr}) = (k\sqrt{m})\,\mathbb{E}(\sigma_i(R_r/\sqrt{m})) = k\sqrt{m} \int_A^B y\, p_r(y)\, dy \tag{33}$$

and using Lemma 4.2 gives the desired result. ☐

This turns out to be an elliptic integral which cannot be expressed in terms of elementary functions [8]. However, it is possible to produce a good lower bound, which is tight in the case $r = 1$:

9

Lemma 4.4.

$$\mathbb{E}(\|R_r\|_{tr}) \geq k\sqrt{m}\sqrt{1 - r\left(1 - \frac{64}{9\pi^2}\right)} \tag{34}$$

with equality when $r = 1$.

Proof. See Appendix B. □

4.2 Random quantum states

Knowing the expected value of the trace norm immediately allows us to say something about the expected distinguishability of an ensemble of random quantum states, for a quite general notion of randomness.

Theorem 4.5. *Let $\mathcal{E}$ be an ensemble of n equiprobable d-dimensional quantum states $\{|\psi_i\rangle\}$ with $n/d \to r \in (0, \infty)$ as $n, d \to \infty$, and let the components of $|\psi_i\rangle$ in some basis be i.i.d. complex random variables with mean 0 and variance $1/d$. Then*

$$\mathbb{E}(P^{pgm}(\mathcal{E})) \geq \begin{cases} \frac{1}{r}\left(1 - \frac{1}{r}\left(1 - \frac{64}{9\pi^2}\right)\right) & \text{if } n \geq d \\ 1 - r\left(1 - \frac{64}{9\pi^2}\right) & \text{otherwise} \end{cases} \tag{35}$$

and in particular $\mathbb{E}(P^{pgm}(\mathcal{E})) > 0.720$ when $n \leq d$.

Proof. The matrix $R = \sqrt{nd}\,S(\mathcal{E})$ fulfils the criteria for the Marčenko-Pastur law (4.1), as its entries are complex random variables with mean 0 and variance 1. We therefore have

$$\mathbb{E}(P^{pgm}(\mathcal{E})) \geq \mathbb{E}\left(\frac{1}{n}\|S(\mathcal{E})\|_{tr}^2\right) \geq \frac{1}{n}\mathbb{E}(\|S(\mathcal{E})\|_{tr})^2 = \frac{1}{n^2 d}\mathbb{E}(\|R\|_{tr})^2 \tag{36}$$

and plugging in the lower bound on the expected trace norm of R from Lemma 4.4 gives the required result. □

We can immediately apply this result to the distinguishability of random quantum states uniformly distributed on the complex unit sphere in d dimensions. A uniformly random quantum state may be produced by creating a vector v, each of whose components are complex Gaussians (say $v_i \sim \tilde{N}(0, 1/d)$), and normalising the result. By the law of large numbers, as $d \to \infty$, the norm of the resulting vector will approach 1, so the normalisation step becomes unnecessary. (This can be formalised and is known as Poincaré's lemma [16].) Therefore, an ensemble of uniformly random states meets the criteria for Theorem 4.5, so we can lower bound its expected distinguishability.

In fact, in this case, we may exploit the concentration of measure effects characteristic of high-dimensional spaces to show that for high d *almost all* ensembles of $n \leq d$ states are quite distinguishable. As with the recent paper [20], our tool will be Levy's Lemma [17]:

Lemma 4.6. (Levy's Lemma) [17]
Given a function $f : \mathbb{S}^d \mapsto \mathbb{R}$ defined on the d-dimensional real hypersphere $\mathbb{S}^d$, and a point
p on the hypersphere chosen uniformly at random,

$$\Pr[|f(p) - \mathbb{E}(f)| \geq \epsilon] \leq 2 \exp\left(\frac{-2C(d+1)\epsilon^2}{\eta^2}\right) \tag{37}$$

where η is the Lipschitz constant of f, $\eta = \sup_{x,y} |f(x) - f(y)|/\|x - y\|_2$, and C is a positive
constant that may be taken to be $1/(18\pi^3)$.

