Quantum Mechanical Algorithms for the Nonabelian Hidden Subgroup Problem

Michelangelo Grigni* Emory University
Leonard J. Schulman† Caltech
Monica Vazirani‡ U. C. San Diego and U. C. Berkeley
Umesh Vazirani§ U. C. Berkeley

November 1, 2000

1 Introduction

1.1 Overview

The hidden subgroup problem is at present the keystone problem in quantum computation. We are given a function $f\colon G \to S$, with the property that f is constant on cosets of an unknown subgroup $H \subseteq G$, and distinct on distinct cosets. Here f is given as an oracle or as an efficient classical program, and S is an arbitrary set. The problem is to determine the hidden subgroup H. (A closely related problem, the "stabilizer problem", was formulated by Kitaev [6].)

The difficulty of the task depends on the type of group G. The abelian case can be effectively computed with a quantum computer by repetition of coset state preparation and Fourier sampling – the "standard method" developed by Simon [11] and Shor [10]. In particular this method is the heart of Shor's solution of the discrete logarithm and factoring problems.

The status of the non-abelian hidden subgroup problems is one of the most fundamental open problems in quantum algorithms. In particular, the graph automorphism problem may be formulated as hidden subgroup problems over the symmetric group S_n (see Section 4.3). It is natural to generalize the standard method for the abelian hidden subgroup to nonabelian groups. Fourier transforms over nonabelian groups are defined in terms of the irreducible complex representations of the group. There are efficient quantum circuits for computing these transforms for some groups of interest such as the symmetric group (see for example [1, 3]). However, since the dimension of these irreducible representations is in general greater than one, the Fourier transform is not unique, and is defined only up to a unitary change of basis for each irreducible. The Fourier sampling step in the standard method now yields the name of an irreducible representation ρ, together with the indices i, j of the entry within that irreducible. The main question, then, is whether the statistics of a sample from the Fourier transform of a coset state reveal sufficient information about the hidden subgroup, to allow for efficient reconstruction. One would hope that this approach is robust, in the sense that the answer to this question should not depend on the arbitrary choice of basis within each irreducible. Our main result is that with respect to a random choice of basis, the Fourier sampling statistics reveal, in general, an exponentially small amount of information about the hidden subgroup. It is still possible that a clever choice of basis with in each irreducible can solve the hidden subgroup problem.

Given how algebraically arbitrary this basis choice is, this seems somewhat unlikely. Ideally, one might hope to go beyond the standard method, which is essentially the only current method giving exponential speed ups of quantum over the best known classical computation.

*Supported in part by NSF grant CCR-9820931.
†Supported in part by NSF CAREER grant CCR-9876172.
‡Supported in part by an NSF Mathematical Sciences Postdoctoral Fellowship.
§Supported in part by NSF grant CCR-9800024 and Darpa Grant F30602-00-2-0601..

Our lower bound on the runtime of the standard method, for subgroups of a group G, depends upon two parameters: the size of the hidden subgroup H (naturally the problem becomes easy if H is very large), and $c(G)$, the number of conjugacy classes in G. We give a lower bound showing that approximately $\left(\frac{\sqrt{|G|}}{|H|\sqrt{c(G)}}\right)^{1/3}$ rounds of Fourier sampling are required before the standard method can identify H.

For the special case of hidden subgroups of order 2 in S_n, this yields a lower bound of approximately $(k!)^{1/6}$ repetitions of Fourier sampling in order to determine the correct non-identity element of H, where k is the number of transpositions in this element. Hallgren, Russell and Ta-Shma [5] independently obtained a similar bound for the weak form of the standard method, where only the name of the irreducible representation ρ is measured, and the indices i, j are ignored.

On the positive side, Hallgren, Russell and Ta-Shma [5] showed that the weak form of the standard method for abelian groups efficiently finds hidden normal subgroups in non-abelian groups. We introduce a measure of nonabelianness of a group — the maximum number of conjugates of any subgroup in the group. We show that there is a polynomial time algorithm for the HSP for any almost abelian group. This is a new class of groups for which a quantum algorithm can solve the HSP efficiently. We illustrate this case by studying the particular example of the semidirect product $C_3 \rtimes C_m$) for large m (here C_k is the cyclic group with k elements).

