The complexity class PDQP
SPUR Final Paper: Summer 2013

Mitchell Lee
Mentor: Adam Bouland
Project suggested by: Adam Bouland, Scott Aaronson

July 31, 2013

Abstract

Quantum computers are believed to be strictly more computationally powerful than classical computers, but not so much more powerful that they can solve NP-complete problems efficiently. This is because of the result that a quantum computer takes $\Theta(N^{1/2})$ time to search an unstructured N-element list for a particular marked item, as opposed to $\mathsf{poly}(\log N)$ time for a nondeterministic computer. On the other hand, many seemingly innocuous modifications of quantum mechanics increase the power of quantum computers drastically enough that they can solve NP-complete problems efficiently [3]. This paper defines a model of computation slightly more powerful than quantum computation, but only slightly so. In particular, we show that by allowing "non-collapsing measurements," we can solve efficiently problems such as Graph Isomorphism and Approximate Shortest Vector which are believed to be intractable for quantum computers. We can also search an unstructured N-element list for a particular marked item in $\tilde{O}(N^{1/3})$ time, but no faster than $\Omega(N^{1/4})$.

Contents

1 Introduction

Quantum computers are believed to be strictly more computationally powerful than classical computers, but not so much more powerful that they can solve NP-complete problems efficiently. In particular, it is known that BQP, the class of languages recognizable in polynomial time by a quantum computer [8], does not contain NP "relative to an oracle," which means that there is some problem $\mathcal{O}$ for which $\mathsf{BQP}^{\mathcal{O}} \not\supset \mathsf{NP}^{\mathcal{O}}$. (For more information about the terminology, see [5, pp. 72-76].) On the other hand, many seemingly innocuous modifications of quantum mechanics—for example, allowing nonlinear transformations, non-unitary transformations, postselection, or measurement statistics based on the pth power of the amplitudes for $p \neq 2$—increase the power of quantum computers drastically enough that they can solve NP-complete problems efficiently [3].

In [2], Aaronson defines the complexity class DQP, which is informally the class of languages recognizable efficiently using a computational model that allows examining the entire history of a quantum system. He shows that $\mathsf{BQP} \subset \mathsf{DQP}$, and also that SZK, the set of languages admitting statistical zero-knowledge proofs, is contained in DQP. SZK contains important problems such as Graph Isomorphism and Approximate Shortest Vector not known to be in BQP. Since $\mathsf{SZK} \neq \mathsf{BQP}$ relative to an oracle [1], this yields the result that $\mathsf{DQP} \neq \mathsf{BQP}$ relative to an oracle. The best known upper bound on DQP is $\mathsf{DQP} \subset \mathsf{EXP}$.

Here, we define a related complexity class PDQP, which is informally the class of all problems solvable using a quantum machine that is allowed to make both ordinary quantum measurements and "non-collapsing" measurements, which do not change the state of the system. Like DQP, the complexity class PDQP also contains both SZK and BQP. It is trivial to prove the upper bound $\mathsf{PDQP} \subset \mathsf{EXP}$; we improve this trivial upper bound to $\mathsf{PDQP} \subset \mathsf{PP}^{\mathsf{PP}}$, where PP is the class of languages that can be recognized by a randomized polynomial-time Turing machine with error probability less than $\frac{1}{2}$.

We also demonstrate the striking result that in the computational model of PDQP, there is an algorithm that searches an unstructured list of N elements for a particular marked item in $\tilde{O}(N^{1/3})$ time (where the $\tilde{O}$ hides a factor polynomial in $\log N$), and that any algorithm that searches an unstructured list of N elements takes at least $\Omega(N^{1/4})$ time. This contrasts with the classical computational model, in which searching an unstructured list of N elements takes $\tilde{\Theta}(N)$ time, and with the quantum computational model, in which searching an unstructured list of N elements takes $\Theta(N^{1/2})$ time. We conclude that PDQP does not contain NP relative to an oracle. Aaronson showed that DQP also does not contain NP relative to an oracle, but there is an error in his proof. We describe the error in Appendix B. An open question is whether there is a hierarchy of computational models for which the kth allows searching in $\tilde{O}(N^{1/k})$ time.

2 Preliminaries

2.1 Qubits, the quantum circuit model, and BQP

A *quantum bit*, or *qubit*, is the indivisible unit of information in quantum computing. It is described by a two-state quantum system. Like a classical bit, a quantum bit can be in one of two distinguishable states 0 or 1 (written $|0\rangle$ and $|1\rangle$), but it can also be in the state

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$$

for complex numbers α, β with $|\alpha|^2 + |\beta|^2 = 1$. This is a phenomenon known as *superposition*. That is, the state of a qubit is described by a unit vector in the Hilbert space $\mathbb{C}^2 = \mathrm{Span}\{|0\rangle, |1\rangle\}$. If we *measure* a qubit

in the state $|\psi\rangle$, we will see $|0\rangle$ with probability $|\alpha|^2$ and $|1\rangle$ with probability $|\beta|^2$, and the state of the qubit will *collapse* to the result of the measurement. That is, if the measurement result is $|0\rangle$, the new state of the qubit will be $|0\rangle$, and if the measurement result is $|1\rangle$, the new state of the qubit is $|1\rangle$.

Transformations on a qubit are described by unitary transformations on the underlying Hilbert space $\mathbb{C}^2$, or, in other words, 2×2 unitary matrices. An example is the Hadamard gate

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

defined by

$$H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$
$$H|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

A system of ℓ qubits is described by a 2^ℓ-state system. Like a system of ℓ classical bits, it can be in one of 2^ℓ distingiushable states, one for each element of $\{0,1\}^\ell$. It can more generally be in the superposition

$$\sum_{x \in \{0,1\}^\ell} \alpha_x |x\rangle ,$$

where α_x is a complex number for each x, and $\sum_{x\in\{0,1\}^\ell} |\alpha_x|^2 = 1$. That is, the state of the system is described by a unit vector in the Hilbert space $(\mathbb{C}^2)^{\otimes \ell}$, where the tensor product is over $\mathbb{C}^1$. This space has an orthonormal basis whose elements are $|x\rangle$, for $x \in \{0,1\}^\ell$, called the *computational basis*.

If we measure the kth qubit of the system, the result is $|0\rangle$ with probability

$$p_0 = \sum_{\substack{x\in\{0,1\}^\ell \\ x_k=0}} |\alpha_x|^2,$$

where the sum is over all x whose kth bit is 0. The result is $|1\rangle$ with probability

$$p_1 = \sum_{\substack{x\in\{0,1\}^\ell \\ x_k=1}} |\alpha_x|^2,$$

where the sum is over all x whose kth bit is 1. Then, the qubit collapses to the result of the measurement, meaning that the resulting state is

$$\frac{1}{\sqrt{p_0}} \sum_{\substack{x\in\{0,1\}^\ell \\ x_k=0}} \alpha_x |x\rangle$$

or

$$\frac{1}{\sqrt{p_1}} \sum_{\substack{x\in\{0,1\}^\ell \\ x_k=1}} \alpha_x |x\rangle$$

accordingly to whether the measurement result is $|0\rangle$ or $|1\rangle$.

A transformation on n qubits is described by a unitary operator on $(\mathbb{C}^2)^{\otimes \ell}$. The simplest unitary transformations are *2-local*, which means that they only act on one or two qubits. That is, they are described by the tensor product of a unitary operation U on one or two qubits with the identity operation on the rest of the qubits.

