On the Query Complexity of Quantum Learners

Jorge Castro
Dept. LSI, Universitat Politècnica de Catalunya
Campus Nord, 08034 Barcelona, Spain, castro@lsi.upc.edu

Abstract

This paper introduces a framework for quantum exact learning via queries, the so-called quantum protocol. It is shown that usual protocols in the classical learning setting have quantum counterparts. A combinatorial notion, the general halving dimension, is also introduced. Given a quantum protocol and a target concept class, the general halving dimension provides a lower bound on the number of queries that a quantum algorithm needs to learn. For usual protocols, this lower bound is also valid even if only involution oracle teachers are considered. The general halving dimension also approximates the query complexity of ordinary randomized learners. From these bounds we conclude that any quantum polynomially query learnable concept class must be also polynomially learnable in the classical setting.

1 Introduction

In recent years many efforts have been done in order to understand the power of quantum devices that can get partial information about an unknown function making some type of oracle calls, see for instance [4, 6, 9, 10, 11, 12, 13]. A main goal of this research is to know how many queries (or oracle calls) a quantum device requires to find out some characteristic of the hidden function , and how it relates to the number of oracle interactions a classical algorithm needs. For example, Deutsch and Jozsa [9] show that exponentially fewer black-box oracle calls (also called membership queries) are needed in the quantum model in order to determine whether the hidden function is constant or it is balanced between outputs 0 and 1.

Quantum concept learning can be seen as an special case of this type of research where the goal of the algorithm is to figure out which the hidden function is. Here several results are known. Bshouty and Jackson [8] define a quantum version of the PAC model introduced by Valiant [15], and provide a quantum learning algorithm for DNF that do not require membership queries, a type of queries used by its classical counterpart. Servedio and Gortler [12] show several bounds on the number of oracle calls required to learn on the quantum PAC setting and on the more demanding scenario of exact learning from membership queries. For these both specific learning settings they conclude that are not possible exponential improvements in the number of oracle interactions. Finally, Hunziker *et al* [11] show a general technique for quantum exact learning from membership and restricted equivalence queries (equivalence queries without counterexamples) that is shown to provide, in a couple of specific cases, more efficient learning algorithms (in terms of number of queries) than is possible classically.

This paper has two goals. The first one is to introduce a general framework for quantum exact learning via queries which sets when a class of queries can be considered to define a learning game played by quantum devices. We note that, as far as we know, the only queries that have been used in the literature have been membership and restricted equivalences [11, 12]. This contrast with the classical setting where a rich variety of queries have been considered, see for instance Angluin [1]. The second goal is to study the number of queries (or query complexity) required by exact learners. We ideally want to obtain lower and upper bounds on the query complexity that shall be valid under any choice of queries defining the learning game.

According to the first goal, we introduce the quantum protocol concept, a notion that allows us to define a learning game played by quantum machines where popular queries from the classical setting, as membership, equivalences, subset and others defined in [1] have natural quantum counterparts. Specific quantum protocols for these queries are presented. Learning games defined by quantum protocols for membership and membership and restricted equivalences agree with learning settings present in [12] and [11].

With respect to the second goal, we define a combinatorial function, the general halving dimension, GHdim, having some nice features. In the quantum learning scenario, we show a lower bound for the query complexity in terms of GHdim that is valid for any quantum protocol and for any target concept class. This lower bound extends the previous one in [12] for the specific protocol of membership queries. In the classical learning model, we prove that GHdim approximates the query complexity of randomized learners. This characterization extends the previous ones provided by Simon [14] for the specific ordinary protocols of membership and membership and equivalence queries. From these two results we conclude that, fixed an arbitrary set of queries, any quantum polynomially query learnable concept class is also polynomially learnable in the ordinary setting. This was only known for the specific case of membership queries [12].

1.1 Organization

We review basic definitions concerning classical exact learning and quantum computation in Section 2. The quantum exact learning framework is defined in Section 3, where concepts of quantum protocol (Section 3.1), quantum teacher (Section 3.2) and answering scheme (Section 3.4) are introduced. The query complexity of exact learners is analyzed in Section 4. First, we introduce the general having dimension (Section 4.1). We show a lower bound for the quantum query complexity in terms of GHdim (Theorem 8). Finally, we prove that GHdim approximates the query complexity of ordinary randomized learners (Theorems 10 and 12). The equivalence of the classical and quantum exact learners with respect to the polynomial learnability is shown in Section 5 (Theorem 13).

2 Preliminaries

2.1 Basic Definitions

Given a complex number α, we denote by $\alpha^\star$ its complex conjugate and by $|\alpha|$ its module. For complex vectors v and w, the l_2-norm (Euclidean norm) of v is expressed by $\|v\|$, the l_1-norm by $\|v\|_1$ and the inner product of v and w by $\langle v|w\rangle$. Note that $\|v\| = \langle v|v\rangle^{1/2}$. Abusing notation, we also denote the cardinality of a set A by $|A|$. For $b, d \in \{0,1\}$ we write $b \oplus d$ to denote $b+d \pmod 2$. For n-bit strings $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ we write $x \oplus y$ to denote $(x_1 \oplus y_1, \ldots, x_n \oplus y_n)$

The set of all Boolean functions defined on $\{0,1\}^n$ is denoted by B_n. A *concept* f is a function of B_n. Equivalently, a concept f can be viewed as the subset $\{x \in \{0,1\}^n \mid f(x) = 1\}$. A *concept class* C is a subset of B_n.

