Electronic Colloquium on Computational Complexity, Report No. 57 (2010)

A Full Characterization of Quantum Advice

Scott Aaronson* MIT

Andrew Drucker† MIT

Abstract

We prove the following surprising result: given any quantum state ρ on n qubits, there exists a local Hamiltonian H on poly (n) qubits (e.g., a sum of two-qubit interactions), such that any ground state of H can be used to simulate ρ on all quantum circuits of fixed polynomial size. In terms of complexity classes, this implies that $\mathsf{BQP/qpoly} \subseteq \mathsf{QMA/poly}$, which supersedes the previous result of Aaronson that $\mathsf{BQP/qpoly} \subseteq \mathsf{PP/poly}$. Indeed, we can exactly characterize quantum advice, as equivalent in power to *untrusted* quantum advice combined with trusted *classical* advice.

Proving our main result requires combining a large number of previous tools—including a result of Alon et al. on learning of real-valued concept classes, a result of Aaronson on the learnability of quantum states, and a result of Aharonov and Regev on 'QMA_+ super-verifiers'—and also creating some new ones. The main new tool is a so-called *majority-certificates lemma*, which is closely related to boosting in machine learning, and which seems likely to find independent applications. In its simplest version, this lemma says the following. Given any set S of Boolean functions on n variables, any function $f \in S$ can be expressed as the pointwise majority of $m = O(n)$ functions $f_1, \ldots, f_m \in S$, such that each f_i is the unique function in S compatible with $O(\log |S|)$ input/output constraints.

Contents

1 Introduction **2**
- 1.1 Our Quantum Information Result . . . 3
- 1.2 Impact on Quantum Complexity Theory . . . 4
- 1.3 Proof Overview . . . 6
- 1.4 Majority-Certificates Lemma in Context . . . 7
- 1.5 Organization of the Paper . . . 8

2 The Majority-Certificates Lemma **8**

3 Extension to Real Functions **10**
- 3.1 Background from Learning Theory . . . 10
- 3.2 The Safe Winnowing Lemma . . . 12
- 3.3 The Real-Valued Majority-Certificates Lemma . . . 13

*Email: aaronson@csail.mit.edu. This material is based upon work supported by the National Science Foundation under Grant No. 0844626. Also supported by a DARPA YFA grant, a Sloan Fellowship, and the Keck Foundation.

†Email: adrucker@mit.edu. This work was done while the author was supported by an Akamai Presidential Graduate Fellowship.

ISSN 1433-8092

4 Application to Quantum Advice **16**
4.1 Bestiary of Quantum Complexity Classes 16
4.2 Characterizing Quantum Advice 19
4.3 The Complexity of Preparing Quantum Advice States 21

5 Further Implications for Quantum Complexity Theory **22**

6 Open Problems **25**

7 Acknowledgments **25**

8 Appendix: Untrusted Oracles **27**

9 Appendix: Isolatability and Learnability **28**

10 Appendix: Winnowing of p-Concept Classes **30**

1 Introduction

How much classical information is needed to specify a quantum state of n qubits?

This question has inspired a rich and varied set of responses, in part because it can be interpreted in many ways. If we want to specify a quantum state ρ *exactly*, then of course the answer is 'an infinite amount,' since amplitudes in quantum mechanics are continuous. A natural compromise is to try to specify ρ *approximately*, i.e., to give a description which yields a state $\widetilde{\rho}$ whose statistical behavior is close to that of ρ under every measurement. (This statement is captured by the requirement that ρ and $\widetilde{\rho}$ are close under the so-called *trace distance* metric.) But it is not hard to see that even for this task, we still need to use an exponential (in n) number of classical bits.

This fact can be viewed as a disappointment, but also as an opportunity, since it raises the prospect that we might be able to encode massive amounts of information in physically compact quantum states: for example, we might hope to store 2^n classical bits in n qubits. But an obvious practical requirement is that we be able to retrieve the information reliably, and this rules out the hope of significant 'quantum compression' of classical strings, as shown by a landmark result of Holevo [20] from 1973. Consider a sender Alice and a recipient Bob, with a one-way quantum channel between them. Then Holevo's Theorem says that, if Alice wants to encode an n-bit classical string x into an m-qubit quantum state ρ_x, in such a way that Bob can retrieve x (with probability 2/3, say) by measuring ρ_x, then Alice must take $m \geq n - O(1)$ (or $m \geq n/2 - O(1)$, if Alice and Bob share entanglement). In other words, for this communication task, quantum states offer essentially no advantage over classical strings. In 1999, Ambainis et al. [12] generalized Holevo's result as follows: even if Bob wants to learn only a *single bit* x_i of $x = x_1 \ldots x_n$ (for some $i \in [n]$ unknown to Alice), and is willing to destroy the state ρ_x in the process of learning that bit, Alice still needs to send $m = \Omega(n)$ qubits for Bob to succeed with high probability.

These results say that the exponential descriptive complexity of quantum states cannot be effectively harnessed for classical data storage, but they do not bound the number of practically meaningful 'degrees of freedom' in a quantum state used for purposes other than storing data. For example, a quantum state could be useful for computation, or it could be a physical system worthy of study in its own right. The question then becomes, what useful information *can* we give about an n-qubit state using a 'reasonable' number (say, $\text{poly}(n)$) of classical bits?

One approach to this question is to identify special subclasses of quantum states for which a faithful approximation can be specified using only $\text{poly}(n)$ bits. This has been done, for example, with matrix product states [29] and 'tree states' [1]. A second approach is to try to describe an *arbitrary* n-qubit state ρ concisely, in such a way that the state $\widetilde{\rho}$ recovered from the description is close to ρ with respect to some natural subclass of *measurements*. This has been done for specific classes like the 'pretty good measurements' of Hausladen and Wootters [19]. A more ambitious goal in this vein, explored by Aaronson in two previous works [2, 5] and continued in the present paper, is to give a description of an n-qubit state ρ which yields a state $\widetilde{\rho}$ that behaves approximately like ρ with respect to all (binary) measurements performable by quantum circuits of 'reasonable' size—say, of size at most n^c, for some fixed $c > 0$. Then if c is taken large enough, $\widetilde{\rho}$ is arguably 'just as good' as ρ for practical purposes.

Certainly we can achieve this goal using $2^{n^{c+O(1)}}$ bits: simply give approximations to the measurement statistics for every size-n^c circuit. However, the results of Holevo [20] and Ambainis et al. [12] suggest that a much more succinct description might be possible. This hope was realized by Aaronson [2], who gave a description scheme in which an n-qubit state can be specified using $\text{poly}(n)$ classical bits. There is a significant catch in Aaronson's result, though: the encoder Alice and decoder Bob both need to invest exponential amounts of computation.

In a subsequent paper [5], Aaronson gave a closely-related result which significantly reduces the computational requirements: now Alice can generate her message in polynomial time (for fixed c). Also, while Bob cannot necessarily construct the state $\widetilde{\rho}$ efficiently on his own, if he is presented with such a state (by an untrusted prover, say), Bob can *verify* the state in polynomial time. The catch in this result is a weakened approximation guarantee: Bob cannot use $\widetilde{\rho}$ to predict the outcomes of *all* the measurements defined by size-n^c circuits, but only *most* of them (with respect to a samplable distribution used by Alice in the encoding process). Aaronson [2, 5] conjectured that the tradeoff between this result and the previous one revealed an inherent limit to quantum compression.

1.1 Our Quantum Information Result

The main result of this paper is that Aaronson's conjecture was false: one really can get the best of both worlds, and simulate an arbitrary quantum state ρ on all small circuits, using a different state $\widetilde{\rho}$ that is easy to recognize. Indeed, we can even take $\widetilde{\rho}$ to be the *ground state of a local Hamiltonian*: that is, the unique pure state $\widetilde{\rho} = |\psi\rangle\langle\psi|$ on $\text{poly}(n)$ qubits that is compatible with $\text{poly}(n)$ local constraints, each involving a constant number of qubits. In a sense, then, this paper completes a 'trilogy' of which [2, 5] were the first two installments.

Here is a formal statement of our result.

Theorem 1 *Let $c, \varepsilon > 0$, and let ρ be any n-qubit quantum state. Then there exists a 2-local Hamiltonian H on $\text{poly}\left(n, \frac{1}{\varepsilon}\right)$ qubits with unique ground state $|\psi\rangle\langle\psi|$, and a transformation $C \longrightarrow C'$ of quantum circuits, computable in time $\text{poly}(n, 1/\varepsilon)$ given H, such that the following holds: $|C'(|\psi\rangle\langle\psi|) - C(\rho)| \leq \varepsilon$ for any measurement C definable by a quantum circuit of size n^c. (Here $C(\rho)$ is the probability that C accepts ρ.)*

In other words, the ground states of local Hamiltonians are 'universal quantum states' in a very non-obvious sense. For example, suppose you own a quantum software store, which sells quantum states ρ that can be fed as input to quantum computers. Then our result says that *ground states of local Hamiltonians are the only kind of state you ever need to stock*. What makes this surprising is that being a good piece of quantum software might entail satisfying an exponential

number of constraints: for example, if ρ is supposed to help a customer's quantum computer Q evaluate some Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$, then $Q(\rho, x)$ should output $f(x)$ for *every* input $x \in \{0,1\}^n$. By contrast, any k-local Hamiltonian H can be described as a set of at most $\binom{n}{k} = O(n^k)$ constraints.

One can also interpret Theorem 1 as a statement about communication over quantum channels. Suppose Alice (who is computationally unbounded) has a classical description of an n-qubit state ρ. She would like to describe ρ to Bob (who is computationally bounded), at least well enough for Bob to be able to *simulate* ρ on all quantum circuits of some fixed polynomial size. However, Alice cannot just send ρ to Bob, since her quantum communication channel is noisy and there is a chance that ρ might get corrupted along the way. Nor can she send a faithful classical description of $|\psi\rangle$, since that would require an exponential number of bits. Our result provides an alternative: Alice can send a different quantum state σ, of $\text{poly}(n)$ qubits, together with a $\text{poly}(n)$-bit classical string x. Then, Bob can use x to *verify* that σ can be used to accurately simulate ρ on all small measurements.

We believe Theorem 1 makes a significant contribution to the study of the effective information content of quantum states. It does, however, leave open whether a quantum state of n qubits can be efficiently encoded *and* decoded in polynomial time, in a way that is 'good enough' to preserve the measurement statistics of measurements defined by circuits of fixed polynomial size. This remains an important problem for future work.

1.2 Impact on Quantum Complexity Theory

The questions addressed in this paper, and our results, are naturally phrased and proved in terms of complexity classes. In recent years, researchers have defined quantum complexity classes as a way to study the 'useful information' embodied in quantum states. One approach is to study the power of nonuniform *quantum advice*. The class $\mathsf{BQP/qpoly}$, defined by Nishimura and Yamakami [25], consists of all languages decidable in polynomial time by a quantum computer, with the help of a $\text{poly}(n)$-qubit advice state that depends only on the input length n. This class is analogous to the classical class $\mathsf{P/poly}$. To understand the role of quantum information in determining the power of $\mathsf{BQP/qpoly}$, a useful benchmark of comparison is the class $\mathsf{BQP/poly}$ of decision problems efficiently solvable by a quantum computer with $\text{poly}(n)$ bits of *classical* advice. It is open whether $\mathsf{BQP/qpoly} = \mathsf{BQP/poly}$.

A second approach studies the power of quantum *proof systems*, by analogy with the classical class NP. Kitaev (unpublished, 1999) defined the complexity class now called QMA, for 'Quantum Merlin-Arthur'. This is the class of decision problems for which a 'yes' answer can be proved by exhibiting a *quantum witness state* (or *quantum proof*) $|\psi\rangle$, on $\text{poly}(n)$ qubits, which is then checked by a skeptical polynomial-time quantum verifier. A natural benchmark class is QCMA (for 'Quantum Classical Merlin-Arthur'), defined by Aharonov and Naveh [8]. This is the class of decision problems for which a 'yes' answer can be checked by a *quantum* verifier who receives a *classical* witness. Here the natural open question is whether $\mathsf{QMA} = \mathsf{QCMA}$.

In this paper we prove a new upper bound on $\mathsf{BQP/qpoly}$:

Theorem 2 $\mathsf{BQP/qpoly} \subseteq \mathsf{QMA/poly}$.

Previously Aaronson showed in [2] that $\mathsf{BQP/qpoly} \subseteq \mathsf{PP/poly}$, and showed in [5] that $\mathsf{BQP/qpoly}$ is contained in the 'heuristic' class $\mathsf{HeurQMA/poly}$; Theorem 2 supersedes both of these earlier results.

Theorem 2 says that one can always replace polynomial-size quantum advice by polynomial-size *classical* advice, together with a polynomial-size quantum *witness* (or equivalently, *untrusted* quantum advice). Indeed, we can *characterize* the class $\mathsf{BQP/qpoly}$, as equal to the subclass of $\mathsf{QMA/poly}$ in which the quantum witness state $|\psi_n\rangle$ can only depend on the input length n.[1]

Using Theorem 2, we also obtain several other results for quantum complexity theory:

(1) Without loss of generality, every quantum advice state can be taken to be the ground state of some local Hamiltonian H. (This essentially follows by combining our $\mathsf{BQP/qpoly} \subseteq \mathsf{QMA/poly}$ result with the result of Kitaev that Local Hamiltonians is QMA-complete.)

(2) It is open whether for every local Hamiltonian H on n qubits, there exists a quantum circuit of size $\mathrm{poly}(n)$ that prepares a ground state of H. It is easy to show that an affirmative answer would imply $\mathsf{QMA} = \mathsf{QCMA}$. As a consequence of Theorem 2, we can show that an affirmative answer would also imply $\mathsf{BQP/qpoly} = \mathsf{BQP/poly}$—thereby establishing a previously-unknown connection between quantum proofs and quantum advice.

(3) In the full version of this paper, we generalize Theorem 2 to show that $\mathsf{QCMA/qpoly} \subseteq \mathsf{QMA/poly}$.

(4) In the full version, we also use our new characterization of $\mathsf{BQP/qpoly}$ to prove a quantum analogue of the Karp-Lipton Theorem [23]. Recall that the Karp-Lipton Theorem says that if $\mathsf{NP} \subset \mathsf{P/poly}$, then the polynomial hierarchy collapses to the second level. Our 'Quantum Karp-Lipton Theorem' says that if $\mathsf{NP} \subset \mathsf{BQP/qpoly}$ (that is, NP-complete problems are efficiently solvable with the help of quantum advice), then $\Pi_2^{\mathsf{P}} \subseteq \mathsf{QMA}^{\mathsf{PromiseQMA}}$. As far as we know, this is the first nontrivial result to derive unlikely consequences from a hypothesis about quantum machines being able to solve NP-complete problems in polynomial time.

