CHICAGO JOURNAL OF THEORETICAL COMPUTER SCIENCE 2010, Article 13, pages 1–15
http://cjtcs.cs.uchicago.edu/

Single-Query Learning from Abelian and non-Abelian Hamming Distance Oracles

David A. Meyer James Pommersheim

Received: November 20, 2009; published: August 5, 2010.

Abstract: We study the problem of identifying an n-bit string using a single quantum query to an oracle that computes the Hamming distance between the query and hidden strings. The standard action of the oracle on a response register of dimension r is by powers of the cycle $(1 \dots r)$, all of which, of course, commute. We introduce a new model for the action of an oracle—by general permutations in S_r—and explore how the success probability depends on r and on the map from Hamming distances to permutations. In particular, we prove that when $r = 2$, for even n the success probability is 1 with the right choice of the map, while for odd n the success probability cannot be 1 for any choice. Furthermore, for small odd n and $r = 3$, we demonstrate numerically that the image of the optimal map generates a non-abelian group of permutations.

1 Introduction

Suppose we wish to identify an n-bit string a by querying an oracle that computes the Hamming distance of any query x from a. Previous work has shown that if the oracle returns the Hamming distance modulo 4, there is a quantum algorithm that identifies a with probability 1, using only a single query (as shown in [17]; see also [16]). On the other hand, if the oracle returns the Hamming distance modulo 2, there is no algorithm, either classical or quantum mechanical, that can identify a with probability greater than $1/2^{n-1}$, using any number of queries. [1] In the latter case, we can think of the oracle adding the Hamming distance into a two dimensional response register (so its remainder modulo 2 is all that matters), or we can think of the oracle adding a single bit—the least significant bit of the Hamming distance—into a two

[1] This follows from the fact that the weight of a modulo 2 partitions the set of n-bit strings into two subsets of size 2^{n-1}, with each element having even Hamming distance from the elements in the same subset and odd Hamming distance from the elements in the other subset.

DOI: 10.4086/cjtcs.2010.013

dimensional response register. The latter point of view might lead us to believe that the difficulty stems from the oracle returning only a single bit, compared to the two bits that it returns when it computes the Hamming distance modulo 4.

Our first, possibly surprising, result demonstrates that when n is even this belief is wrong—there is a quantum algorithm that takes a single bit from the Hamming distance computed by the oracle and identifies a with probability 1 using a single query. Knowing such an algorithm exists, our second result is perhaps equally surprising: when n is odd the original belief is at least partially correct—there is no probability 1 algorithm for finding a using any single bit of the Hamming distance. By "any single bit of the Hamming distance" we mean any function

$$g_a(x) = h\big(\mathsf{dist}(a,x)\big), \tag{1.1}$$

where $h : \{0,\dots,n\} \to \{0,1\}$. Both of these results involve *learning* (or failing to *learn*) an element from a set of binary functions of x, indexed by a, so they can be understood as solutions to problems in *computational learning theory* where the set is a *concept class* and its elements are *concepts* [1].

Combining our new results with the previous ones leads us to make two observations: the probability of correctly learning a depends on (1) the dimension of the response register and (2) how the oracle's response acts on this register. The first observation suggests generalizing the notion of concepts, which are binary functions, to Y-valued functions, for sets Y other than $\{0,1\}$.[2] The second observation motivates the main conceptual contribution of this paper—a new model for the action of quantum (and reversible classical) *non-abelian* oracles—the permutation model. In this model, we fix a response register $\mathbb{C}^R$, where R is a finite set, and assign to each possible reponse $y \in Y$ a permutation $\sigma_y \in S_R$ of the set R. More precisely, to implement an oracle which computes the (classical) function $f : X \to Y$, we are free to choose any map $\sigma : Y \to S_R$, and given this choice, the oracle acts on $\mathbb{C}^X \otimes \mathbb{C}^R$ by

$$\mathcal{O}(f)|x,b\rangle = |x,\sigma_{f(x)}(b)\rangle$$

When the oracle acts in the standard way, by adding the function value it computes into the response register, $\sigma(Y) \subseteq C_r \leq S_R$, where C_r is the cyclic group with $r = |R|$ elements, and is thus abelian. But this need not be the case: $\sigma(Y)$ can generate a non-abelian subgroup of S_R when $r > 2$, and for some problems the optimal solution has this property.

Our final set of results addresses the problem of *maximizing* the probability of success for odd n within this permutation model. We emphasize that although in this paper we study only Hamming distance oracles, *any* non-trivial oracle can be set up to have a non-abelian action, and this can improve the probability of success relative to an abelian action, as it does for the oracles we consider.

Finally, we note a similarity between the Hamming distance oracles considered in this paper and the game of Mastermind, a connection which was actually one of the authors' original motivations. In Mastermind, players make queries which receive two responses, one of which is equivalent to the Hamming distance between the query string and a hidden string [21].

[2] Limited versions of this generalization have been considered previously. See, for example, [2],[27].

2 Background

Most quantum algorithms include one or more calls to a subroutine or oracle that evaluates some function at the argument passed to it. In some cases, like Shor's algorithm [26] and the various quantum algorithms for hidden subgroup problems [22], the range of this function is a large set Y (so that, for example, the function can take distinct values on distinct cosets of the hidden subgroup). In others, like Grover's algorithm [12], the range of the function is only $\{0,1\}$.