Some other interesting previous work on the positive side is that of Ettinger and Hoyer [4], showing how to obtain sufficient statistical information about the hidden subgroup of a dihedral group with polynomial query complexity; leaving the question of whether there is an efficient reconstruction algorithm.

1.2 The Fourier transform and the standard method for hidden subgroup computation

We first recall some basic group representation theory [9]. Given a group G, a matrix representation is a group homomorphism $\rho: G \to GL(d_\rho, \mathbb{C})$, where $GL(d, \mathbb{C})$ is the group of invertible $d \times d$ complex matrices. A finite group G has a finite list of inequivalent irreducible representations $\{\rho\}$, which we henceforth call its *irreps*. Without loss of generality we may assume the irreps are unitary. The sum of the squares of irrep dimensions $\sum_\rho d_\rho^2$ equals $|G|$, the order of the group.

To every group element g we associate a complex vector of dimension $|G|$, indexed by triples ρ, i, j where ρ is an irrep and $1 \le i, j \le d_\rho$ indicate an entry of the matrix ρ. The vector associated with g has value $\frac{\sqrt{d_\rho}\rho_{ij}(g)}{\sqrt{|G|}}$ in the ρ, i, j entry.

The Fourier transform over G is the extension of this mapping by linearity to the vector space $\mathbb{C}^G$ of complex linear combinations of group elements. This linear mapping (whose matrix we will denote F) is unitary; this fact is a consequence of the orthogonality relations for group representations.

The trivial representation is the 1-dimensional homomorphism which assigns to every group element the number 1. For a subset S of G, define $|S\rangle = \frac{1}{\sqrt{|S|}}\sum_{g\in S}|g\rangle$ and $\rho(S) = \rho(|S\rangle) = \frac{1}{\sqrt{|S|}}\sum_{g\in S}\rho(g)$. The orthogonality relations imply that $\rho(G)$ is $\sqrt{|G|}$ when ρ is the trivial representation, and a zero matrix otherwise. (As mentioned above, the Fourier transform has a scalar factor $\sqrt{d_\rho/|G|}$, so this corresponds to the fact that the Fourier transform of the unit norm uniform superposition on G, is 1 on the trivial representation and 0 elsewhere.)

$\mathbb{C}^G$ has an additional structure beyond its vector space structure: it is also an algebra over $\mathbb{C}$, using the product which is the extension of the group product by linearity. This structure is preserved by the Fourier transform, simply because each irrep is a group homomorphism. This is what is often known, for abelian groups (where each irrep is 1-dimensional), as the "convolution-multiplication" property of the Fourier transform.

In the "standard method" for the hidden subgroup problem we begin by forming the uniform superposition over a random coset gH of the hidden subgroup H: in other words, we form[1] the uniform distribution over

[1] To form this mixture of superpositions, we first form the uniform-amplitudes superposition $\frac{1}{\sqrt{|G|}}\sum_{g\in G}|g,0\rangle$, we then compute f obtaining the superposition $\frac{1}{\sqrt{|G|}}\sum_{g\in G}|g, f(g)\rangle$. We then measure $f(g)$, which determines the coset gH. The result is the superposition $\frac{1}{\sqrt{|H|}}\sum_{h\in H}|gh\rangle$ for a uniformly random g.

vectors $|gH\rangle$. First suppose that we know g (or at least gH), then we have the pure superposition $|gH\rangle$. We then apply the Fourier transform to this superposition, obtaining the vector

$$\frac{1}{\sqrt{|G||H|}} \sum_{\rho,i,j} \sqrt{d_\rho} \sum_{h\in H} \rho_{ij}(gh) \, |\rho, i, j\rangle \, .$$

This gives rise to the probability distribution

$$P_{gH}(|\rho,i,j\rangle) = \frac{d_\rho}{|G||H|} \left| \sum_{h\in H} \rho_{ij}(gh) \right|^2 = \frac{d_\rho}{|G|} |\rho(gH)_{ij}|^2.$$

Since we actually do not know g, and g is distributed uniformly, we sample ρ, i, j with the probability $P_H(|\rho,i,j\rangle) = \frac{1}{|G|}\sum_{g\in G} P_{gH}(|\rho,i,j\rangle)$.