Finally, using our result, we are able to explain a previously-mysterious aspect of a 2000 paper of Watrous [31]. Watrous gave the best-known example of a problem in QMA that is not *obviously* in QCMA—that is, for which quantum proofs actually seem to help.[2] This problem is called Group Non-Membership, and is defined as follows: Arthur is given a finite *black-box group* G and a subgroup $H \leq G$ (specified by their generators), as well as an element $x \in G$. His task is to verify that $x \notin H$. It is known that, as a black-box problem, this problem is not in MA. But Watrous showed that Group Non-Membership is in QMA, since Merlin can always persuade Arthur that $x \notin H$ by giving him the following quantum proof:

$$|H\rangle = \frac{1}{\sqrt{|H|}} \sum_{h \in H} |h\rangle .$$

Arthur's verification procedure consists of two tests. In the first test, Arthur *assumes* that Merlin sent $|H\rangle$, and then uses $|H\rangle$ to decide whether $x \in H$. The test is a simple, beautiful illustration of the power of quantum algorithms. The second test in Watrous's protocol confirms that Merlin really sent $|H\rangle$ or at least, a state which is 'equivalent' for purposes of the first test. This second test and its analysis are considerably more involved, and seem less 'natural'.

Using our results, we see that a slightly weaker version of Watrous's result can be derived in an almost automatic way from his first test, as follows. If we assume that the black-box group

[1] We call this restricted class $\mathsf{YQP/poly}$; in another notation it would be $\mathsf{OQMA/poly} \cap \mathsf{coOQMA/poly}$ (where the O stands for 'oblivious').

[2] Aaronson and Kuperberg [6], however, give evidence that this problem might be in QCMA, under conjectures related to the Classification of Finite Simple Groups.

$H = H_n$ is fixed for each input length, then GROUP NON-MEMBERSHIP is in BQP/qpoly, by letting $|H_n\rangle$ as above be the trusted advice for length n and using Watrous' first test as the BQP/qpoly algorithm. Then Theorem 2 (which can be readily adapted to the black-box setting) tells us that Group Non-Membership is in QMA/poly as well.

1.3 Proof Overview

We now give an overview of the proof Theorem 2, that BQP/qpoly $\subseteq$ QMA/poly. As we will explain, our proof rests on a new idea we call the 'majority-certificates' technique, which is not specifically quantum and which seems likely to find other applications.

We begin with a language $L \in$ BQP/qpoly and, for $n > 0$, a poly(n)-size quantum circuit $Q(x, \xi)$ that computes $L(x)$ with high probability when given the 'correct' advice state $\xi = \rho_n$ on poly(n) qubits. The challenge, then, is to force Merlin to supply a witness state ρ' that behaves like ρ_n on every input $x \in \{0,1\}^n$.

Every potential advice state ξ defines a function $f_\xi : \{0,1\}^n \to [0,1]$, by $f_\xi(x) := \Pr[Q(x, \xi) = 1]$. For each such ξ, let $\widehat{f_\xi}(x) := [f_\xi(x) \geq 1/2]$ be the Boolean function obtained by rounding f_ξ. As a simplification, suppose that Merlin is restricted to sending an advice state ξ for which $f_\xi(x) \notin (1/3, 2/3)$: that is, an advice state which renders a 'clear opinion' about every input x. (This simplification helps to explain the main ideas, but does not follow the actual proof.) Let S be the set of all Boolean functions $f : \{0,1\}^n \to \{0,1\}$ that are expressible as $\widehat{f_\xi}$ for some such advice state ξ. Then S includes the 'target function' $f^* := L_n$ (the restriction of L to inputs of length n), as well as a potentially-large number of other functions. However, we claim S is not *too* large: $|S| \leq 2^{\text{poly}(n)}$. This bound on the 'effective information content' of quantum states was derived previously by Aaronson [2, 5], building on the work of Ambainis et al. [12].

One might initially hope that, just by virtue of the size bound on S, we could find some set of poly(n) values

$$(x_1, f^*(x_1)), \ldots, (x_k, f^*(x_t))$$

which *isolate* f^* in S—that is, which differentiate f^* from all other members of S. In that case, the trusted classical advice could simply specify those values, as 'tests' for Arthur to perform on the quantum state sent by Merlin. Alas, this hope is unfounded in general. For consider the case where f^* is the identically-zero function, and S consists of f^* along with the 'point function' f_y (which equals 1 on y and 0 elsewhere), for all $y \in \{0,1\}^n$. Then f^* can only be isolated in S by specifying its value at *every* point!

Luckily, this counterexample leads us to a key observation. Although f is not isolatable in S by a small number of values, each point function f_y *can* be isolated (by its value at y), and moreover, f_y is quite 'close' to f. In fact, if we choose any three distinct strings x, y, z, then $f^* = \text{MAJ}(f_x, f_y, f_z)$. (Of course if f^* were the identically-zero function, it could be easily specified with classical advice! But f^* could have been any function in this example.)

This suggests a new, more indirect approach to our general problem: we try to express f as the pointwise majority vote

$$f^*(x) \equiv \text{MAJ}(f_1(x), \ldots, f_m(x)),$$

of a small number ($m = O(n)$, say) of *other* functions $f_1, \ldots, f_m$ in S, where each f_i is isolatable in S by specifying at most $k = O(\log |S|)$ of its values. Indeed, we will show this can *always* be done. We call this key result the *majority-certificates lemma*; we will say more about its proof and its relation to earlier work in Section 1.4.

With this lemma in hand, we can solve our (artificially simplified) problem: in the QMA/poly protocol for L, we use certificates which isolate $f_1, \ldots, f_m \in S$ as above as the classical advice

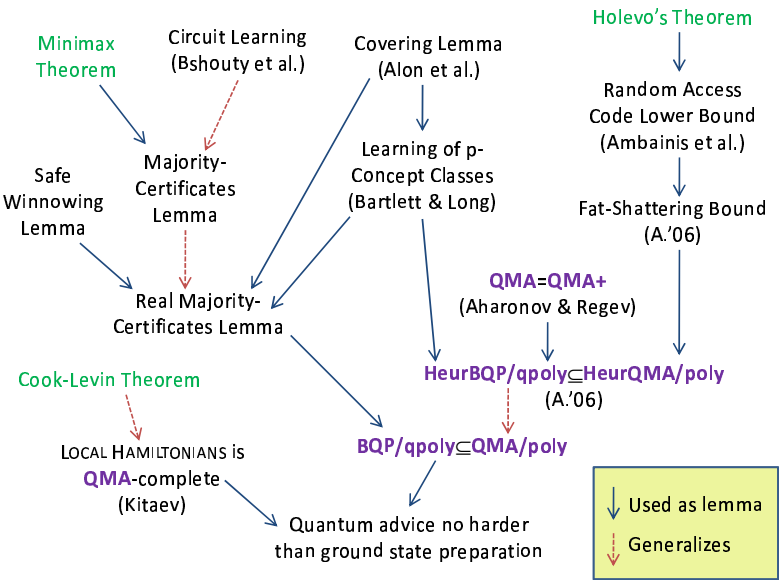

Figure 1: Dependency structure of our proof that quantum advice states can be expressed as ground states of local Hamiltonians.

for Arthur. Arthur requests from Merlin each of the m states $\xi_1, \ldots, \xi_m$ such that $f_i = f_{\xi_i}$, and verifies that he receives appropriate states by checking them against the certificates. This involves multiple measurements of each ξ_i—and an immediate difficulty is that, since measurements are irreversible in quantum mechanics, the process of verifying the witness state might also destroy it. However, we get around this difficulty by appealing to a result of Aharonov and Regev [10]. This result essentially says that a QMA protocol in which Arthur is granted the (physically unrealistic) ability to perform 'non-destructive measurements' on his witness state, can be efficiently simulated by an ordinary QMA protocol.

To build intuition, we will begin (in Section 2) by proving the majority-certificates lemma for Boolean functions, as described above. However, to remove the artificial simplification we made and prove Theorem 2, we will need to generalize the lemma substantially, to a statement about possibly-infinite sets of real-valued functions $f : \{0,1\}^n \rightarrow [0,1]$. In the general version, the hypothesis that S is finite and not too large gets replaced by a more subtle assumption: namely, an upper bound on the so-called *fat-shattering dimension* of S. To prove our generalization, we use powerful results of Alon et al. [11] and Bartlett and Long [13] on the learnability of real-valued functions. We then use a bound on the fat-shattering dimension of real-valued functions defined by quantum states (from Aaronson [5], building on Ambainis et al. [12]). Figure 1 shows the overall dependency structure of the proof.

1.4 Majority-Certificates Lemma in Context

The majority-certificates lemma is closely related to the seminal notion of *boosting* [27] from computational learning theory. Boosting is a broad topic with a vast literature, but a common 'generic' form of the boosting problem is as follows: we want to learn some target function f^*, given sample data of the form $(x, f^*(x))$. We assume we have a *weak learning algorithm* $A^{f^*,\mathcal{D}}$, with the

property that, for any probability distribution $\mathcal{D}$ over inputs x, with high probability A finds a hypothesis $f \in \mathcal{F}$ which predicts $f^*(x)$ 'reasonably well' when $x \sim \mathcal{D}$. The task is to 'boost' this weak learner into a *strong* learner B^{f^*}. The strong learner should output a collection of functions $f_1, \ldots, f_m \in \mathcal{F}$, such that a (possibly-weighted) majority vote over $f_1(x), \ldots, f_m(x)$ predicts $f^*(x)$ 'extremely well.' It turns out [27, 18] that this goal can be achieved in a very general setting.

Our majority-certificates lemma has strengths and weaknesses compared to boosting. Our assumptions are much milder than those of boosting: rather than needing a weak learner, we assume only that the hypothesis class S is 'not too large.' Also, we represent our target function f^* *exactly* by $\mathrm{MAJ}(f_1, \ldots, f_m)$, not just approximately. On the other hand, we do not give an efficient algorithm to *find* our majority-representation. Also, the f_i's are not 'explicitly given': we only give a way to *recognize* each f_i, under the assumption that the function purporting to be f_i is in fact drawn from the original hypothesis class.

The proof of our lemma also has similarities to boosting. As an analogue of a 'weak learner', we show that for every distribution $\mathcal{D}$, there exists a function $f \in S$ which agrees with the target function f^* on most $x \sim \mathcal{D}$, *and* which is isolatable in S by specifying $O(\log |S|)$ queries. Using the Minimax Theorem, we then nonconstructively 'boost' this fact into the desired majority-representation of f^*. We note that Nisan used the Minimax Theorem for boosting in a similar way, in his alternative proof of Impagliazzo's 'hard-core set theorem' (see [21]).

The majority-certificates lemma is also reminiscent of Bshouty et al.'s algorithm [15], for learning small circuits in the complexity class $\mathsf{ZPP}^{\mathsf{NP}}$. Our lemma lacks the algorithmic component of this earlier work, but unlike Bshouty et al., we do not require the functions being learned to come with any succinct labels (such as circuit descriptions).

1.5 Organization of the Paper

In Section 2, we prove the Boolean majority-certificates-lemma. In Section 3, we give our real-valued generalization of this lemma, and in Section 4 we use it to prove Theorem 2, and state some consequences for quantum complexity theory. Theorem 1 is proved in Section 4.3. Section 5 contains some further results for quantum complexity, and the Appendices provide some additional applications of and perspectives on the majority-certificates lemma.

2 The Majority-Certificates Lemma

A *Boolean concept class* is a family of sets $\{S_n\}_{n \geq 1}$, where each S_n consists of Boolean functions $f : \{0,1\}^n \to \{0,1\}$ on n variables. Abusing notation, we will often use S to refer directly to a set of Boolean functions on n variables, with the quantification over n being understood.

By a *certificate*, we mean a partial Boolean function $C : \{0,1\}^n \to \{0,1,*\}$. The *size* of C, denoted $|C|$, is the number of inputs x such that $C(x) \in \{0,1\}$. A Boolean function $f : \{0,1\}^n \to \{0,1\}$ is *consistent* with C if $f(x) = C(x)$ whenever $C(x) \in \{0,1\}$. Given a set S of Boolean functions and a certificate C, let $S[C]$ be the set of all functions $f \in S$ that are consistent with C. Say that a function $f \in S$ is *isolated in* S by the certificate C if $S[C] = \{f\}$.

We now prove a lemma that represents one of the main tools of this paper.

Lemma 3 (Majority-Certificates Lemma) *Let S be a set of Boolean functions $f : \{0,1\}^n \to \{0,1\}$, and let $f^* \in S$. Then there exist $m = O(n)$ certificates $C_1, \ldots, C_m$, each of size $k = O(\log |S|)$, and functions $f_1, \ldots, f_m \in S$, such that*

(i) $S[C_i] = \{f_i\}$ all $i \in [m]$;

(ii) $\text{MAJ}(f_1(x), \ldots, f_m(x)) = f^*(x)$ *for all* $x \in \{0,1\}^n$.

Proof. Our proof of Lemma 3 relies on the following claim.

Claim 4 *Let* $\mathcal{D}$ *be any distribution over inputs* $x \in \{0,1\}^n$. *Then there exists a function* $f \in S$ *such that*

(i) f *is isolatable in* S *by a certificate* C *of size* $k = O(\log |S|)$;

(ii) $\Pr_{x \sim \mathcal{D}}[f(x) \neq f^*(x)] \leq \frac{1}{10}$.

Lemma 3 follows from Claim 4 by a boosting-type argument, as follows. Consider a two-player game where:

- Alice chooses a certificate C of size k that isolates some $f \in S$, and
- Bob simultaneously chooses an input $x \in \{0,1\}^n$.

Alice wins the game if $f(x) = f^*(x)$. Claim 4 tells us that for every mixed strategy of Bob (i.e., distribution $\mathcal{D}$ over inputs), there exists a pure strategy of Alice that succeeds with probability at least 0.9 against $\mathcal{D}$. Then by the Minimax Theorem, there exists a mixed strategy for Alice—that is, a probability distribution $\mathcal{C}$ over certificates—that allows her to win with probability at least 0.9 against *every* pure strategy of Bob.

Now suppose we draw $C_1, \ldots, C_m$ independently from $\mathcal{C}$, isolating functions $f_1, \ldots, f_m$ in S. Fix an input $x \in \{0,1\}^n$; then by the success of Alice's strategy against x, and applying a Chernoff bound,

$$\Pr_{f_1, \ldots, f_m \sim \pi}[\text{MAJ}(f_1(x), \ldots, f_m(x)) \neq f^*(x)] < \frac{1}{2^n},$$

provided we choose $m = O(n)$ suitably. But by the union bound, this means there must be a *fixed* choice of $C_1, \ldots, C_m$ such that $\text{MAJ}(f_1, \ldots, f_m) \equiv f^*(x)$, where each f_i is isolated in S by C_i. This proves Lemma 3, modulo the Claim. ■

Proof of Claim 4. By symmetry, we can assume without loss of generality that f^* is the identically-zero function. Given the mixed strategy $\mathcal{D}$ of Bob, we construct the certificate C as follows. Initially C is empty: that is, $C(x) = *$ for all $x \in \{0,1\}^n$. In the first stage, we draw $t = O(\log |S|)$ inputs $x_1, \ldots, x_t$ independently from $\mathcal{D}$. For any $f : \{0,1\}^n \rightarrow \{0,1\}$, let

$$w_f := \Pr_{x \sim \mathcal{D}}[f(x) = 1].$$

Now suppose f is such that $w_f > 0.1$. Then

$$\Pr_{x_1, \ldots, x_t \sim \mathcal{D}}[f(x_1) = 0 \wedge \cdots \wedge f(x_t) = 0] < 0.9^t \leq \frac{1}{|S|},$$

provided $t \geq \log_{10/9} |S|$. So by the union bound, there must be a *fixed* choice of $x_1, \ldots, x_t$ that kills off every $f \in S$ such that $w_f > 0.1$—that is, such that $f(x_1) = \cdots = f(x_t) = 0$ implies $w_f \leq 0.1$. Fix that $x_1, \ldots, x_t$, and set $C(x_i) := 0$ for all $i \in [t]$.