In the latter cases, the problem of identifying the function can be recognized as a problem in computational learning theory [18]: The set of possible functions $\mathcal{C} \subseteq \{0,1\}^X$, where X is the domain of the function, is the *concept class*; each function $c : X \to \{0,1\}$ is a *concept*; and $c^{-1}(1) \subseteq X$ is the *extension* of the concept c. *Concept learning* is the process by which a student (the learner) identifies (or approximates) a target concept c from a concept class $\mathcal{C}$. In active learning the student can query a teacher for information about the target concept. Asking a teacher if $x \in X$ is in the extension of c is equivalent to passing x to a subroutine or oracle that evaluates c at its argument.

Many natural concept learning problems for which quantum algorithms have been found—including Grover's [12] UNSTRUCTURED SEARCH problem; Bernstein and Vazirani's [6], and Barg and Zhou's [3], SIMPLEX CODE DECODING problem; and Hunziker, *et al.*'s [18] BATTLESHIP and MAJORITY problems—are highly symmetric. In each of these $|\mathcal{C}| = |X|$ and there is an abelian group G acting transitively on $\mathcal{C}$ and on X, satisfying $(g \cdot c)(g \cdot x) = c(x)$ for all $c \in \mathcal{C}$, $x \in X$, and $g \in G$.

In this paper we consider problems which have this symmetry for $G = X = \mathbb{Z}_2^n$. Each involves a specific function of the Hamming distance between an unknown n-bit string $a \in \mathbb{Z}_2^n$ and a query string $x \in \mathbb{Z}_2^n$, $\mathsf{dist}(a,x) = |\{i \mid a_i \neq x_i\}|$; this is invariant under the action of G since $\mathsf{dist}(g+a, g+x) = \mathsf{dist}(a,x)$. Now, until it is composed with a binary function as in (1.1), $\mathsf{dist}(a,\cdot) : X \to \{0,\dots,n\} = Y$ does not define a traditional concept (except in the trivial case $n = 1$), so it is useful to define a *Y-valued concept class* to be a set of functions $\mathcal{C} \subseteq Y^X$. We extend our use of "learning problems" to include these cases. To do so, we introduce the notion of an *(n,r)-Hamming distance oracle*, which accepts queries $x \in \mathbb{Z}_2^n$ and then acts on an r-dimensional response register according to some function of the Hamming distance $\mathsf{dist}(a,x)$, for some fixed $a \in \mathbb{Z}_2^n$.[3] We will formalize this notion precisely in Section 5.

Our goal is to optimize single-query learning from such Hamming distance oracles, i. e., to maximize the probability of correctly identifying a after a single call to the subroutine that computes the function. Since we assume a uniform prior distribution on a, we consider only quantum algorithms that begin with an *equal superposition query*,[4] i. e., that pass to the oracle a state of the form $|\eta^0\rangle \otimes \psi = H^{\otimes n}|0\dots0\rangle \otimes \psi$, where H is the Hadamard transformation $\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)/\sqrt{2}$ and $\psi \in \mathbb{C}^r$. If $\mathcal{O}(a) : (\mathbb{C}^2)^{\otimes n} \otimes \mathbb{C}^r \to (\mathbb{C}^2)^{\otimes n} \otimes \mathbb{C}^r$ denotes the action of the oracle with parameter a, the problem reduces to identifying which of the 2^n states $\mathcal{O}(a)|\eta^0\rangle \otimes \psi$ is returned by the oracle. An optimal solution to this problem can be obtained by a complete von Neumann measurement [14],[13],[20],[15]; equivalently, we want to maximize

$$\sum_{a=0}^{2^n-1} \sum_{b=0}^{r-1} \left| \langle a,b|U\mathcal{O}(a)|\eta^0\rangle \otimes \psi \right|^2, \tag{2.1}$$

[3] It is also natural to consider problems with $X = \mathbb{Z}_k^n$ [17], in which the Hamming distance is defined by the same formula, but in this paper we restrict our attention to $k = 2$.

[4] In fact, we conjecture that for problems with transitive group actions and uniform priors, the optimal solutions always include one that begins with an equal superposition query.

over all unitary maps $U \in U(2^n r)$ and states $\psi \in \mathbb{C}^r$. We remark that the literature cited establishes the fact a complete von Neumann measurement (instead of a more general POVM) is sufficient for optimality in the case that the state vectors $\mathcal{O}(a)|\eta^0\rangle \otimes \psi$ are linearly independent. For the case of linearly dependent state vectors, an additional argument is needed (see Section 6).

3 Using a different bit of the Hamming distance

We begin by considering the problem of learning an n-bit string from an oracle that returns the *second* least significant bit of the Hamming distance of a query, rather than the least significant bit as in [17]. To be precise, let n be a natural number, and for any $a \in \mathbb{Z}_2^n$, define a function $f_a : \mathbb{Z}_2^n \to \{0,1\}$ by

$$f_a(x) = \begin{cases} 0 & \text{if } \mathsf{dist}(a,x) \equiv 0,1 \pmod 4; \\ 1 & \text{if } \mathsf{dist}(a,x) \equiv 2,3 \pmod 4. \end{cases}$$

Thus $f_a(x)$ is the second least significant bit of the Hamming distance between a and x. Set $b_1(d)$ to be the second least significant bit of a nonnegative integer d, so $f_a(x) = b_1\big(\mathsf{dist}(a,x)\big)$. Define $\mathcal{C}_n$ to be the concept class $\{f_a \mid a \in \mathbb{Z}_2^n\}$.