This is useful for us because a state matrix is precisely such a point on a hypersphere:

Lemma 4.7. *Let $\mathcal{E}$ be an ensemble of n equiprobable d-dimensional quantum states picked*
uniformly at random. Then, for large d, the state matrix $S(\mathcal{E})$ defines a point picked uni-
formly at random on the sphere in nd complex dimensions (equivalently, the real sphere
$\mathbb{S}^{2nd-1}$ in 2nd dimensions).

Proof. As noted previously, by the properties of quantum states distributed uniformly at
random, for high d the elements of $S(\mathcal{E})$ will be complex Gaussians with mean 0 and variance
$1/nd$. The lemma follows. $\qquad\square$

Lemma 4.8. *Let S be an $n \times d$ matrix with $\|S\|_2 = 1$, and define $f(S) = \frac{1}{n}\|S\|_{tr}^2$. Then*
the Lipschitz constant η of f satisfies $\eta \leq 2$.

Proof. See Appendix C. $\qquad\square$

Plugging this function f and this value of η into Levy's Lemma gives the following
theorem.

Theorem 4.9. *Let $\mathcal{E}$ be an ensemble of n d-dimensional quantum states picked uniformly*
at random. Set $p = \mathbb{E}(P^{pgm}(\mathcal{E})) = \frac{1}{r}\left(1 - \frac{1}{r}\left(1 - \frac{64}{9\pi^2}\right)\right)$ if $n \geq d$, and $p = 1 - r\left(1 - \frac{64}{9\pi^2}\right)$
otherwise. Then

$$\Pr[P^{pgm}(\mathcal{E}) \leq p - \epsilon] \leq 2 \exp\left(\frac{-C(2nd+1)\epsilon^2}{2}\right) \tag{38}$$

where $C = 1/(18\pi^3)$.

Figure 1 shows numerical evidence that ensembles $\mathcal{E}$ of quantum states picked uniformly
at random appear to have a value of $P^{pgm}(\mathcal{E})$ close to this lower bound, even when the states
are (relatively) low-dimensional.

5 Application to oracle identification

The *oracle identification problem* may be defined as follows [1]. Given an unknown n-bit
Boolean function $f : \{0,1\}^n \mapsto \{0,1\}$ (the *oracle*), picked uniformly at random from a
known set F of functions, identify f with the minimum number of uses of f. Set $N = |F|$
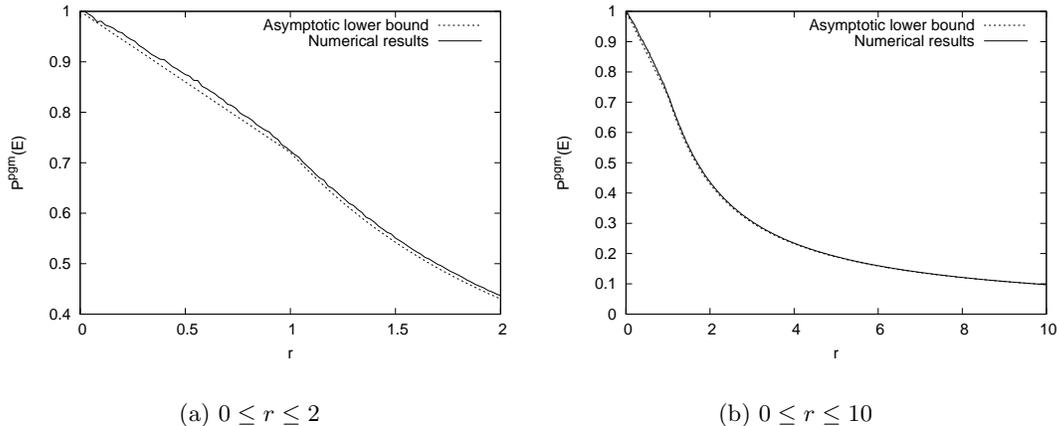
Figure 1: Asymptotic bound on $P^{pgm}(\mathcal{E})$ vs. numerical results (averaged over 10 runs) for ensembles of $n = 50r$ 50-dimensional uniformly random states.