The success of this method depends on how much statistical information about H is present in this distribution. In particular: do a polynomial number of samples suffice to identify H with high probability? In the following $\chi_\rho(g)$ denotes the *character* of ρ at g, which is simply the trace of $\rho(g)$.

Lemma 1 *$\rho(H) = \frac{1}{\sqrt{|H|}}\sum_{h\in H}\rho(h)$ is $\sqrt{|H|}$ times a projection matrix, and* $\mathrm{rank}(\rho(H)) = \frac{1}{|H|}\sum_{h\in H}\chi_\rho(h)$.

Proof: Restricted to H, ρ decomposes into the direct sum of several irreps $\sigma_1, ..., \sigma_k$. $\rho(H)$ is the direct sum of $\sigma_i(H)$; as discussed above $\sigma_i(H)$ is $\sqrt{|H|}$ if σ_i is the trivial representation of H, and zero otherwise. □

A certain amount of information about H is given just by sampling ρ, and ignoring the matrix indices i and j. We refer to this as

the "weak form" of the standard method. In the normal case this more limited information is already enough, and in fact no further information is available in the indices. For general subgroups further information is present in the indices, and in the "strong form" of the method, these are sampled as well; we will discuss this issue below. First we show that, when we Fourier sample the unit norm uniform superposition on gH, (i.e. sample from the probability distribution defined by the Fourier transform of this superposition), the probability $P_{gH}(|\rho\rangle)$ of sampling ρ is independent of g.

Lemma 2 [2] *The probability of measuring ρ is the same for the uniform superposition on the coset gH (or Hg), as for the superposition on H.*

Proof: $\rho(gH) = \rho(g)\rho(H)$ and $\rho(g)$ is unitary. □

Corollary 3 $P_{gH}(|\rho\rangle) = P_H(|\rho\rangle) = \frac{d_\rho}{|G|}\sum_{h\in H}\chi_\rho(h) = \frac{|H|d_\rho}{|G|}\mathrm{rank}(\rho(H))$.

Corollary 4 *The probability of sampling ρ is the same for the subgroup H as it is for a conjugate subgroup $g^{-1}Hg$.*

2 Normal H

Hallgren, Russell, and Ta-Shma [5] showed that the weak form of the standard method quickly obtains enough information to identify hidden normal subgroups. This section briefly describe the argument.

Recall that in the standard method we sample from the Fourier transform of the uniform superposition over a random coset gH of the hidden subgroup H. We first show that when we restrict attention to normal subgroups, all the information about H is present in the label of the sampled irrep. By the last lemma, the probability of sampling ρ is independent of the particular coset gH: so we will examine the uniform superposition on H.

By not using $f(g)$ to affect the subsequent computation, we are discarding some potentially useful information. No proposal, however, exists for taking advantage of this information.

[2]For methods that measure ρ but discard i and j, Lemma 2 implies that there is no loss in discarding $f(g)$ as well. In particular we may discard $f(g)$ when G is commutative.

Lemma 5 *If H is a normal subgroup, $\rho(H)$ is a nonnegative scalar multiple of the identity I, nonzero if and only if $H \subseteq Ker(\rho)$.*

□

By the above lemma, ρ has non-zero probability of being sampled if and only if H is in its kernel. The task of reconstructing H from a sequence of such samples is just that of intersecting the kernels of the sampled irreps. In order to distinguish H from all other contending subgroups H', it suffices to make the probability of mistaking H' for H, less than inverse in the number of contending subgroups.

Next note that no group G has more than $2^{\lg^2 |G|}$ subgroups. Finally, one must show that the large L_2 distance between the uniform superpositions on two contending subgroups H and H', translates to a large L_1 distance between their Fourier transforms; this again uses the above lemma. In combination these arguments give us:

Theorem 6 *$O(\log^2 |G|)$ repetitions of Fourier sampling suffice to identify, with high probability, a normal subgroup of a group G.*

3 "Almost Abelian" Groups

3.1 Algorithm

The case of normal subgroups was one way of extending the abelian-G method. Another extension is to consider the case in which the intersection of the normalizers of all subgroups of G, is large. We will call this intersection N_G. (Thus $N_G = \cap_H N(H)$ where the intersection ranges over subgroups H of G.) For abelian groups, of course, $N_G = G$. There are two equivalent ways of characterizing this case, both of them in terms of a "small" parameter k which characterizes the size of N_G. In order for our algorithm to run in polynomial time (in $n = \log |G|$), k should be $\exp(O(\log^{1/2} n))$.