A *quantum circuit* C is described by integers ℓ, T and a sequence $C = (U_1, M_1, U_2, M_2, \cdots, U_T, M_T)$, where each U_i is a 2-local unitary transformation on ℓ qubits (called a *quantum gate*), and each M_i is a measurement of zero or more of the ℓ qubits.

[1]When writing elements of the tensor product $(\mathbb{C}^2)^{\otimes \ell}$, we will often omit the tensor product symbol. For example, we write the two-qubit state $|0\rangle \otimes |0\rangle$ as $|0\rangle|0\rangle$ or even simply $|00\rangle$.

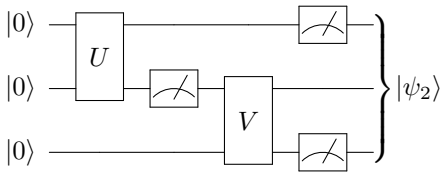

Figure 1: An example circuit on three qubits. The meter symbol represents a measurement.

If $C = (U_1, M_1, \cdots, U_T, M_T)$ is a quantum circuit, then define the sequence $\{|\psi_t\rangle\}_{t=0}^{T}$ of quantum states as follows: Let

$$|\psi_0\rangle = |0\rangle^{\otimes \ell} := \underbrace{|0 \cdots 0\rangle}_{\ell},$$

and for $t > 0$, let $|\psi_t\rangle$ be the resulting state when the measurement M_t is applied to $U_t |\psi_{t-1}\rangle$. Since measurement is in general not deterministic, each $|\psi_t\rangle$ is a random variable depending on $|\psi_{t-1}\rangle$.

Now, let $\mathcal{Q}$ be an oracle that takes as input the integers ℓ, T and the quantum circuit C, and outputs a sample from the distribution on $\{0,1\}^{\ell}$ defined by the measurement of all qubits of the final state $|\psi_T\rangle$. BQP is then defined as the class of all languages that can be recognized by a deterministic Turing machine with one query to $\mathcal{Q}$, with error probability at most $\frac{1}{3}$.

Two things are worth noting about this definition. First, an algorithm cannot specify a general 2-local unitary operator, because a unitary operator depends on continuous parameters. Instead, we usually make the assumption that the gates $U_1, \cdots, U_T$ come from some finite collection $\mathcal{U}$ of 2-qubit unitary operators, called the *gate set*. As long $\mathcal{U}$ is *universal*, which means that every unitary operator on two qubits can be approximated arbitrarily well by the composition of a sequence of gates from $\mathcal{U}$, the actual set $\mathcal{U}$ does not matter [9]. Second, the *principle of deferred measurement* [11] implies that disallowing the "intermediate" measurements $M_1, \cdots, M_T$ does not affect the complexity class BQP at all.

2.2 Hidden-variable theories

The formal definition of the complexity class DQP is based on the notion of a hidden-variable theory. A hidden-variable theory is an interpretation of quantum mechanics in which a quantum system is described by both a state vector and a definite state (called the "hidden variable"), which determines the result of measurements on the system. When a transformation is applied to the system, the state vector evolves by a unitary linear transformation, like in ordinary quantum mechanics, and the hidden variable evolves stochastically according to the state vector and the unitary linear transformation. According to the Kochen-Specker theorem [12], it is impossible for the hidden variable to determine a result for all possible measurements on the system. Therefore, in what follows, we will only ever measure the quantum system in some fixed basis.

Suppose that our quantum system is described by a Hilbert space with N basis states $|1\rangle, \cdots, |N\rangle$. Then, the hidden variable has one of the values $1, \cdots, N$. The hidden-variable theory specifies the probabilities that the hidden variable changes from i to j given that the state density matrix [11] was ρ and was transformed by the unitary U. More precisely, a hidden variable theory $\mathcal{T}$ is specified by a stochastic matrix $S_{\mathcal{T}}(\rho, U)$ for every density matrix ρ and unitary transformation U of dimension N, which indicates how the hidden variable evolves when the state density matrix transforms from ρ to $U\rho U^{\dagger}$. If $\mathcal{T}$ is understood from context, then we simply write $S(\rho, U)$. If the state $\rho = |\psi\rangle \langle\psi|$ is pure, then we use the notation $S(|\psi\rangle, U) := S(\rho, U)$. The hidden-variable theory must be consistent with the predictions of quantum mechanics, which is to say that the probability that the hidden variable is equal to i when the state vector is ρ is equal to ρ_{ii}. This

means that the stochastic matrix $S = S(\rho, U)$ must satisfy

$$(U\rho U^\dagger)_{jj} = \sum_{i=1}^{n} (\rho)_{ii}(S)_{ij}.$$

Other "reasonable" properties that we might expect a hidden-variable theory to have, for example that

$$S(\rho, WV) = S(\rho, V)S(V\rho V^\dagger, W),$$

need not be satisfied.

Sometimes, the hidden-variable theory is described instead by the matrix $P = P(\rho, U)$ of joint probabilities, defined by $(P)_{ij} = (\rho)_{ii}(S)_{ij}$. The matrix S is then recovered by

$$S(\rho, U) = \lim_{\epsilon \to 0^+} \frac{(P(\rho_\epsilon, U))_{ij}}{(\rho_\epsilon)_{ii}}$$

where $\rho_\epsilon = (1 - \epsilon)\rho + \epsilon I$ and I is the maximally mixed state. The function $P(\rho, U)$ only defines a hidden-variable theory if this limit actually exists.

The hidden-variable theory is called *local* if unitary transformations on some subsystem A of the system do not affect the value of the hidden variable on a separate subsystem B. A stronger property is *indifference*, which is the property that if U is block-diagonal, then $S(\rho, U)$ is block-diagonal with the same block structure or some refinement thereof. It is called *commutative* if the order of unitaries applied to separate subsystems is irrelevant. A theorem of Bell states that no hidden-variable theory satisfies both locality and commutativity. The theory is called *robust* if for every polynomial $q(N)$, there is a polynomial $p(N)$ such that perturbing the unitary U and density matrix ρ by at most $\frac{1}{p(N)}$ in the infinity norm changes the matrix $P(\rho, U)$ by at most $\frac{1}{q(N)}$ in the infinity norm. An example of a robust indifferent hidden variable theory is the flow theory $\mathcal{FT}$ defined in [2], which is based on network flows.

2.3 Total variation distance and trace distance

The *total variation distance* (or Kolmogorov distance) measures the ability to distinguish between two probability distributions. If μ and λ are two probability distributions on a finite sample space S, then their total variation distance is

$$d_{TV}(\mu, \lambda) := \frac{1}{2} \sum_{x \in S} |\mu(x) - \lambda(x)|.$$

It can also be defined by

$$d_{TV}(\mu, \lambda) = \sup_{A \subset S} (\mu(A) - \lambda(A)).$$

The total variation distance between two distributions is bounded between 0 and 1. A total variation distance of 0 indicates that the two distributions are identical, and a total variation distance of 1 indicates that the two distributions are disjoint. If X and Y are two random variables, we write $d_{TV}(X, Y)$ for the total variation distance between the distributions of X and Y.