and $D = 2^n$. Clearly, classical computation cannot identify f with fewer than $\log_2 N$ queries in the worst case (as each query may reduce the search space by at most half). However, quantum computation can sometimes do better. On a quantum computer, we can encode the oracle as an n qubit unitary operator U_f, defined by the action $U_f|x\rangle \mapsto (-1)^{f(x)}|x\rangle$. Now if the uniform superposition $\frac{1}{2^{n-1}} \sum_{x=0}^{2^n-1} |x\rangle$ is input to the oracle, the following *oracle state* will be be produced:

$$|\psi_f\rangle = \frac{1}{2^{n-1}} \sum_{x=0}^{2^n-1} (-1)^{f(x)}|x\rangle \tag{39}$$

In some cases, a single quantum query to U_f may be enough to identify f with certainty. This will be the case if $\langle\psi_f|\psi_g\rangle = 0$ for all $f \neq g$ (although this is not a necessary condition). The satisfaction of this orthogonality condition may be expected to be a rare event, and is certainly impossible when $N > D$. However, if we are content with a small probability of error, the situation is better: we will show here that, in particular, *almost all* sets of $N = D$ oracles may be distinguished almost certainly in a constant number of quantum queries.

The oracle identification problem was introduced and studied by Ambainis et al [1], who (among other results) developed a hybrid quantum-classical algorithm for the random oracle case with which we concern ourselves here. However, the upper bound they obtained in the case where $N = D$ is only $O(\log_2 N)$ queries, which is no better than classical computation.

Lemma 5.1. *Let $\mathcal{E}$ be an ensemble of N D-dimensional oracle states corresponding to Boolean functions picked uniformly at random (call these* random oracle states*). Then the rescaled state matrix $\sqrt{ND}\,S(\mathcal{E})$ defines a point picked uniformly at random on the ND-dimensional hypercube $\{-1, 1\}^{ND}$.*

Proof. Each component of each state will be $\pm 1/\sqrt{ND}$, with equal probability of each. $\quad\square$

12

$\sqrt{ND}\,S(\mathcal{E})$ therefore meets the required conditions for the Marčenko-Pastur law (4.1), so we may say immediately

Lemma 5.2. *Let $\mathcal{E}$ be an ensemble of N D-dimensional random oracle states, and set $r = N/D$. Then*

$$\mathbb{E}(P^{pgm}(\mathcal{E})) \geq \begin{cases} \frac{1}{r}\left(1 - \frac{1}{r}\left(1 - \frac{64}{9\pi^2}\right)\right) & \text{if } N \geq D \\ 1 - r\left(1 - \frac{64}{9\pi^2}\right) & \text{otherwise} \end{cases} \tag{40}$$

and in particular $\mathbb{E}(P^{pgm}(\mathcal{E})) \geq 0.720$ when $N \leq D$.

Like the sphere, the high-dimensional hypercube exhibits the concentration of measure phenomenon, and we can write down a similar result to Levy's Lemma [17]:

Lemma 5.3. (Concentration of measure on the cube) [17]
Given a function $f : \{-1, 1\}^d \mapsto \mathbb{R}$ defined on a d-dimensional hypercube, and a point p on the hypercube chosen uniformly at random,

$$\Pr[|f(p) - \mathbb{E}(f)| \geq \epsilon] \leq 2\exp\left(\frac{-2\epsilon^2}{d\eta^2}\right) \tag{41}$$

where η is the Lipschitz constant of f with respect to the Hamming distance, $\eta = \sup_{x,y} |f(x) - f(y)|/d(x,y)$.

Lemma 5.4. *Let H be a point on the nd-dimensional hypercube written down as an $n \times d$ $\{-1,1\}$-matrix, and let $f(H) = \frac{1}{n^2d}\|H\|_{tr}^2$. Then the Lipschitz constant η of f satisfies $\eta \leq 4/nd$.*

