Pretty simple bounds on quantum state discrimination

Ashley Montanaro*

November 7, 2019

Abstract

We show that the quantum measurement known as the pretty good measurement can be used to identify an unknown quantum state picked from any set of n mixed states that have pairwise fidelities upper-bounded by a constant below 1, given $O(\log n)$ copies of the unknown state, with high success probability in the worst case. If the unknown state is promised to be pure, there is an explicit measurement strategy which solves this worst-case quantum state discrimination problem with $\tilde{O}(\|G\|)$ copies, where G is the Gram matrix of the states.

1 Introduction

A fundamental task in quantum information theory is *quantum state discrimination*. Here we will be concerned with the following variant of this problem: We are given an unknown state ρ picked from a known set $S = \{\rho_i\}$ of mixed states, where $|S| = n$, and our task is to identify ρ with the lowest possible worst-case probability δ of failure. That is, we want to find a quantum measurement (POVM), described by a set of positive semidefinite operators μ_i with $\sum_i \mu_i = I$, such that $\max_i 1 - \operatorname{tr} \mu_i \rho_i$ is minimised. This task has been termed "minimax" quantum state discrimination [9], though here we will refer to it as *worst-case* quantum state discrimination. We will also consider the closely related question where we are given $\delta > 0$ in advance, and would like to determine the number k of copies of ρ that are required to achieve failure probability δ by performing a measurement on $\rho^{\otimes k}$.

Quantum state discrimination has been a topic of intensive study within quantum information theory (see [4, 7, 2] for reviews), although the majority of works consider the setting where each state ρ_i is produced with a known probability p_i. However, in the context of a quantum algorithm which should have a low worst-case probability of failure, worst-case discrimination is often the most natural setting.

In the worst-case setting, it was shown by Harrow and Winter [10] that, if all states in S have pairwise fidelities $F(\rho_i, \rho_j) := \|\sqrt{\rho_i}\sqrt{\rho_j}\|_1$ upper-bounded by F (where $0 < F < 1$), then the worst-case state discrimination problem can be solved with $O(\log(n/\delta)/\log(1/F))$ copies of ρ. This result is nonconstructive, and the proof proceeds via a minimax theorem. That is, existence of a measurement that solves the worst-case state discrimination problem is shown, without describing the measurement explicitly.

Here we show that there is an explicit measurement, the pretty good measurement [6, 13, 12] ("PGM", defined below) which achieves a similar scaling of the number of copies:

Theorem 1. *If $F(\rho_i, \rho_j) \leq 1 - \epsilon$ for all pairs of distinct states $\rho_i, \rho_j \in S$, then the worst-case state discrimination problem can be solved with failure probability δ by applying the PGM to $O(\log(n/\delta)/\epsilon)$ copies of ρ.*

*School of Mathematics, University of Bristol, UK; ashley.montanaro@bristol.ac.uk.

Next we show that, in the case of pure states, this result can sometimes be improved to a scaling of the number of copies required which does not depend on n. In the case where all states in S are pure, write $\rho_i = |\psi_i\rangle\langle\psi_i|$, and let G be the Gram matrix of these vectors, i.e. $G_{ij} = \langle\psi_i|\psi_j\rangle$. Then:

Theorem 2. *If all states in S are pure, then the worst-case state discrimination problem can be solved with success probability at least $\|G\|^{-1}$ by applying the PGM to one copy of ρ. If additionally $\operatorname{tr}\rho_i\rho_j \leq 1-\epsilon$ for all pairs of distinct states in S, then there is an explicit measurement strategy which, applied to*

$$O((\|G\|/\epsilon)(\log 1/\delta)\log(\|G\|/\delta)) = \widetilde{O}(\|G\|/\epsilon)$$

copies of ρ, solves the worst-case state discrimination problem with failure probability δ.

In Theorem 2 and throughout, we use $\widetilde{O}(f(\|G\|,\epsilon,\delta))$ to denote $O(f(\|G\|,\epsilon,\delta)\operatorname{polylog}(\|G\|,1/\epsilon,1/\delta))$.

Thus the operator norm of G bounds the success probability and the number of copies required to solve the worst-case quantum state discrimination problem. To gain some intuition for this result, note that if all the states in S are orthogonal, G is the identity matrix, so $\|G\| = 1$; whereas if all the states in S are equal, $G_{ij} = 1$ and $\|G\| = n$.