In the second stage, our goal is just to isolate some *particular* function $f \in S[C]$. We do this recursively as follows. If $|S[C]| = 1$ then we are done. Otherwise, there exists an input x such that $f(x) = 0$ for some $f \in S[C]$ and $f(x) = 1$ for other $f \in S[C]$. If setting $C(x) := 0$ decreases

$|S[C]|$ by at least a factor of 2, then set $C(x) := 0$; otherwise set $C(x) := 1$. Since $S[C]$ can halve in size at most $\log_2 |S|$ times, this procedure terminates after at most $\log_2 |S|$ steps with $|S[C]| = 1$.

The end result is a certificate C of size $O(\log |S|)$, which isolates a function f in S for which $w_f \leq 1/10$. We have therefore found a pure strategy for Alice that fails with probability at most $1/10$ against $\mathcal{D}$, as desired. ■

3 Extension to Real Functions

In this section, we extend the majority-certificates lemma from Boolean functions to real-valued functions $f : \{0,1\}^n \rightarrow [0,1]$. We will need this extension for the application to quantum advice in Section 4. In proving our extension we will have to confront several new difficulties. Firstly, the concept classes S that we want to consider can now contain a *continuum* of functions—so Lemma 3, which assumed that S was finite and constructed certificates of size $O(\log |S|)$, is not going to work. In Section 3.1, we review notions from computational learning theory, including fat-shattering dimension and ε-covers, which (combined with results of Alon et al. [11] and Bartlett and Long [13]) can be used to get around this difficulty. Secondly, it is no longer enough to isolate a function $f_i \in S$ that we are interested in; instead we will need to 'safely' isolate f_i, which roughly speaking means that (i) f_i is consistent with some certificate C, and (ii) any $f \in S$ that is even *approximately* consistent with C is close to f_i. In Section 3.2, we prove a 'safe winnowing lemma' that can be used for this purpose. Finally, in Section 3.3, we put the pieces together to prove a real-valued majority-certificates lemma.

3.1 Background from Learning Theory

A *p-concept class* S is a family of functions $f : \{0,1\}^n \rightarrow [0,1]$ (as usual, quantification over all n is understood). Given functions $f, g : \{0,1\}^n \rightarrow [0,1]$ and a subset of inputs $X \subseteq \{0,1\}^n$, we will be interested in three measures of the distance between f and g restricted to X:

$$\Delta_\infty(f,g)[X] := \max_{x \in X} |f(x) - g(x)|,$$

$$\Delta_2(f,g)[X] := \sqrt{\sum_{x \in X} (f(x) - g(x))^2},$$

$$\Delta_1(f,g)[X] := \sum_{x \in X} |f(x) - g(x)|.$$

For convenience, we define $\Delta_\infty(f,g) := \Delta_\infty(f,g)[\{0,1\}^n]$, and similarly for $\Delta_2(f,g)$ and $\Delta_1(f,g)$. Also, given a distribution $\mathcal{D}$ over $\{0,1\}^n$, define

$$\Delta_1(f,g)\langle\mathcal{D}\rangle := \mathop{\mathrm{E}}_{x \sim \mathcal{D}} [|f(x) - g(x)|].$$

Finally we will need the notions of coverability and fat-shattering dimension.

Definition 5 (Coverability) *Let S be a p-concept class. The subset $C \subseteq S$ is an ε-cover for S if for all $f \in S$, there exists a $g \in C$ such that $\Delta_\infty(f,g) \leq \varepsilon$. We say S is coverable if for all $\varepsilon > 0$, there exists an ε-cover for S of size $2^{\mathrm{poly}(n,1/\varepsilon)}$.*

Definition 6 (Fat-Shattering Dimension) *Let S be a p-concept class and $\varepsilon > 0$ be a real number. We say the set $A \subseteq \{0,1\}^n$ is ε-shattered by S if there exists a function $r : A \rightarrow [0,1]$ such that for all $2^{|A|}$ Boolean functions $g : A \rightarrow \{0,1\}$, there exists a p-concept $f \in S$ such that for all $x \in A$, we have $f(x) \leq r(x) - \varepsilon$ whenever $g(x) = 0$ and $f(x) \geq r(x) + \varepsilon$ whenever $g(x) = 1$. Then the ε-fat-shattering dimension of S, or $\mathrm{fat}_\varepsilon(S)$, is the size of the largest set ε-shattered by S. We say S is bounded-dimensional if $\mathrm{fat}_\varepsilon(S) \leq \mathrm{poly}(n, 1/\varepsilon)$ for all $\varepsilon > 0$.*

The following central result was shown by Alon et al. [11] (see also [22]).

Theorem 7 ([11]) *Every p-concept class S has an ε-cover of size $\exp\left[O\left((n + \log 1/\varepsilon)\,\mathrm{fat}_{\varepsilon/4}(S)\right)\right]$. So in particular, if S is bounded-dimensional then S is coverable.*

Building on the work of Alon et al. [11], Bartlett and Long [13] then proved the following:

Theorem 8 ([13]) *Let S be a p-concept class and $\mathcal{D}$ be a distribution over $\{0,1\}^n$. Fix an $f : \{0,1\}^n \rightarrow [0,1]$ (not necessarily in S) and an error parameter $\alpha > 0$. Suppose we form a set $X \subseteq \{0,1\}^n$ by choosing m inputs independently with replacement from $\mathcal{D}$. Then there exists a positive constant K such that, with probability at least $1 - \delta$ over X, any hypothesis $h \in S$ that minimizes $\Delta_1(h, f)[X]$ also satisfies*

$$\Delta_1(h, f)\langle\mathcal{D}\rangle \leq \alpha + \inf_{g \in S} \Delta_1(g, f)\langle\mathcal{D}\rangle,$$

provided

$$m \geq \frac{K}{\alpha^2}\left(\mathrm{fat}_{\alpha/5}(S)\log^2\frac{1}{\alpha} + \log\frac{1}{\delta}\right).$$

Theorem 8 has the following corollary, which is similar to Corollary 2.4 of Aaronson [5], but more directly suited to our purposes here.[3]

Corollary 9 *Let S be a p-concept class and $\mathcal{D}$ be a distribution over $\{0,1\}^n$. Fix an $f \in S$ and an error parameter $\varepsilon > 0$. Suppose we form a set $X \subseteq \{0,1\}^n$ by choosing m inputs independently with replacement from $\mathcal{D}$. Then there exists a positive constant K such that, with probability at least $1-\delta$ over X, any hypothesis $h \in S$ that satisfies $\Delta_\infty(h, f)[X] \leq \varepsilon$ also satisfies $\Delta_1(h, f)\langle\mathcal{D}\rangle \leq 11\varepsilon$, provided*

$$m \geq \frac{K}{\varepsilon^2}\left(\mathrm{fat}_\varepsilon(S)\log^2\frac{1}{\varepsilon} + \log\frac{1}{\delta}\right).$$

Proof. Let S^* be the p-concept class consisting of all functions $g : \{0,1\}^n \rightarrow [0,1]$ for which there exists an $f \in S$ such that $\Delta_\infty(g, f) \leq \varepsilon$. Fix an $f \in S$ and a distribution $\mathcal{D}$, and let X be chosen as in the statement of the corollary. Suppose we choose a hypothesis $h \in S$ such that $\Delta_\infty(h, f)[X] \leq \varepsilon$. Then there exists a function $g \in S^*$ such that $g(x) = h(x)$ for all $x \in X$. This g is simply obtained by setting $g(x) := h(x)$ if $x \in X$ and $g(x) := f(x)$ otherwise. In particular, note that $\Delta_1(h, g)[X] = 0$, which means that h minimizes the functional $\Delta_1(h, g)[X]$ over all hypotheses in S (and indeed in S^*). By Theorem 8, this implies that with probability at least $1 - \delta$ over X,

$$\Delta_1(h, g)\langle\mathcal{D}\rangle \leq \alpha + \inf_{u \in S^*} \Delta_1(u, g)\langle\mathcal{D}\rangle = \alpha$$

[3] It would also be possible to apply the bound from [5] 'off-the-shelf,' but at the cost of a worse dependence on $1/\varepsilon$.

for all $\alpha > 0$, provided we take

$$m \geq \frac{K}{\alpha^2}\left(\mathrm{fat}_{\alpha/5}\left(S^*\right)\log^2\frac{1}{\alpha}+\log\frac{1}{\delta}\right).$$

Here we have used the fact that $g \in S^*$, and hence

$$\inf_{u\in S^*}\Delta_1\left(u,g\right)\langle\mathcal{D}\rangle = 0.$$

So by the triangle inequality,

$$\begin{aligned}\Delta_1\left(h,f\right)\langle\mathcal{D}\rangle &\leq \Delta_1\left(h,g\right)\langle\mathcal{D}\rangle + \Delta_1\left(g,f\right)\langle\mathcal{D}\rangle \\ &\leq \alpha + \Delta_\infty\left(g,f\right) \\ &\leq \alpha+\varepsilon.\end{aligned}$$

Next, we claim that $\mathrm{fat}_{\alpha/5}\left(S^*\right) \leq \mathrm{fat}_{\alpha/5-\varepsilon}\left(S\right)$. The reason is simply that, if a given set β-fat-shatters S^*, then it must also $(\beta-\varepsilon)$-fat-shatter S by the triangle inequality. Setting $\alpha := 10\varepsilon$ now yields the desired statement. ■

3.2 The Safe Winnowing Lemma

To prove the real-valued majority-certificates lemma, the first step is to prove a so-called 'safe winnowing lemma.' This lemma says intuitively that, given any set S of real-valued functions with a small ε-cover (or equivalently, with polynomially-bounded fat-shattering dimension), it is possible to find a set of $k = \mathrm{poly}\left(n\right)$ constraints $\left|f\left(x_1\right)-a_1\right| \leq \epsilon, \ldots, \left|f\left(x_k\right)-a_k\right| \leq \epsilon$ that are *essentially* compatible with one and only one function $f \in S$. Here 'essentially' means that (i) any function that satisfies the constraints is close to f in L_∞-norm, and (ii) f itself not only satisfies the constraints, but does so with a 'margin to spare.'

Lemma 10 (Safe Winnowing Lemma) *Let S be a set of functions $f : \{0,1\}^n \rightarrow [0,1]$. Fix a function $f^* \in S$ and subset $Y \subseteq \{0,1\}^n$. For some parameter $\varepsilon > 0$, let C be a finite ε-cover for S. Then there exists an $f \in S$, as well as a subset $Z \subseteq \{0,1\}^n$ of size at most $k = \log_2|C|$, such that:*

(i) Every $g \in S$ that satisfies $\Delta_\infty\left(f,g\right)\left[Y\cup Z\right] \leq \frac{\varepsilon}{5k}$ also satisfies $\Delta_\infty\left(f,g\right) \leq 3\varepsilon$.

(ii) $\Delta_\infty\left(f,f^*\right)\left[Y\right] \leq \varepsilon/5$.

Proof. Let $\delta := \frac{\varepsilon}{5k}$. We construct (f,Z) by an iterative procedure. Initially let $S_0 := S$, let $f_0 := f^*$, and let $Z_0 := Y$. We will form new sets $S_1, S_2, \ldots$ by repeatedly adding constraints of the form $f\left(x\right) \leq \alpha$ or $f\left(x\right) \geq \alpha$ for various x, α, maintaining the invariant that $f_t \in S_t$. At iteration t, suppose there exists a function $g \in S_{t-1}$ such that $\Delta_\infty\left(f_{t-1},g\right)\left[Y\cup Z_{t-1}\right] \leq \delta$, but nevertheless $\left|f_{t-1}\left(z_t\right)-g\left(z_t\right)\right| > 3\varepsilon$ for some input z_t. Then first set $Z_t := Z_{t-1}\cup\{z_t\}$ (i.e., add z_t into our set of inputs, if it is not already there). Let $v := \frac{1}{2}\left[f_{t-1}\left(z_t\right)+g\left(z_t\right)\right]$, let A be the set of all functions $h \in S_{t-1}$ such that $h\left(z_t\right) < v$, and let B be the set of all $h \in S_{t-1}$ such that $h\left(z_t\right) \geq v$. Also, for any given set M, let $M^\Diamond := M \cap C$. Then clearly $\min\left\{\left|A^\Diamond\right|,\left|B^\Diamond\right|\right\} \leq \left|S_{t-1}^\Diamond\right|/2$. If $\left|A^\Diamond\right| < \left|B^\Diamond\right|$, then set $S_t := A$; otherwise set $S_t := B$. Then set $f_t := f_{t-1}$ if $f_{t-1} \in S_t$ and $f_t := g$ otherwise.

Since $\left|S_t^{\Diamond}\right|$ can halve at most $k = \log_2 |C|$ times, it is clear that after $T \leq k$ iterations we have $\left|S_T^{\Diamond}\right| \leq 1$. Set $f := f_T$ and $Z := Z_T$. Then by the triangle inequality,

$$\Delta_{\infty}(f, f^*)[Y] \leq T\delta \leq \frac{\varepsilon}{5},$$

and also

$$|f(z_t) - f_t(z_t)| \leq (T-t)\,\delta < \frac{\varepsilon}{5}$$

for all $t \in [T]$. So suppose by contradiction that there still exists a function $g \in S_T$ such that $\Delta_{\infty}(f, g)[Y \cup Z] \leq \delta$ but $|f(x) - g(x)| > 3\varepsilon$ for some x, and consider functions $p, q \in C$ in the cover such that $\Delta_{\infty}(f, p) \leq \varepsilon$ and $\Delta_{\infty}(g, q) \leq \varepsilon$. Then $p, q \in S_T^{\Diamond}$ but $p \neq q$, which contradicts the fact that $\left|S_T^{\Diamond}\right| \leq 1$. Also notice that for all $g \in S$, if $\Delta_{\infty}(f, g)[Y \cup Z] \leq \delta$ then $g \in S_T$. Thus $\Delta_{\infty}(f, g)[Y \cup Z] \leq \delta$ implies $\Delta_{\infty}(f, g) \leq 3\varepsilon$ as desired. ■

Note that Lemma 10 is still interesting in the special case $Y = \varnothing$, in which case f^* is irrelevant, and the problem reduces to finding a Z such that every $g \in S$ that satisfies $\Delta_{\infty}(f, g)[Z] \leq \frac{\varepsilon}{5k}$ also satisfies $\Delta_{\infty}(f, g) \leq 3\varepsilon$. In Appendix 10, we will develop the theory of 'winnowability' of p-concept classes for its own sake. We show there that the condition $\Delta_{\infty}(f, g)[Z] = O(\varepsilon/k)$ can be improved to $\Delta_1(f, g)[Z] = O(\varepsilon)$. On the other hand, the proof becomes more involved, and we no longer know how to incorporate f^* and Y. We also show that the condition $\Delta_{\infty}(f, g)[Z] = O(\varepsilon/k)$ *cannot* be improved to $\Delta_{\infty}(f, g)[Z] = O(\varepsilon)$ or even $\Delta_2(f, g)[Z] = O(\varepsilon)$.