2.2 Classical Exact Learning

In query learning two players, the *learner* and the *teacher*, play a game. The learner is a (classical) randomized algorithm and the teacher is an oracle function. Some concept class C (the *target* concept class) is known to both players and the teacher chooses a concept in C (the target concept) that is unknown to the learner. The goal of the learner is to find out what concept is, asking the teacher some type of queries.

A query is a question that the learner poses to the teacher. The most popular in the literature are *membership queries* and *equivalence queries*. Other type of queries, as subset, superset and restricted equivalence, have been defined, see [1]. In general, the setting of the learning game is complete when the learning *protocol* is defined. The protocol is the agreement about which the admissible queries are and, for each target concept, which the possible answers for such queries are. Answers provide a property of the target. A teacher is valid for the target concept f and the protocol P if it replies to each query q choosing one of the admissible answers in P for q and f.

A concept class C is learnable with k queries under protocol P if there exists a randomized learning algorithm L such that for any $f \in C$ and for any valid teacher T that answers with respect to f using P, with probability at least $2/3$ the learner L outputs a circuit h such that $h(x) = f(x)$ for all $x \in \{0,1\}^n$ after at most k interactions with T. For a class $C \subseteq B_n$ and a protocol P, the *query complexity*, is the smallest $k \geq 0$ such that C is learnable with k queries under P.

2.3 Quantum Computation

Detailed descriptions of quantum Turing machines and quantum oracle computations are provided in [6, 7]. In spite of assuming the reader is familiar with basic aspects of quantum computers, we provide below a short summary of essential elements.

To each quantum Turing machine M corresponds a inner-product vector space S. The vectors of S are finite complex linear combinations of configurations (or instantaneous descriptions) of M. The elements of S are also so-called *superpositions* and the complex coefficients defining a vector of S are called *amplitudes*. The inner-product is defined by given an orthonormal basis for S, the vectors of this basis are the configurations of M. The time evolution operator of a quantum Turing machine M is fixed by a unitary matrix U_M, which defines a linear operator on S that conserves the distance. As any unitary operator on S has a unique unitary extension in the Hilbert space defined as the Cauchy closure of S, we can think that U_M is an unitary transformation in a Hilbert space. This is necessary to show that computations performed by quantum Turing machines are consistent with axioms of quantum mechanics.

At step j of the computation of M, the time evolution operator U_M is applied to a superposition of configurations (a vector $|v_j\rangle$ of S). The initial superposition $|v_0\rangle$ is the linear combination of configurations having all amplitudes value 0 except the only one corresponding to the initial configuration of the machine that has value 1.

A quantum Turing machine M finishes at step t if the corresponding superposition $|v_t\rangle$ only has nonzero amplitudes on final configurations (those whose state is a final one) and previous superpositions $|v_j\rangle$ where $j < t$ give amplitude zero to each final configuration. Let us assume that M finishes at step t and that $|v_t\rangle = \sum_x \alpha_x |x\rangle$ is the corresponding superposition. Now the machine M chooses to be in a single configuration rather than in a superposition of configurations making an *observation* (or *measurement*). The superposition is then changed so that a single configuration has amplitude 1 and all others are zero. Formally, the observation operation provides configuration $|x\rangle$ with probability $|\alpha_x|^2$. Note that $\sum_x |\alpha_x|^2 = 1$ because $|v_t\rangle$ has norm 1 (it is obtained by applying a unitary operator to an initial $|v_0\rangle$ superposition that has norm 1).

2.3.1 Oracle Quantum Turing Machine

We follow definitions in [6]. An oracle quantum Turing machine has an special query tape (that has to accomplish some rules of behavior, see [6]) and two distinguished internal states: a pre-query state p_1 and a post-query state p_2. A query is executed whenever the machine enters the pre-query state. In this case, it applies a fixed unitary operator U to the current contents $|q\rangle$ of the query tape, replacing it by $U|q\rangle$. In order to ensure that a single machine cycle ought not to make infinite changes in the tape, we require that $U|q\rangle$ have amplitude zero on all but finitely many basis vectors. The use of this kind of unitary oracles still provide unitary time evolution for, in other aspects, well-defined quantum Turing machines. Another natural restriction one may wish to impose upon U is that it be an involution, $U^2 = I$, so that the effect of an oracle call can be undone by a further call on the same oracle. This may be crucial to allow proper interference to take place.

3 Quantum Exact Learning

The learning game is similar to classical one but now the learner is a quantum algorithm and the teacher is a quantum oracle function. The game is completely defined when the learning protocol is provided.

3.1 Quantum protocols

We review first the classical protocol notion introduced in [2, 3]. A classical protocol specifies which the admissible queries are and, for each query, which the valid answers are. Queries belong to a finite set Q and answers are from a finite set A. Formally, a classical protocol P is a subset of $Q \times A$ that satisfies the two requirements listed below. Their justification can be found in [2, 3].

- To each element (q, a) of P corresponds a subset of B_n, so-called *consistent set*. Fixed the query q, the Boolean functions of this set are said to be consistent with answer a. It is also assumed that two different answers to query q define different consistent sets.
- Completeness: Given any query q of Q and any function f in B_n there exists an answer a such that (q, a) is an element of P and function f is consistent with a.