Proof. See Appendix C. □

Plugging this value of η into Lemma 5.3 gives

Theorem 5.5. *Let $\mathcal{E}$ be an ensemble of N D-dimensional random oracle states. Set $p = \mathbb{E}(P^{pgm}(\mathcal{E})) = \frac{1}{r}\left(1 - \frac{1}{r}\left(1 - \frac{64}{9\pi^2}\right)\right)$ if $N \geq D$, and $p = 1 - r\left(1 - \frac{64}{9\pi^2}\right)$ otherwise, where $r = N/D$. Then*

$$\Pr[P^{pgm}(\mathcal{E}) \leq p - \epsilon] \leq 2\exp\left(\frac{-2ND\epsilon^2}{16}\right) \tag{42}$$

and we have our desired result: with 1 query, all but an exponentially small fraction of the possible sets of N N-dimensional random oracle states may be distinguished with a constant probability bounded away from 1/2 (in fact, to get a probability of success greater than 1/2, we may take $r = N/D$ to be as high as ~ 1.66). A constant number of repetitions allows this probability to be boosted to be arbitrarily high.

6 Discussion

This work can be seen as part of an overall programme of understanding the behaviour of random quantum states [21, 20, 11, 22].

There is a fundamental correspondence between the mixed state obtained from an equal mixture of uniformly random pure states, and that produced by starting with a larger system in a uniformly random pure state, and tracing out part of the system. Consider a d-dimensional state

$$\rho_{n,d} = \frac{1}{n} \sum_{i=1}^{n} |\psi_i\rangle\langle\psi_i| \tag{43}$$

where each state in the set $\mathcal{E} = \{|\psi_i\rangle\}$ is picked uniformly at random. We can think of $\rho_{n,d}$ as being produced from the following dn-dimensional state (which we consider to live in a Hilbert space $\mathcal{H}_d \otimes \mathcal{H}_n$) by tracing out the second subsystem:

$$|v\rangle = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} |v_k\rangle|k\rangle = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \sum_{l=0}^{d-1} \alpha_{kl}|l\rangle|k\rangle \tag{44}$$

for some coefficients α_{kl}. As mentioned previously, the α_{kl} will be approximately normally distributed as $\tilde{N}(0, 1/d)$. So, because of the normalisation factor at the front of the sum, the overall state $|v\rangle$ has coefficients which are normally distributed and scaled as $\tilde{N}(0, 1/dn)$. Therefore, this state is picked from the uniform distribution on the unit sphere in $\mathbb{C}^{dn}$. Popescu, Short and Winter [20] obtained an upper bound on the expected trace distance of such a state $\rho_{n,d}$ from the maximally mixed state I/d, and used this to show that for $n \gg d$, $\rho \approx I/d$.

Because the non-zero eigenvalues of the Gram matrix of (rescaled) states in $\mathcal{E}$ are the same as the eigenvalues of $\rho_{n,d}$ [15], this paper can be seen as obtaining a similar result to [20] for the *fidelity* of $\rho_{n,d}$ with the maximally mixed state, via quite different methods. However, the bound is tighter for n close to d, and the notion of "randomness" of the states $\{|\psi_i\rangle\}$ is more general (which is simply a side-effect of relying on the powerful Marčenko-Pastur law).

Acknowledgements

Appendices

A   The PGM is close to optimal

Theorem 2.1. (Barnum, Knill) [3] $P^{pgm}(\mathcal{E}) \geq P^{opt}(\mathcal{E})^2$.

Proof. Consider an arbitrary POVM R consisting of measurement operators $\{R_i\}$, and an arbitrary ensemble $\mathcal{E}$ of renormalised states $\{|\psi_i'\rangle\}$, with a priori probabilities p_i, where as before $|\psi_i'\rangle = \sqrt{p_i}|\psi_i\rangle$ and $\rho = \sum_{i=1}^{n} |\psi_i'\rangle\langle\psi_i'|$. Assume wlog that $R_i = |\mu_i\rangle\langle\mu_i|$ for some vectors $|\mu_i\rangle$, as the optimal measurement will always be of this form [7]. Then

$$P^R(\mathcal{E}) \;=\; \sum_{i=1}^{n}\langle\psi_i'|R_i|\psi_i'\rangle = \sum_{i=1}^{n}|\langle\psi_i'|\mu_i\rangle|^2 = \sum_{i=1}^{n}|\langle\psi_i'|\rho^{-1/4}\rho^{1/4}|\mu_i\rangle|^2 \tag{45}$$