3.3 The Real-Valued Majority-Certificates Lemma

We are finally ready to generalize Lemma 3 to the case of real-valued functions.

Lemma 11 (Real Majority-Certificates) *Let S be a p-concept class, let $f^* \in S$, and let $\varepsilon > 0$. Then for some $m = O\left(n/\varepsilon^2\right)$, there exist functions $f_1, \ldots, f_m \in S$, sets $X_1, \ldots, X_m \subseteq \{0,1\}^n$ each of size $k = O\left(\left(n + \frac{\log^2 1/\varepsilon}{\varepsilon^2}\right)\mathrm{fat}_{\varepsilon/48}(S)\right)$, and an $\alpha = \Omega\left(\frac{\varepsilon}{(n+\log 1/\varepsilon)\,\mathrm{fat}_{\varepsilon/48}(S)}\right)$ for which the following holds. All $g_1, \ldots, g_m \in S$ that satisfy $\Delta_{\infty}(f_i, g_i)[X_i] \leq \alpha$ for $i \in [m]$ also satisfy $\Delta_{\infty}(f^*, g) \leq \varepsilon$, where*

$$g(x) := \frac{g_1(x) + \cdots + g_m(x)}{m}.$$

Proof. Let

$$\begin{aligned}
\beta &:= \frac{\varepsilon}{48}, \\
t &:= C\left(n + \log\frac{1}{\beta}\right)\mathrm{fat}_{\beta}(S), \\
\alpha &:= \frac{0.4\beta}{t},
\end{aligned}$$

where C is a suitably large constant. Also, let S_{fin} be a finite α-cover for S: that is, a finite subset $S_{\mathrm{fin}} \subseteq S$ such that for all $f \in S$, there exists a $g \in S_{\mathrm{fin}}$ such that $\Delta_{\infty}(f, g) \leq \alpha$.[4] Given f and X, let $S[f, X]$ be the set of all $g \in S$ such that $\Delta_{\infty}(f, g)[X] \leq \alpha$.

[4] We will need S_{fin} for the technical reason that the basic Minimax Theorem only works with finite strategy spaces.

Now consider a two-player game where Alice chooses a function $f \in S_{\text{fin}}$ and a set $X \subseteq \{0,1\}^n$ of size k, and Bob simultaneously chooses an input $x \in \{0,1\}^n$. Alice's *penalty* in this game (the number she is trying to minimize) equals

$$\sup_{g \in S[f,X]} |f^*(x) - g(x)|.$$

We claim that there exists a mixed strategy for Alice—that is, a probability distribution $\mathcal{P}$ over (f,X) pairs—that gives her an expected penalty of at most $\varepsilon/2$ against every pure strategy of Bob.

Let us see why the lemma follows from this claim. Fix an input $x \in \{0,1\}^n$, and suppose Alice draws $(f_1,X_1),\ldots,(f_m,X_m)$ independently from $\mathcal{P}$. Then for all $i \in [m]$,

$$\mathop{\mathrm{E}}_{(f_i,X_i)\sim\mathcal{P}}\left[\sup_{g \in S[f,X]} |f^*(x) - g(x)|\right] \leq \frac{\varepsilon}{2}.$$

Thus, letting $z_1,\ldots,z_m$ be independent random variables in $[0,1]$, each with expectation at most $\varepsilon/2$, the expression

$$\Pr_{(f_1,X_1),\ldots,(f_m,X_m)\sim\mathcal{P}}\left[\exists g_1 \in S[f_1,X_1],\ldots,g_m \in S[f_m,X_m] \;:\; \left|f^*(x) - \frac{g_1(x)+\cdots+g_m(x)}{m}\right| > \varepsilon\right]$$

is at most $\Pr[z_1+\cdots z_m > \varepsilon m]$ by the triangle inequality. This, in turn, is less than

$$2\exp\left(-\frac{2(\varepsilon m)^2}{m}\right) < 2^{-n}$$

by Hoeffding's inequality, provided we choose $m = O\left(n/\varepsilon^2\right)$ suitably. By the union bound, this means that there must be a fixed choice of $f_1,\ldots,f_m$ and $X_1,\ldots,X_m$ such that

$$\left|f^*(x) - \frac{g_1(x)+\cdots+g_m(x)}{m}\right| \leq \varepsilon$$

for all $g_1 \in S[f_1,X_1],\ldots,g_m \in S[f_m,X_m]$ and all inputs $x \in \{0,1\}^n$ simultaneously, as desired.

We now prove the claim. By the Minimax Theorem, our task is equivalent to the following: given any mixed strategy $\mathcal{D}$ of Bob, find a *pure* strategy of Alice that achieves a penalty of at most $\varepsilon/2$ against $\mathcal{D}$. In other words, given any distribution $\mathcal{D}$ over inputs $x \in \{0,1\}^n$, we want a fixed function $f \in S_{\text{fin}}$, and a set $X \subseteq \{0,1\}^n$ of size k, such that

$$\mathop{\mathrm{E}}_{x\sim\mathcal{D}}\left[\sup_{g \in S[f,X]} |f^*(x) - g(x)|\right] \leq \frac{\varepsilon}{2}.$$

We construct this (f,X) pair as follows.

In the first stage, we let Y be a set, of size at most

$$M := \frac{K}{\beta^2}\left(\text{fat}_\beta(S)\log^2\frac{1}{\beta} + \log\frac{1}{\delta}\right),$$

formed by choosing M inputs independently with replacement from $\mathcal{D}$. Here $\beta = \varepsilon/48$ as defined earlier, $\delta = 1/2$, and K is the constant from Corollary 9. Then by Corollary 9, with probability at least $1-\delta = 1/2$ over the choice of Y, any $g \in S$ that satisfies $\Delta_\infty(f^*,g)[Y] \leq \beta$ also satisfies

$\Delta_1(f^*, g)\langle\mathcal{D}\rangle \leq 11\beta$. So there must be a *fixed* choice of Y with that property. Fix that Y, and let S' be the set of all $g \in S$ such that $\Delta_\infty(f^*, g)[Y] \leq \beta$.

In the second stage, our goal is just to winnow S' down to a particular function f. More precisely, we want to find an $f \in S' \cap S_{\text{fin}}$, and a set $X \subseteq \{0,1\}^n$ containing Y, such that any $g \in S$ that satisfies $\Delta_\infty(f, g)[X] \leq \alpha$ also satisfies $\Delta_\infty(f, g) \leq 11\beta$.

We find this (f, X) pair as follows. By Theorem 7, the class S' has a 4β-cover of size

$$N = \exp\left[O\left(\left(n + \log\frac{1}{4\beta}\right)\text{fat}_\beta(S')\right)\right] \leq \exp\left[O\left(\left(n + \log\frac{1}{\beta}\right)\text{fat}_\beta(S)\right)\right].$$

Let $t := \log_2 N$. Then by Lemma 10, there exists a function $u \in S'$, as well as a subset $Z \subseteq \{0,1\}^n$ of size at most t, such that:

(i) $\Delta_\infty(u, f^*)[Y] \leq 0.8\beta$.

(ii) Every $g \in S'$ that satisfies $\Delta_\infty(u, g)[Y \cup Z] \leq \frac{0.8\beta}{t}$ also satisfies $\Delta_\infty(u, g) \leq 12\beta$.

Let $X := Y \cup Z$, and observe that

$$\begin{aligned}|X| &= O\left(\frac{1}{\beta^2}\text{fat}_\beta(S)\log^2\frac{1}{\beta} + \left(n + \log\frac{1}{\beta}\right)\text{fat}_\beta(S)\right)\\ &= O\left(\left(n + \frac{\log^2 1/\varepsilon}{\varepsilon^2}\right)\text{fat}_{\varepsilon/48}(S)\right)\end{aligned}$$

as desired. Now let f be a function in S_{fin} such that $\Delta_\infty(f, u) \leq \alpha$. Let us check that f has the properties we want. First,

$$\begin{aligned}\Delta_\infty(f^*, f)[Y] &\leq \Delta_\infty(f^*, u)[Y] + \Delta_\infty(u, f)[Y]\\ &\leq 0.8\beta + \alpha\\ &< 0.9\beta,\end{aligned}$$

hence $f \in S'$ as desired. Next, any $g \in S$ that satisfies $\Delta_\infty(f, g)[X] \leq \alpha$ also satisfies

$$\begin{aligned}\Delta_\infty(f^*, g)[Y] &\leq \Delta_\infty(f^*, f)[Y] + \Delta_\infty(f, g)[Y]\\ &\leq 0.9\beta + \alpha\\ &< \beta,\end{aligned}$$

hence $g \in S'$, hence $\Delta_1(f^*, g)\langle\mathcal{D}\rangle \leq 11\beta$. So any $g \in S$ that satisfies $\Delta_\infty(f, g)[X] \leq \alpha$ satisfies

$$\begin{aligned}\Delta_\infty(u, g)[Z] &\leq \Delta_\infty(u, f)[Z] + \Delta_\infty(f, g)[Z]\\ &\leq 2\alpha\\ &= \frac{0.8\beta}{t},\end{aligned}$$

hence $\Delta_\infty(u, g) \leq 12\beta$ (since such a g must belong to S'), hence

$$\begin{aligned}\Delta_\infty(f, g) &\leq \Delta_\infty(f, u) + \Delta_\infty(u, g)\\ &\leq \alpha + 12\beta\\ &\leq 13\beta.\end{aligned}$$

To conclude,

$$\begin{aligned}\mathop{\mathrm{E}}_{x\sim\mathcal{D}}\left[\sup_{g\in S[f,X]}\left|f^*(x)-g(x)\right|\right] &\leq \Delta_1\left(f^*,f\right)\langle\mathcal{D}\rangle+\sup_{g\in S[f,X]}\Delta_\infty\left(f,g\right)\\ &\leq 11\beta+13\beta\\ &=\frac{\varepsilon}{2}\end{aligned}$$

as desired. This proves the claim and hence the lemma. ■

4 Application to Quantum Advice

In this section, we use the real-valued majority-certificates lemma to prove Theorems 1 and 2, as well as several other results.

4.1 Bestiary of Quantum Complexity Classes

Given a language $L\subseteq\{0,1\}^*$, let $L:\{0,1\}^*\rightarrow\{0,1\}$ be the characteristic function of L. We now give a formal definition of the class $\mathsf{BQP/qpoly}$.

Definition 12 *A language L is in $\mathsf{BQP/qpoly}$ if there exists a polynomial-time quantum algorithm A and polynomial p such that for all n, there exists an advice state ρ_n on $p(n)$ qubits such that $A(x,\rho_n)$ outputs $L(x)$ with probability $\geq 2/3$ for all $x\in\{0,1\}^n$.*

Closely related to quantum advice are *quantum proofs*. We now recall the definition of QMA (Quantum Merlin-Arthur), a quantum version of NP.

Definition 13 *A language L is in QMA if there exists a polynomial-time quantum algorithm A and polynomial p such that for all $x\in\{0,1\}^n$:*

(i) If $x\in L$ then there exists a witness ρ_x on $p(n)$ qubits such that $A(x,\rho_x)$ accepts with probability $\geq 2/3$.

(ii) If $x\notin L$ then $A(x,\rho)$ accepts with probability $\leq 1/3$ for all ρ.

We will actually need a generalization of QMA, which was called QMA_+ by Aharonov and Regev [9].[5]

Definition 14 *A language L is in QMA_+ if there exists a polynomial-time algorithm A, which takes $x\in\{0,1\}^n$ as input and produces quantum circuits $C_{x,1},\ldots,C_{x,m}$ and rational numbers $r_{x,1},\ldots,r_{x,m}$ as output, as well as polynomials p,q such that for all $x\in\{0,1\}^n$:*

(i) If $x\in L$ then there exists a witness ρ_x on $p(n)$ qubits such that $\left|\Pr\left[C_{x,i}(\rho_x)\ \ accepts\right]-r_{x,i}\right|\leq 1/q(n)$ for all $i\in[m]$.

(ii) If $x\notin L$ then for all ρ, there exists an $i\in[m]$ such that $\left|\Pr\left[C_{x,i}(\rho)\ \ accepts\right]-r_{x,i}\right|>5/q(n)$.

[5] Aharonov and Regev actually defined QMA_+ in a slightly more general way. However, the definition below is all we need; note that all these classes turn out to equal QMA anyway.

Aharonov and Regev [9] made the following extremely useful observation, which we prove for completeness.

Theorem 15 ([10]) $\mathsf{QMA}_+ = \mathsf{QMA}$.

Proof. $\mathsf{QMA} \subseteq \mathsf{QMA}_+$ is obvious. For the other direction, let $L \in \mathsf{QMA}_+$, and fix an input $x \in \{0,1\}^n$, quantum circuits $C_{x,1}, \ldots, C_{x,m}$, rational numbers $r_{x,1}, \ldots, r_{x,m}$, and polynomials p, q. Then consider the following QMA verification procedure. Given a witness state σ on $K = O\left(q(n)^2 \log q(n)\right)$ registers:

(1) Choose $i \in [m]$ uniformly at random.

(2) For $k := 1$ to K, apply $C_{x,i}$ to the k^{th} register of σ.

(3) If the fraction α of invocations that accepted satisfies $|\alpha - r_{x,i}| \leq 2/q(n)$, then accept. Otherwise reject.

Let $P_i(\sigma)$ be the probability that the above procedure accepts, conditioned on choosing $i \in [m]$ in step (i).

Completeness is easy: an honest Merlin can send Arthur the product state $\rho_x^{\otimes K}$. Then provided we take K sufficiently large, $P_i\left(\rho_x^{\otimes K}\right) < 1/q(n)^2$ for all $i \in [m]$ by a Chernoff bound.