A basic feature of total variation distance is that it satisfies the triangle inequality: if X, Y, Z are three random variables (or equivalently, probability distributions), then $d_{TV}(X, Z) \leq d_{TV}(X, Y) + d_{TV}(Y, Z)$. Another basic feature of total variation distance is that it cannot be "created" [13, p. 8]: if X and Y are two random variables and A is any randomized process[2], then

$$d_{TV}(A(X), A(Y)) \leq d_{TV}(X, Y).$$

[2]Formally, a randomized process is a function that takes as input a variable and outputs a random variable.

The quantum-mechanical analogue of total variation distance is *trace distance*. The trace distance between two density operators ρ_1, ρ_2 on the same state space measures the ability to distinguish between the two states ρ_1, ρ_2. It is equal to the maximum over all measurements M of

$$d_{TV}(M(\rho_1), M(\rho_2)),$$

where $M(\rho_i)$ is the outcome of measurement M on ρ_i. Equivalently, it is the maximum over all unitaries U of $\mathrm{Tr}\,|U(\rho_1 - \rho_2)U^\dagger|$, where $|\cdot|$ is the entrywise absolute value. We often write the trace distance between ρ_1, ρ_2 as $\frac{1}{2}\|\rho_1 - \rho_2\|_{tr}$.

3 The complexity class DQP

The complexity class DQP (Dynamical Quantum Polynomial Time) is the class of all problems solvable efficiently in the dynamic quantum model of computation. The basic idea is that a dynamic quantum computer is allowed to see the whole history of a hidden variable through some quantum computation (and postprocess it classically), as opposed to a quantum computer which can only see the final value of the hidden variable. In a vague sense, it can measure the state multiple times without collapsing it.

More formally, suppose that $U_1, \cdots, U_T$ are unitary transformations on ℓ qubits, each specified by a sequence of gates from some finite universal gate set $\mathcal{U}$. Then, a *history* of the hidden variable is a sequence $(v_0, \cdots, v_T)$ of computational basis states, with $v_0 = |0\rangle^{\otimes \ell}$. For any hidden-variable theory $\mathcal{T}$, the rule

$$\Pr[v = (v_0, \cdots, v_T)] = \prod_{k=0}^{T-1} (S_{\mathcal{T}}(U_k \cdots U_1 |0\rangle^{\otimes \ell}, U_{k+1}))_{v_k v_{k+1}}$$

defines a Markov distribution on histories. The oracle $\mathcal{O}(\mathcal{T})$ takes as input the unitaries $(U_1, \cdots, U_T)$, specified by sequences of gates from $\mathcal{U}$, and outputs a sample from this distribution.

Now, we are ready to define the complexity class DQP. The computational model is a deterministic classical polynomial-time Turing machine A that is allowed one oracle query to $\mathcal{O}(\mathcal{T})$. A language L is in DQP if there is such a Turing machine A, such that for *any* robust indifferent hidden-variable theory $\mathcal{T}$, the machine A correctly decides, with probability at least $2/3$, whether a string of length n is in L, for all sufficiently large n. It follows from the principle of deferred measurement that $\mathsf{DQP} \supset \mathsf{BQP}$, because viewing the entire history of a quantum system is at least as powerful as observing it only at the end of a computation [2]. It is important that there is one machine A that works for all robust indifferent hidden-variable theories $\mathcal{T}$.

4 The complexity class PDQP

Let $\mathcal{Q}_P$ be an oracle that takes as input a quantum circuit $C = (U_1, M_1, U_2, M_2, \cdots, U_T, M_T)$. Similarly to BQP, define the sequence $\{|\psi_t\rangle\}_{t=0}^T$ of quantum states by

$$|\psi_0\rangle = |0\rangle^{\otimes \ell} := \underbrace{|0 \cdots 0\rangle}_{\ell},$$

and for $t > 0$, $|\psi_t\rangle$ is the resulting state when the measurement M_t is applied to $U_t |\psi_{t-1}\rangle$. The oracle $\mathcal{Q}_P$ samples the sequence $\{|\psi_t\rangle\}_{t=0}^T$ (note that the random variables $|\psi_t\rangle$ are not independent), measures $|\psi_t\rangle$ for every t, and outputs the $T+1$ measurement results. The output is an element of $(\{0,1\}^\ell)^{T+1}$. We can think of the $T+1$ measurement samples as the results of *non-collapsing* measurements on the state vector, which give information about the state without changing it. We can also think of them as the history of a hidden variable which evolves according to the product theory $\mathcal{PT}$, defined by

$$(P_{\mathcal{PT}}(\rho, U))_{ij} = \rho_{ii}(U \rho U^\dagger)_{jj}.$$

(It is worth noting that the product theory is not indifferent.)

PDQP (Product Dynamical Quantum Polynomial-Time) is then defined as the class of all languages that can be recognized by a deterministic Turing machine with one query to $\mathcal{Q}_P$, with error probability at most $\frac{1}{3}$. It contains BQP, because the oracle $\mathcal{Q}_P$ gives more information than $\mathcal{Q}$ (which only returns a measurement sample from $|\psi_t\rangle$). The constant $\frac{1}{3}$ is arbitrary: we can decrease the error probability arbitrarily close to 0 by repetition, which can be accomplished by packing multiple copies of a quantum circuit into a single call to $\mathcal{Q}_P$.

It turns out that neither DQP nor PDQP is affected by the choice of universal gate set $\mathcal{U}$ (Appendix A).

5 SZK $\subset$ PDQP

Recall that SZK is the class of languages admitting statistical zero-knowledge proofs. (The precise definition can be found in [13]; it is not important here.) It includes important problems such as Graph Isomorphism and Approximate Shortest Vector not known to be in BQP.

Aaronson showed in [2] that SZK $\subset$ DQP. We introduce the new result that SZK $\subset$ PDQP as well. It is enough to prove that Statistical Difference, a problem shown in [13] to be SZK-complete, is in PDQP. The statistical difference problem is to determine, for two functions $P_0, P_1 : \{0,1\}^n \to \{0,1\}^m$ specified by classical circuits, whether the distributions of $P_0(X), P_1(X)$ for uniformly random X are close or far. Here, two distributions are "close" if their total variation distance is less than $\frac{1}{3}$ and they are "far" if their total variation distance is more than $\frac{2}{3}$.

By the Polarization Lemma of Sahai and Vadhan [13, Lemma 3.3], we can assume that the distributions $P_0(X)$ and $P_1(X)$ have total variation distance less than 2^{-n^c} or more than $1-2^{-n^c}$, for some large constant c. For now, assume that the distributions have total variation distance equal to either 1 or 0.

Prepare the state

$$\frac{1}{2^{(n+1)/2}} \sum_{b\in\{0,1\}, x\in\{0,1\}^n} |b\rangle\,|x\rangle\,|P_b(x)\rangle\,.$$

Now, measure the third register. If the distributions P_0, P_1 have total variation distance 1, then the resulting state of the first two registers will be of the form $|b\rangle\,|\psi\rangle$ for some b and $|\psi\rangle$. On the other hand, if they have total variation distance 0, the state of the first two registers will be an equal superposition $\frac{1}{\sqrt{2}}(|0\rangle\,|\psi_0\rangle + |1\rangle\,|\psi_1\rangle)$ where $|\psi_1\rangle$ and $|\psi_2\rangle$ have unit norm. We can distinguish the two cases by now repeatedly performing non-collapsing measurements of the value of the first register. If P_0, P_1 have total variation distance 1, then all of these measurements will give the same value b; if P_0 and P_1 have total variation distance 0, then each of these measurements will independently give 0 with probability $\frac{1}{2}$ and 1 with probability $\frac{1}{2}$. We can distinguish the two cases with high probability. Furthermore, it makes no difference that the total variation distances are merely exponentially close to 0 or 1 rather than actually being equal to 0 and 1.