Lemma 3.1. *If $n \not\equiv 1 \pmod 4$ then $|\mathcal{C}_n| = 2^n$. If $n \equiv 1 \pmod 4$ then $f_a = f_{\bar{a}}$, where $\bar{a}$ is the bitwise complement of $a \in \mathbb{Z}_2^n$, so there are only 2^{n-1} concepts in the class.*

Proof. Suppose $f_{a'} = f_a$ and $\mathsf{dist}(a,a') = d$. Since $b_1(d) = b_1\big(\mathsf{dist}(a',a)\big) = f_{a'}(a) = f_a(a) = b_1\big(\mathsf{dist}(a,a)\big) = b_1(0) = 0$, we must have $d \equiv 0$ or $1 \pmod 4$. If $a' \neq a$ there is a bit at which a' differs from a. Let x be the bit string obtained from a by complementing this bit. Then $b_1\big(\mathsf{dist}(a,x)\big) = b_1(1) = 0$ so $b_1\big(\mathsf{dist}(a',x)\big) = b_1(d-1) = 0$, so we can conclude that $d \equiv 1 \pmod 4$. Now suppose there were a bit at which a' agreed with a. Let y be the bit string obtained from a by complementing this bit. Then $b_1\big(\mathsf{dist}(a,y)\big) = b_1(1) = 1$ and $b_1\big(\mathsf{dist}(a',y)\big) = b_1(d+1)$, which would imply that $d \equiv 0 \pmod 4$, a contradiction. So if $a' \neq a$ but $f_{a'} = f_a$, there can be no bit at which a' agrees with a, which means $a' = \bar{a}$ and $n \equiv 1 \pmod 4$. □

As we explained in the previous sections, we are interested in analyzing the probability of correctly identifying the hidden bit string a using only a single query to the oracle. Classically, it is not hard to see that when the f_a are distinct, any strategy yields a worst-case success probability of at most $2/2^n = 1/2^{n-1}$, the number of possible oracle responses divided by the number of concepts. In contrast, we next show that for even n, this learning problem can be solved quantum mechanically with probability 1 using a single query.

Theorem 3.2. *Let n be even. Then the learning problem defined by $\mathcal{C}_n$ can be solved with probability 1 using a single quantum query.*

We will prove Theorem 3.2 by giving an explicit algorithm below. To show that the algorithm is correct we will need two lemmas. For $x \in \mathbb{Z}_2^n$, define $\hat{x} \in \mathbb{Z}_2^n$ by:

$$\hat{x} = \begin{cases} x & \text{if wt(x) is even;} \\ \bar{x} & \text{if wt(x) is odd.} \end{cases}$$

Here the *weight* of x, $\mathsf{wt}(x) = \mathsf{dist}(0,x)$. Note that if n is even, then the function $x \mapsto \hat{x}$ is a permutation of $\mathbb{Z}_2^n$.

Lemma 3.3. *Let n be a natural number and let $a,x \in \mathbb{Z}_2^n$. Then*

$$a \cdot x + \mathsf{wt}(a)\mathsf{wt}(x) \equiv a \cdot \hat{x} \pmod 2.$$

Proof. If $\mathsf{wt}(x)$ is even, then $\hat{x} = x$, and the congruence is easily seen to hold. If $\mathsf{wt}(x)$ is odd, then $\hat{x} = \bar{x}$, and the congruence follows from the identity $a \cdot x + a \cdot \bar{x} = \mathsf{wt}(a)$. □

Lemma 3.4. *Let n be a natural number and let $a,x \in \mathbb{Z}_2^n$. Then*

$$(-1)^{b_1(\mathsf{dist(a,x)})} = (-1)^{b_1(\mathsf{wt(a)})}(-1)^{b_1(\mathsf{wt(x)})}(-1)^{a\cdot\hat{x}}.$$

Proof. First note that for any integer d,

$$b_1(d) \equiv \frac{d(d-1)}{2} \pmod 2.$$

Since $\mathsf{dist}(a,x) = \mathsf{wt}(a) + \mathsf{wt}(x) - 2(a \cdot x)$, this implies

$$b_1\big(\mathsf{dist}(a,x)\big) \equiv \frac{\big(\mathsf{wt}(a) + \mathsf{wt}(x) - 2(a\cdot x)\big)\big(\mathsf{wt}(a) + \mathsf{wt}(x) - 2(a\cdot x) - 1\big)}{2} \pmod 2.$$

Expanding the numerator on the right hand side of this congruence, and dropping multiples of 4, gives

$$b_1(\mathsf{dist}(a,x)) \equiv \frac{\mathsf{wt}(a)^2 - \mathsf{wt}(a) + \mathsf{wt}(x)^2 - \mathsf{wt}(x)}{2} + \mathsf{wt}(a)\mathsf{wt}(x) + a \cdot x \pmod 2.$$

Using Lemma 3.3, we can replace $\mathsf{wt}(a)\mathsf{wt}(x) + a\cdot x$ with $a \cdot \hat{x}$ and the result follows. □

Proof of Theorem 3.2. We take the oracle to act on $(\mathbb{C}^2)^{\otimes n} \otimes \mathbb{C}^2$ in the standard way,

$$\mathcal{O}(a) : |x\rangle|b\rangle \mapsto |x\rangle|b + f_a(x)\rangle,$$

although it is $b_1\big(\mathsf{dist}(a,x)\big)$ that is being added modulo 2 into the response register, not $\mathsf{dist}(a,x)$. The following quantum algorithm identifies a with probability 1, applying $\mathcal{O}(a)$ only once.

Algorithm A.