In the soundness case, suppose by way of contradiction that there exists a state σ such that

$$\mathop{\mathrm{E}}_{i \in [m]} \left[P_i(\sigma)\right] < \frac{2}{q(n)^2}.$$

Then by Markov's inequality, $P_i(\sigma) < 2/q(n)$ for each *particular* $i \in [m]$. Now let σ_k be the k^{th} register of σ, and let $\rho := \frac{1}{K}(\sigma_1 + \cdots + \sigma_K)$. Then by linearity of expectation,

$$\begin{aligned}
\left|\Pr\left[C_{x,i}(\rho) \text{ accepts}\right] - r_{x,i}\right| &= \left|\mathop{\mathrm{E}}_{k \in [K]} \left[\Pr\left[C_{x,i}(\sigma_k) \text{ accepts}\right]\right] - r_{x,i}\right| \\
&\leq \frac{2}{q(n)} + P_i(\sigma) \\
&\leq \frac{4}{q(n)},
\end{aligned}$$

which contradicts the assumption that there exists an $i \in [m]$ such that

$$\left|\Pr\left[C_{x,i}(\rho) \text{ accepts}\right] - r_{x,i}\right| > \frac{5}{q(n)}.$$

The theorem now reduces to the standard fact that QMA protocols can be amplified to any desired $1/\operatorname{poly}(n)$ soundness gap. ■

To state our results, it will be helpful to have the further notion of *untrusted advice*, which is like advice in that it depends only on the input length n, but like a witness in that it cannot be trusted. This notion has been studied before: Chakaravarthy and Roy [16] and Fortnow and Santhanam [17] defined the complexity class ONP ('Oblivious NP'), which is like NP except that the witness can depend only on the input length. Independently, Aaronson [5] defined the complexity class

YP,[6] which is easily seen to equal $\mathsf{ONP} \cap \mathsf{coONP}$. We will adopt the 'Y' notation in this paper, because it is much easier to write $\mathsf{YQP/poly}$ (for example) than $\mathsf{OQMA/poly} \cap \mathsf{coOQMA/poly}$.

We now give a formal definition of YP, as well as a slight variant called YP^*.

Definition 16 *A language L is in YP if there exist polynomial-time algorithms A, B and a polynomial p such that:*

(i) For all n, there exists an advice string $y_n \in \{0,1\}^{p(n)}$ such that $A(x, y_n) = 1$ for all $x \in \{0,1\}^n$.

(ii) If $A(x, y) = 1$, then $B(x, y) = L(x)$.

L is in YP^* if moreover A ignores x, depending only on y.

Clearly $\mathsf{P} \subseteq \mathsf{YP}^* \subseteq \mathsf{YP} \subseteq \mathsf{P/poly} \cap \mathsf{NP} \cap \mathsf{coNP}$. Also, Aaronson [5] showed that $\mathsf{ZPP} \subseteq \mathsf{YP}$. We will be interested in the natural quantum analogues of YP and YP^*:

Definition 17 *A language L is in YQP if there exist polynomial-time quantum algorithms A, B and a polynomial p such that:*

(i) For all n, there exists an advice state ρ_n on $p(n)$ qubits such that $A(x, \rho_n)$ accepts with probability $\geq 2/3$ for all $x \in \{0,1\}^n$.

(ii) If $A(x, \rho)$ accepts with probability $\geq 1/3$, then $B(x, \rho)$ outputs $L(x)$ with probability $\geq 2/3$.

L is in YQP^* if moreover A ignores x, depending only on ρ.

Clearly $\mathsf{BQP} \subseteq \mathsf{YQP}^* \subseteq \mathsf{YQP} \subseteq \mathsf{BQP/qpoly} \cap \mathsf{QMA} \cap \mathsf{coQMA}$. By direct analogy to QMA_+, we can define the following generalizations of YQP and YQP^*:

Definition 18 *A language L is in YQP_+ if there exists a polynomial-time algorithm A, which takes $x \in \{0,1\}^n$ as input and produces quantum circuits $C_{x,1}, \ldots, C_{x,m}$ and rational numbers $r_{x,1}, \ldots, r_{x,m}$ as output; a polynomial-time quantum algorithm B; and polynomials p, q such that:*

(i) For all n, there exists an advice state ρ_n on $p(n)$ qubits such that $|\Pr[C_{x,i}(\rho_n) \text{ accepts}] - r_{x,i}| \leq 1/q(n)$ for all $i \in [m]$ and $x \in \{0,1\}^n$.

(ii) If $|\Pr[C_{x,i}(\rho) \text{ accepts}] - r_{x,i}| \leq 5/q(n)$ for all $i \in [m]$, then $B(x, \rho)$ outputs $L(x)$ with probability $\geq 2/3$.

L is in YQP^*_+ if moreover A ignores x.

Then we have the following direct counterpart to Theorem 15:

Theorem 19 $\mathsf{YQP}_+ = \mathsf{YQP}$ *and* $\mathsf{YQP}^*_+ = \mathsf{YQP}^*$.

[6] YP stands for 'Yoda Polynomial-Time,' a nomenclature that seems to make neither more nor less sense than 'Arthur-Merlin.'

Proof. For $\mathsf{YQP} \subseteq \mathsf{YQP}_+$ and $\mathsf{YQP}^* \subseteq \mathsf{YQP}^*_+$, we simply take $m = 1$ and take $q(n)$ to be a constant. For $\mathsf{YQP}_+ \subseteq \mathsf{YQP}$, the simulation procedure is essentially the same as in the proof of Theorem 15. Namely, let $L \in \mathsf{YQP}_+$, and fix an input $x \in \{0,1\}^n$, quantum circuits $C_{x,1}, \ldots, C_{x,m}$ generated by an algorithm A, rational numbers $r_{x,1}, \ldots, r_{x,m}$, polynomials p, q, and an algorithm B. Then given a witness state σ on $K = O\left(q(n)^2 \log q(n)\right)$ registers, the YQP algorithm A' does the following:

(1) Choose $i \in [m]$ uniformly at random.

(2) For $k := 1$ to K, apply $C_{x,i}$ to the k^{th} register of σ.

(3) If the fraction α of invocations that accepted satisfies $|\alpha - r_{x,i}| \leq 2/q(n)$, then accept. Otherwise reject.

Likewise, let σ_k be the k^{th} register of σ. Then the YQP algorithm B' chooses $k \in [K]$ uniformly at random, runs $B(x, \sigma_k)$, and outputs the result. One can check that conditions (i) and (ii) in the definition of YQP are both satisfied, albeit with $1 - 1/q(n)^2$ and $1 - 2/q(n)^2$ in place of $2/3$ and $1/3$ (which is not an important difference, because of amplification). The proof of $\mathsf{YQP}^*_+ \subseteq \mathsf{YQP}^*$ is the same, except that both A and A' ignore the input x when generating the $C_{x,i}$'s and $r_{x,i}$'s. ■

4.2 Characterizing Quantum Advice

Fix a polynomial-size quantum circuit Q. For a given advice state ρ, let $f_\rho(x) := \Pr[Q \text{ accepts } x, \rho]$. Let S be the p-concept class consisting of f_ρ for all $p(n)$-qubit mixed states ρ. Then Aaronson [5] proved the following.

Theorem 20 ([5]) $\mathrm{fat}_\gamma(S) = O\left(p(n)/\gamma^2\right)$.

We now prove the following characterization of $\mathsf{BQP/qpoly}$, which immediately implies (and strengthens) Theorem 2:

Theorem 21 $\mathsf{BQP/qpoly} = \mathsf{YQP/poly}$.

Proof. One direction ($\mathsf{YQP/poly} \subseteq \mathsf{BQP/qpoly}$) is obvious, since untrusted quantum advice and trusted classical advice can both be simulated by trusted quantum advice. We prove that $\mathsf{BQP/qpoly} \subseteq \mathsf{YQP/poly}$. It suffices to show that $\mathsf{BQP/qpoly} \subseteq \mathsf{YQP}_+\mathsf{/poly}$, since $\mathsf{YQP} = \mathsf{YQP}_+$ by Theorem 19. Let $L \in \mathsf{BQP/qpoly}$, let Q be a quantum algorithm that decides L with completeness and soundness errors $1/5$, and let $x \in \{0,1\}^n$ be the input. Also, let $f_\xi(z) := \Pr[Q(z, \xi) \text{ accepts}]$, where ξ is a $p(n)$-qubit quantum advice state for Q. Then by definition, there exists a 'true' advice state ρ_n such that

$$|f_{\rho_n}(z) - L(z)| \leq 0.2$$

for all $z \in \{0,1\}^n$. Let S be the p-concept class consisting of f_ξ for all $p(n)$-qubit mixed states ξ. Then Theorem 20 implies that $\mathrm{fat}_\gamma(S) = O\left(p(n)/\gamma^2\right)$ for all $\gamma > 0$. Set $\gamma := 1/480$. Then by Lemma 11, for some $m = O(n)$, there exist $p(n)$-qubit mixed states $\rho[1], \ldots, \rho[m]$, sets $X_1, \ldots, X_m \subseteq \{0,1\}^n$ each of size $k = O(n \cdot p(n))$, and an $\alpha = \Omega\left(\frac{1}{n \cdot p(n)}\right)$ for which the following holds:

(*) *All $p(n)$-qubit states $\sigma[1], \ldots, \sigma[m]$ that satisfy $\Delta_\infty\left(f_{\rho[i]}, f_{\sigma[i]}\right)[X_i] \leq 5\alpha$ for $i \in [m]$ also satisfy $\Delta_\infty(f_{\rho_n}, f_\sigma) \leq 0.1$, where $\sigma := \frac{1}{m}(\sigma[1] + \cdots + \sigma[m])$.*

Our $\mathsf{YQP}_+/\mathsf{poly}$ simulation is now the following. The classical /poly advice encodes the sets $X_1, \ldots, X_m$, as well as a rational approximation $r_{i,z}$ to $f_{\rho[i]}(z)$ for each $i \in [m]$ and $z \in X_i$. The untrusted quantum advice ρ'_n consists of m registers of $p(n)$ qubits each; in the honest case, ρ'_n is simply $\rho[1] \otimes \cdots \otimes \rho[m]$. Let $\sigma[i]$ be the i^{th} register of ρ'_n. Then given the advice, the YQP_+ machine A outputs a circuit $C_{i,z}$ that runs $Q(z, \sigma[i])$ and outputs the result, for each $i \in [m]$ and $z \in X_i$. The machine B chooses $i \in [m]$ uniformly at random, then runs $Q(x, \sigma[i])$ and outputs the result.

We are interested in the difference between $\Pr[C_{i,z}(\rho'_n) \text{ accepts}]$ and $r_{i,z}$. In the honest case,

$$\Pr\left[C_{i,z}\left(\rho'_n\right) \text{ accepts}\right] = \Pr[Q(z, \rho[i]) \text{ accepts}] = f_{\rho[i]}(z)$$

for all i, z. Moreover, we can easily arrange each $r_{i,z}$ to be within α of $f_{\rho[i]}(z)$, by using $O(\log n)$ bits to specify each $r_{i,z}$. For the soundness case, suppose

$$\left|\Pr\left[C_{i,z}\left(\rho'_n\right) \text{ accepts}\right] - r_{i,z}\right| \leq 5\alpha$$

for all $i \in [m]$ and $z \in X_i$. Then by (*), we have $\Delta_\infty(f_{\rho_n}, f_\sigma) \leq 0.1$. Notice that by linearity of expectation,

$$\Pr[B \text{ accepts}] = \mathop{\mathrm{E}}_{i \in [m]}[\Pr[Q(x, \sigma[i]) \text{ accepts}]] = f_\sigma(x),$$

and that this holds regardless of what entanglement might be present among the m registers $\sigma[1], \ldots, \sigma[m]$. Hence

$$\begin{aligned} |\Pr[B \text{ accepts}] - L(x)| &\leq |\Pr[B \text{ accepts}] - f_{\rho_n}(x)| + |f_{\rho_n}(x) - L(x)| \\ &\leq 0.1 + 0.2 \end{aligned}$$

which is less than $1/3$ as desired, and $L \in \mathsf{YQP}_+/\mathsf{poly} = \mathsf{YQP}/\mathsf{poly}$. ■

Theorem 21 actually yields the stronger result that $\mathsf{BQP}/\mathsf{qpoly} \subseteq \mathsf{YQP}^*/\mathsf{poly}$, since the machine A had no dependence on the input x. We therefore have $\mathsf{BQP}/\mathsf{qpoly} = \mathsf{YQP}^*/\mathsf{poly} = \mathsf{YQP}/\mathsf{poly}$: the two definitions of YQP collapse in the presence of polynomial-size classical advice. Since we never needed the assumption that the $\mathsf{BQP}/\mathsf{qpoly}$ machine computes a *language* (i.e., a total Boolean function), another strengthening we can easily observe is $\mathsf{PromiseBQP}/\mathsf{qpoly} = \mathsf{PromiseYQP}/\mathsf{poly}$.

If we prefer, we can interpret Theorem 21 as a statement about quantum communication protocols rather than quantum complexity classes. The following theorem makes this connection more precise.

Theorem 22 *Suppose that Alice, who is computationally unbounded, has a classical description of an n-qubit quantum state ρ. She wants to send ρ to Bob, who is limited to BQP computations. Alice has at her disposal a noiseless one-way classical channel to Bob, as well as a noisy one-way quantum channel. Then for all m and $\varepsilon > 0$, there exists a protocol whereby*

(i) Alice sends Bob a classical string y of $\mathrm{poly}(n, m, 1/\varepsilon)$ bits, as well as a state σ of $\mathrm{poly}(n, m, 1/\varepsilon)$ qubits.

(ii) Bob receives y together with a possibly-corrupted version $\widetilde{\sigma}$ of σ.

(iii) *If $\widetilde{\sigma} = \sigma$, then for any measurement E performed by a circuit with at most m gates, Bob can perform another measurement $f_y(E)$ on $\widetilde{\sigma}$, and then output a number $\beta \in [0,1]$ such that $|\beta - \mathrm{Tr}(E\rho)| \leq \varepsilon$ with $1 - 1/\exp(n)$ probability. Here $f_y(E)$ can be computed in polynomial time given y together with a description of E.*

(iv) *For every $\widetilde{\sigma}$ and every such measurement E, with $1 - 1/\exp(n)$ probability Bob outputs either 'FAIL' or else a $\beta \in [0,1]$ such that $|\beta - \mathrm{Tr}(E\rho)| \leq \varepsilon$.*

Proof. This is just a direct translation of Theorem 21 to the communication setting. The string y plays the role of the trusted classical advice, the state $\widetilde{\sigma}$ plays the role of the untrusted quantum advice, the measurement E plays the role of the input $x \in \{0,1\}^n$, and Bob plays the role of the verifier. To get $1 - 1/\exp(n)$ success probability, we amplify the protocol $O(n)$ times, which just makes y and $\widetilde{\sigma}$ polynomially longer. ■

4.3 The Complexity of Preparing Quantum Advice States

If we combine Theorem 21 with known QMA-completeness results, we can obtain a striking consequence for quantum complexity theory. Namely, *the preparation of quantum advice states can always be reduced to the preparation of ground states of local Hamiltonians*—despite the fact that quantum advice states involve an exponential number of constraints, while ground states of local Hamiltonians involve only a polynomial number. In particular, if ground states of local Hamiltonians can be prepared by polynomial-size circuits, then we have not only $\mathsf{QMA} = \mathsf{QCMA}$, but also $\mathsf{BQP/qpoly} = \mathsf{BQP/poly}$. The following theorem makes this connection precise.