Theorem 2 can be applied, for example, to random pure states picked from a variety of distributions. It often holds that, for random states in d dimensions and with $n = O(d)$, $\|G\| = O(1)$ with high probability. Indeed, this holds for any states whose amplitudes with respect to an arbitrary basis are close to iid random variables with suitably bounded 4th moments [15, 3]. Examples are Haar-random pure states and states of the form $\frac{1}{\sqrt{d}}\sum_{i=1}^{d} z_i|i\rangle$, where z_i is uniformly randomly chosen from $\{\pm 1\}$. Upper bounds were proven on the probability of failure of discriminating these ensembles of states in [14] in the case where there is a uniform probability distribution on the states in S. Theorem 2 extends this to a worst-case setting.

2 Definitions

The pretty good measurement [6, 13, 12] (PGM), also known as the square-root measurement [11], is defined as follows: for each state ρ_i we introduce a measurement operator $\mu_i = \Sigma^{-1/2}\rho_i\Sigma^{-1/2}$, where $\Sigma := \sum_i \rho_i$ and the inverse is taken on the support of Σ. This is a valid POVM as

$$\sum_i \mu_i = \sum_i \Sigma^{-1/2}\rho_i\Sigma^{-1/2} = \Sigma^{-1/2}\left(\sum_i \rho_i\right)\Sigma^{-1/2} = I,$$

where the identity operator is with respect to the span of the states in the support of S. Note that often this measurement is defined in terms of states ρ_i normalised by some a priori probabilities p_i, but here these are not used.

Write $\rho_i = \sum_j \lambda_{ij}|\psi_{ij}\rangle\langle\psi_{ij}|$ for the eigendecomposition of ρ, and let G be the Gram matrix of the weighted states $\{\sqrt{\lambda_{ij}}|\psi_{ij}\rangle\}$. G has a natural block structure in terms of i. If we define the vectors $|\mu_{ij}\rangle = \Sigma^{-1/2}\sqrt{\lambda_{ij}}|\psi_{ij}\rangle$ and the positive semidefinite matrix $P_{ik,jl} = \sqrt{\lambda_{jl}}\langle\mu_{ik}|\psi_{jl}\rangle$, then

$$\begin{aligned}
(P^2)_{ik,jl} &= \sum_{r,s}\sqrt{\lambda_{ik}}\lambda_{rs}\sqrt{\lambda_{jl}}\langle\psi_{ik}|\Sigma^{-1/2}|\psi_{rs}\rangle\langle\psi_{rs}|\Sigma^{-1/2}|\psi_{jl}\rangle \\
&= \sqrt{\lambda_{ik}}\sqrt{\lambda_{jl}}\langle\psi_{ik}|\Sigma^{-1/2}\left(\sum_{r,s}\lambda_{rs}|\psi_{rs}\rangle\langle\psi_{rs}|\right)\Sigma^{-1/2}|\psi_{jl}\rangle \\
&= G_{ik,jl}.
\end{aligned}$$

Thus the probability that the PGM outputs i on input ρ_j is

$$\operatorname{tr} \mu_i \rho_j = \operatorname{tr} \left(\sum_k |\mu_{ik}\rangle\langle\mu_{ik}| \right) \left(\sum_l \lambda_{jl} |\psi_{jl}\rangle\langle\psi_{jl}| \right) = \sum_{k,l} \lambda_{jl} |\langle \mu_{ik}|\psi_{jl}\rangle|^2 = \|P^{(ij)}\|_2^2 = \|\sqrt{G}^{(ij)}\|_2^2,$$

where we use $P^{(ij)}$ to denote the (i,j)'th block of P and $\|M\|_2^2 := \sum_{i,j} |M_{ij}|^2$. Write

$$P_E(S) := \max_i 1 - \operatorname{tr} \mu_i \rho_i = \max_i 1 - \|\sqrt{G}^{(ii)}\|_2^2$$

for the worst-case probability of error when the PGM is used.