$$\leq \;\; \sum_{i=1}^{n}\langle\psi_i'|\rho^{-1/2}|\psi_i'\rangle\langle\mu_i|\rho^{1/2}|\mu_i\rangle \tag{46}$$

$$\leq \;\; \sqrt{\left(\sum_{i=1}^{n}\langle\psi_i'|\rho^{-1/2}|\psi_i'\rangle^2\right)\left(\sum_{j=1}^{n}\langle\mu_j|\rho^{1/2}|\mu_j\rangle^2\right)} \tag{47}$$

$$\leq \;\; \sqrt{\sum_{i=1}^{n}\langle\psi_i'|\rho^{-1/2}|\psi_i'\rangle^2} = \sqrt{P^{pgm}(\mathcal{E})} \tag{48}$$

The first and second inequalities are Cauchy-Schwarz inequalities, and the third follows because the vectors $\{\rho^{1/2}|\mu_i\rangle\}$ can easily be seen to define an ensemble with density matrix ρ:

$$\sum_{i=1}^{n}\rho^{1/2}|\mu_i\rangle\langle\mu_i|\rho^{1/2} = \rho^{1/2}\left(\sum_{i=1}^{n}|\mu_i\rangle\langle\mu_i|\right)\rho^{1/2} = \rho \tag{49}$$

and we therefore have $\sum_{i=1}^{n}\langle\mu_i|\rho^{1/2}|\mu_i\rangle^2 \leq 1$, as this is the probability of success of the measurement R applied to this ensemble. $\qquad\square$

B   Proof of Lemma 4.4

In this appendix we will prove a lemma which immediately implies Lemma 4.4. See [8] for the facts used about elliptic integrals and hypergeometric series.

Lemma B.1. *Let $0 \leq r \leq 1$ and $A = 1 - \sqrt{r}$, $B = 1 + \sqrt{r}$. Then*

$$\int_{A}^{B} \sqrt{(y^2 - A^2)(B^2 - y^2)}\, dy \geq r\pi\sqrt{1 - r\left(1 - \frac{64}{9\pi^2}\right)} \tag{50}$$

with equality at $r = 0$, $r = 1$.

Proof. We have

$$f(r) = \int_A^B \sqrt{(y^2 - A^2)(B^2 - y^2)}\, dy \tag{51}$$

$$= \frac{B}{3}\left((A^2 + B^2)E\left(\frac{\sqrt{B^2 - A^2}}{B^2}\right) - 2A^2 K\left(\frac{\sqrt{B^2 - A^2}}{B^2}\right)\right) \tag{52}$$

$$= \frac{2(1 + \sqrt{r})}{3}\left((1 + r)E\left(\frac{2r^{1/4}}{1 + \sqrt{r}}\right) - (1 - \sqrt{r})^2 K\left(\frac{2r^{1/4}}{1 + \sqrt{r}}\right)\right) \tag{53}$$

where $K(r)$ and $E(r)$ are the complete elliptic integrals of the first and second kind, respectively:

$$K(r) = \int_0^1 \frac{dx}{\sqrt{(1 - x^2)(1 - r^2 x^2)}}\,,\ E(r) = \int_0^1 \frac{\sqrt{1 - r^2 x^2}}{\sqrt{1 - x^2}}\, dx \tag{54}$$

Note that $f(r)$ may be evaluated explicitly for $r = 0$ and $r = 1$, giving 0 and $8/3$ respectively. Now we may apply a standard change of variables (Landen's transformation) to both elliptic integrals, giving

$$f(r) = \frac{2(1 + \sqrt{r})}{3}\left(\frac{1 + r}{1 + \sqrt{r}}\left(2E(\sqrt{r}) - (1 - r)K(\sqrt{r})\right) - (1 - \sqrt{r})^2(1 + \sqrt{r})K(\sqrt{r})\right)$$

$$= \frac{4}{3}\left((1 + r)E(\sqrt{r}) - (1 - r)K(\sqrt{r})\right) \tag{55}$$

We now move to the representation of $K(r)$ and $E(r)$ as hypergeometric series, which are defined as follows (using the notation $a^{\bar{n}} = a(a + 1)\cdots(a + n - 1)$).