Theorem 23 *Let Q be a polynomial-size quantum circuit that takes an advice state ρ_n. Then there exists another polynomial-size quantum circuit Q' with the following property. For all n and $\varepsilon > 0$, there exists a 2-local Hamiltonian H on $\mathrm{poly}(n, 1/\varepsilon)$ qubits, such that for all ground states $|\phi\rangle$ of H and inputs $x \in \{0,1\}^n$,*

$$\left|\Pr\left[Q' \text{ accepts } x, |\phi\rangle\right] - \Pr\left[Q \text{ accepts } x, \rho_n\right]\right| \leq \varepsilon.$$

Furthermore, Q' can be efficiently generated given Q together with a description of H.

Proof. Kempe, Kitaev, and Regev [24] proved that the 2-Local Hamiltonians problem is QMA-complete. Furthermore, examining their proof, we find that it yields the following stronger result. Let V be a QMA verification procedure with completeness and soundness errors δ. Then there exists a 2-local Hamiltonian H, as well as a polynomial-time 'recovery procedure' R, such that if $|\phi\rangle$ is any ground state of H, then with $\Omega(1/\mathrm{poly}(n))$ probability, $R(|\phi\rangle)$ outputs a state $|\varphi\rangle$ such that $\Pr[V \text{ accepts } |\varphi\rangle] \geq 1 - \delta$. To prove the stronger result: consider a ground state of H, which Kempe et al. show to be a *history state* of the form

$$|\phi\rangle = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} |t\rangle |\phi_t\rangle .$$

Then R can simply measure the clock register $|t\rangle$, postselect on obtaining the outcome $t = 1$, and then retrieve $|\varphi\rangle$ from the computation register $|\phi_1\rangle$.

Indeed, we can strengthen the above result further, to increase R's success probability from $\Omega(1/\mathrm{poly}(n))$ to $1 - \delta$. To do so, we simply increase the number of steps T by a $1/\delta$ factor, then

put additional terms in H, to impose the constraint that the computation should do nothing for the first $(1-\delta)T$ steps (leaving $|\phi_1\rangle$ unchanged), and only apply V during the final δT steps.

Now let Q be a polynomial-size quantum circuit that takes advice state ρ_n, and let (A, B) be the YQP/poly checking algorithm (with error parameter δ) from Theorem 21. Then by the above, there exists a 2-local Hamiltonian H on $\text{poly}(n, 1/\delta)$ qubits, as well as a polynomial-time algorithm R, such that

(i) If $|\phi\rangle$ is any ground state of H, then with at least $1-\delta$ probability, $R(|\phi\rangle)$ outputs a state $|\varphi\rangle$ such that $\Pr[A \text{ accepts } |\varphi\rangle] \geq 1-\delta$.

(ii) This $|\varphi\rangle$ satisfies $|B_\varphi(x) - Q_{\rho_n}(x)| \leq \delta$ for all $x \in \{0,1\}^n$, where $B_\varphi(x) := \Pr[B \text{ accepts } x, |\varphi\rangle]$ and $Q_{\rho_n}(x) := \Pr[Q \text{ accepts } x, \rho_n]$.

We can now combine R and B into a single algorithm Q', such that $\left|Q'_\phi(x) - Q_{\rho_n}(x)\right| \leq 2\delta$ for all $x \in \{0,1\}^n$. Setting $\delta := \varepsilon/2$ then yields the corollary. ■

Let us make two remarks about Theorem 23. First, as a 'free byproduct,' we get that

$$\left|\Pr\left[Q' \text{ accepts } x, |\phi\rangle\right] - \Pr[Q \text{ accepts } x, \rho_n]\right| \leq 2\varepsilon$$

for all $|\phi\rangle$ that are ε-*close* in trace distance to a ground state of H. Second, there is nothing special here about 2-Local Hamiltonians. So far as we know, *all* existing QMA-completeness reductions have the property we needed for Theorem 23: namely, the property that any ground state of the new instance can be transformed into a QMA witness for the original instance, with $\Omega(1/\text{poly}(n))$ success probability. As one example, Aharonov et al. [7] showed that even finding the ground state energy of a nearest-neighbor Hamiltonian on the line is QMA-complete, provided the line is composed of qudits with $d \geq 12$. We can combine their result with Theorem 21 to show that for all $L \in$ BQP/qpoly, there exists a nearest-neighbor qudit Hamiltonian H on the line, such that any ground state of H is a valid quantum advice state for L.

Proof of Theorem 1. Fix $c, \varepsilon > 0$, and let ρ be the n-qubit state in Theorem 1. Let $Q(C, \xi)$ be an efficiently constructible polynomial-size quantum circuit that takes a description of a quantum measurement circuit C of size n^c, as well as a quantum state ξ of n qubits, and that outputs the measurement result $C(\xi)$.

Fix $\rho_n := \rho$. Let H be the 2-local Hamiltonian given by Theorem 23, with ground state $|\psi\rangle$, and let $Q'(C, \xi)$ be the circuit in Theorem 23, which is efficiently constructible given Q and H. Then, if we define the measurement C' as $C'(\xi) := Q'(C, \xi)$, we have

$$\left|C'(|\psi\rangle\langle\psi|) - C(\rho)\right| = \left|Q'(C, |\psi\rangle\langle\psi|) - Q(C, \rho)\right| \leq \varepsilon.$$

■

5 Further Implications for Quantum Complexity Theory

In this section, we use the BQP/qpoly = YQP/poly theorem to harvest two more results about quantum complexity classes. The first is an 'exchange theorem' stating that QCMA/qpoly $\subseteq$ QMA/poly: in other words, *one can always simulate quantum advice together with a classical witness by classical advice together with a quantum witness.* This is a straightforward generalization

of Theorem 21. The second result is a 'Quantum Karp-Lipton Theorem,' which states that if $\mathsf{NP} \subset \mathsf{BQP/qpoly}$ (that is, NP-complete problems are efficiently solvable by quantum computers with quantum advice), then $\Pi_2^{\mathsf{P}} \subseteq \mathsf{QMA}^{\mathsf{PromiseQMA}}$, which one can think of as 'almost as bad' as a collapse of the polynomial hierarchy. This result makes essential use of Theorem 21, and is a good illustration of how that theorem can be applied in quantum complexity theory.

Theorem 24 (Exchange Theorem) $\mathsf{QCMA/qpoly} \subseteq \mathsf{QMA/poly}$.

Proof. The proof is almost the same as that of Theorem 21. Let $L \in \mathsf{QCMA/qpoly}$. Then there exists a polynomial-time quantum verifier Q, a family of polynomial-size advice states $\{\rho_n\}_n$, and a polynomial p such that for all inputs $x \in \{0,1\}^n$:

- $x \in L \implies \exists w \in \{0,1\}^{p(n)} \ \Pr\left[Q\left(x, w, \rho_n\right) \text{ accepts}\right] \geq 2/3$.
- $x \notin L \implies \forall w \in \{0,1\}^{p(n)} \ \Pr\left[Q\left(x, w, \rho_n\right) \text{ accepts}\right] \leq 1/3$.

Now consider the following promise problem: given x and w as input, as well as a constant $c \in [0,1]$, decide whether $\Pr\left[Q\left(x, w, \rho_n\right) \text{ accepts}\right]$ is at most $c - 1/10$ or at least $c + 1/10$, promised that one of these is the case. (Equivalently, *estimate* the probability within an additive error $\pm 1/10$.) This problem is clearly in $\mathsf{PromiseBQP/qpoly}$, since we can take ρ_n as the advice. So by Theorem 21, the problem is in $\mathsf{PromiseYQP/poly}$ as well.

We claim this implies $L \in \mathsf{QMA/poly}$. For our $\mathsf{QMA/poly}$ verifier can take the $\mathsf{PromiseYQP/poly}$ advice string a_n as its advice, and a state of the form $\sigma \otimes |w\rangle \langle w|$ as its witness. It can then do the following:

(1) Using a_n, check that σ is a valid witness for the $\mathsf{PromiseYQP/poly}$ protocol, and reject if not.

(2) Using σ, check that $\Pr\left[Q\left(x, w, \rho_n\right) \text{ accepts}\right] \geq 2/3$ (under the promise that it is either at least $2/3$ or at most $1/3$).

■

Indeed, let $\mathsf{YQ{\cdot}QCMA}$ denote the complexity class where a BQP verifier receives a classical witness that depends on the input, as well as a quantum witness that depends only on the input size n. Then we can *characterize* $\mathsf{QCMA/qpoly}$ as equal to $\mathsf{YQ{\cdot}QCMA/poly}$, similarly to how we characterized $\mathsf{BQP/qpoly}$ as equal to $\mathsf{YQP/poly}$.

We now use Theorem 21 to prove an analogue of the Karp-Lipton Theorem for quantum advice.

Theorem 25 (Quantum Karp-Lipton Theorem) *If* $\mathsf{NP} \subset \mathsf{BQP/qpoly}$, *then* $\Pi_2^{\mathsf{P}} \subseteq \mathsf{QMA}^{\mathsf{PromiseQMA}}$.

Proof. By Theorem 21, the hypothesis implies $\mathsf{NP} \subset \mathsf{YQP/poly} = \mathsf{YQP}^*\mathsf{/poly}$. So let Q be a $\mathsf{YQP}^*\mathsf{/poly}$ algorithm to decide SAT, which takes an input x, a trusted advice string a, and an untrusted advice state ρ. Let a^* be the correct value of the advice string.

Now consider an arbitrary language $L \in \Pi_2^{\mathsf{P}}$, which is defined by a polynomial-time predicate $R\left(x, y, z\right)$ like so:

- $x \in L \iff \forall y \exists z \ R\left(x, y, z\right)$.

Using Q, we can create a pair of quantum algorithms $Q_1\left(a, \rho\right)$, $Q_2\left(\rho, x, y\right)$ with the following properties:

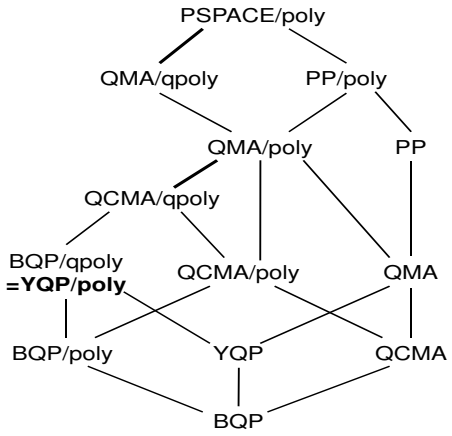

Figure 2: Containments among complexity classes related to quantum proofs and advice, in light of this paper's results. The containments $\mathsf{QMA/qpoly} \subseteq \mathsf{PSPACE/poly}$ and $\mathsf{QCMA/qpoly} \subseteq \mathsf{PP/poly}$ were shown previously by Aaronson [4]. This paper shows that $\mathsf{BQP/qpoly} \subseteq \mathsf{QMA/poly}$, and indeed $\mathsf{BQP/qpoly} = \mathsf{YQP/poly}$, where YQP is like QMA except that the quantum witness can depend only on the input length n. It also shows that $\mathsf{QCMA/qpoly} \subseteq \mathsf{QMA/poly}$.

(P1) There exists a ρ such that $\Pr\left[Q_1\left(a^*, \rho\right) \text{ accepts}\right] \geq 2/3$.

(P2) If $\Pr\left[Q_1\left(a^*, \rho\right) \text{ accepts}\right] \geq 1/3$, then for all x, y pairs, $\Pr\left[Q_2\left(\rho, x, y\right) \text{ accepts}\right] \geq 2/3$ if there exists a z such that $R\left(x, y, z\right)$ holds, and $\Pr\left[Q_2\left(\rho, x, y\right) \text{ accepts}\right] \leq 1/3$ otherwise.

Using standard amplification and NP self-reducibility, we can then strengthen property (P2) to the following, for some quantum algorithm $Q_2'\left(\rho, x, y\right)$:

(P2') If $\Pr\left[Q_1\left(a^*, \rho\right) \text{ accepts}\right] \geq 1/3$, then for all x, y pairs, $Q_2'\left(\rho, x, y\right)$ *outputs* a z such that $R\left(x, y, z\right)$ holds with probability at least $2/3$, whenever such a z exists.

Now let $U\left(a, \rho, x, y\right)$ be a quantum algorithm that does one of the following, both with equal probability:

- Runs $Q_1\left(a, \rho\right)$, and accepts if and only if it rejects.
- Runs $Q_2'\left(\rho, x, y\right)$, and accepts if and only if $R\left(x, y, Q_2'\left(\rho, x, y\right)\right)$ holds.

Then we claim that

(A1) $x \in L \implies \exists a, \rho \left[\Pr\left[Q_1\left(a, \rho\right) \text{ accepts}\right] \geq 2/3\right] \wedge \left[\forall \sigma, y \ \Pr\left[U\left(a, \sigma, x, y\right) \text{ accepts}\right] \geq 1/3\right]$.

(A2) $x \notin L \implies \forall a, \rho \left[\Pr\left[Q_1\left(a, \rho\right) \text{ accepts}\right] \leq 1/2\right] \vee \left[\exists \sigma, y \ \Pr\left[U\left(a, \sigma, x, y\right) \text{ accepts}\right] \leq 1/4\right]$.

It is clear that this claim implies $L \in \mathsf{QMA}^{\mathsf{PromiseQMA}}$. (The crucial point here is that U does *not* take the existentially-quantified advice state ρ as input—and therefore, the QMA machine does

not need to pass a quantum state to the PromiseQMA oracle, which would be illegal. This is why we needed the BQP/qpoly = YQP*/poly theorem for this result.)

We now prove the claim. First suppose $x \in L$. Then there exists an advice string $a = a^*$ with the following properties:

(B1) There exists a ρ such that $\Pr[Q_1(a, \rho) \text{ accepts}] \geq 2/3$. (By (P1).)

(B2) For all σ, y pairs, either $\Pr[Q_1(a, \sigma) \text{ rejects}] \geq 2/3$, or $\Pr[R(x, y, Q_2'(\sigma, x, y)) \text{ holds}] \geq 2/3$. (By (P2') and $x \in L$.)

By (B2), we have $\forall \sigma, y \ \Pr[U(a, \sigma, x, y) \text{ accepts}] \geq 1/3$. This proves (A1).

Next suppose $x \notin L$. Then given an advice string a, suppose there exists a ρ such that $\Pr[Q_1(a, \rho) \text{ accepts}] > 1/2$. Set $\sigma := \rho$, and choose a y for which there is *no* z such that $R(x, y, z)$ holds. Then

- $\Pr[Q_1(a, \sigma) \text{ rejects}] < 1/2$. (By assumption.)
- $\Pr[R(x, y, Q_2'(\sigma, x, y)) \text{ holds}] = 0$. (By $x \notin L$.)