1. Initialize the state to $|0\ldots0\rangle|0\rangle \in (\mathbb{C}^2)^{\otimes n} \otimes \mathbb{C}^2$.

2. Apply the unitary transformation $H^{\otimes n} \otimes HX$, where $X = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. This produces the state

$$|\eta^0\rangle|-\rangle = \frac{1}{2^{n/2}} \sum_{x \in \mathbb{Z}_2^n} |x\rangle|-\rangle,$$

where $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$.

3. Let D be the diagonal matrix acting on $(\mathbb{C}^2)^{\otimes n}$ by $D|x\rangle = (-1)^{b_1(\mathsf{wt(x)})}|x\rangle$. Apply $D \otimes I$, producing the state
$$\frac{1}{2^{n/2}} \sum_{x \in \mathbb{Z}_2^n} (-1)^{b_1(\mathsf{wt(x)})}|x\rangle|-\rangle.$$

4. Apply the oracle $\mathcal{O}(a)$. This produces the state
$$\frac{1}{2^{n/2}} \sum_{x \in \mathbb{Z}_2^n} (-1)^{b_1(\mathsf{wt(x)})}(-1)^{b_1(\mathsf{dist(a,x)})}|x\rangle|-\rangle.$$
By Lemma 3.4, this equals
$$\frac{1}{2^{n/2}} \sum_{x \in \mathbb{Z}_2^n} (-1)^{b_1(\mathsf{wt(a)})}(-1)^{a\cdot\hat{x}}|x\rangle|-\rangle.$$

5. Let P be the permutation matrix acting on $(\mathbb{C}^2)^{\otimes n}$ by $P|x\rangle = |\hat{x}\rangle$. Applying $P \otimes I$ yields
$$\frac{1}{2^{n/2}} \sum_{x \in \mathbb{Z}_2^n} (-1)^{b_1(\mathsf{wt(a)})}(-1)^{a\cdot\hat{x}}|\hat{x}\rangle|-\rangle,$$
which is equal to
$$\frac{(-1)^{b_1(\mathsf{wt(a)})}}{2^{n/2}} \sum_{x \in \mathbb{Z}_2^n} (-1)^{a\cdot x}|x\rangle|-\rangle,$$
since $x \mapsto \hat{x}$ is a bijection.

6. Apply $H^{\otimes n} \otimes I$. This produces the state $(-1)^{b_1(\mathsf{wt(a)})}|a\rangle|-\rangle$.

7. Now measure the query register (the $(\mathbb{C}^2)^{\otimes n}$ tensor factor) and observe a with probability 1.

□

4 Concept classes that cannot be learned with a single query

Theorem 3.2 cannot be extended to odd $n > 1$; there is no single equal superposition query probability 1 quantum learning algorithm for the concept class $\mathcal{C}_n$ in this case. In fact, when $n > 1$ is odd, there is no concept class defined by any function of the Hamming distance that is perfectly learnable with a single equal superposition quantum query. To see this, we begin with the following lemma:

Lemma 4.1. *Let $\mathcal{C}$ be a concept class of size M over a set X of size N. Suppose that there is a probability 1 learning algorithm using a single equal superposition quantum query. Identifying concepts with bitstrings indexed by X, there exists an integer $d \geq N/2$ such that any two distinct concepts of $\mathcal{C}$ have Hamming distance d. If $M = N > 2$ is even, then the quantum learning matrix, which has entries $L_{xc} = (-1)^{c(x)}$ for $x \in X$, $c \in \mathcal{C}$, is a Hadamard matrix.*

Proof. Suppose that there is a single query learning algorithm with equal superposition query

$$\frac{1}{\sqrt{N}}\sum_{x\in X}|x\rangle\otimes\psi,$$

for some unit vector $\psi\in\mathbb{C}^2$. If $\lambda=\psi^\dagger X\psi$ then $-1\le\lambda\le 1$.[5] Let A be the matrix whose columns, indexed by concepts, contain the state of the system after querying the oracle. Then $B=A^\dagger A$ is a matrix whose rows and columns are both indexed by concepts, with elements

$$B_{cc'}=\frac{1}{N}\sum_{x\in X}\begin{cases}1 & \text{if } c(x)=c'(x);\\ \lambda & \text{if } c(x)\neq c'(x).\end{cases}$$

Thus $NB_{cc'}=\big(N-\mathsf{dist}(c,c')\big)+\lambda\,\mathsf{dist}(c,c')$. Since the algorithm succeeds with probability 1, we must have $B_{cc'}=0$ for distinct concepts $c\neq c'$. In this case

$$d=\mathsf{dist}(c,c')=\frac{N}{1-\lambda}\geq\frac{N}{2},$$

where the inequality follows from $\lambda\geq -1$.

Now suppose that $M=N>2$ is even. Note that the the concepts of $\mathcal{C}$ form a code of distance d. Hence if $d>N/2$, then the Plotkin bound [23] implies that

$$M\leq 2\left\lfloor\frac{d}{2d-N}\right\rfloor.$$

Since N is even, $2d-N\geq 2$, and it follows that $M<N$ unless $d=N$, in which case $M\leq 2$. Thus we must have $d=N/2$ so the columns of L are orthogonal. That is, if $M=N>2$ is even, the quantum learning matrix is a Hadamard matrix. □

We now use Lemma 4.1 to prove the general result:

Theorem 4.2. *Let $n>1$ be odd. Suppose that $\mathcal{E}_n=\{g_a\mid a\in\mathbb{Z}_2^n\}$, where the functions $g_a:\mathbb{Z}_2^n\to\mathbb{Z}_2$ have the property that $g_a(x)$ depends only on the Hamming distance $\mathsf{dist}(a,x)$. If $|\mathcal{E}_n|=2^n$, then the learning problem defined by $\mathcal{E}_n$ cannot be solved with probability 1 using a single equal superposition quantum query.*

Note that if $|\mathcal{E}_n|\neq 2^n$, then a is not determined by g_a. Thus, in general, when n is odd, the bitstring a cannot be learned with probability 1 in a single quantum query from any binary-valued function of the Hamming distance.