Abusing language we will also say that answer a of q is consistent with function f when f is consistent with (q, a).

For technical convenience, we will impose two extra requirements to a classical protocol P in order to be a valid *quantum protocol*. Before stating the new requirements, let us see an example. Quantum membership queries (or quantum black-box oracle calls) have been frequently used in the literature, see for instance [4, 6, 9, 10, 11, 12]. A quantum black-box oracle for function f in B_n transforms $(x, b) \in \{0,1\}^n \times \{0,1\}$ to $(x, b \oplus f(x))$. Thus, in the corresponding protocol the set of queries and the set of answers are $\{0,1\}^n \times \{0,1\}$. Valid answers to query (x, b) are (x, b') where $b' \in \{0,1\}$. So, tuples of the protocol are $((x, b), (x, b'))$ for all x in $\{0,1\}^n$ and for all b and b' in $\{0,1\}$. Consistent Boolean functions with answer (x, b') to query (x, b) are those functions that evaluates to $b' \oplus b$ on x. It is easy to see that this protocol satisfies the two requirements above.

Note that under this protocol we can think that queries (x, b) and (y, d) are equivalent whenever $x = y$. Intuitively, the valid answers of query $(x, 0)$ define the same set of consistent function sets as the answers of query $(x, 1)$. Observe also that, for any valid answer a, the consistent function sets defined by answer a to queries $(x, 0)$ and $(x, 1)$ are different. Moreover, if $x \neq y$ queries (x, b) and (y, d) do not share any answer.

We want that any quantum protocol preserve these properties pointed out for the membership case. The reason will be clear later. In general, given a classical protocol $P \subset Q \times A$, it defines an equivalence relation on

Q: queries q_i and q_j are related if their respective sets of consistent function sets defined by their (respective) valid answers coincide. If queries q_i and q_j belongs to the same equivalence class we say that they are equivalent ones. The equivalence class of query q is denoted by $[q]$ and the set of equivalence classes by $[Q]$.

The extra requirements that a classical protocol has to accomplish in order to be a quantum one impose some compatible behavior of the protocol with respect to the equivalence relation just defined. Formally, they are the following.

- If q_i and q_j are non-equivalent queries then they do not share any answer.
- If a is valid answer for two different queries q_i and q_j then the consistent sets of (q_i, a) and (q_j, a) are different.

One may wonder what the reason is to define protocols accepting different queries to have the same role because they define the same equivalence class. It is straightforward to see that this is useless in the classical learning scenario. However, accepting this behavior for a quantum protocol, it will allow us to define teachers that as quantum oracles, in addition to be unitary operators are also involutions, a property that one may wish to impose to a quantum oracle to allow proper interference to take place, as we have noted in Section 2.3.1.

Example 1. Several examples of quantum protocols:

Membership query protocols.

Equivalence queries protocol. Fixed a hypothesis class H, where H is a subset of B_n, queries and answers are tuples (h, x, b) belonging to $H \times \{0,1\}^n \times \{0,1\}$. Valid answers to query (h, x, b) are $(h, x \oplus y, b)$ for any $y \in \{0,1\}^n$ and $(h, x, 1 \oplus b)$. The consistent set corresponding to answer $(h, x \oplus y, b)$ are those Boolean functions f such that $f(y) \neq h(y)$. The consistent set of answer $(h, x, 1 \oplus b)$ has only a single element, the function h. Note that queries (h, x, b) and (g, z, d) are equivalent whenever $h = g$. It is straightforward to see that this defines a quantum protocol.

Restricted equivalence queries protocol.

Subset query protocol. Queries and answers are tuples (h, x, b) belonging to $H \times \{0,1\}^n \times \{0,1\}$. Valid answers to query (h, x, b) are $(h, x \oplus y, b)$ for any $y \in \{0,1\}^n$ such that $h(y) = 1$ and $(h, x, 1 \oplus b)$. The consistent set corresponding to answer $(h, x \oplus y, b)$ are those Boolean functions f such that $f(y) = 0$. The consistent set of answer $(h, x, 1 \oplus b)$ are those functions

f such that $f \geq h$. Note that queries (h, x, b) and (g, z, d) are equivalent whenever $h = g$. This defines a quantum protocol.

3.2 Quantum Teachers

Let $P \subset Q \times A$ be a quantum protocol. We associate to the set of queries Q a Hilbert space S_Q defined as follows. Vectors of S_Q are superpositions of query vectors $|q\rangle$ where q is a query of Q. The inner product of S_Q is the one defined by considering the set of query vectors $\{|q\rangle \mid q \in Q\}$ as an orthonormal basis. In a similar way, we also define a Hilbert space S_A corresponding to the set of answers A.

Let f be a Boolean function. A *quantum teacher* for f under protocol P is a unitary operator T transforming each basis query vector $|q\rangle$ to a superposition in S_A of valid answers according to P that are consistent with f. Quantum teacher T for f is said to be a *permutation teacher* whether it transforms each basis query $|q\rangle$ to a consistent basis answer $|a\rangle$. When $S_Q = S_A$ and the quantum teacher operator T holds that $T^2 = I$, we say that T is an *involution teacher*. Involution teachers shall correspond with involution oracle gates.