3 Worst-case bounds for mixed states

It was shown in [10], based on a bound of [5], that $O(\log n)$ copies are required to identify a state picked from an arbitrary set of n states with pairwise fidelities bounded above by a constant $F < 1$. The result of [10] states that there exists a measurement that achieves this complexity, without describing that measurement explicitly. The reason is that the result of [5] is stated in terms of a known probability distribution on the states, and [10] uses a minimax theorem to lift this to a worst-case bound. Here we show that the PGM itself achieves such a bound. The proof is directly analogous to that of [5] in terms of considering Hilbert-Schmidt norms of off-diagonal blocks of $\sqrt{G}$, though it proceeds via a slightly different route.

Lemma 3.

$$P_E(S) \leq \sum_{i \neq j} F(\rho_i, \rho_j).$$

Proof. Set $\Lambda = \sum_i |i\rangle\langle i| \otimes G^{(ii)}$, $\Delta = \sum_{i \neq j} |i\rangle\langle j| \otimes G^{(ij)}$, such that $G = \Lambda + \Delta$. We are interested in upper-bounding

$$P_E(S) = \max_i \sum_{j \neq i} \|\sqrt{G}^{(ij)}\|_2^2 \leq \sum_{i,j:i\neq j} \|\sqrt{G}^{(ij)}\|_2^2 \leq \|\sqrt{G} - \sqrt{\Lambda}\|_2^2,$$

where the second inequality holds because $\sqrt{\Lambda}^{(ij)} = 0$ for $i \neq j$. For any unitarily invariant norm $\|\cdot\|$ and any positive operators A, B, we have [8, Theorem X.1.3]

$$\|\sqrt{A} - \sqrt{B}\| \leq \|\sqrt{|A-B|}\|.$$

Hence

$$\|\sqrt{G} - \sqrt{\Lambda}\|_2^2 = \|\sqrt{\Lambda + \Delta} - \sqrt{\Lambda}\|_2^2 \leq \|\sqrt{|\Delta|}\|_2^2 = \|\Delta\|_1 \leq \sum_{i \neq j} \|G^{(ij)}\|_1 = \sum_{i \neq j} F(\rho_i, \rho_j),$$

where $\|\cdot\|_1$ denotes the trace norm. The last equality follows from

$$G^{(ij)} = \sum_{k,l} \sqrt{\lambda_{ik}}\sqrt{\lambda_{jl}} \langle \psi_{ik}|\psi_{jl}\rangle |k\rangle\langle l| = \left(\sum_k \sqrt{\lambda_{ik}} |k\rangle\langle\psi_{ik}| \right) \left(\sum_l \sqrt{\lambda_{jl}} |\psi_{jl}\rangle\langle l| \right),$$

which implies that

$$\|G^{(ij)}\|_1 = \left\| \left(\sum_k \sqrt{\lambda_{ik}} |\psi_{ik}\rangle\langle\psi_{ik}| \right) \left(\sum_l \sqrt{\lambda_{jl}} |\psi_{jl}\rangle\langle\psi_{jl}| \right) \right\|_1 = \|\sqrt{\rho_i}\sqrt{\rho_j}\|_1 = F(\rho_i, \rho_j)$$

by unitary invariance of the trace norm and orthonormality of the states $\{|\psi_{ik}\rangle\}$ for each i. □

Lemma 3 implies that if $F(\rho_i, \rho_j) \leq 1/(3n^2)$ for all $i \neq j$, then $P_E(S) \leq 1/3$. This can be seen as a generalisation of a folklore result proven by Ambainis and de Wolf [1], albeit with a somewhat worse constant. The result of [1] was only shown for pure states, but states that if $F(\rho_i, \rho_j) \leq 1/n^2$ for all $i \neq j$, then $P_E(S) \leq 1/3$.

Theorem 1 (restated). *If $F(\rho_i, \rho_j) \leq 1 - \epsilon$ for all pairs of distinct states $\rho_i, \rho_j \in S$, then the worst-case state discrimination problem can be solved with failure probability δ by applying the PGM to $O(\log(n/\delta)/\epsilon)$ copies of ρ.*

Proof. Let $S' = \{\rho_i^{\otimes k} : i \in \{1, \ldots, n\}\}$; then by Lemma 3,

$$P_E(S') \leq \sum_{i \neq j} F(\rho_i^{\otimes k}, \rho_j^{\otimes k}) \leq n(n-1) \max_{i \neq j} F(\rho_i, \rho_j)^k \leq n^2(1-\epsilon)^k \leq n^2 e^{-k\epsilon}$$

so it is sufficient to take $k = \lceil (2/\epsilon) \ln(n/\delta) \rceil$ to achieve failure probability at most δ. □