Proof. Since $g_a(x)$ depends only on the Hamming distance $\mathsf{dist}(a,x)$, there exists a function $h:\{0,\ldots,n\}\to\{0,1\}$ such that $g_a(x)=h\big(\mathsf{dist}(a,x)\big)$.

Suppose that the learning problem defined by $\mathcal{E}_n$ can be solved with probability 1 using a single quantum query. Then by Lemma 4.1, the quantum learning matrix L, with elements $L_{xa}=(-1)^{g_a(x)}$, is a

[5] This X is the bit-flip matrix defined in step 2 of Algorithm A, not the set over which the concept class is defined.

Hadamard matrix. Consider the inner product of the two rows of L corresponding to the queries $y = 0^n$ and $z = 1^2 0^{n-2}$. Since L is a Hadamard matrix,

$$\sum_{a \in \mathbb{Z}_2^n} (-1)^{g_a(y)} (-1)^{g_a(z)} = 0.$$

In half of the terms of this sum, those for which the bits a_0 and a_1 differ, $\mathsf{dist}(a,y) = \mathsf{dist}(a,z)$. Then $g_a(y) = g_a(z)$, and hence each of these terms contributes $+1$ to the sum. In the other half of the terms, those for which $a_0 = a_1$, each term must contribute -1 to the sum, so $g_a(y) \equiv g_a(z) + 1 \pmod 2$. But $a_0 = a_1$ implies $\mathsf{dist}(a,y) = \mathsf{dist}(a,z) \pm 2$. It follows that for any $d \in \{0,\ldots,n-1\}$, $h(d) \neq h(d+2)$. Hence for some $s \in \{0,1,2,3\}$, $h(d) = b_1(d+s)$ for all $d \in \{0,\ldots,n\}$. Thus under the assumption that the concept class can be learned with probability 1 from a single quantum query, we have shown that h is a translate of b_1.

It remains to show that if n is odd, taking h to be a translate of b_1 leads to a matrix L that is not a Hadamard matrix. One easily sees that for such a function h, there is a sign $\varepsilon = \pm 1$ such that

$$(-1)^{h(n-d)} = \varepsilon (-1)^{h(d)}$$

for all d. It follows that any two rows of L corresponding to complementary values of x are equal up to sign. Hence L is not a Hadamard matrix. □

When $n \equiv 3$ (mod 4), the concept class $\mathcal{C}_n$ we introduced in the previous section satisfies the hypotheses of Theorem 4.2, so it cannot be learned with probability 1 from a single quantum query. When $n \equiv 1$ (mod 4), Lemma 3.1 tells us that the concept class has only 2^{n-1} concepts so Theorem 4.2 does not apply to learning the concept classe $\mathcal{C}_n$ in this case. We already know in this case that a cannot be identified with probability greater than $1/2$ with any number of queries, since $f_a = f_{\bar{a}}$. Using Algorithm A (with appropriate minor modifications), however, a single query determines a up to complementation, so the *concept class* $\mathcal{C}_n$ can be learned with a single quantum query.

Notice that we did not use the fact that n is odd to reach the conclusion that h is a translate of b_1. This means that for even n, the Hamming distance concept class $\mathcal{C}_n$ is essentially the only one that can be learned with probability 1 using a single query. More precisely, we have:

Corollary 4.3. *When n is even, b_1 (and translates) are the only functions of Hamming distance that yield a concept class learnable with probability* 1 *using a single quantum query.*

5 The permutation model

The results of the previous section demonstrate that an n-bit string a cannot be learned with probability 1 using a single quantum query to an $(n,2)$-Hamming distance oracle, when n is odd. A natural question, then, is:

> What is the largest probability with which a can be learned using a single quantum query to an $(n,2)$-Hamming distance oracle?

Furthermore, although previous work has shown that a can be learned with probability 1 from an $(n,4)$-Hamming distance oracle [17], neither that work nor our results to this point address the potential for learning with a 3-dimensional response register. So there is a second natural question:

> What is the largest probability with which a can be learned using a single quantum query to an $(n,3)$-Hamming distance oracle?

Before answering these questions, we reconsider the formulation of oracle algorithms.

To allow comparison with the classical query complexity of oracle (learning) problems, the action of the oracle in a quantum algorithm must be the linear extension of a classical reversible operation. In Deutsch's [9] and Deutsch and Jozsa's [10] original quantum algorithms for oracle problems, the oracle acts on $(\mathbb{C}^2)^{\otimes n} \otimes \mathbb{C}^2$ by

$$\mathcal{O}(c)|x,b\rangle = |x,b+c(x)\rangle, \tag{5.1}$$

where the sum is computed modulo 2, but the second register is initialized to $|0\rangle$, so the action has the effect of simply writing the function value computed by the oracle into that register. Similarly, in quantum algorithms for hidden subgroup problems [19] the oracle computes a function that is constant on cosets of the hidden subgroup, and takes distinct values on distinct cosets, so it acts on $\mathbb{C}^N \otimes \mathbb{C}^r$, where N is the size of the group and r is the number of distinct cosets, by (5.1), where the sum is computed modulo r. Again the second register is initialized to $|0\rangle$ so this also has the effect of merely writing the function value into that register.