Example 2. Classical deterministic teachers for membership, equivalence, subset and other popular queries trivially define corresponding permutation teachers in the quantum setting. Note that they are also involutions teachers.

3.3 Query Complexity

A superposition $|\phi\rangle$ of an oracle quantum machine is said to be a *query superposition* if there is a configuration with nonzero amplitude in $|\phi\rangle$ whose state is the pre-query one. Let P be a quantum protocol. A concept class $C \subseteq B_n$ is learnable under protocol P with m query superpositions if there exists an oracle quantum Turing machine L –so-called learner– such that for any target function f in C and for any quantum teacher T for f under P:

1. L^T gets a final superposition and with probability at least $2/3$, outputs a circuit for f.
2. The computation of L^T yields at most m query superpositions.

For target class C and quantum protocol P we define the *quantum query complexity*, $\mathrm{QC}(C, P)$, as the smallest m such that C is learnable with m query superpositions under P. We note that this query complexity notion is consistent with the definition given in Beals *et al* [4] (see also Servedio *et al* [12]) for quantum networks.

3.4 Answering Schemes

Let $P \subseteq Q \times A$ be a quantum protocol. A subset $\mathcal{T}$ of P is said to be an answering scheme if:

1. For any query $q \in Q$ there is exactly one answer a such that (q, a) belongs to $\mathcal{T}$.
2. If (q_i, a_i) and (q_j, a_j) are tuples of $\mathcal{T}$ and q_i and q_j are equivalent queries then (q_i, a_i) and (q_j, a_j) define the same consistent set of Boolean functions.

The following lemma is an immediate consequence of the quantum protocol and the answering scheme definitions.

Lemma 3. *Answers of an answering scheme are all different.*

Thus, observe that answering schemes extend naturally to unitary transformations from S_Q to S_A and they can be considered as quantum oracle functions. However, for an answering scheme $\mathcal{T}$ it is possible that there is no function in B_n consistent with all tuples in $\mathcal{T}$. This contrast with the quantum teacher notion introduced above where there is always a consistent Boolean function with all teacher answers. As we will see later, answering schemes have an adversary role in our arguments in Section 4.2.

Let L be a quantum learner under protocol P and let $\mathcal{T}$ be a answering scheme of P. We consider the computation of L when oracle calls are solved according to $\mathcal{T}$ and we denote by $L^{\mathcal{T}}$ the resulting quantum oracle machine. Let $|\phi_i\rangle$ be the superposition of $L^{\mathcal{T}}$ at time i and let $w_q(|\phi_i\rangle)$ be the sum of squared amplitude modules of configurations in superposition $|\phi_i\rangle$ which are querying q; i.e. $w_q(|\phi_i\rangle) = \sum_c |\alpha_c|^2$ where the sum extends over configurations c querying q and α_c denotes the amplitude of c in $|\phi_i\rangle$. We refer to $w_q(|\phi_i\rangle)$ as the query magnitude of q in $|\phi_i\rangle$. We naturally extend this concept to query classes: $w_{[q]}(|\phi_i\rangle)$ is the sum of query magnitudes $w_{q'}(|\phi_i\rangle)$ where q' is any query equivalent to q.

For the specific case of membership queries Bennet *et al* (Theorem 3.3 in [6]) showed that the final outcome of L's computations cannot depend very much on the oracle's answers to queries of little magnitude. We extend this result to any quantum protocol in Theorem 5 below. We give a fully detailed proof for two reasons. First, we think that it is a non-trivial extension of the original theorem statement. Second, as we point out later, there is an incorrect assert in the proof given in [6]. We show first the following lemma where we assume some arbitrary underlying protocol.

Lemma 4. *Let $|\phi\rangle$ be a valid superposition of $L^{\mathcal{T}}$. Let $G \subseteq [Q]$ be a set of query classes and let $\tilde{\mathcal{T}}$ be any answering scheme that agrees with $\mathcal{T}$ on any query q such that $[q] \notin G$. Let U and $\tilde{U}$ be, respectively, the unitary time operators of $L^{\mathcal{T}}$ and $L^{\tilde{\mathcal{T}}}$. Then, $\|U|\phi\rangle - \tilde{U}|\phi\rangle\|^2 \leq 4\sum_{[q]\in G} w_{[q]}(|\phi\rangle)$.*

Proof. Let $|E\rangle = U|\phi\rangle - \tilde{U}|\phi\rangle$ be the error vector. Assume that $|\phi\rangle = \sum_{c\in I^G}\alpha_c c + |\varphi\rangle$ where I^G is the set of configurations querying some query equivalent to those defined by G and $|\varphi\rangle$ is a superposition of configurations with no query in G. Then,

$$\begin{aligned}\|E\|^2 \; = \; & \sum_{c,d\in I^G}\alpha_c\alpha_d^*\langle Uc|Ud\rangle + \sum_{c,d\in I^G}\alpha_c\alpha_d^*\langle \tilde{U}c|\tilde{U}d\rangle - \\ & \sum_{c,d\in I^G}\alpha_c\alpha_d^*\langle Uc|\tilde{U}d\rangle - \sum_{c,d\in I^G}\alpha_c\alpha_d^*\langle \tilde{U}c|Ud\rangle.\end{aligned}$$