1. $(G : N_G) = k$

2. For every subgroup H of G, the number of conjugate subgroups of H is at most k.

(1) ⇒ (2): Immediate from $N_G \subseteq N(H) \subseteq G$ and the fact that there is a $1-1$ correspondence between left cosets $aN(H)$, and conjugate subgroups aHa^{-1}.

(1) ⇒ (2): Follows from the equation $N(AB) = N(A) \cap N(B)$, where A and B are subgroups of G, and AB is the group they generate (i.e., all products of elements of A and B). ($\subseteq$ is immediate; the converse is proven by induction on word length.)

The basic method to identify the unknown H for "almost abelian" G begins again with the observation that $N_G \subseteq N(H) \subseteq G$. Now, N_G is normal in G. For, let $g \in N_G$ and $a \in G$. We wish to show that $aga^{-1} \in N(H)\ \forall H$. Fix H, then for any $h \in H$, $(aga^{-1})h(aga^{-1})^{-1} = aga^{-1}hag^{-1}a^{-1}$. Now $a^{-1}ha \in a^{-1}Ha$, and since $g \in N_G \subseteq N(a^{-1}Ha)$, it follows that $ga^{-1}hag^{-1} \in a^{-1}Ha$. But then $aga^{-1}hag^{-1}a^{-1} \in H$ as desired.

$N(H)$ is unknown, but $N(H)/N_G$ is a subgroup of G/N_G, and we have the time to examine all possibilities. So, we consider each subgroup containing N_G in turn, and for each one, we run the normal-subgroups algorithm; when we happen upon $N(H)$ we with high probability pick out H.

Of course, along the way we may also select subgroups of H. But we will not (except with low probability) pick out any group which is not a subgroup of H. So at the end we can simply report the largest group we find.

3.2 Example: extensions of groups

One way to construct an almost abelian group is by extending one abelian group A by another B. We say G is an extension of A by B if A is normal in G and $G/A \simeq B$.

Here we'll consider the special case when G is the semidirect product of A by B, written $G = A \rtimes B$. In other words, A is a normal subgroup of G, B is isomorphic to a subgroup of G, $AB = G$ and $A \cap B = \{1\}$. (Classifying other extensions is a question in group cohomology.) The representation theory of $G = A \rtimes B$ is well understood in terms of that of A and B.

To define the semidirect product, we need a homomorphism $\theta : B \to \mathrm{Aut}(A)$. Then the group structure of $G = AB$ is defined by the identity

$$bab^{-1} = (\theta(b))(a).$$

(Since construction of G from A and B requires specification of θ, one can more carefully write $G = A \rtimes_\theta B$. This is unnecessary when A and B are specified as particular subgroups of a given G.)

We remark that θ need not be injective or surjective. In fact, it will be convenient for us to have $\ker(\theta)$ be large, because $\ker(\theta) \subseteq Z(G) \subseteq N_G$ so this provides us with a large N_G.

Basic example of a semidirect product: the dihedral group $D_n = \langle x, y \mid x^2 = y^n = 1; xyx^{-1} = y^{-1} \rangle = C_n \rtimes C_2$ (where C_n denotes a cyclic group of order n). The homomorphism θ sends the nontrivial element $x \in C_2$ to the map $y \mapsto y^{-1}$, because $xyx^{-1} = y^{-1}$.

In the context of "almost abelian" groups we are interested in the following example: $G = C_3 \rtimes C_m$ where m is a power of two. Let a be a generator for the C_3 subgroup, and b a generator for the C_m subgroup. We have $aba^{-1} = a^2$.

Let ω be a primitive 3'rd root of unity and ζ a primitive m'th root of unity.

G has m one-dimensional representations. These correspond to the trivial character of C_3. Each is indexed by $0 \leq k < m$, and sends (a^i, b^j) to ζ^{kj}.