Since [1] has the result that SZK $\not\subset$ BQP relative to an oracle, we have PDQP $\neq$ BQP relative to an oracle.

6 Search in $\tilde{O}(N^{1/3})$ time

Suppose that we are given query access to a function $f : \{0,1\}^n \to \{0,1\}$ such that the preimage $f^{-1}(1)$ contains exactly one element, x. In the classical randomized computational model, we can find x in $O(N)$ time, where $N = 2^n$. In the quantum computational model, we can find x in $O(N^{1/2})$ time using Grover's search algorithm [10]. In [2, Theorem 10], Aaronson shows a DQP procedure that finds x in $\tilde{O}(N^{1/3})$ time with $O(N^{1/3})$ queries to the function f. Here, the $\tilde{O}$ hides a factor polynomial in $\log N$.

The main result of this section is the following.

Theorem 6.1. *Suppose, in the definition of* PDQP*, that the unitaries $U_1, \cdots, U_T$ are now allowed to query f. That is, we are given access to the n-qubit gate U_f defined by*

$$U_f\,|y\rangle = (-1)^{f(y)}\,|y\rangle$$

for all $y \in \{0,1\}^n$. Then there is an algorithm to find the value of x that uses $O(N^{1/3})$ queries and $\tilde{O}(N^{1/3})$ time.

Proof. Prepare the uniform superposition of all basis states, and apply $O(N^{1/3})$ iterations of Grover's algorithm [10] to obtain the state

$$\alpha |x\rangle + \beta \sum_{y \in \{0,1\}^n} |y\rangle$$

with

$$\alpha = \frac{1}{\sqrt{2^{n/3} + 2^{-n/3+1} + 1}}$$
$$\beta = 2^{-n/3}\alpha.$$

Now, apply the identity operation $10N^{\frac{1}{3}} \log N$ times. We then claim that with high probability, the marked item x will appear at least twice in our measurement outcomes and no other item will appear more than once. Indeed, the marked item x has a probability at least $\frac{1}{10N^{1/3}}$ of turning up as a measurement outcome after each application of the identity, so it occurs at least twice with probability more than $1 - (\log N + 1)e^{-\log N} = 1 - o(1)$. Furthermore, the hidden variable has probability at most N^{-1} of visiting any particular value besides x at each step, so the probability that any item besides x occurs twice is, by the union bound, at most $N\frac{1}{N^2}\binom{10N^{1/3}\log N}{2} = o(1)$. Therefore, to find x, we simply look at which value appears most often in the hidden variable history. □

Note that if we are willing to use an enormous amount of *time*, we can search in the PDQP model using only one *query*: apply just one iteration of Grover's algorithm and then repeatedly apply the identity operation exponentially many times, and output the value that the hidden variable visits the most often.

7 Lower bounds for search

It turns out that no algorithm can search much faster than the one described above. In particular,

Theorem 7.1. *Suppose, in the definition of* PDQP*, that the unitaries $U_1, \cdots, U_T$ are now allowed to query f. Then any algorithm to find the value of x uses $\Omega(N^{1/4})$ time.*

The following lemma is essential: it bounds the total variation distance between two Markov distributions.

Lemma 7.1. *Suppose that $T \geq 1$, and that $v = (v_0, \cdots, v_T)$ is a random variable governed by a Markov distribution. That is, for all $1 \leq i \leq T$, v_i is independent of $v_0, \cdots, v_{i-2}$ conditioned on a particular value of v_{i-1}. Let $w = (w_0, \cdots, w_T)$ be another random variable governed by a Markov distribution. If $d_{TV}(\cdot, \cdot)$ denotes the total variation distance between random variables, then*

$$d_{TV}(v, w) \leq 2 \sum_{i=1}^{T} d_{TV}((v_{i-1}, v_i), (w_{i-1}, w_i)).$$

Proof. We proceed by induction on T. The base case $T = 1$ is trivial. For $T > 1$, since w_T depends only on w_{T-1} (by the Markov property), it is equal to $A(w_{T-1})$ for some randomized process A; let $w'_T := A(v_{T-1})$ be a variable that depends on v_{T-1} in exactly the same way that w_T depends on w_{T-1}. Then, define the random variable $v' = (v_0, \cdots, v_{T-1}, w'_T)$. By the triangle inequality,

$$d_{TV}(v, w) \leq d_{TV}(v, v') + d_{TV}(v', w). \tag{1}$$

Applying the same randomized process to two random variables cannot increase their total variation distance [13]. We can generate random variables identically distributed to v and v' by applying a suitable randomized process to (v_{T-1}, v_T) and (v_{T-1}, w'_T). We can also generate random variables identically

distributed to v' and w by applying a suitable randomized process to $(v_0, \cdots, v_{T-1})$ and $(w_0, \cdots, w_{T-1})$. Therefore, the right hand side of (1) is bounded above by

$$d_{TV}((v_{T-1}, v_T), (v_{T-1}, w'_T)) + d_{TV}((v_0, \cdots, v_{T-1}), (w_0, \cdots, w_{T-1})).$$

By the triangle inequality,

$$\begin{aligned} d_{TV}((v_{T-1}, v_T), (v_{T-1}, w'_T)) &\leq d_{TV}((v_{T-1}, v_T), (w_{T-1}, w_T)) + d_{TV}((w_{T-1}, w_T), (v_{T-1}, w'_T)) \\ &= d_{TV}((v_{T-1}, v_T), (w_{T-1}, w_T)) + d_{TV}(v_{T-1}, w_{T-1}) \\ &\leq 2d_{TV}((v_{T-1}, v_T), (w_{T-1}, w_T)). \end{aligned}$$

Putting all of this together,

$$d_{TV}(v, w) \leq 2d_{TV}((v_{T-1}, v_T), (w_{T-1}, w_T)) + d_{TV}((v_0, \cdots, v_{T-1}), (w_0, \cdots, w_{T-1})).$$

The result follows from induction. □

Lemma 7.2. *The trace distance between two pure states $|\psi\rangle\langle\psi|$ and $|\phi\rangle\langle\phi|$ is less than or equal to the 2-norm $\||\psi\rangle - |\phi\rangle\|_2$.*

Proof. The trace distance between $|\psi\rangle\langle\psi|$ and $|\phi\rangle\langle\phi|$ is equal to $\sqrt{1 - |\langle\psi|\phi\rangle|^2}$ [11, p. 415], and the 2-norm $\||\psi\rangle - |\phi\rangle\|_2$ is $\sqrt{2 - 2\mathrm{Re}(\langle\psi|\phi\rangle)}$. The inequality follows from $|\langle\psi|\phi\rangle| \leq 1$. □

From the hybrid argument of Bennett, Bernstein, Brassard, and Vazirani [7], we have the following lemma:

Lemma 7.3. *For any time t, if there are no measurements made before time t, we have*

$$\sum_{x=0}^{N-1} \||\psi_t\rangle - |\psi_t(x)\rangle\|_2^2 \leq 4Q^2.$$

Proof of Theorem 7.1. Since it is always possible to copy measured qubits, we can assume that qubits which are measured in an intermediate step of the algorithm are never directly modified again. Now, assume that the algorithm uses ℓ qubits and applies unitaries $U_1, \cdots, U_T$, each of which is either a (controlled) query to the search function f or a gate from the finite universal gate set $\mathcal{U}$. Measurements might be applied between the operators $U_1, \cdots, U_T$.