In this expression, by orthogonality the first two summands are both equal to $\sum_{[q]\in G} w_{[q]}(|\phi\rangle)$. For the last two summands observe that all scalar products are zero except for those configurations c and d such that $Uc = \tilde{U}d$. Fixed a configuration c_0 there is at most one d_0 where this equality happens because the answers in a answering scheme are all different, see Lemma 3. Let J be the set of pairs (c_0, d_0) having this property. Then,

$$\sum_{c,d\in I^G}\alpha_c\alpha_d^*\langle Uc|\tilde{U}d\rangle + \sum_{c,d\in I^G}\alpha_c\alpha_d^*\langle \tilde{U}c|Ud\rangle = \sum_{(c_0,d_0)\in J}\alpha_{c_0}\alpha_{d_0}^* + \sum_{(c_0,d_0)\in J}\alpha_{d_0}\alpha_{c_0}^* =$$

$$\sum_{(c_0,d_0)\in J} 2\mathrm{Re}(\alpha_{c_0}\alpha_{d_0}^*) \leq \sum_{(c_0,d_0)\in J} 2|\alpha_{c_0}||\alpha_{d_0}^*| \leq \sum_{(c_0,d_0)\in J} |\alpha_{c_0}|^2 + |\alpha_{d_0}|^2 \leq 2\sum_{[q]\in G} w_{[q]}(|\phi\rangle).$$

Thus, $\|E\|^2 \leq 4\sum_{[q]\in G} w_{[q]}(|\phi\rangle$. □

We note that the proof of Theorem 3.3 in [6] states (see first line in the last paragraph of the proof) that $\|E\|^2 = 2\sum_{[q]\in G} w_{[q]}(|\phi\rangle$, that is a better characterization than the inequality given by Lemma 4. However, we provide a counterexample for this equality under the membership query protocol (which is the protocol considered in [6]) in Appendix A.

Theorem 5. *Let $|\phi_i\rangle$ be the superposition of $L^{\mathcal{T}}$ at time i. Let $\varepsilon > 0$. Let $F \subseteq \{0,\ldots,t-1\} \times [Q]$ be a set of time-query class pairs such that*

$\sum_{(i,[q])\in F} w_{[q]}(|\phi_i\rangle) \leq \frac{\varepsilon^2}{4t}$. *For each* i, *let* $\tilde{\mathcal{T}}_i$ *be any answering scheme that agrees with* $\mathcal{T}$ *on any query* q *such that* $(i,[q]) \notin F$. *Let* $|\tilde{\phi}_t\rangle$ *be the time* t *superposition that* L *will get if the answer to each query instance* $(i,[q]) \in F$ *is modified according to* $\tilde{\mathcal{T}}_i$. *Then,* $\||\phi_t\rangle - |\tilde{\phi}_t\rangle\| < \varepsilon$.

Proof. Let U and U_i, for $i = 0,\ldots,t-1$, be respectively the unitary time operators of $L^{\mathcal{T}}$ and $L^{\tilde{\mathcal{T}}_i}$. We define $|E_i\rangle$ to be the error vector in the ith step caused by replacing answering scheme T by $\tilde{\mathcal{T}}_i$. Then

$$|E_i\rangle = U|\phi_i\rangle - U_i|\phi_i\rangle.$$

So we have

$$\begin{aligned}|\phi_t\rangle &= U|\phi_{t-1}\rangle = U_{t-1}|\phi_{t-1}\rangle + |E_{t-1}\rangle = \ldots = \\ & U_{t-1}\ldots U_0|\phi_0\rangle + \textstyle\sum_{i=0}^{t-1} U_{t-1}\ldots U_{i+1}|\mathrm{E}_i\rangle.\end{aligned}$$

Since all of the U_i are unitary, $\|U_{t-1}\ldots U_{i+1}|\mathrm{E}_i\rangle\| = \||E_i\rangle\|$. By Lemma 4 we know the squared norm of each error vector $|E_i\rangle$ is bounded by $4\sum_{[q]\in F_i} w_y(|\phi_i\rangle)$ where set F_i is $\{[q] \mid (i,[q]) \in F\}$. Therefore,

$$\begin{aligned}\||\phi_t\rangle - |\tilde{\phi}_t\rangle\|^2 &\leq \left(\textstyle\sum_{i=0}^{t-1} \||E_i\rangle\|\right)^2 \leq \\ t\left(\textstyle\sum_{i=0}^{t-1} \||E_i\rangle\|^2\right) &\leq 4t\textstyle\sum_{(i,[q])\in F} w_{[q]}(|\phi_i\rangle) \leq \varepsilon^2\end{aligned}$$

□

4 The Query Complexity of Exact Learners

4.1 The General Halving Dimension

Let $C \subseteq B_n$ be a concept class and let P be a quantum protocol. We associate the parameter $\mathrm{ghdim}(V,P)$ to each subset V of C with $|V| > 1$. This parameter is the smallest non-negative integer d satisfying the following predicate: for any answering scheme $\mathcal{T}$ from P there exists a a subset $S \subseteq \mathcal{T}$ of cardinality d such that at most half of the functions in V are consistent whith all tuples in S. When there is no integer d satisfying the predicate, $\mathrm{ghdim}(V,P)$ is defined to be ∞. The *general halving dimension* of C under P, $\mathrm{GHdim}(C,P)$, is the maximum of parameters $\mathrm{ghdim}(V,P)$. Thus,

$$\mathrm{GHdim}(C,P) = \max\{\mathrm{ghdim}(V,P) \mid V \subseteq C \wedge |V| > 1\}.$$

The general halving dimension has two ancestors. One is the general dimension concept introduced in [2], where it is shown to be a nice characterization of the query complexity of deterministic learners in the ordinary learning scenareo. The other one is the halving complexity notion defined by Simon [14], that approximates the query complexity of randomized learners in the classical setting. We prove below several bounds for the query complexity in terms of the general halving dimension as much for quantum protocols as for classical ones.