G has $m/2$ two-dimensional representations. These correspond to the character χ of C_3 for which $\chi(a) = \omega$. The $m/2$ representations, indexed by $0 \leq k < m/2$, are

$$\Phi_k(a) = \begin{pmatrix} \omega & 0 \\ 0 & \omega^2 \end{pmatrix}$$
$$\Phi_k(b) = \begin{pmatrix} 0 & \zeta^{2k} \\ 1 & 0 \end{pmatrix}.$$

Observe b^2 acts by the scalar matrix $\begin{pmatrix} \zeta^{2k} & 0 \\ 0 & \zeta^{2k} \end{pmatrix}$.

It is easy to modify the standard quantum fourier transform circuits to compute the fourier transform for G.

4 General H

Up to now we have focussed on the extent to which information about H can be detected from the measuring just the name of the irrep, ρ. Of course we can actually measure more, namely the row i and column j within ρ. It is possible that this contributes substantially to our power. In particular, conjugate subgroups give rise to identical distributions on irreps and so cannot be told apart without measuring the matrix indices within the irreps. In this section, we establish limits on what further information can be obtained from the row and column labels.

4.1 Rows provide no information

In this section we show that there is no point in measuring the row i. (Whether row or column depends on whether the group acts on the left or right. Here we suppose the group acts on the left.) This is because, conditional on measuring ρ and j, the distribution on i is independent of H (actually it's always uniform); we now show how this is due to the fact that in the standard method we average over random cosets gH.

For a particular coset gH, the probability of sampling the entry i, j of ρ is proportional to the norm squared of $\rho(gH)_{ij}$. Thus the probability of sampling entry i, j is the norm squared of the $|G|$-dimensional vector $(\rho(gH)_{ij})_{g \in G}$ with entries indexed by g. Since $\rho(gH) = \rho(g)\rho(H)$, this vector is a linear combination of the $|G|$-dimensional vectors $(\rho(g)_{ik})_{g \in G}$, with coefficients $\rho(H)_{kj}$. By the orthogonality relations, the d_ρ vectors $(\rho(g)_{ik})_{g \in G}$ are orthonormal, and therefore the norm squared of the $|G|$-dimensional vector $(\rho(gH)_{ij})_{g \in G}$ is equal to the norm squared of the j-th column of $\rho(H)$, and independent of i.

If we keep track of the leading constants, this argument shows:

Theorem 7 $P_H(|\rho, i, j\rangle) = \frac{1}{|G|} \left|\rho(H)_j\right|_2^2$.

4.2 Random basis

The Fourier transform is uniquely defined only up to a change of basis within each irrep; for abelian groups all irreps are one-dimensional so there is no ambiguity in the definition of the transform, but for nonabelian groups there is an arbitrary choice of basis to be made within each irrep. How much statistical information is available by measuring the matrix entries i, j, in addition to the irrep ρ, may in general be basis dependent. In this section, we show that if we choose a random basis for each irrep, then the additional information available is negligible, provided that the subgroup H is sufficiently small and the group G is sufficiently non-abelian. To formalize this we compare the actual distribution $P_H = P_H(|\rho, j\rangle)$ on irreps and columns (we omit the row i thanks to the previous section), with the averaged distribution $\overline{P_H} = \overline{P_H}(|\rho, j\rangle) = \int P_H(|A^{-1}\rho A, j\rangle)dA = \frac{1}{d_\rho}P_H(|\rho\rangle)$, where dA is the Haar measure for unitary matrices. (The last equality follows from the orthogonality relations for representations.) Let $\alpha = \frac{\sqrt{|G|}}{|H|\sqrt{c(G)}}$ where $c(G)$ is the number of conjugacy classes in G.

Theorem 8 *Let $\varepsilon^3 \geq \frac{1}{\alpha}\frac{54}{2-\sqrt{3}}\ln\frac{4|G|}{\delta}$. Then with probability at least $1-\delta$ (over the choice of random basis for the Fourier transform), $\left|P_H - \overline{P_H}\right|_1 \leq \varepsilon$.*

Proof: Given an irrep ρ in a particular basis, the probability of sampling the j-th column of ρ is $\frac{d_\rho}{|G|}\left|\rho(H)_j\right|_2^2$ (where $\rho(H)_j$ is the j-th column of $\rho(H)$). Suppose we instead choose a random basis for ρ, which we do by replacing ρ by the isomorphic irrep $A^{-1}\rho A$ for A chosen with the Haar distribution in the unitary group. Let A_j be the j-th column of A, which is a vector chosen uniformly from the unit sphere. Then the probability of measuring the j-th column in this modified irrep is $\frac{d_\rho}{|G|}\left|\rho(H)A_j\right|_2^2$. Our task is to show that for sufficiently large α, this is with high probability close to $\frac{1}{d_\rho}P_H(|\rho\rangle)$.