Cleve, *et al.*

[6] noticed that the success probability of Deutsch's original algorithm could be improved to 1 by initializing the response register in the state $|-\rangle$, thereby taking advantage of the action (5.1) when $b=1$ as well as when $b=0$, to "kick back" a phase of $(-1)^{c(x)}$ [8]. Algorithm A does the same thing. This application, as opposed to the application on the response register initialized to $|0\rangle$, emphasizes that $\mathcal{O}(c)$ acts as a map on $\{0,1\}$—the (labels of the) computational basis vectors of the $\mathbb{C}^2$ response register—and is a classical reversible operation for each of the possible values of $c(x)$: 0 acts as the identity and 1 acts to exchange 0 and 1. That is, the oracle response, both classically and quantum mechanically, can be thought of as an element of S_2, the permutations of a two element set—it is either the identity, (1), or the other element of S_2, the transposition (12) (using *cycle notation* [7] for permutations of the elements of R, which we label $\{1,\ldots,r\}$). From this point of view, the action of an $(n,2)$-Hamming distance oracle depends on a map $\{0,\ldots,n\} \to S_2$: Simply adding the Hamming distance into the response register would be the map $d \mapsto (12)^d$, while the Algorithm A oracle action comes from the map $d \mapsto (12)^{b_1(d)}$.

But this implies a novel conceptualization of the action of an oracle when $r>2$, as it can be for (n,r)-Hamming distance oracles, namely that the action should depend on a map $\sigma : \{0,\ldots,n\} \to S_r$ which takes each function value computed by the oracle and associates to it a permutation of a *response set* R with $|R|=r$. In a quantum algorithm, R is identified with the computational basis of the tensor factor used as the response register. The map σ can be more complicated than $d \mapsto (12\ldots r)^d$, i. e., addition of the Hamming distance modulo r. This simple action can be characterized an *abelian oracle* since the range of σ is contained in a cyclic subgroup of S_r. It allows a to be identified with probability 1 when $r=4$ [17], but in other cases there is no reason to think that it is the optimal action. In general we

[6] And Tapp, according to a note in [8], and most likely others as well.

should consider *non-abelian oracles*, ones for which the image of σ contains noncommuting permutations of R. More precisely, we define the action of an oracle on $(\mathbb{C}^2)^{\otimes n} \otimes \mathbb{C}^r$ by

$$\mathcal{O}_\sigma(a)|x,b\rangle = |x, \sigma_{\mathsf{dist(a,x)}}(b)\rangle, \tag{5.2}$$

and let

$$p_n(r) = \max_{\substack{\sigma:\{0,\ldots,n\}\to S_r \\ \psi\in\mathbb{C}^r,\, U\in U(2^n r)}} \sum_{a=0}^{2^n-1}\sum_{b=0}^{r-1}\Big|\langle a,b|U\mathcal{O}_\sigma(a)|\eta^0\rangle\otimes\psi\Big|^2. \tag{5.3}$$

Using this notation, Hunziker and Meyer's result [17] shows that $p_n(r) = 1$ for $r \geq 4$, Theorem 3.2 shows that $p_{2j}(r) = 1$ for $r \geq 2$, and Lemma 3.1 and Theorem 4.2 show that $p_{2j-1}(2) < 1$, for j any natural number. Furthermore, the two questions above can be phrased as: What are $p_{2j-1}(2)$ and $p_{2j-1}(3)$, respectively?

6 Numerical optimization results

We are considering learning algorithms that send a single equal superposition query $|\eta^0\rangle\otimes\psi$ to the oracle. If the states $\{\mathcal{O}(c)|\eta^0\rangle\otimes\psi \mid c\in\mathcal{C}\}$ are linearly independent, then the optimal measurement to distinguish them, i. e., to identify c, is the *square root measurement*, as Sasaki, *et al.* noted [24] using early results in quantum state discrimination (Appendix A in [15]).[7] Thus we have the following:

Proposition 6.1. *Let $\mathcal{C}$ be a Y-valued concept class of size M over a set X of size N. Fix a response set R and an assignment σ of a permutation of R to each $y\in Y$. Also fix the initial state ψ of the response register $\mathbb{C}^R$. Let B be the Nr by M matrix with columns indexed by the concepts $c\in\mathcal{C}$, and with column c the state $\mathcal{O}(c)|\eta^0\rangle\otimes\psi$. Suppose that the columns of B are linearly independent, and that the diagonals of $G = B^\dagger B$ are constant. Let $\sqrt{G}$ denote the positive semi-definite square root of G. Then the optimal single-query quantum algorithm using the equal superposition query $|\eta^0\rangle\otimes\psi$ succeeds with probability* $\mathrm{diag}(\sqrt{G})$.

This proposition justifies the main step in the following numerical method.

Method B.

1. Input n and r.
2. Repeat Steps **3** and **4** below for all possible assignments $\sigma : \{0,\ldots,n\}\to S_r$.
3. For $\psi\in\mathbb{C}^r$ a unit vector, Proposition 6.1 allows us to calculate the maximal success probability $M(\psi)$ of a single-query quantum algorithm using the query $|\eta^0\rangle\otimes\psi$.
4. Numerically maximize $M(\psi)$ over all unit vectors $\psi\in\mathbb{C}^r$.