4 Improved bounds for pure states

We now find an alternative bound which is only good for pure (or not too mixed) states, but which further improves on [10] by not having any (explicit) dependence on n. The bound states that, for any set of pure states whose pairwise fidelities are bounded above by $1-\epsilon$, where $\epsilon > 0$ is a constant, the state discrimination problem can be solved with $\widetilde{O}(\|G\|)$ copies of the unknown state, where $\|\cdot\|$ is the operator norm. The bound is based on the following technical lemma:

Lemma 4. *For all i,*

$$\|G\|^{-1} \operatorname{tr} \rho_i^2 \leq \operatorname{tr} \mu_i \rho_i \leq \|G^{-1}\| \operatorname{tr} \rho_i^2.$$

Proof. We have

$$\lambda_{\min}(\sqrt{G})\sqrt{G} \leq G \leq \lambda_{\max}(\sqrt{G})\sqrt{G}$$

in a positive semidefinite sense, where $\lambda_{\min}(G)$, $\lambda_{\max}(G)$ are the minimal and maximal eigenvalues of G. This inequality is preserved under projections and taking the 2-norm. So

$$\|G\| \operatorname{tr} \mu_i \rho_i = \lambda_{\max}(\sqrt{G})^2 \|\sqrt{G}^{(ii)}\|_2^2 \geq \|G^{(ii)}\|_2^2 = \operatorname{tr} \rho_i^2,$$

which is the lower bound of the lemma, and

$$\|G^{-1}\|^{-1} \operatorname{tr} \mu_i \rho_i = \lambda_{\min}(\sqrt{G})^2 \|\sqrt{G}^{(ii)}\|_2^2 \leq \|G^{(ii)}\|_2^2 = \operatorname{tr} \rho_i^2,$$

which is the upper bound. □

If one assumes a uniform distribution on the states in S and that they are pure, the lower bound in Lemma 4 is a corollary of [14, Lemma 2].

Theorem 2 (restated). *If all states in S are pure, then the worst-case state discrimination problem can be solved with success probability at least $\|G\|^{-1}$ by applying the PGM to one copy of ρ. If additionally $\operatorname{tr} \rho_i \rho_j \leq 1 - \epsilon$ for all pairs of distinct states in S, then there is an explicit measurement strategy which, applied to*

$$O((\|G\|/\epsilon)(\log 1/\delta) \log(\|G\|/\delta)) = \widetilde{O}(\|G\|/\epsilon)$$

copies of ρ, solves the worst-case state discrimination problem with failure probability δ.

Proof. The first part is immediate from Lemma 4. For the second part, apply the PGM separately to $k = \lceil \|G\| \ln(2/\delta) \rceil$ copies of ρ, obtaining outcomes $i_1, \ldots, i_k$. The probability that the outcome corresponding to ρ is not among the outcomes obtained is at most

$$(1 - \|G\|^{-1})^k \leq e^{-\|G\|^{-1}k} \leq \delta/2.$$

Then copies of ρ are tested via $l = \lceil \ln(2k/\delta)/\epsilon \rceil$ uses of each of the accept/reject measurements that project onto ρ_{i_j}, for each outcome i_j. If all measurements in the j'th group accept, then the protocol outputs i_j. If no such group of measurements all accept, the protocol outputs "fail".

Each accept/reject measurement accepts ρ_{i_j} with certainty. By the fidelity constraint, the probability that all the measurements for a given j accept ρ if $\rho \neq \rho_{i_j}$ is at most $(1-\epsilon)^l \leq e^{-\epsilon l} \leq \delta/(2k)$. By a union bound, the probability that any group of measurements incorrectly all accepts is at most $\delta/2$. Thus the probability that the whole protocol fails is at most δ. The overall number of copies of ρ used is at most $k(l+1) = O((\|G\|/\epsilon)(\log 1/\delta)\log(\|G\|/\delta))$ as claimed. □

Note the following additional points about Theorem 2:

1. For pure states, Theorem 1 is a corollary of Theorem 2. Letting G be the Gram matrix of the states $\{|\psi_i\rangle^{\otimes k}\}$,
$$\|G\| \leq 1 + (n-1)\max_{i \neq j} |G_{ij}| = 1 + (n-1)\max_{i \neq j} |\langle \psi_i | \psi_j \rangle|^k.$$
If $|\langle \psi_i | \psi_j \rangle| \leq 1 - \epsilon$ for all $i \neq j$, this quantity becomes arbitrarily close to 1 for sufficiently large $k = O((\log n)/\epsilon)$.