4.2 A Lower Bound for the Quantum Query Complexity

Lemma 6. *Let us assume GHdim$(C, P) > l \geq 1$. There exists a set of concepts $V \subseteq C$ with $|V| > 1$ and an answering scheme $\mathcal{T}$ such that for any tuple $(q, a) \in \mathcal{T}$ less than $\frac{|V|}{l}$ concepts from V do not satisfy (q, a).*

Proof. For the sake of contradiction suppose that for each subset V of C with $|V| > 1$ and for any answering scheme $\mathcal{T}$ there exists a tuple $(q, a) \in \mathcal{T}$ such that at least $\frac{|V|}{l}$ concepts from V do not satisfy (q, a). Fix a subset of concepts $V = V_0$ and let $\mathcal{T}$ be an answering scheme. Thus, it corresponds to V_0 a tuple $(q_0, a_0) \in \mathcal{T}$ such that at least $\frac{|V_0|}{l}$ concepts from V_0 do not satisfy (q_0, a_0). Let V_1 the subset of V_0 consistent with (q_0, a_0). By assumption, $|V_1| \leq |V|(1 - 1/l)$. We repeat this process with V_1 instead of V_0 and so on and so forth. After l iterations we get a subset V_l of V with $|V_l| \leq |V|/2$. This implies that $\mathrm{ghdim}(V, P) \leq l$. $\square$

Let l be such that $1 \leq l < \mathrm{GHdim}(C, P)$ and let V and $\mathcal{T}$ be respectively the subset of C and the answering scheme promised by Lemma 6. Inspired by Servedio *et al* [12], we define the *difference matrix* M as the $|V| \times |Q|$ zero/one matrix where rows are indexed by concepts in V, columns are indexed by queries in Q, and $M_{f,q} = 1$ iff the Boolean function f is not consistent with the answer a of q in $\mathcal{T}$. By our choice of V and $\mathcal{T}$, each column of M has less than $\frac{|V|}{l}$ ones. Thus, the l_1 matrix norm of M is $\|M\|_1 < \frac{|V|}{l}$. The following lemma, which is a generalization of Lemma 6 from [12], shows that no quantum learning algorithm L with small query complexity can effectively distinguish many concepts in V.

Lemma 7. *Let L be a quantum learner with query complexity m. Let $\varepsilon > 0$. There are a set $W \subseteq V$ and quantum teachers T_f for concepts f in W such that:*

1. $|W| > |V|(1 - \frac{8m^2}{l\varepsilon^2})$

2. *If $|\phi^{T_f}\rangle$ denotes the final superposition of L^{T_f} then, for any pair of concepts f and g of W, it holds $\||\phi^{T_f}\rangle - |\phi^{T_g}\rangle\| < \varepsilon$.*

Proof. Let $\mathcal{T}$ be the answering scheme promised by Lemma 6. We define a permutation teacher T_f for each $f \in V$ in the following way. Teacher T_f answers to query q with the answer a such that $(q, a) \in \mathcal{T}$ whenever f is consistent with (q, a). Otherwise, any consistent basis answer is chosen in such a way that equivalent queries have equivalent answers. Note that such permutation teacher can always be constructed and it defines a valid answering scheme.

Let $|\phi_i^{\mathcal{T}}\rangle$ be the i-th query superposition of $L^{\mathcal{T}}$. Let $w(|\phi_i^{\mathcal{T}}\rangle) \in \mathbb{R}^{|Q|}$ be the $|Q|$-dimensional vector which has entries indexed by queries $q \in Q$ and which has $w_q(|\phi_i^{\mathcal{T}}\rangle)$ as its q-th entry.

Let $w_f(|\phi_i^{\mathcal{T}}\rangle)$ be the sum of all query magnitudes $w_q(|\phi_i^{\mathcal{T}}\rangle)$ where query q is such that f is not consistent with its corresponding tuple $(q, a) \in \mathcal{T}$. Note that $w_f(|\phi_i^{\mathcal{T}}\rangle)$ is the magnitude in superposition $|\phi_i^{\mathcal{T}}\rangle$ of those queries where answering schemes T_f and $\mathcal{T}$ are different. Moreover, observe that $Mw(|\phi_i^{\mathcal{T}}\rangle) \in \mathbb{R}^{|V|}$ is a $|V|$-dimensional vector whose f-th entry is precisely $w_f(|\phi_i^{\mathcal{T}}\rangle)$. Since $\|M\|_1 < \frac{|V|}{l}$ and $\|w(|\phi_i^{\mathcal{T}}\rangle)\|_1 \leq 1$ we have that $\|Mw(|\phi_i^{\mathcal{T}}\rangle)\|_1 < \frac{|V|}{l}$, i. e. $\sum_{f \in V} w_f(|\phi_i^{\mathcal{T}}\rangle) < \frac{|V|}{l}$. Hence

$$\sum_{i=1}^{m} \sum_{f \in V} w_f(|\phi_i^{\mathcal{T}}\rangle) < \frac{m|V|}{l}. \tag{1}$$