Let $v(x) = (v_0(x), v_1(x), \cdots, v_T(x))$ be the non-collapsing measurement results when the marked item is x, so that $v_i(x)$ is sampled immediately before the application of U_{i+1}. Let $v = (v_0, \cdots, v_T)$ be the non-collapsing measurement results when there is no marked item. In general, both $v(x)$ and v are random variables. Since the postprocessing step can distinguish the distributions of v and $v(x)$ with success probability 2/3,

$$d_{TV}(v, v(x)) \geq \frac{1}{3}$$

for all x. On the other hand, each v and $v(x)$ is a Markov process. Therefore, by Lemma 7.1,

$$d_{TV}(v, v(x)) \leq 2\sum_{i=1}^{T} d_{TV}((v_{i-1}, v_i), (v_{i-1}(x), v_i(x))).$$

Now, we bound the term $d_{x,i} := d_{TV}((v_{i-1}, v_i), (v_{i-1}(x), v_i(x)))$. Since it is possible to defer measurements in a quantum circuit to a later stage [11, p. 186], we can assume that all intermediate measurements that occurred before the application of U_i occurred immediately before the sampling of v_i. Suppose that

these measurements were applied to the first k qubits of the state. Let $|\phi\rangle$ and $|\phi(x)\rangle$ be the state vectors immediately before these measurements. Then, we decompose:

$$|\phi\rangle = \sum_{s\in\{0,1\}^k} \alpha_s |s\rangle |\phi_s\rangle$$
$$|\phi(x)\rangle = \sum_{s\in\{0,1\}^k} \beta_s |s\rangle |\phi_s(x)\rangle$$

Possible values for (v_{i-1}, v_i) and $(v_{i-1}(x), v_i(x))$ can be written in the form (st_1, st_2), where s is a k-bit string and t_1, t_2 are $\ell - k$-bit strings.

Assume for now that U_i does not contain a query to f. Then, since it does not affect the first k qubits, it can be decomposed into the sum

$$\sum_{s\in\{0,1\}^k} |s\rangle V_s \langle s|$$

for some unitaries V_s. The transformation U_i can be thought of as applying the unitary V_s to the last $\ell - k$ qubits if the (measured) first k qubits are equal to s. Then, the probability that $(v_{i-1}, v_i) = (st_1, st_2)$ is equal to

$$|\alpha_s|^2 |\langle t_1|\phi_s\rangle|^2 |\langle t_2| V_s |\phi_s\rangle|^2,$$

and the probability that $(v_{i-1}(x), v_i(x)) = (st_1, st_2)$ is equal to

$$|\beta_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2| V_s |\phi_s(x)\rangle|^2.$$

Therefore, the total variation distance $d_{x,i}$ is by the triangle inequality

$$\begin{aligned}
d_{x,i} &= \frac{1}{2} \sum_{s,t_1,t_2} \left| |\alpha_s|^2 |\langle t_1|\phi_s\rangle|^2 |\langle t_2| V_s |\phi_s\rangle|^2 - |\beta_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2| V_s |\phi_s(x)\rangle|^2 \right| \\
&\le \frac{1}{2} \sum_{s,t_1,t_2} \left(\left| |\alpha_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2| V_s |\phi_s(x)\rangle|^2 - |\beta_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2| V_s |\phi_s(x)\rangle|^2 \right| \right) \\
&\quad + \frac{1}{2} \sum_{s,t_1,t_2} \left(|\alpha_s|^2 \left| |\langle t_1|\phi_s\rangle|^2 |\langle t_2| V_s |\phi_s\rangle|^2 - |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2| V_s |\phi_s\rangle|^2 \right| \right) \\
&\quad + \frac{1}{2} \sum_{s,t_1,t_2} \left(|\alpha_s|^2 \left| |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2| V_s |\phi_s\rangle|^2 - |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2| V_s |\phi_s(x)\rangle|^2 \right| \right) \\
&=: \frac{1}{2}(S_1 + S_2 + S_3)
\end{aligned}$$

where S_1, S_2, S_3 are the three sums written above, which range over $s \in \{0,1\}^k$ and $t_1, t_2 \in \{0,1\}^{\ell-k}$. Now, we have:

$$\begin{aligned}
S_1 &:= \sum_{s,t_1,t_2} \left(\left| |\alpha_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2| V_s |\phi_s(x)\rangle|^2 - |\beta_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2| V_s |\phi_s(x)\rangle|^2 \right| \right) \\
&= \sum_s \left| |\alpha_s|^2 - |\beta_s|^2 \right| \left(\sum_{t_1,t_2} |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2| V_s |\phi_s(x)\rangle|^2 \right) \\
&= \sum_s \left| |\alpha_s|^2 - |\beta_s|^2 \right| \\
&\le \left\| |\phi\rangle\langle\phi| - |\phi(x)\rangle\langle\phi(x)| \right\|_{tr} \\
&\le 2 \left\| |\phi(x)\rangle - |\phi\rangle \right\|_2 .
\end{aligned}$$

Additionally,

$$\begin{aligned}
S_2 &:= \sum_{s,t_1,t_2} \left(|\alpha_s|^2 \left| |\langle t_1|\phi_s\rangle|^2 |\langle t_2| V_s |\phi_s\rangle|^2 - |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2| V_s |\phi_s\rangle|^2 \right| \right) \\
&= \sum_{s,t_1} \left(|\alpha_s|^2 \left| |\langle t_1|\phi_s\rangle|^2 - |\langle t_1|\phi_s(x)\rangle|^2 \right| \right) \\
&\leq \sum_{s,t_1} \left(\left| |\alpha_s|^2 |\langle t_1|\phi_s\rangle|^2 - |\beta_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 \right| \right) + \sum_{s,t_1} \left(\left| |\alpha_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 - |\beta_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 \right| \right) \\
&= \sum_{s,t_1} \left(\left| |\alpha_s|^2 |\langle t_1|\phi_s\rangle|^2 - |\beta_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 \right| \right) + \sum_{s} \left(\left| |\alpha_s|^2 - |\beta_s|^2 \right| \right) \\
&\leq 2 \left\| |\phi\rangle\langle\phi| - |\phi(x)\rangle\langle\phi(x)| \right\|_{tr} \\
&\leq 4 \left\| |\phi(x)\rangle - |\phi\rangle \right\|_2 .
\end{aligned}$$

Finally,

$$\begin{aligned}
S_3 &= \sum_{s,t_1,t_2} \left(|\alpha_s|^2 \left| |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2| V_s |\phi_s\rangle|^2 - |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2| V_s |\phi_s(x)\rangle|^2 \right| \right) \\
&= \sum_{s,t_2} \left(|\alpha_s|^2 \left| |\langle t_2| V_s |\phi_s\rangle|^2 - |\langle t_2| V_s |\phi_s(x)\rangle|^2 \right| \right) \\
&\leq \sum_{s,t_2} \left(\left| |\alpha_s|^2 |\langle t_2| V_s |\phi_s\rangle|^2 - |\beta_s|^2 |\langle t_2| V_s |\phi_s(x)\rangle|^2 \right| \right) \\
&\quad + \sum_{s,t_2} \left(\left| |\alpha_s|^2 |\langle t_2| V_s |\phi_s(x)\rangle|^2 - |\beta_s|^2 |\langle t_2| V_s |\phi_s(x)\rangle|^2 \right| \right) \\
&= \sum_{s,t_2} \left(\left| |\alpha_s|^2 |\langle t_2| V_s |\phi_s\rangle|^2 - |\beta_s|^2 |\langle t_2| V_s |\phi_s(x)\rangle|^2 \right| \right) + \sum_{s} \left(\left| |\alpha_s|^2 - |\beta_s|^2 \right| \right) \\
&\leq 2 \left\| |\phi\rangle\langle\phi| - |\phi(x)\rangle\langle\phi(x)| \right\|_{tr} \\
&= 4 \left\| |\phi(x)\rangle - |\phi\rangle \right\|_2
\end{aligned}$$