[7] The introduction of this approach into the context of concept learning may be found in [18].

Using this method we obtain the following numerical results:

First, let $n = 3$.

For $r = 2$, we find $p_3(2) \approx 0.800$. This is achieved using the permutations $\sigma_0 = \sigma_2 = \sigma_3 = (1)$ and $\sigma_1 = (12)$. It can also be achieved using the permutations $\sigma_0 = \sigma_1 = \sigma_2 = (1)$ and $\sigma_3 = (12)$.

When $r = 3$, this improves to $p_3(3) \approx 0.974$. Here a best permutation assignment is $\sigma_0 = (1)$, $\sigma_1 = (12)$, $\sigma_2 = (132)$, and $\sigma_3 = (123)$. (There are several other assignments of permutations that yield the same success probability.)

Second, let $n = 5$.

When $r = 2$, we find $p_5(2) \approx 0.721$. This is achieved using the permutations $\sigma_0 = \sigma_3 = \sigma_4 = \sigma_5 = (1)$ and $\sigma_1 = \sigma_2 = (12)$.

When $r = 3$, this improves to $p_5(3) \approx 0.955$. Here the best permutation assignment is $\sigma_0 = (1)$, $\sigma_1 = (123)$, $\sigma_2 = (132)$, $\sigma_3 = (12)$, $\sigma_4 = (1)$, and $\sigma_5 = (123)$. The optimum initialization for the response register is approximately

$$|1\rangle - 0.1065i|2\rangle + 1.1064i|3\rangle,$$

normalized to have unit length.

Note that Proposition 6.1 requires the columns of the matrix B to be linearly independent. In cases that the the columns of B are are linearly dependent, Proposition 6.1 does not tell us what to do, and Method B may not succeed in finding the optimal solution. In our problem it turns out that certain assignments of permutations lead to matrices B with linearly dependent columns. One suspects that these assignments are not as good as the assignments for which B has full rank, but this is not guaranteed by Proposition 6.1. In particular, when the rank of B is low, it is generally true that it is impossible to distinguish these states with high probability:[8]

Lemma 6.2. *Suppose ψ_i, $i \in \{1, \ldots, n\}$ are pure states contained in a k-dimensional subspace W. Then any n-valued measurement for identifying i succeeds with probability at most k/n.*

Proof. Let ρ_i be the density matrix corresponding to ψ_i. Let Π_W denote projection onto W. Then $\rho_i \leq \Pi_W$ for all i. Hence, if $\{X_i\}$ is any measurement, we can bound the success probability of this measurement as follows:

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{Tr}(X_i\rho_i) \leq \frac{1}{n}\sum_{i=1}^{n} \mathrm{Tr}(X_i\Pi_w) = \frac{1}{n}\mathrm{Tr}(\Pi_W) = \frac{k}{n}.$$

[8] This is a broadly applicable result that may well exist in the literature, but we have been unable to find it.

□

Lemma 6.2 suffices to guarantee that cases in which B has linearly dependent columns yield success probabilities that are smaller than the ones presented in the list above. When $n = 3$ (for both $r = 2$ and $r = 3$), we find that a given assignment of permutations either leads to a matrix B that is full rank (rank 8) for a generic choice of ψ, or has rank at most 5. In this latter case, Lemma 6.2 implies that the success probability is at most $5/8$, which is smaller that the probabilities shown above in the full rank case. When $n = 5$, B has either full rank (rank 32), or rank at most 22, which implies a success probability of at most $22/32$ in the linearly dependent case. Again, this is smaller than the numbers reported above for the linearly independent case.[9]

7 Conclusions

We have introduced a novel generalization for the action of oracles in quantum (and reversible classical) algorithms: the permutation model. For n-bit Hamming distance oracles the action is specified by a choice of map $\sigma : \{0,\ldots,n\} \to S_r$ when the response register has dimension r. The standard additive action of the oracle is described by the map $\sigma(d) = (1\ldots r)^d$. Algorithm A in Theorem 3.2 demonstrates the striking improvement possible by an oracle that acts by some other map of Hamming distances to permutations: for $r = 2$ the success probability of learning from a single query to an oracle that acts by the additive action is $1/2^{n-1}$, while for any even n it is 1 for an oracle that acts by $\sigma(d) = (12)^{b_1(d)}$, and for $n = 3$ and $n = 5$ it is approximately 0.800 and 0.721, respectively, using the actions listed in §6.

Allowing a larger response register, namely $r = 3$, improves the latter two probabilities to approximately 0.974 and 0.955, respectively. In general, $p_n(r)$ is a nondecreasing function of r. One might guess that if there is enough room in the response register to encode each possible function value $y \in Y$ as a distinct permutation of $\{1,\ldots,r\}$, then adding additional dimensions to the response register would not improve the success probability. This guess would mean that $p_n(r)$ would be constant for $r! \geq n+1$. This is not the case, however, as the $n = 3$ results show: $p_3(3) < 1$ while $p_3(4) = 1$.

As this counterexample indicates, the permutation model raises a host of new questions. We close by listing a few more: Is there some dimension for the response register above which $p_n(r)$ is constant? Perhaps $n+1$? What happens to $p_{2j-1}(r)$ as $j \to \infty$ for fixed r? Does it decrease to $1/2$? Or to something larger? What constitutes a good, or optimal, choice of permutations and initial response register state?