Let $K = \sum_\rho d_\rho$. We upper bound K by applying the Cauchy-Schwartz inequality to the two vectors $(1)_\rho$ and $(d_\rho)_\rho$. The norm squared of the first vector is simply $c(G)$, the number of irreps of G. Now $K \leq (\sum_\rho 1)^{1/2}(\sum_\rho d_\rho^2)^{1/2} = c(G)^{1/2}|G|^{1/2}$.

To bound the L_1 distance between the vectors P_H and $\overline{P_H}$ we first bound the probability p_T of sampling an irrep ρ with $\mathrm{rank}(\rho(H))$ bounded above by T. After that we will bound the L_1 distance conditional on having sampled an irrep with $\mathrm{rank}(\rho(H)) > T$.

By Corollary 3, $P_H(|\rho\rangle) = \frac{|H|d_\rho}{|G|}\mathrm{rank}(\rho(H))$. So $p_T = \sum_{\rho:\ \mathrm{rank}(\rho(H))\leq T} P_H(|\rho\rangle) \leq \frac{|H|T}{|G|}\sum_{\rho:\ \mathrm{rank}(\rho(H))\leq T} d_\rho$ $\leq \frac{|H|T}{|G|}\sum_\rho d_\rho = \frac{|H|TK}{|G|}$ which from the preceding argument is bounded above by $\frac{|H|T\sqrt{c(G)}}{\sqrt{|G|}} = T/\alpha$. We choose $T = \varepsilon\alpha/3$, giving $p_T \leq \varepsilon/3$.

Now for the high rank case, consider the distribution on j conditional on having observed an irrep ρ with $\mathrm{rank}(\rho(H)) > T$. We need to show that with probability at least $1-\delta$ (in terms of the random choice of A), this distribution is within L_1 distance at most $2\varepsilon/3$ of the uniform distribution. We will show slightly more: with probability at least $1-\delta$, for all ρ and all j, $\left|\rho(H)A_j\right|_2^2$ deviates from its expectation by at most a $2\varepsilon/3$ fraction. (Since we are concerned here only with fractional error we have suppressed the leading scale factor of the projection $\frac{1}{\sqrt{|H|}}\rho(H)$.)

What we are considering is the following process: a unit vector is chosen uniformly in $\mathbb{C}^{d_\rho}$, then projected onto a fixed subspace of dimension $t > T$; by appropriate change of basis we can without loss of generality suppose that the subspace is spanned by the first t basis vectors of $\mathbb{C}^{d_\rho}$. Let s be the probability that the squared length of the projected vector differs from its expectation t/d_ρ by a fraction greater than than $2\varepsilon/3$. Since we will apply a union bound over all ρ and j, it suffices to show that $s \leq \delta/|G|$. To begin with, note that, due to the isometric correspondence between the unit spheres in $\mathbb{C}^{d_\rho}$ and $\mathbb{R}^{2d_\rho}$, the problem is equivalent to the same problem in real spaces of twice the dimensions, namely projection of the unit sphere in $\mathbb{R}^{2d_\rho}$ onto a $2t$-dimensional subspace. Let M denote the projection matrix; in the appropriate basis it is diagonal, with $2t$ 1's on the diagonal.

We analyze the uniform sampling from the unit sphere indirectly, approximating it by the process of sampling a vector v from the spherically symmetric, $2d_\rho$-dimensional unit variance Gaussian distribution. Let the projection of v be $v' = Mv$. (Note that v' is distributed according to a $2t$-dimensional Gaussian distribution of variance t/d_ρ.) Then $v'/\left|v\right|_2$ has the same distribution as $\frac{1}{\sqrt{|H|}}\rho(H)A_j$ (with the understanding that pairs of real coordinates in the first vector form individual complex coordinates in the second). The probability s that $|v'/|v|_2|_2$ deviates from its expectation by fraction $2\varepsilon/3$ is bounded by the sum of the probabilities that $|v|_2$ and $|v'|_2$ deviate from their expectations by fraction $\varepsilon/3$.