Let us define the subset of concepts $W = \{f \in V \mid \sum_{i=1}^{m} w_f(|\phi_i^{\mathcal{T}}\rangle) \leq \varepsilon^2/8m\}$. By inequality 1, we have $|V \backslash W| < \frac{8m^2|V|}{l\varepsilon^2}$. Finally, for any $f \in W$, Theorem 5 implies that $\||\phi_m^{T_f}\rangle - |\phi_m^{\mathcal{T}}\rangle\| < \varepsilon/2$. □

We can prove now a lower bound for the quantum query complexity

Theorem 8. *Let P be a quantum protocol and let C be a target concept class. The learning query complexity of C under P holds that*

$$QC(C, P) \geq \frac{\sqrt{GHdim(C, P)}}{32}.$$

Proof. It is trivial if $\mathrm{GHdim}(C, P)$ is at most 1. Otherwise, let $l = \mathrm{GHdim}(C, P) - 1$, let $m = \mathrm{QC}(C, P)$ and fix $\varepsilon = 1/8$. For the sake of contradiction assume that $m \leq \sqrt{l}/32$ and let L be a quantum learner achieving this query complexity. By applying Lemma 7 there are at least two target concepts f and g and, respectively, valid teachers T_f and T_g

such that the final superposition $|\phi^{T_f}\rangle$ (respectively, $|\phi^{T_g}\rangle$, of L^{T_f} (L^{T_g}) satisfies $\||\phi^{T_f}\rangle - |\phi^{T_g}\rangle\| < 1/8$. It is well known that performing the same observation on two quantum superpositions of Euclidean distance at most δ induces distributions which have total variation distance at most 4δ [7]. So, distributions $\mathcal{D}_f$ and $\mathcal{D}_g$ corresponding respectively to $|\phi^{T_f}\rangle$ and $|\phi^{T_g}\rangle$ are so that $\sum_{h\in B_n} |\mathcal{D}_f(h) - \mathcal{D}_g(h)| < 1/2$. This contradicts that L learns f and g. $\square$

We finally note that as teachers used in the proof of Theorem 8 are permutation teachers. Thus, for popular protocols as the ones in Example 1, the statement of this theorem is also valid even if only involution teachers are considered as valid oracle functions.

4.3 An Upper Bound for the Query Complexity

In this section we provide an upper bound for deterministic learners under classical protocols in terms of the general halving dimension. This immediately yields an upper bound for the quantum query complexity. The results below can be easily proved using arguments similar to those in [3, 2]. Here, P denotes any (classical) protocol.

Lemma 9. *Let $GHdim(C, P) = k$. Then, any subset V of C with $|V| > 1$ accomplish the following predicate. There exist a query q such that for any valid answer a at least $\frac{|V|}{2k}$ concepts from V do not satisfy (q, a).*

(Proof omitted).

From Lemma 9 we get an upper bound for the query complexity.

Theorem 10. *There is a deterministic leaner for the class C under protocol P whose query complexity is bounded by $2\ln|C|\,GHdim(C, P)$.*

(Proof omitted).

Note that Theorem 10 also applies to quantum protocols because any quantum protocol is also a classical one. Moreover, since reversible Turing machines can simulate any deterministic algorithm [5] the upper bound in Theorem 10 also applies to the quantum query complexity.

Corollary 11. *Let P be a quantum protocol. It holds that $QC(C, P) \leq 2\ln|C|\,GHdim(C, P)$*

4.4 The General Halving Dimension and the Query Complexity of Randomized Learners

We show below that the general halving dimension also provides a characterization of the query complexity of randomized learners (under classical protocols) that is slightly better than the characterization given for the quantum query complexity. The results in this section are straightforward extensions of results by Simon [14].

Given a classical protocol P and a target concept class C, Simon defines a *halving game* between two deterministic players and associates a complexity to each halving game, the *halving complexity*. It can be easily shown that GHdim exactly characterizes this complexity, as it is shown in [14] for the *halving dimension* function under the specific protocol of membership and equivalence queries. The halving dimension defined in [14] is nothing more than an incarnation of the general halving dimension for the membership and equivalence queries protocol, as it can be shown following arguments in [3]. Theorem 3.1 in [14] provides a lower bound of the query complexity of randomized learners in terms of the halving complexity. This theorem immediately yields the following lower bound in terms of the general halving dimension –where the constant is different from the one in the original version because Simon defines the query complexity as an expected value–.

Theorem 12. *Any randomized learner for the target class C under protocol P with success probability 2/3 makes at least $\frac{1}{4}GHdim(C, P)$ queries.*

This theorem, jointly with Theorem 10, gives the promised characterization for the query complexity of randomized learners.