Combining these, $\Pr[U(a, \sigma, x, y) \text{ accepts}] < 1/4$. This proves (A2), and hence the claim, and hence the lemma. ■

Previously, Aaronson [3] showed that if PP $\subset$ BQP/qpoly, then the counting hierarchy CH collapses. However, he had been unable to show that NP $\subset$ BQP/qpoly would have unlikely consequences in the uniform world.

6 Open Problems

One open problem is simply to find more applications of the majority-certificates lemma, which seems likely to have uses outside of quantum complexity theory; we mention one application (to 'untrusted oracles') in Appendix 8. Can we improve the parameters of the majority-certificates lemma (the size of the certificates or the number $O(n)$ of certificates), or alternatively, show that the current parameters are essentially optimal? Also, can we prove the real-valued majority-certificates lemma with an error tolerance α that depends only on the desired accuracy ε of the final approximation, not on n or the fat-shattering dimension of S?

On the quantum complexity side, we mention several questions. First, in Theorem 23, is the polynomial blowup in the number of qubits unavoidable? Could one hope for a way to simulate an n-qubit advice state by the ground state of n-qubit local Hamiltonian, or would that have implausible complexity consequences? Second, can we use the ideas in this paper to prove any upper bound on the class QMA/qpoly better than the PSPACE/poly upper bound shown by Aaronson [4]? Third, if NP $\subset$ BQP/qpoly, then does $\mathsf{QMA}^{\mathsf{PromiseQMA}}$ contain not just Π_2^{P} but the entire polynomial hierarchy? Finally, is BQP/qpoly = BQP/poly?

7 Acknowledgments

We thank Sanjeev Arora, Kai-Min Chung, Avinatan Hassidim, John Watrous, and Colin Zheng for helpful discussions, and the anonymous reviewers for their comments.

References

[1] S. Aaronson. Multilinear formulas and skepticism of quantum computing. In *Proc. ACM STOC*, pages 118–127, 2004. quant-ph/0311039.

[2] S. Aaronson. Limitations of quantum advice and one-way communication. *Theory of Computing*, 1:1–28, 2005. quant-ph/0402095. Conference version in Proceedings of CCC'2004.

[3] S. Aaronson. Oracles are subtle but not malicious. In *Proc. IEEE Conference on Computational Complexity*, pages 340–354, 2006. ECCC TR05-040.

[4] S. Aaronson. QMA/qpoly is contained in PSPACE/poly: de-Merlinizing quantum protocols. In *Proc. IEEE Conference on Computational Complexity*, pages 261–273, 2006. quant-ph/0510230.

[5] S. Aaronson. The learnability of quantum states. *Proc. Roy. Soc. London*, 463(2088):3089–3114, 2007. quant-ph/0608142.

[6] S. Aaronson and G. Kuperberg. Quantum versus classical proofs and advice. *Theory of Computing*, 3(7):129–157, 2007. Previous version in Proceedings of CCC 2007. quant-ph/0604056.

[7] D. Aharonov, D. Gottesman, S. Irani, and J. Kempe. The power of quantum systems on a line. *Comm. Math. Phys.*, 287(1):41–65, 2009. Conference version in IEEE FOCS 2007. arXiv:0705.4077.

[8] D. Aharonov and T. Naveh. Quantum NP - a survey. quant-ph/0210077, 2002.

[9] D. Aharonov and O. Regev. A lattice problem in Quantum NP. In *Proc. IEEE FOCS*, pages 210–219, 2003. quant-ph/0307220.

[10] D. Aharonov and O. Regev. Lattice problems in NP intersect coNP. *J. ACM*, 52(5):749–765, 2005. Earlier version in IEEE FOCS 2004.

[11] N. Alon, S. Ben-David, N. Cesa-Bianchi, and D. Haussler. Scale-sensitive dimensions, uniform convergence, and learnability. *J. ACM*, 44(4):615–631, 1997.

[12] A. Ambainis, A. Nayak, A. Ta-Shma, and U. V. Vazirani. Quantum dense coding and quantum finite automata. *J. ACM*, 49:496–511, 2002. Earlier version in ACM STOC 1999, pp. 376-383. quant-ph/9804043.

[13] P. L. Bartlett and P. M. Long. Prediction, learning, uniform convergence, and scale-sensitive dimensions. *J. Comput. Sys. Sci.*, 56(2):174–190, 1998.

[14] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *J. ACM*, 36(4):929–965, 1989.

[15] N. H. Bshouty, R. Cleve, R. Gavaldà, S. Kannan, and C. Tamon. Oracles and queries that are sufficient for exact learning. *J. Comput. Sys. Sci.*, 52(3):421–433, 1996.

[16] V. Chakaravarthy and S. Roy. Oblivious symmetric alternation. In *Proc. Intl. Symp. on Theoretical Aspects of Computer Science (STACS)*, pages 230–241, 2006.

[17] L. Fortnow and R. Santhanam. Fixed-polynomial size circuit bounds, 2006. ECCC TR06-157.

[18] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Sys. Sci.*, 55(1):119–139, 1997.

[19] P. Hausladen and W. K. Wootters. A 'pretty good' measurement for distinguishing quantum states. *J. Modern Optics*, 41(12):2385–2390, 1994.

[20] A. S. Holevo. Some estimates of the information transmitted by quantum communication channels. *Problems of Information Transmission*, 9:177–183, 1973. English translation.

[21] R. Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proc. IEEE FOCS*, pages 538–545, 1995.

[22] S. Kakade and A. Tewari. Learning theory lecture notes, 2008. ttic.uchicago.edu/~tewari/LT_SP2008.html.

[23] R. M. Karp and R. J. Lipton. Turing machines that take advice. *Enseign. Math.*, 28:191–201, 1982.

[24] J. Kempe, A. Kitaev, and O. Regev. The complexity of the Local Hamiltonian problem. *SIAM J. Comput.*, 35(5):1070–1097, 2006. quant-ph/0406180.

[25] H. Nishimura and T. Yamakami. Polynomial time quantum computation with advice. *Inform. Proc. Lett.*, 90:195–204, 2003. ECCC TR03-059, quant-ph/0305100.

[26] N. Sauer. On the density of families of sets. *J. Combinatorial Theory Series A*, 13:145–147, 1972.

[27] R. E. Schapire. The strength of weak learnability. *Machine Learning*, 5(2):197–227, 1990.

[28] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27:1134–1142, 1984.

[29] G. Vidal. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.*, 91, 2003. quant-ph/0301063.

[30] E. Viola. On approximate majority and probabilistic time. In *Proc. IEEE Conference on Computational Complexity*, pages 155–168, 2007. Journal version to appear in Computational Complexity.

[31] J. Watrous. Succinct quantum proofs for properties of finite groups. In *Proc. IEEE FOCS*, pages 537–546, 2000. cs.CC/0009002.

8 Appendix: Untrusted Oracles

In this appendix, we give an interesting consequence of the majority-certificates lemma for classical complexity theory.

When we give a machine an oracle, normally we assume the oracle can be trusted. But it is also natural to consider *untrusted* oracles, which are nevertheless restricted in their computational power. We formalize this notion as follows:

Definition 26 (Untrusted Oracles) *Let $\mathcal{C}$ and $\mathcal{D}$ be complexity classes. Also, given a family $a = \{a_n\}_{n \geq 1}$ of $p(n)$-bit advice strings and a machine V, let $V[a]$ be the language decided by V given a as advice. Then $\mathcal{C}^{\mathsf{Untrusted}\text{-}\mathcal{D}}$ is the class of languages L for which there exists a $\mathcal{C}$ machine U, a $\mathcal{D}$ machine V, and a polynomial p such that for all n:*

(i) There exist $p(n)$-bit advice strings $a_1, \ldots, a_m$ such that $U^{V[a_1],\ldots,V[a_m]}$ decides L.

(ii) $U^{V[a_1],\ldots,V[a_m]}(x)$ outputs either $L(x)$ or 'FAIL', for all inputs $x \in \{0,1\}^n$ and all $p(n)$-bit advice strings $a_1, \ldots, a_m$.

We can now state the consequence.

Theorem 27 *Let $\mathcal{C}$ be a uniform syntactic complexity class, such as* P, NP, *or* EXP. *Then* $\mathcal{C}/\mathsf{poly} \subseteq \left(\mathsf{AC}^0\right)^{\mathsf{Untrusted}\text{-}\mathcal{C}}$.

Proof. Let V be a $\mathcal{C}/\mathsf{poly}$ machine that uses a family $a = \{a_n\}_{n \geq 1}$ of $p(n)$-bit advice strings. Fix an input length n, and let $f_w(x)$ be the output of V on input x and advice string $w \in \{0,1\}^{p(n)}$. Then $S = \{f_w\}_{w \in \{0,1\}^{p(n)}}$ is a Boolean concept class of size $|S| \leq 2^{\mathrm{poly}(n)}$. So by Lemma 3, there exist $m = O(n)$ polynomial-size certificates $C_1, \ldots, C_m$, which isolate functions $f_1, \ldots, f_m \in S$ respectively such that $\mathrm{MAJ}(f_1, \ldots, f_m) = f_{a_n}$. Now, we can easily modify the proof of Lemma 3 to ensure not only that $\mathrm{MAJ}(f_1, \ldots, f_m) = f^*$, but also that

$$f_{a_n}(x) = 1 \Longrightarrow f_1(x) + \cdots + f_m(x) \geq \frac{2m}{3},$$
$$f_{a_n}(x) = 0 \Longrightarrow f_1(x) + \cdots + f_m(x) \leq \frac{m}{3}$$

for all inputs x. To do so, we simply take $m = O(n)$ sufficiently large and redo the Chernoff bound. Furthermore, it is known that Approximate Majority—that is, Majority where the fraction of 1's in the input is bounded away from 1/2 by a constant—can be computed by polynomial-size depth-3 circuits, so in particular, in AC^0 (see Viola [30] for example).

By hardwiring the certificates $C_1, \ldots, C_m$ into the AC^0 circuit, we can produce an AC^0 circuit that first checks whether f_i is consistent with C_i for all $i \in [m]$, outputs 'FAIL' if not, and otherwise outputs $U^{f_1,\ldots,f_m}(x) = f_{a_n}(x)$. ■

If $\mathcal{C}$ is a semantic complexity class, such as BPP or UP, the difficulty is that there might be a $\mathcal{C}/\mathsf{poly}$ machine M and advice string w for which the function f_w is undefined (since M need not decide a language for every w). However, if we force the $\mathsf{Untrusted}\text{-}\mathcal{C}$ oracle to restrict itself to w for which f_w *is* defined, then Theorem 27 goes through for semantic classes as well. Using the *real*-valued majority-certificates lemma that we develop in Section 3, it is possible to remove the assumption that f_w is defined for all w for semantic classes such as BPP.

9 Appendix: Isolatability and Learnability

The following definition abstracts a key notion from the majority-certificates lemma.

Definition 28 (Majority-Isolatability) *A Boolean concept class S is majority-isolatable if for every $f \in S$, there exist $m = \mathrm{poly}(n)$ certificates $C_1, \ldots, C_m$, each of size $\mathrm{poly}(n)$, such that*

(i) $S[C_i]$ is nonempty for all $i \in [m]$, and

(ii) if $f_i \in S[C_i]$ for all $i \in [m]$, then $\mathrm{MAJ}(f_1, \ldots, f_m) = f$, where MAJ *denotes pointwise majority.*

We now show that the majority-isolatability of a Boolean concept class S is equivalent to a large number of other properties of S—including having singly-exponential cardinality, having polynomial VC-dimension, being PAC-learnable using $\text{poly}(n)$ samples, and being 'winnowable.' While we do not need this equivalence theorem elsewhere in the paper, it might be of interest anyway. Note that the equivalence theorem is easily seen to break down for concept classes with infinite input domains.

Definition 29 (VC-dimension) *We say a Boolean concept class S shatters the set $A \subseteq \{0,1\}^n$ if for all $2^{|A|}$ functions $g : A \rightarrow \{0,1\}$, there exists an $f \in S$ whose restriction to A equals g. Then the VC-dimension of S, or $\text{VCdim}(S)$, is the size of the largest set shattered by S.*

Given a distribution $\mathcal{D}$ over $\{0,1\}^n$, we say the Boolean functions $f, g : \{0,1\}^n \rightarrow \{0,1\}$ are $(\mathcal{D}, \varepsilon)$-*close* if

$$\Pr_{x \sim \mathcal{D}} [g(x) = f(x)] \geq 1 - \varepsilon.$$

Definition 30 (Learnability) *S is learnable if for all $f \in S$, distributions $\mathcal{D}$, and $\varepsilon, \delta > 0$, there exists an $m = \text{poly}(n, 1/\varepsilon, \log 1/\delta)$ such that with probability at least $1 - \delta$ over sample points $x_1, \ldots, x_m$ drawn independently from $\mathcal{D}$, every $g \in S$ satisfying $g(x_1) = f(x_1), \ldots, g(x_m) = f(x_m)$ is $(\mathcal{D}, \varepsilon)$-close to f.*

We can also define 'approximability,' which is like learnability except that the choice of training examples can be nondeterministic:

Definition 31 (Approximability) *S is approximable if for all $f \in S$ and distributions $\mathcal{D}$, there exists a certificate C of size $\text{poly}(n, 1/\varepsilon)$ such that every $g \in S[C]$ is $(\mathcal{D}, \varepsilon)$-close to f.*

Finally, let us call attention to a notion that implicitly played a major role in the proof of Lemma 3.

Definition 32 (Winnowability) *S is winnowable if for all nonempty subsets $S' \subseteq S$, there exists a certificate C of size $\text{poly}(n)$ such that $|S'[C]| = 1$.*

We can now prove the equivalence theorem.

Theorem 33 *Let S be a Boolean concept class. Then $|S| \leq 2^{\text{poly}(n)}$ iff $\text{VCdim}(S) \leq \text{poly}(n)$ iff S is learnable iff S is approximable iff S is majority-isolatable iff S is winnowable.*

Proof. $|S| \leq 2^{\text{poly}(n)} \Longrightarrow \text{VCdim}(S) \leq \text{poly}(n)$ follows from the trivial upper bound $\text{VCdim}(S) \leq \log_2 |S|$.

$\text{VCdim}(S) \leq \text{poly}(n) \Longrightarrow |S| \leq 2^{\text{poly}(n)}$ is Sauer's Lemma [26], which implies the relation $|S| \leq 2^{n \, \text{VCdim}(S)}$.

$|S| \leq 2^{\text{poly}(n)} \Longrightarrow$**Learnable** was proved by Valiant [28].

Learnable$\Longrightarrow$**Approximable** is immediate, and **Approximable**$\Longrightarrow \text{VCdim}(S) \leq \text{poly}(n)$ follows from a counting argument (see Blumer et al. [14] for details).

$|S| \leq 2^{\text{poly}(n)} \Longrightarrow$**Majority-Isolatable** was the content of Lemma 3.