$$_2F_1(a, b; c; r) = \sum_{n=0}^{\infty} \frac{a^{\bar{n}} b^{\bar{n}}}{c^{\bar{n}} n!} r^n \tag{56}$$

$$K(r) = (\pi/2)\,_2F_1(1/2, 1/2; 1; r^2)\,,\ E(r) = (\pi/2)\,_2F_1(-1/2, 1/2; 1; r^2) \tag{57}$$

This has the advantage that, by a transformation rule due to Gauss, we can rewrite $f(r)$ as a single hypergeometric series.

$$f(r) = \frac{2\pi}{3}\left((1 + r)\,_2F_1(-1/2, 1/2; 1; r) - (1 - r)\,_2F_1(1/2, 1/2; 1; r)\right) \tag{58}$$

$$= \pi r\,_2F_1(-1/2, 1/2; 2; r) \tag{59}$$

Returning to the original inequality, our task has been simplified to showing that

$$g(r) = {}_2F_1(-1/2, 1/2; 2; r)^2 \geq 1 - r\left(1 - \frac{64}{9\pi^2}\right) \tag{60}$$

16

Evaluating $g(r)$ at 0 and 1 makes it clear that this is equivalent to showing that $g(r)$ is concave for $0 \leq r \leq 1$, which would follow from showing the second derivative $g''(r)$ to be negative in this region. From the rules governing differentiation of hypergeometric series, it is easy to show that

$$g''(r) = \frac{1}{32} \left({}_2F_1(1/2, 3/2; 3; r)^2 - 2 \, {}_2F_1(-1/2, 1/2; 2; r) \, {}_2F_1(3/2, 5/2; 4; r) \right) \quad (61)$$

The following hypergeometric transformation allows this to be simplified.

$$ {}_2F_1(a, b; c; r) = (1 - r)^{c-a-b} \, {}_2F_1(c - a, c - b; c; r) \quad (62)$$

$$\Rightarrow \quad g''(r) = \frac{1}{32} \left((1 - r)^2 \, {}_2F_1(5/2, 3/2; 3; r)^2 \right. \quad (63)$$

$$\left. - 2(1 - r)^2 \, {}_2F_1(5/2, 3/2; 2; r) \, {}_2F_1(3/2, 5/2; 4; r) \right) \quad (64)$$

We will show that ${}_2F_1(5/2, 3/2; 3; r)^2 \leq {}_2F_1(5/2, 3/2; 2; r) \, {}_2F_1(5/2, 3/2; 4; r)$ for all positive r, implying that $g''(r)$ is negative in this region. We write out the two hypergeometric series explicitly:

$$ {}_2F_1(5/2, 3/2; 3; r)^2 \;=\; \sum_{m,n=0}^{\infty} \frac{k_m k_n}{3^{\bar{m}} 3^{\bar{n}}} \;,\; \text{where } k_n = \frac{(5/2)^{\bar{n}}(3/2)^{\bar{n}}}{n!} r^n \quad (65)$$

$$ {}_2F_1(5/2, 3/2; 2; r) \, {}_2F_1(5/2, 3/2; 4; r) \;=\; \sum_{m,n=0}^{\infty} \frac{k_m k_n}{4^{\bar{m}} 2^{\bar{n}}} \quad (66)$$

$$ = \sum_{m,n=0}^{\infty} \frac{k_m k_n}{3^{\bar{m}} 3^{\bar{n}}} \left(\frac{3}{3+m} \right) \left(\frac{2+n}{2} \right) \quad (67)$$

$$ = \sum_{m=0}^{\infty} \frac{k_m^2}{3^{\bar{m}} 3^{\bar{m}}} \left(\frac{6+3m}{6+2m} \right) + \sum_{\substack{m,n=0 \\ m>n}}^{\infty} \frac{k_m k_n}{3^{\bar{m}} 3^{\bar{n}}} \left(\frac{3(2+n)}{2(3+m)} + \frac{3(2+m)}{2(3+n)} \right) \quad (68)$$

$$ \geq \sum_{m=0}^{\infty} \frac{k_m^2}{3^{\bar{m}} 3^{\bar{m}}} + \sum_{\substack{m,n=0 \\ m>n}}^{\infty} \frac{2 k_m k_n}{3^{\bar{m}} 3^{\bar{n}}} = {}_2F_1(5/2, 3/2; 3; r)^2 \quad (69)$$

where elementary methods can be used to show that the bracketed last term in eqn. (68) is at least 2 for any non-negative m and n. This completes the proof of the lemma. $\qquad \square$