We use the following Chernoff bound: if $a_1, ..., a_\tau$ are independent Gaussian random variables each with unit standard deviation, then $P(|\frac{1}{\tau}\sum(a_i^2 - 1)| > \varepsilon) < 2[(1+\varepsilon)^{1/2}e^{-\varepsilon/2}]^\tau$. For $\varepsilon \leq 2$, $(1+\varepsilon)^{1/2}e^{-\varepsilon/2} \leq \exp(-\varepsilon^2\frac{2-\sqrt{3}}{4})$, and therefore $P(|\frac{1}{\tau}\sum(a_i^2 - 1)| > \varepsilon) < 2\exp(-\tau\varepsilon^2\frac{2-\sqrt{3}}{4})$. Since this bound is decreasing in τ, and we are applying it with $\tau = 2t > 2T$, we conclude that $s < 4\exp(-2T(\varepsilon/3)^2\frac{2-\sqrt{3}}{4}) = 4\exp(-\alpha\varepsilon^3\frac{2-\sqrt{3}}{54})$.

In order to ensure that $s \leq \delta/|G|$ it suffices therefore that

$$\alpha\varepsilon^3 \geq \frac{54}{2-\sqrt{3}}\ln\frac{4|G|}{\delta}$$

as assumed. □

Corollary 9 *With probability at least $1-\delta$ (over the choice of random basis for the Fourier transform), $\Omega((\frac{\alpha}{\log(|G|/\delta)})^{1/3})$ repetitions of Fourier sampling are required in order to achieve constant bias in distinguishing any two (a priori equally probable) conjugate subgroups H and H'.*

Proof: The Hellinger distance $D_H(\vec{p},\vec{q}) = \sum(p^{1/2} - q^{1/2})^2$ between two distributions $\vec{p}$ and $\vec{q}$ is additive across independent samples, and obeys the inequalities $|\vec{p}-\vec{q}|_1^2/4 \leq D_H(\vec{p},\vec{q}) \leq |\vec{p}-\vec{q}|_1$. □

As an example consider the symmetric group $G = S_n$; we know that $c(G) = \exp(\Theta(\sqrt{n}))$. If $|H| \leq |G|^{1/2-\gamma}$ for a fixed $\gamma > 0$ then we need exponentially many samples to gain useful information from j.

4.3 Distinguishing $|H| = 2$ from $|H| = 1$

The graph automorphism problem reduces, via polynomial time reductions [7], to determining the size of the automorphism group in the special case where it is known to be at most 2. So, although sampling irreps cannot distinguish conjugate subgroups, one may still hope that this method is useful for distinguishing subgroups that are not conjugate.

Now consider the special case of distinguishing $H = \{e, s\}$ from $H' = \{e\}$. Let $C(s)$ be the conjugacy class of s.

Theorem 10 *The L_1 distance between the distributions on irreps due to Fourier sampling from $|H\rangle$ and $|H'\rangle$, is at most $1/\sqrt{|C(s)|}$.*

The equation $P_H(|\rho\rangle) = \frac{d_\rho}{|G|}\sum_{h\in H}\chi_\rho(h)$ implies that for H', ρ is sampled with probability $d_\rho^2/|G|$; while for H, ρ is sampled with probability $d_\rho(d_\rho + \chi_\rho(s))/|G|$. So the L_1 distance between the distributions is $\frac{1}{|G|}\sum_\rho d_\rho|\chi_\rho(s)|$.

We upper bound this using the Cauchy-Schwartz inequality and the following equalities:

1. $\sum_\rho d_\rho^2 = |G|$.
2. $|Z(s)| \cdot |C(s)| = |G|$.
3. $\sum_\rho(\chi_\rho(s))^2 = |Z(s)|$.

Here $Z(s)$ is the centralizer of s.