Majority-Isolatable$\Longrightarrow |S| \leq 2^{\text{poly}(n)}$ follows from another counting argument: if S is majority-isolatable, then every $f \in S$ is uniquely determined by $\text{poly}(n)$ certificates $C_1, \ldots, C_m$, each of which can be specified using $\text{poly}(n)$ bits.

For $|S| \leq 2^{\text{poly}(n)} \Longrightarrow$**Winnowable**, let $S' \subseteq S$. Then as in the proof of Lemma 3, we can use binary search to winnow S' down to a single function $f \in S'$, which yields a certificate of size at most $\log_2 |S'| \leq \log_2 |S|$.

For **Winnowable**$\Longrightarrow |S| \leq 2^{\text{poly}(n)}$, we prove the contrapositive. Suppose $|S| \geq 2^{t(n)}$ for some superpolynomial function $t(n)$ (at least, for infinitely many n). Then define a subset $S' \subseteq S$ by the following iterative procedure. Initially $S' = S$. Then so long as there exists a certificate C of size at most $t(n)/(2n+2)$ such that $|S'[C]| = 1$, remove the function $f \in S'[C]$ from S', halting only when no more such 'isolating certificates' can be found.

The number of certificates of size k is at most $2^{(n+1)k}$, and a given certificate C can only be chosen once, since thereafter $S'[C]$ is empty. So when the above procedure halts, we are left with a set S' such that $|S'| \geq 2^{t(n)} - 2^{(n+1)t(n)/(2n+2)} > 0$. Furthermore, for every function f remaining in S', there can be no polynomial-size certificate C such that $S'[C] = \{f\}$—for if there were, then we would already have eliminated f in the process of forming S'. Hence S is not winnowable. ■

10 Appendix: Winnowing of p-Concept Classes

In this appendix, we look more closely at the problem solved by Lemma 10 (the 'Safe Winnowing Lemma'), and ask in what senses it is possible to winnow a p-concept class down to 'essentially' just one function. The answer turns out to be interesting, even though we do not need it for our quantum complexity applications.

We first give a definition that abstracts part of what Lemma 10 was trying to accomplish.

Definition 34 (Winnowability) *A p-concept class S is L_1-winnowable if the following holds. For all nonempty subsets $S' \subseteq S$ and $\varepsilon > 0$, there exists a function $f \in S'$, a set $X \subseteq \{0,1\}^n$ of size $\text{poly}(n, 1/\varepsilon)$, and a $\delta = \text{poly}(\varepsilon)$ such that every $g \in S'$ that satisfies $\Delta_1(f,g)[X] \leq \delta$ also satisfies $\Delta_\infty(f,g) \leq \varepsilon$. Likewise, S is L_2-winnowable if $\Delta_2(f,g)[X] \leq \delta$ implies $\Delta_\infty(f,g) \leq \varepsilon$, and L_∞-winnowable if $\Delta_\infty(f,g)[X] \leq \delta$ implies $\Delta_\infty(f,g) \leq \varepsilon$.*

Clearly L_∞-winnowability implies L_2-winnowability implies L_1-winnowability. The following lemma will imply that every set of functions with a small cover is L_1-winnowable.

Lemma 35 (L_1-Winnowing Lemma) *Let S be a set of functions $f : \{0,1\}^n \rightarrow [0,1]$. For some parameter $\varepsilon > 0$, let C be a finite ε-cover for S. Then there exists an $f \in S$, as well as a subset $X \subseteq \{0,1\}^n$ of size $O\left(\frac{1}{\varepsilon}\log|C|\right)$, such that every $g \in S$ that satisfies $\Delta_1(f,g)[X] \leq 0.4\varepsilon$ also satisfies $\Delta_\infty(f,g) \leq 2\varepsilon$.*

Proof. We will consider functions $P : S \rightarrow [0,1]$, which we think of as assigning a probability weight $P(g)$ to each function $g \in S$. In particular, given an $f \in S$ and a subset of inputs $X \subseteq \{0,1\}^n$, define

$$P_{f,X}(g) := \exp\left(-\Delta_1(f,g)[X]\right).$$

Clearly $P_{f,X}(f) = 1$. Our goal will be to find $f \in S$ and $X \subseteq \{0,1\}^n$, with $|X| = O\left(\frac{1}{\varepsilon}\log|C|\right)$, such that every $g \in S$ that satisfies $P_{f,X}(g) \geq e^{-0.4\varepsilon}$ also satisfies $\Delta_\infty(f,g) \leq 2\varepsilon$. Supposing we have found such an (f,X) pair, the lemma is proved.

Consider the progress measure

$$M_{f,X} := \sum_{h \in C} P_{f,X}(h).$$

Clearly $M_{f,X} \leq |C|$ for all (f, X). We claim, furthermore, that $M_{f,X} \geq \exp(-\varepsilon |X|)$ for all (f, X). For since C is an ε-cover for S, there always exists an $h \in C$ such that $\Delta_1(f, h)[X] \leq \varepsilon |X|$, and that h alone contributes at least $\exp(-\varepsilon |X|)$ to $M_{f,X}$.

We will construct (f, X) by an iterative process. Initially f is arbitrary and X is the empty set, so $P_{f,X}(g) = 1$ for all g, and $M_{f,X} = |C|$. Now, suppose there exists a $g \in S$ such that $P_{f,X}(g) \geq e^{-0.4\varepsilon}$, as well as an input y such that $|f(y) - g(y)| > 2\varepsilon$. As a first step, let $Y := X \cup \{y\}$ (that is, add y into our set of inputs). Then the crucial claim is that either $M_{f,Y}$ or $M_{g,Y}$ is a $1 - \Omega(\varepsilon)$ factor smaller than $M_{f,X}$. This means in particular that, by replacing X with Y (increasing $|X|$ by 1), and possibly also replacing f with g, we can decrease $M_{f,X}$ by a $1 - \Omega(\varepsilon)$ factor compared to its previous value. Since $\exp(-\varepsilon |X|) \leq M_{f,X} \leq |C|$, it is clear that $M_{f,X}$ can decrease in this way at most

$$O\left(\log_{1+\varepsilon} \frac{|C|}{\exp(-\varepsilon |X|)}\right)$$

times. Setting the above expression equal to $|X|$ and solving, we find that the process must terminate when $|X| = O\left(\frac{1}{\varepsilon} \log |C|\right)$, returning an (f, X) pair with the properties we want.

We now prove the crucial claim. The first step is to show that either

$$M_{f,Y} = \sum_{h \in C} P_{f,X}(h)\, e^{-|f(y) - h(y)|}$$

or else

$$M' := \sum_{h \in C} P_{f,X}(h)\, e^{-|g(y) - h(y)|}$$

is at most

$$\frac{1 + e^{-\varepsilon}}{2} M_{f,X}.$$

For since $|f(y) - g(y)| > 2\varepsilon$, either $|f(y) - h(y)| > \varepsilon$ or $|g(y) - h(y)| > \varepsilon$ by the triangle inequality. So for every y, either $e^{-|f(y) - h(y)|} < e^{-\varepsilon}$ or $e^{-|g(y) - h(y)|} < e^{-\varepsilon}$. This in turn means that either $M_{f,Y}$ or M' must have at least half its terms (as weighted by the $P_{f,X}(h)$'s) shrunk by an $e^{-\varepsilon}$ factor.

If $M_{f,Y} < \frac{1+e^{-\varepsilon}}{2} M_{f,X}$ then we are done. So suppose instead that $M' < \frac{1+e^{-\varepsilon}}{2} M_{f,X}$. Then

$$\begin{aligned}
M_{g,Y} &= \sum_{h \in C} P_{g,X}(h)\, e^{-|g(y) - h(y)|} \\
&\leq M' \max_{h \in C} \frac{P_{g,X}(h)}{P_{f,X}(h)} \\
&= M' \max_{h \in C} \frac{\exp(-\Delta_1(g, h)[X])}{\exp(-\Delta_1(f, h)[X])} \\
&\leq M' \exp(\Delta_1(f, g)[X]) \\
&= \frac{M'}{P_{f,X}(g)} \\
&< \frac{\frac{1+e^{-\varepsilon}}{2} M_{f,X}}{e^{-0.4\varepsilon}} \\
&< \left(1 - \frac{\varepsilon}{20}\right) M_{f,X}
\end{aligned}$$

and we are done. ■

Recall that S is *coverable* if for all $\varepsilon > 0$, there exists an ε-cover for S of size $2^{\text{poly}(n,1/\varepsilon)}$. We can now prove the following equivalence theorem.

Theorem 36 *A p-concept class S is coverable if and only if it is L_1-winnowable.*

Proof. For **Coverable$\Longrightarrow$L$_1$-Winnowable**: fix a subset $S' \subseteq S$ and an $\varepsilon > 0$. Let C be an $\varepsilon/2$-cover for S' of size $2^{\text{poly}(n,1/\varepsilon)}$. Then by Lemma 35, there exists an $f \in S'$, as well as a subset $X \subseteq \{0,1\}^n$ of size $O\left(\frac{1}{\varepsilon}\log|C|\right) = \text{poly}(n, 1/\varepsilon)$, such that every $g \in S'$ that satisfies $\Delta_1(f,g)[X] \leq \varepsilon/5$ also satisfies $\Delta_\infty(f,g) \leq \varepsilon$.

For **L$_1$-Winnowable$\Longrightarrow$Coverable**, we prove the contrapositive. Suppose there exists a function $t(n, 1/\varepsilon)$, superpolynomial in either n or $1/\varepsilon$, such that S has no ε-cover of size $2^{t(n,1/\varepsilon)}$ (at least, for infinitely many n or $1/\varepsilon$). Let $p = \text{poly}(n, 1/\varepsilon)$ and $\delta = \text{poly}(\varepsilon)$. Given a function f and subset $X \subseteq \{0,1\}^n$, let $L[f,X]$ be the set of all functions g such that $\Delta_1(f,g)[X] \leq \delta$. Then our goal is to construct a subset $S' \subseteq S$ for which there is no pair (f,X) such that

- $f \in S'$,
- $X \subseteq \{0,1\}^n$ is a set of inputs with $|X| = p$, and
- $g \in S' \cap L[f,X]$ implies $\Delta_\infty(f,g) \leq \varepsilon$.

Let $W := \lceil 2p/\delta \rceil$. Also, call a set B of functions $f : \{0,1\}^n \rightarrow [0,1]$ a *sliver* if there exists a set $X \subseteq \{0,1\}^n$ with $|X| = p$, as well a function $a : X \rightarrow [W]$, such that

$$f \in B \Longleftrightarrow f(x) \in \left[\frac{a(x)-1}{W}, \frac{a(x)}{W}\right] \ \forall x \in X.$$

Then define a subset $S' \subseteq S$ by the following iterative procedure. Initially $S' = S$. Then so long as there exists a sliver B such that $S' \cap B$ is nonempty, together with a function $f_B \in S$ such that

$$g \in S' \cap B \Longrightarrow \Delta_\infty(f_B, g) \leq \varepsilon,$$

remove B from S' (that is, set $S' := S' \setminus B$). Halt only when no more such slivers B can be found.

As a first observation, the total number of slivers is at most $(2^n W)^p = 2^{\text{poly}(n,1/\varepsilon)}$. Thus, the above procedure must halt after at most $2^{\text{poly}(n,1/\varepsilon)}$ iterations.

As a consequence, we claim that S' must be nonempty after the procedure has halted. For suppose not. Then the sequence of functions f_B chosen by the procedure would form an ε-cover for S of size $2^{\text{poly}(n,1/\varepsilon)}$—since for all $g \in S$, we would simply need to find a sliver B containing g that was removed by the procedure; then f_B would satisfy $\Delta_\infty(f_B, g) \leq \varepsilon$. But this contradicts the assumption that no such ε-cover exists.

Finally, we claim that once the procedure halts, there can be no $f \in S'$ and set X of p inputs such that $\Delta_\infty(f,g) \leq \varepsilon$ for all $g \in S' \cap L[f,X]$. For suppose to the contrary that such an (f,X) pair existed. It is not hard to see that for every (f,X), there exists a sliver B that contains f and is contained in $L[f,X]$. But then $S' \cap B$ would be nonempty, and (B,f) would satisfy the condition $g \in S' \cap B \Longrightarrow \Delta_\infty(f,g) \leq \varepsilon$. So B (or some other sliver containing f) would already have been eliminated in the process of forming S'. ■

A natural question is whether Lemma 35 and Theorem 36 would also hold with L_2-winnowability or L_∞-winnowability in place of L_1-winnowability. The next theorem shows, somewhat surprisingly, that the use of the L_1 norm was essential.

Theorem 37 *There exists a p-concept class S that is coverable, but not L_2-winnowable or L_∞-winnowable.*

Proof. We prove a stronger statement: there exists a *finite* p-concept class S, of size $|S| \leq 2^{\text{poly}(n)}$, that is not L_2-winnowable (and as a direct consequence, not L_∞-winnowable either). To prove this, it suffices to find a set S with $|S| \leq 2^{\text{poly}(n)}$, as well as a constant $\varepsilon > 0$, for which the following holds. For all $f \in S$, subsets $X \subseteq \{0,1\}^n$ of size less than $2^n - n^2$, and constants δ depending on ε, there exists a $g \in S$ such that $\Delta_2(f,g)[X] \leq \delta$ but $\Delta_\infty(f,g) > \varepsilon$ (at least, for all sufficiently large n).

Let ε be any constant in $(0,1)$, and let S be the class of all functions $f : \{0,1\}^n \rightarrow [0,1]$ of the form

$$f(x) = \frac{a_x}{n},$$

where the a_x's are nonnegative integers satisfying

$$\sum_{x \in \{0,1\}^n} a_x = n^2.$$

Then clearly $|S| \leq (2^n)^{n^2}$, since we can form any $f \in S$ by starting from the identically-0 function, then choosing n^2 inputs x (with repetition) on which to increment f by $1/n$.

Now let $f \in S$, and let $X \subseteq \{0,1\}^n$ have size $|X| < 2^n - n^2$. Then we can 'corrupt' f to create a new function $g \in S$ as follows. Let Z be a set of n inputs $x \in \{0,1\}^n$ on which $f(x) > 0$ (note that such a Z must exist, since $\sum_x f(x) = n$ but $f(x) \leq 1$ for all x). By the pigeonhole principle, there exists a $y \in \{0,1\}^n \setminus X$ such that $f(y) = 0$. Fix that y, and define

$$g(x) := \begin{cases} 1 & \text{if } x = y \\ f(x) - 1/n & \text{if } x \in Z \\ f(x) & \text{otherwise.} \end{cases}$$

Clearly $g \in S$ and

$$\Delta_2(f,g)[X] = \sqrt{\sum_{x \in Z \cap X} \frac{1}{n^2}} \leq \frac{1}{\sqrt{n}}.$$

On the other hand, we have $f(y) = 0$ and $g(y) = 1$, so $\Delta_\infty(f,g) = 1$. Therefore S is not L_2-winnowable. ■

ECCC ISSN 1433-8092

http://eccc.hpi-web.de