5 Polynomial Learnability

We assume in this section some fixed underlying protocol. In order to discuss the polynomial learnability, we need to extend the concept class notion used until now. In this section a concept class C will be the union of former concept classes, i.e. $C = \cup_n C_n$ where C_n is a subset of B_n. We also need a fixed length notion l defined on concepts in C. For instance, the length can be the circuit size. In this case, the length of concept f, denoted by $l(f)$, is the length of the minimum circuit description for function f. We assume that length notions are so that at most 2^l concepts from C have length less than l.

Fixed a concept class $C = \cup_n C_n$ and a length notion l, a learner L for C and l is an algorithm that accomplish the following predicate. For each n and for any target concept $f \in C_n$, given as inputs $l(f)$ and n and provided that a valid teacher answers the queries according to f, the algorithm L learns f. Moreover, L is a polynomial query learner when its query complexity is bounded by a polynomial on $l(f)$ and n. A concept class is polynomially query learnable when it has a polynomial query learner.

The following theorem can be shown from Theorems 8 and 10 by using standard arguments.

Theorem 13. *If C is quantum polynomially learnable, then C also has a deterministic polynomial query learner.*

References

[1] D. Angluin. Queries and concept learning. *Machine Learning*, 2:319–342, 1988.

[2] J. L. Balcázar, J. Castro, and D. Guijarro. A general dimension for exact learning. In *Proceedings of the 14th Annual Conference on Computational Learning Theory*, volume 2111 of *LNAI*, pages 354–367. Springer, 2001.

[3] J. L. Balcázar, J. Castro, and D. Guijarro. A new abstract combinatorial dimension for exact learning via queries. *J. Comput. Syst. Sci.*, 64(1):2–21, 2002.

[4] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[5] C. H. Bennett. Logical reversibility of computation. *IBM Journal of Research and Development*, 17:525–532, 1973.

[6] C. H. Bennett, E. Bernstein, G. Brassard, and U. V. Vazirani. Strengths and weaknesses of quantum computing. *SIAM J. Comput.*, 26(5):1510–1523, 1997.

[7] E. Bernstein and U. V. Vazirani. Quantum complexity theory. *SIAM J. Comput.*, 26(5):1411–1473, 1997.

[8] N. H. Bshouty and J. C. Jackson. Learning DNF over the uniform distribution using a quantum example oracle. *SIAM Journal on Computing*, 28(3):1136–1153, 1999.

[9] D. Deutsch and R. Jozsa. Rapid solution of problems by quantum computation. *Proc Roy Soc Lond A*, 439:553–558, 1992.

[10] L. K. Grover. A fast quantum mechanical algorithm for database search. In *STOC*, pages 212–219, 1996.

[11] M. Hunziker, D. A. Meyer, J. Park, J. Pommersheim, and M. Rothstein. The geometry of quantum learning. *arXiv:quant-ph/0309059*, 2003. To appear in Quantum Information Processing.

[12] R. A. Servedio and S. J. Gortler. Equivalences and separations between quantum and classical learnability. *SIAM J. Comput.*, 33(5):1067–1092, 2004.

[13] D. R. Simon. On the power of quantum computation. *SIAM J. Comput.*, 26(5):1474–1483, 1997.

[14] H. U. Simon. How many queries are needed to learn one bit of information? *Annals of Mathematics and Artificial Intelligence*, 39:333–343, 2003.

[15] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, Nov. 1984.

A Appendix

Let $|\phi_0\rangle$ be the superposition

$$\frac{1}{\sqrt{2^{n+1}}} \sum_{x \in \{0,1\}^n} (|x0\rangle - |x1\rangle).$$

Note that $|\phi_0\rangle$ can be easily obtained from $|0^{n+1}\rangle$ by applying first the operator $H^{\otimes n+1}$ and then $I^{\otimes n} \otimes Z$, where H, Z and I are respectively the Hadamard, Pauli-Z, and Identity unitary operators. Let f be the singleton Boolean function that only evaluates to one on 0^n. If a membership query to f is done on superposition $|\phi_0\rangle$ we obtain as a result the superposition

$$|\phi_1\rangle = \frac{1}{\sqrt{2^{n+1}}} \left(|0^n 1\rangle - |0^n 0\rangle + \sum_{x \neq 0^n} |x0\rangle - |x1\rangle \right).$$

Assume now the membership query on $|\phi_0\rangle$ is done to the null function instead of f. This oracle call does not have any effect, so the superposition

$|\phi_1'\rangle$ after the query is equal to $|\phi_0\rangle$. Note that according Definition 3.2 in [6], $q_{0^n0}(|\phi_0\rangle) + q_{0^n1}(|\phi_0\rangle)$ is $\frac{1}{2^n}$. On the other hand, the error vector $|E\rangle = |\phi_1\rangle - |\phi_1'\rangle$ is

$$\frac{2}{\sqrt{2^{n+1}}}(|0^n1\rangle - |0^n0\rangle)$$

whose squared norm is $\frac{1}{2^{n-2}}$. This number is greater than $2(q_{0^n0}(|\phi_0\rangle + q_{0^n1}(|\phi_0\rangle))$, in contrast with the assert in the first line of the last paragraph of the proof of Theorem 3.3 in [6].

Note that for $n = 1$, $\varepsilon = 1/\sqrt{2}$, $F = \{(0, 00), (0, 01)\}$ and $T = 1$ and following the lines above, a counterexample for the theorem statement can be easily obtained.