2. It is not possible to obtain a similar result to Theorem 2 for arbitrary mixed states (i.e. an upper bound only in terms of $\|G\|$), as can be seen by considering the states $\rho_i = I/n$. In this case one can calculate that $G = J \otimes (I/n)$, where $J_{ij} = 1$, $i, j \in \{1, \ldots, n\}$; so $\|G\| = 1$, but the maximal worst-case success probability that can be achieved is $1/n$.

3. A case where Theorem 2 is quite weak is a set of states whose pairwise inner products are all equal to some constant $c \in (0, 1)$. Then $\|G\| = 1 + c(n-1)$, so Theorem 2 states that the state discrimination problem could be solved with $\widetilde{O}(n)$ copies. It is obvious that this could be improved to $\widetilde{O}(\log n)$ copies, as taking k copies maps $c \mapsto c^k$. But in this case we can explicitly calculate that $(\sqrt{G}_{ii})^2 \geq 1 - c - O(1/n)$, so only $O(1)$ copies are required to achieve a high probability of success [14].

Acknowledgements

I would like to thank Aram Harrow, Andreas Winter and Jon Tyson for comments on previous versions. I acknowledge support from the QuantERA ERA-NET Cofund in Quantum Technologies implemented within the European Union's Horizon 2020 Programme (QuantAlgo project), EPSRC Early Career Fellowship EP/L021005/1, and EPSRC grant EP/R043957/1. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 817581). No new data were created during this study.

References

[1] A. Ambainis and R. de Wolf. How low can approximate degree and quantum query complexity be for total boolean functions? *Computational Complexity*, 23(2):305–322, 2014. arXiv:1206.0717.

[2] J. Bae and L.-C. Kwek. Quantum state discrimination and its applications. *J. Phys. A: Math. Gen.*, 48(8):083001, 2015. arXiv:1707.02571.

[3] Z. D. Bai. Methodologies in spectral analysis of large dimensional random matrices, a review. *Statist. Sinica*, 9(3):611–677, 1999.

[4] S. Barnett and S. Croke. Quantum state discrimination. *Advances in Optics and Photonics*, 1(2):238–278, 2009. arXiv:0810.1970.

[5] H. Barnum and E. Knill. Reversing quantum dynamics with near-optimal quantum and classical fidelity. *J. Math. Phys.*, 43(5):2097–2106, 2002. quant-ph/0004088.

[6] V. P. Belavkin. Optimal multiple quantum statistical hypothesis testing. *Stochastics*, 1:315–345, 1975.

[7] J. Bergou. Discrimination of quantum states. *Journal of Modern Optics*, 57(3):160–180, 2010.

[8] R. Bhatia. *Matrix Analysis*. Springer-Verlag, 1997.

[9] G. D'Ariano, M. Sacchi, and J. Kahn. Minimax quantum-state discrimination. *Phys. Rev. A*, 72:032310, 2005. quant-ph/0504048.

[10] A. Harrow and A. Winter. How many copies are needed for state discrimination? *IEEE Trans. Inform. Theory*, 58(1):1–2, 2012. quant-ph/0606131.

[11] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. Wootters. Classical information capacity of a quantum channel. *Phys. Rev. A*, 54(3):1869–1876, 1996.

[12] P. Hausladen and W. Wootters. A "pretty good" measurement for distinguishing quantum states. *J. Mod. Opt.*, 41(12):2385–2390, 1994.

[13] A. S. Holevo. On asymptotically optimal hypothesis testing in quantum statistics. *Theory of Prob. and its Appl.*, 23:411–415, 1979.

[14] A. Montanaro. On the distinguishability of random quantum states. *Comm. Math. Phys.*, 273(3):619–636, 2007. quant-ph/0607011.

[15] Y. Yin, Z. Bai, and P. Krishnaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields*, 78(4):509–521, 1988.