C Lipschitz constants

This appendix contains derivations of the Lipschitz constants of the functions used for the concentration of measure results.
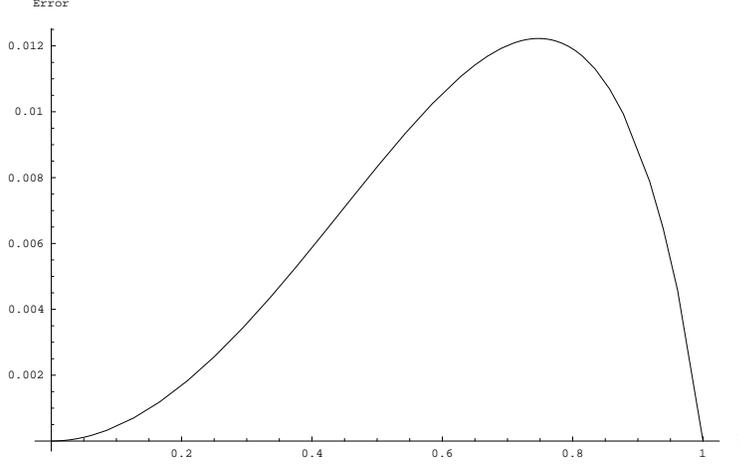
Figure 2: Error in approximation to elliptic integral (50) for $0 \le r \le 1$.

Lemma 4.8. *Let S be an $n \times d$ matrix with $\|S\|_2 = 1$, and define $f(S) = \frac{1}{n}\|S\|_{tr}^2$. Then the Lipschitz constant η of f satisfies $\eta \le 2$.*

Proof. Let $k = \min(n, d)$. We have

$$\eta = \sup_{S,T} \frac{|f(S) - f(T)|}{\|S - T\|_2} = \sup_{S,T} \frac{|\|S\|_{tr}^2 - \|S\|_{tr}^2|}{n\|S - T\|_2} \tag{70}$$

$$= \sup_{S,T} \left(\frac{\|S\|_{tr} + \|T\|_{tr}}{n}\right) \frac{|\|S\|_{tr} - \|S\|_{tr}|}{\|S - T\|_2} \tag{71}$$

$$\le \sup_{S,T} \left(\frac{\|S\|_{tr} + \|T\|_{tr}}{n}\right) \frac{\|S - T\|_{tr}}{\|S - T\|_2} \tag{72}$$

$$\le \sup_{S,T} \frac{\sqrt{k}\,(\|S\|_{tr} + \|T\|_{tr})}{n} \le 2k/n \le 2 \tag{73}$$

The first inequality is a triangle inequality, and the second two are derived from

$$\|S\|_{tr} = \sum_{i=1}^{k} \sigma_i(S) \le \sqrt{k \sum_{i=1}^{k} \sigma_i^2(S)} \le \sqrt{k}\|S\|_2 \tag{74}$$

which in turn uses a Cauchy-Schwarz inequality. $\square$

Lemma 5.4. *Let S be a point on the nd-dimensional hypercube written down as an $n \times d$ $\{-1, 1\}$-matrix, and let $f(S) = \frac{1}{n^2 d}\|S\|_{tr}^2$. Then the Lipschitz constant η of f (with respect to the Hamming distance) satisfies $\eta \le 4/nd$.*

Proof. The proof is very similar to that of Lemma 4.8. As before, let $k = \min(n, d)$. We have