Therefore,

$$d_{x,i} \leq \frac{1}{2}(S_1 + S_2 + S_3) \leq 5 \left\| |\phi(x)\rangle - |\phi\rangle \right\|_2 .$$

On the other hand, if U_i is a query to f, then it only applies a local phase of -1 to some of the probability amplitudes of $|\phi\rangle$ and $|\phi_x\rangle$. Therefore, the same argument still shows that $d_{x,i} \leq 5 \left\| |\phi(x)\rangle - |\phi\rangle \right\|_2$.

By the Cauchy-Schwarz inequality and Lemma 7.3,

$$\begin{aligned}
\frac{1}{N} \sum_{x=0}^{N-1} d_{x,i} &\leq 5 \cdot \frac{1}{N} \sum_{x=0}^{N-1} \left\| |\phi(x)\rangle - |\phi\rangle \right\|_2 \\
&\leq 5 \sqrt{\frac{1}{N} \sum_{x=0}^{N-1} \left\| |\phi(x)\rangle - |\phi\rangle \right\|_2^2} \\
&\leq \frac{10Q}{\sqrt{N}}
\end{aligned}$$

for all i. Therefore, there is some x for which

$$d_{TV}(v, v(x)) \leq 2 \sum_{i=1}^{T} d_{x,i} \leq \frac{20TQ}{\sqrt{N}}.$$

On the other hand, $d_{TV}(v, v(x)) \geq \frac{1}{3}$ for all x, so

$$\frac{20TQ}{\sqrt{N}} \geq \frac{1}{3},$$

and the running time of the algorithm is at least $T + Q = \Omega(N^{1/4})$. □

An important corollary of Theorem 7.1 is that, relative to an oracle, $\mathsf{NP} \not\subset \mathsf{PDQP}$. This follows from the well-known "diagonalization method" of Baker, Gill, and Solovay [6].

8 An upper bound on PDQP

Adleman, DeMarrais, and Huang showed in [4] that $\mathsf{BQP} \subset \mathsf{PP}$, where PP is the class of languages that can be recognized by a randomized Turing machine with error probability less than $\frac{1}{2}$. This is the best known classical upper bound on BQP. The proof involves writing the acceptance probability of a quantum circuit as an exponentially large sum of terms (an expression originating from the Feynman path integral formulation of quantum mechanics), each of which can be efficiently computed in isolation. By sampling a random term of this sum, it is possible, using standard techniques, to determine whether the sum is greater than $\frac{2}{3}$ or less than $\frac{1}{3}$, with probability strictly greater than $\frac{1}{2}$.

The acceptance probability of a PDQP circuit as an exponentially large sum of terms, each of which is the quotient of two exponentially large sums. Unfortunately, due to the division, it is not possible to directly adapt the proof of $\mathsf{BQP} \subset \mathsf{PP}$ to this setting. However, we can prove the weaker result that $\mathsf{PDQP} \subset \mathsf{PP}^{\mathsf{PP}}$, where $\mathsf{PP}^{\mathsf{PP}}$ is the class of languages that can be recognized by a randomized Turing machine with error probability less than $\frac{1}{2}$, where the machine has an oracle that can solve any problem in PP in constant time. A consequence is that $\mathsf{PDQP} \subset \mathsf{PSPACE}$. Our proof involves an alternative definition of PDQP.

8.1 Alternative definition of PDQP

If B is a partition of $\{0,1\}^\ell$ and U is a unitary operator on $(\mathbb{C}^2)^{\otimes \ell}$, then we say that U respects the block structure B if $U_{ij} = 0$ whenever i and j are in different parts of B. If ρ is a density operator and U is a unitary that respects the block structure B, then the stochastic matrix $S_{\mathcal{PT}_B}(\rho, U)$ is formed by applying the product theory $\mathcal{PT}$ separately on each block of B. More precisely, let $\sim$ be the equivalence relation on $\{1, \cdots, n\}$ defined by $i \sim j$ if and only if i and j are in the same block of B. Then,

$$(S_{\mathcal{PT}_B}(\rho, U))_{ij} = \begin{cases} \frac{(U\rho U^\dagger)_{jj}}{\sum_{k \sim j} (U\rho U^\dagger)_{kk}} & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

where the sum over k ranges over all k with $k \sim j$. If $|\psi\rangle$ is a state vector, define $S_{\mathcal{PT}_B}(|\psi\rangle, U) := S_{\mathcal{PT}_B}(|\psi\rangle\langle\psi|, U)$.

Suppose that $\mathcal{V} = (U_1, \cdots, U_T)$ are unitary operators on ℓ qubits, and $\mathcal{B} = (B_1, \cdots, B_T)$ are partitions of $\{0, \cdots, 2^\ell - 1\}$ such that for every i, B_{i+1} is a refinement of B_i, and U_i respects the block structure B_i. Then they define a probability distribution $\Omega = \Omega_{\mathcal{PT}}(\mathcal{V}, \mathcal{B})$ over hidden variable histories $v = (v_0, \cdots, v_T)$ by

$$\Omega_{(v_0, \cdots, v_T)} = \prod_{k=1}^{T} (S_{\mathcal{PT}_{B_k}}(U_{k-1} \cdots U_1 |0\rangle^{\otimes \ell}, U_k))_{v_{k-1} v_k}.$$

The oracle $\mathcal{Q}_B$ takes as input the unitaries $U_1, \cdots, U_T$ specified by sequences of gates from some finite universal gate set $\mathcal{U}$. It also takes as input the partitions $B_1, \cdots, B_T$, specified by polynomial-time computable functions $b_1, \cdots, b_T : \{0,1\}^\ell \to \{0,1\}^m$ satisfying the property that x and y are in the same part of the partition B_i if and only if $b_i(x) = b_i(y)$. It outputs a sample from the distribution $\Omega_{\mathcal{PT}}(\mathcal{V}, \mathcal{B})$. Then, let PDQP' be the class of all languages that can be recognized by a polynomial-time Turing machine with one query to $\mathcal{Q}_B$, with error probability at most $\frac{1}{3}$.

Lemma 8.1. $\mathsf{PDQP}' = \mathsf{PDQP}$.

Proof. We first demonstrate a procedure for converting oracle queries to $\mathcal{Q}_B$ to oracle queries to $\mathcal{Q}_P$. Suppose that $B_1, \cdots, B_T$ are specified by polynomial-time computable functions $b_1, \cdots, b_T : \{0,1\}^\ell \to \{0,1\}^m$ (so that x, y are in the same part of the partition B_i if and only if $b_i(x) = b_i(y)$). Now, add an extra T registers of m qubits each, which start in the state $|0 \cdots 0\rangle$. Create a quantum circuit with the same unitary operators $U_1, \cdots, U_T$, but before applying the unitary U_i, apply a unitary that writes the value $|b_i(x)\rangle$ to the ith register when the first ℓ qubits are $|x\rangle$. Then measure the ith register. The effect is that the non-collapsing measurement results will never jump from one part of B_i to a different part, which is exactly what is desired.