[9] An alternative approach to dealing with the case of linearly dependent state vectors would be to use a result of [4] which states that in general the success probability of the optimal measurement is bounded above by the square root of the success probability of the square root measurement.

References

[1] D. ANGLUIN: Computational learning theory: Survey and selected bibliography. In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*, pp. 351–369, New York, 1992. ACM. 2

[2] D. ANGLUIN AND D. K. SLONIM: Randomly fallible teachers: learning monotone dnf with an incomplete membership oracle. *Machine Learning*, 14:7–26, 1994. 2

[3] A. BARG AND S. ZHOU: A quantum decoding algorithm for the simplex code. In *Proceedings of the 36th Annual Allerton Conference on Communication, Control and Computing*, pp. 359–365. UIUC, 1998. Monticello, IL, 23–25 September 1998. 3

[4] H. BARNUM AND E. KNILL: Reversing quantum dynamics with near-optimal quantum and classical fidelity. *J. Math. Phys.*, 43:2097–2106, 2002. 12

[5] E. BERNSTEIN AND U. VAZIRANI: Quantum complexity theory. In *Proceedings of the 25th ACM Symposium on Theory of Computing*, pp. 11–20, New York, 1993. ACM Press. San Diego, CA, 16–18 May 1993.

[6] E. BERNSTEIN AND U. VAZIRANI: Quantum complexity theory. *SIAM J. Comput.*, 26:1411–1473, 1997. 3

[7] A. L. CAUCHY: *Exercises d'analyse et de physique mathématique*. Volume 3 of *Tome*. Paris, 1844. 9

[8] R. CLEVE, A. EKERT, C. MACCHIAVELLO, AND M. MOSCA: Quantum algorithms revisited. *Proc. Roy. Soc. Lond. A*, 454:339–354, 1998. 9

[9] D. DEUTSCH: Quantum theory, the church-turing principle and the universal quantum computer. *Proc. Roy. Soc. Lond. A*, 400:97–117, 1985. 9

[10] D. DEUTSCH AND R. JOZSA: Rapid solution of problems by quantum computation. *Proc. Roy. Soc. Lond. A*, 439:553–558, 1992. 9

[11] L. K. GROVER: A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing*, pp. 212–219, New York, 1996. ACM. Philadelphia, PA, 22–24 May 1996.

[12] L. K. GROVER: Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.*, 79:325–328, 1997. 3

[13] R. S. KENNEDY H. P. YUEN AND M. LAX: On optimal quantum receivers for digital signal detection. *Proc. IEEE Lett.*, 58:1770–1773, 1970. 3

[14] C. W. HELSTROM: Detection theory and quantum mechanics. *Inform. Control*, 10:254–291, 1967. 3

[15] C. W. HELSTROM: *Quantum Detection and Estimation Theory*. Academic, New York, 1976. 3, 10

[16] T. HOGG: Highly structured searches with quantum computers. *Phys. Rev. Lett.*, 80:2473–2476, 1998. 1

[17] M. HUNZIKER AND D. A. MEYER: Quantum algorithms for highly structured search problems. *Quantum Inform. Processing*, 1:145–154, 2002. 1, 3, 4, 9, 10

[18] M. HUNZIKER, D. A. MEYER, J. PARK, J. POMMERSHEIM, AND M. ROTHSTEIN: The geometry of quantum learning. to appear in Quantum Inform. Processing. 3, 10

[19] R. JOZSA: Quantum factoring, discrete logarithms, and the hidden subgroup problem. *Computing Science Eng.*, 3:34–43, 2001. 9

[20] R. S. KENNEDY: On the optimal receiver for the m-ary pure state problem. *MIT Res. Lab. Electron. Quart. Prog. Rep.*, 110:142–146, July 15 1973. 3

[21] D. E. KNUTH: The computer as master mind. *J. Recreational Math.*, 9:1–6, 1976-77. 2

[22] C. MOORE, D. ROCKMORE, A. RUSSELL, AND L. J. SCHULMAN: The power of strong fourier sampling: quantum algorithms for affine groups and hidden shifts. *SIAM J. Comput.*, 37:938–958, 2007. 3

[23] M. PLOTKIN: Binary codes with specified minimum distance. *IRE Trans. Inform. Theory*, 6:445–450, 1960. 7

[24] M. SASAKI, K. KATO, M. IZUTSU, AND O. HIROTA: Quantum channels showing superadditivity in classical capacity. *Phys. Rev. A*, 58:146–158, 1998. 10

[25] P. W. SHOR: Algorithms for quantum computation: discrete logarithms and factoring. In S. GOLDWASSER, editor, *Proceedings of the 35th Symposium on Foundations of Computer Science*, Los Alamitos, CA, 1994. IEEE Computer Society Press. Sante Fe, NM, 20-22 November 1994.

[26] P. W. SHOR: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26:1484–1509, 1997. 3

[27] R. H. SLOAN AND G. TURÁIN: Learning with queries but incomplete information. In *Proceedings of the Seventh Annual ACM Workshop on Computational Learning Theory*, pp. 237–245, New York, 1994. ACM. 2

AUTHORS

David A. Meyer
Project in Geometry and Physics
Department of Mathematics
University of California/San Diego
La Jolla, CA 92093-0112
dmeyer [at] math [dot] ucsd [dot] edu

James Pommersheim
Project in Geometry and Physics
Department of Mathematics
University of California/San Diego
La Jolla, CA 92093-0112
Department of Mathematics
Reed College, Portland, OR 97202-8199
jamie [at] reed [dot] edu