18

$$\eta \quad = \quad \sup_{S,T} \frac{|f(S) - f(T)|}{d(S,T)} = \sup_{S,T} \frac{1}{n^2 d} \frac{|\, \|S\|_{tr}^2 - \|S\|_{tr}^2 \,|}{d(S,T)} \tag{75}$$

$$\leq \quad \sup_{S,T} \left(\frac{\|S\|_{tr} + \|T\|_{tr}}{n^2 d} \right) \frac{\|S - T\|_{tr}}{\frac{1}{2}\|S - T\|_1} \tag{76}$$

$$\leq \quad \sup_{S,T} \frac{2\sqrt{k}\,(\|S\|_{tr} + \|T\|_{tr})}{n^2 d} \leq 4k/n^2 d \leq 4/nd \tag{77}$$

where, extending inequality (74), we use $\|S\|_{tr} \leq \sqrt{k}\|S\|_2 \leq \sqrt{k}\|S\|_1$. $\qquad\square$

References

[1] A. Ambainis, K. Iwama, A. Kawachi, H. Masuda, R. H. Putra, S. Yamashita (2004). Quantum identification of boolean oracles. Proc. STACS 04, LNCS 2996, pp. 105116. quant-ph/0403056

[2] Z. D. Bai (1999). Methodologies in spectral analysis of large dimensional random matrices, a review. Statist. Sinica 9, pp. 611-677.

[3] H. Barnum, E. Knill (2002). Reversing quantum dynamics with near-optimal quantum and classical fidelity. J. Math. Phys. 43, pp. 20972106. quant-ph/0004088

[4] H. Buhrman, R. Cleve, J. Watrous, R. de Wolf (2001). Quantum fingerprinting. Phys. Rev. Lett. 87, 167902. quant-ph/0102001

[5] E. B. Davies (1978). Information and quantum measurement. IEEE Trans. Inform. Theory 24, pp. 596-599.

[6] Y. C. Eldar, G. D. Forney, Jr. (2001). On quantum detection and the square-root measurement. IEEE Trans. Inform. Theory 47, pp. 858872. quant-ph/0005132

[7] Y. C. Eldar, A. Megretski, G. Verghese (2003). Designing optimal quantum detectors via semidefinite programming. IEEE Trans. Inform. Theory 49, pp. 1007-1012. quant-ph/0205178

[8] I. S. Gradshteyn, I. M. Ryzhik. Table of integrals, series and products (1980). Academic Press, New York.

[9] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, W. Wootters (1996). Classical information capacity of a quantum channel. Phys. Rev. A 54, pp. 1869-1876.

[10] P. Hausladen, W. Wootters (1994). A "pretty good" measurement for distinguishing quantum states. J. Mod. Opt. 41, 2385.

[11] P. Hayden, D. W. Leung, A. Winter (2006). Aspects of generic entanglement. Commun. Math. Phys. 265, 95117. quant-ph/0407049

[12] C. W. Helstrom (1976). Quantum Detection and Estimation Theory. Academic Press, New York.

[13] A. S. Holevo (1973). Bounds for the quantity of information transmittable by a quantum communications channel. Problemy Peredachi Informatsii 9, no. 3, pp. 3-11. English translation: Problems of Information Transmission 9, pp. 177-183.

[14] R. A. Horn, C. R. Johnson (1985). Matrix Analysis. Cambridge University Press, Cambridge.

[15] R. Jozsa, J. Schlienz (2000). Distinguishability of states and von Neumann entropy. Phys. Rev. A 62 012301. quant-ph/9911009

[16] M. Ledoux (1996). Isoperimetry and Gaussian analysis. Ecole d'Eté de Probabilités de St.-Flour 1994. Lecture Notes in Math. 1648, pp. 165-294.

[17] M. Ledoux (2001). The concentration of measure phenomenon. AMS Mathematical Surveys and Monographs 89, American Mathematical Society.

[18] V. A. Marčenko, L. A. Pastur (1967). Distributions of eigenvalues of some sets of random matrices. Math. USSR-Sb. 1, pp. 507-536.

[19] M. A. Nielsen, I. L. Chuang (2000). Quantum computation and quantum information. Cambridge University Press, Cambridge.

[20] S. Popescu, A. J. Short, A. Winter (2005). Entanglement and the foundations of statistical mechanics. quant-ph/0511225

[21] W. K. Wootters (1990). Random quantum states. Found. Phys. 20, 1365.

[22] K. Zyczkowski, H. Sommers (2005). Average fidelity between random quantum states. Phys. Rev. A 71, 032313. quant-ph/0311117