To convert a query $C = (U_1, M_1, \cdots, U_T, M_T)$ to $\mathcal{Q}_P$ to a query to $\mathcal{Q}_B$, we first assume, as in the proof of Theorem 7.1, that measured qubits are never modified again. Keep the unitaries $U_1, \cdots, U_T$ and let B_i be the partition of $\{0,1\}^\ell$ induced by the measurements $M_1, \cdots, M_{i-1}$. By the principle of deferred measurement, $\Omega_{\mathcal{V},\mathcal{B}}$ is the same distribution that we would have seen had we queried $\mathcal{Q}_P$ instead. □

8.2 $\mathsf{PDQP} \subset \mathsf{PP}^{\mathsf{PP}}$

We are now ready to prove that $\mathsf{PDQP} \subset \mathsf{PP}^{\mathsf{PP}}$. First, we prove that a #P oracle can compute exponentially large sums, when the summands are written in such a way that numerical precision is not an issue. Here, a #P oracle accepts as input a polynomial-time computable function $p : \{0,1\}^M \to \{0,1\}$ and returns the number of $x \in \{0,1\}^M$ with $p(x) = 1$.

Lemma 8.2. *Let $\mathbb{Z}[1/5, i]$ be the ring of all complex numbers that can be written in the form $\frac{a+bi}{5^c}$, where a, b, c are integers. Suppose that a function $f : \{0,1\}^m \to \mathbb{Z}[1/5, i]$ is polynomial-time computable, where a number $\frac{a+bi}{5^c}$ in $\mathbb{Z}[1/5]$ is specified by the triple $(a, b, 5^c)$. (In particular, c must be bounded by a polynomial.) Then, using a #P oracle, it is possible to compute the value of the sum*

$$\sum_{x \in \{0,1\}^m} f(x)$$

in polynomial time.

Proof. Since we can sum complex numbers by summing their real and imaginary parts separately, it suffices to consider the case where f is real.

For $k = 0, 1, 2, \cdots$, using the #P oracle, find the number a_k of x for which the denominator of $f(x)$ is greater than 5^k. For a polynomially sized k, a_k will be 0, and 5^k will be a common denominator for all the $f(x)$. Therefore, the values of the function $g(x) : \{0,1\}^m \to \mathbb{Z}$ defined by $g(x) = 5^k f(x)$ are all integers. The function g can also be computed in polynomial time.

For $\ell = 0, 1, 2, \cdots$, using the #P oracle, find the number b_ℓ of x for which $g(x) < -2^\ell$. For some polynomially sized ℓ, $b_\ell = 0$, and $h(x) := 2^\ell + g(x)$ will be nonnegative. The function $h(x)$ is also polynomial-time computable.

Now, find some polynomially-sized M for which 2^M is an upper bound on $h(x)$. Using the #P oracle, compute the number of pairs (x, y) where $x \in \{0,1\}^m$ and $0 \leq y < 2^M$ for which $y < h(x)$. This number is exactly equal to

$$S = \sum_{x \in \{0,1\}^m} h(x).$$

We can then recover

$$\sum_{x \in \{0,1\}^m} f(x) = \sum_{x \in \{0,1\}^m} \frac{1}{5^k}(h(x) - 2^\ell) = \frac{1}{5^k}(S - 2^{\ell+m}).$$

□

Theorem 8.1. $\mathsf{PDQP} \subset \mathsf{PP}^{\mathsf{PP}}$.

Proof. Since $\mathsf{PP}^{\mathsf{PP}} = \mathsf{PP}^{\#\mathsf{P}}$ [5], it is enough to prove $\mathsf{PDQP} \subset \mathsf{PP}^{\#\mathsf{P}}$.

We use the alternative formulation PDQP' given by Lemma 8.1. It is enough to be able to simulate any PDQP' computation in $\mathsf{PP}^{\#\mathsf{P}}$. Since the universal gate set used in the computation does not matter (Appendix A), we may assume by a result of Shi [14, Theorem 1.2] that all the components U_i are Controlled-NOT gates or one of the two single qubit gates

$$\exp\left(i\frac{X}{2}\cos^{-1}\left(\frac{3}{5}\right)\right) = \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix}$$

$$\exp\left(i\pi\frac{Y}{4}\right) = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}.$$

First, we show, given the unitaries $U_1, \cdots, U_T$ and the partitions $B_1, \cdots, B_T$ specified by hash functions $b_1, \cdots, b_T$, a method to compute the probability Ω_v of any particular history $v := (v_0, \cdots, v_T)$, where $\Omega = \Omega_{\mathcal{PT}}(\mathcal{U}, \mathcal{B})$, in polynomial time using the $\#\mathsf{P}$ oracle with an absolute error of at most $\epsilon := \frac{1}{8}2^{-\ell T}$.

To compute the probabilities, we use the formula

$$\Omega_v = \prod_{k=1}^{T} (S_{\mathcal{PT}_{B_k}}(U_{k-1}\cdots U_1 |0\rangle^{\otimes \ell}, U_k))_{v_{k-1}v_k}.$$

Our strategy is to compute each factor of this product *exactly* as a rational number. Fix k with $1 \le k \le T$. Let $|\psi\rangle = U_k U_{k-1}\cdots U_1 |0\rangle^{\otimes \ell}$, and let $\alpha_x = \langle x|\psi\rangle$ be the amplitudes of $|\psi\rangle$. Then, we can write α_x as the sum

$$\alpha_x = \sum_{j_k \in \{0,1\}^\ell} \cdots \sum_{j_2 \in \{0,1\}^\ell} \sum_{j_1 \in \{0,1\}^\ell} \prod_{t=1}^{k} (U_t)_{j_{t-1} j_t},$$

where $j_0 = 0^\ell$. Each term of this sum is computable individually in polynomial time, and it is an element of the ring $\mathbb{Z}[1/5, i]$.

If v_{k-1} and v_k are in different parts of B_k, then

$$(S_{\mathcal{PT}_{B_k}}(U_{k-1}\cdots U_1 |0\rangle^{\otimes \ell}, U_k))_{v_{k-1}v_k} = 0.$$

Otherwise,

$$(S_{\mathcal{PT}_{B_k}}(U_{k-1}\cdots U_1 |0\rangle^{\otimes \ell}, U_k))_{v_{k-1}v_k} = \frac{|\alpha_{v_k}|^2}{\sum_{x=0}^{2^\ell - 1} \delta_{b_k(x), b_k(v_k)} |\alpha_x|^2}$$

where δ is the Kronecker delta. Since every entry of every unitary in our gate set is in $\mathbb{Z}[1/5, i]$, each $|\alpha_x|^2 = \alpha_x \overline{\alpha_x}$ can be written as an exponentially large sum of elements of $\mathbb{Z}[1/5, i]$, and then computed using the $\#\mathsf{P}$ oracle (by Lemma 8.2). (Here we use the fact that the product of two sums can be expanded out into a single sum.) Also, the entire sum

$$\sum_{x=0}^{2^{\ell-1}} \delta_{b_k(x), b_k(v_k)} |\alpha_x|^2$$

can be written as an exponentially large sum of elements of $\mathbb{Z}[1/5, i]$, and then computed using the $\#\mathsf{P}$ oracle (by Lemma 8.2). Using the result of these computations, it is possible to compute the probability Ω_v in the form of a rational number whose numerator and denominator both have polynomially many bits, and then we can divide the numerator and denominator to within an absolute error of ϵ in polynomial time.