(1) is basic. (2) follows by considering the action of G on itself by conjugation, since under this action $Z(s)$ is the stabilizer of s and $C(s)$ is the orbit of s. (3), which generalizes (1), holds for the following reason. Recall that the unitary character table of G has conjugacy classes labeling columns, irreps labeling rows, and the (ρ, s) entry is $\sqrt{\frac{|C(s)|}{|G|}}\chi_\rho(s)$. Now since each column is unit norm, $1 = \sum_\rho\frac{|C(s)|}{|G|}(\chi_\rho(s))^2$. With (2)

this shows (3). Now we can apply Cauchy-Schwartz.

$$\begin{aligned} \left| |F\,|H\rangle\,|^2 - |F\,|H'\rangle\,|^2 \right|_1 &= \frac{1}{|G|} \sum_\rho d_\rho |\chi_\rho(s)| \\ &\leq \frac{1}{|G|} [\sum_\rho d_\rho^2]^{1/2} [\sum_\rho (\chi_\rho(s))^2]^{1/2} \\ &= (\frac{|Z(s)|}{|G|})^{1/2} = |C(s)|^{-1/2}. \end{aligned}$$

□

There are examples in which it is challenging to compute s even though the conjugacy class $C(s)$ is known. Observe that in such cases this quantum algorithm has at most a quadratic advantage over the simple probabilistic strategy of checking whether $f(s') = f(e)$ for a random conjugate s'.

We apply Theorem 10 in the case that s is an involution in $G = S_n$, i.e. s is a product of some k disjoint transpositions. In this case $|C(s)| = \frac{n!}{2^k k!(n-2k)!}$; as a convenient lower bound on this quantity, count only those conjugates which transpose odd elements with even elements, of which there are $\binom{\lceil n/2 \rceil}{k}\binom{\lfloor n/2 \rfloor}{k} k! \geq k!$. So the L_1 distance between the distributions on irreps is at most $(k!)^{-1/2}$. In the graph automorphism application k can be proportional to n, in which case this is exponentially small. A similar bound was independently obtained by Hallgren, Russell and Ta-Shma [5].

Finally we combine this bound with Theorem 8. Note that $\alpha = \frac{\sqrt{n!}}{2 \cdot 2^{\Theta(\sqrt{n})}} \in \exp(\frac{1}{2} n \ln n - O(n))$, so:

Corollary 11 *If we apply the standard method, using a random basis, to the graph automorphism problem, then with probability at least* $1 - \delta$ *the* L_1 *distance between the Fourier sampling distribution given that the automorphism group is trivial, and the Fourier sampling distribution given that the automorphism group is of size* 2 *and contains an involution with* k *transpositions, is at most* $\exp(-\frac{1}{6} n \ln n + O(n)) \log^{\frac{1}{3}} \frac{1}{\delta} + (k!)^{-1/2}$.

Just as in Corollary 9, the upper bound on L_1 distance implies a lower bound on the number of samples which must be collected in order to distinguish the hypotheses reliably.

References

[1] Robert Beals. Quantum computation of Fourier transforms over symmetric groups. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, 48–53, 1997.

[2] M. Clausen, Fast Generalized Fourier Transforms, *Theor. Comp. Sci.*, 67, 55–63, 1989.

[3] Persi Diaconis and Daniel Rockmore. Efficient computation of the Fourier transform of finite groups. *J. Amer. Math. Soc.*, 3(2), 297–332, 1990.

[4] Mark Ettinger and Peter Høyer. On quantum algorithms for noncommutative hidden subgroups. LANL e-preprint quant-ph/9807029, 1998.

[5] Sean Hallgren, Alexander Russell, and Amnon Ta-Shma. Normal subgroup reconstruction and quantum computation using group representations. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, 627–635, 2000.

[6] Alexei Kitaev. Quantum measurements and the abelian stabilizer problem. ECCC Report TR96-003, 1996.

[7] Johannes Köbler, Uwe Schöning, and Jacobo Torán. *The Graph Isomorphism Problem: Its Structural Complexity.* Birkhäuser, 1993.

[8] D. Rockmore, Computation of Fourier Transforms on the Symmetric Group, in: E. Kaltofen, S. M. Watt (eds.), *Computers and Mathematics*, Springer, 1989.

[9] Jean-Pierre Serre. *Linear Representations of Finite Groups.* Springer-Verlag, 1977.

[10] Peter Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5), 1484–1509, 1997.

[11] Daniel Simon. On the power of quantum computation. *SIAM Journal on Computing*, 26(5), 1474–1483, 1997.