Therefore, for any particular $v = (v_0, \cdots, v_T)$, we can compute Ω_v using a $\#\mathsf{P}$ oracle. Let V be the space of all hidden-variable histories (it has size $2^{\ell T}$), and define the function $a : V \to \{0, 1\}$ by $a(v) = 1$ if the postprocessing step accepts v, and 0 otherwise. We want to decide whether

$$\sum_{v \in V} \Omega_v a(v)$$

is greater than $\frac{2}{3}$ or less than $\frac{1}{3}$. We can compute any particular term of this sum with absolute error at most ϵ.

Define $p(v) = \Omega_v a(v) - \frac{1}{2}2^{\ell T}$, so that we are now trying to decide whether

$$P = \sum_{v \in V} p(v)$$

is at least $\frac{1}{6}$ or at most $-\frac{1}{6}$. Sample a uniformly random $v \in V$ and calculate an approximation $\tilde{p}(v)$ to $p(v)$ using the #P oracle, where $|\tilde{p}(v) - p(v)| < \epsilon$. Then, output "yes" with probability $\frac{1+\tilde{p}(v)}{2}$. Since $|\tilde{p}(v)| \leq 1$, this is a well-defined probability. The probability that this procedure outputs "yes" is the average

$$\frac{1 + \frac{1}{|V|}\sum_{v \in V} \tilde{p}(v)}{2}.$$

As desired, this is greater than $\frac{1}{2}$ if $P \geq \frac{1}{6}$ and less than $\frac{1}{2}$ if $P \leq -\frac{1}{6}$. □

9 Directions for further research

The results given in this paper leave many questions about the complexity classes DQP and PDQP unanswered.

1. Can we improve the upper bound $\mathsf{PDQP} \subset \mathsf{PP}^{\mathsf{PP}}$ to $\mathsf{PDQP} \subset \mathsf{PP}$? One possible way to approach this problem is to use the alternative formulation of PP as PostBQP [3].
2. We demonstrated a $\tilde{O}(N^{1/3})$-time algorithm for the search problem in the PDQP model, as well as the result that any search algorithm takes $\Omega(N^{1/4})$ time. Is it possible to close the gap between these two bounds?
3. Can we demonstrate a lower bound, superpolynomial in $\log N$, for the running time of a search algorithm in the DQP model? The proof given in [2] of an $\Omega(N^{1/3})$ lower bound is flawed (Appendix B).
4. Is there a hierarchy of computational models for which the kth allows searching in $\tilde{O}(N^{1/k})$ time?

Appendices

A Universal gate set does not matter

We prove that the universal gate set $\mathcal{U}$ used in the definition of PDQP does not matter. (As a side note, the same proof shows that the universal gate set used in the definition of DQP does not matter.) Our proof relies on Lemma 7.1 and the Solovay-Kitaev theorem [9] to show that any computation using a particular universal gate set $\mathcal{U}$ can be done using a different gate set $\mathcal{U}'$ in such a way that the distributions of the histories does not change significantly in total variation distance.

Theorem A.1. *Any universal gate set $\mathcal{U}$ yields the same complexity class* PDQP.

Proof. If A is an operator on a Hilbert space, denote by $\|A\|$ the maximum value of $\|A|\phi\rangle\|_2$ over all ϕ with $\||\phi\rangle\|_2 = 1$.

Lemma A.1. *Suppose that $V_1, \cdots, V_m$ and $V_1', \cdots, V_m'$ are unitary operators. Then,*

$$\|V_1 \cdots V_m - V_1' \cdots V_m'\| \leq \sum_{k=1}^{m} \|V_k - V_k'\|.$$

Proof. By induction, it suffices to prove the statement for $m = 2$. We have

$$\begin{aligned}
\|V_1V_2 - V_1'V_2'\| &= \max_{\||\phi\rangle\|_2=1} \|V_1V_2 |\phi\rangle - V_1'V_2' |\phi\rangle\|_2 \\
&\leq \max_{\||\phi\rangle\|_2=1} (\|V_1V_2 |\phi\rangle - V_1'V_2 |\phi\rangle\|_2 + \|V_1'V_2 |\phi\rangle - V_1V_2 |\phi\rangle\|) \\
&= \max_{\||\phi\rangle\|_2=1} (\|(V_1 - V_1')V_2 |\phi\rangle\|_2 + \|(V_2 - V_2') |\phi\rangle\|_2) \\
&\leq \|V_1 - V_1'\| + \|V_2 - V_2'\| .
\end{aligned}$$

□

If ρ is a density operator and U is a unitary operator on ℓ qubits that respects the block structure B, then define the joint probabilities matrix $P_{\mathcal{PT}_B}(\rho, U)$ by

$$(P_{\mathcal{PT}_B}(\rho, U))_{ij} = \begin{cases} \frac{\rho_{ii}(U\rho U^\dagger)_{jj}}{\sum_{k\sim j}(U\rho U^\dagger)_{kk}} & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}.$$

If $|\psi\rangle$ is a state vector, define $P_{\mathcal{PT}_B}(|\psi\rangle, U) := P_{\mathcal{PT}_B}(|\psi\rangle \langle\psi|, U)$. It is straightforward to show that

$$\|P_{\mathcal{PT}_B}(|\psi\rangle, U) - P_{\mathcal{PT}_B}(|\psi'\rangle, U')\|_1 \leq 2^{2\ell}(\||\psi\rangle - |\psi'\rangle\|_{tr} + \|U - U'\|)$$

whenever $|\psi\rangle, |\psi'\rangle$ are state vectors and U, U' are unitaries.

We use the alternative formulation PDQP' (Lemma 8.1). Suppose that $\mathcal{U}$ and $\mathcal{U}'$ are two universal gate sets, and that $\mathcal{V} = (U_1, \cdots, U_T)$ and $\mathcal{B} = (B_1, \cdots, B_T)$ are a query to the $\mathcal{Q}_B$ oracle, where the operators U_t are specified by sequences of gates from $\mathcal{U}$. It is enough to be able to compute in polynomial time a sequence $\mathcal{V}' = (U_1', \cdots, U_T')$ of unitaries, specified by sequences of gates from $\mathcal{U}'$, such that

$$d_{TV}(\Omega_{\mathcal{PT}}(\mathcal{V}, \mathcal{B}), \Omega_{\mathcal{PT}}(\mathcal{V}', \mathcal{B})) < \frac{1}{8}.$$

Let $\epsilon = 2^{-\ell^2 T - 10}$. Then, by the Solovay-Kitaev theorem [9], it is possible to compute in polynomial time a sequence $\mathcal{V}' = (U_1', \cdots, U_T')$ such that

$$\|U_t - U_t'\| \leq \epsilon$$

for all t. Suppose that $v = (v_0, \cdots, v_T)$ is sampled from $\Omega_{\mathcal{PT}}(\mathcal{V}, \mathcal{B})$, and that $v' = (v_0', \cdots, v_T')$ is sampled from $\Omega_{\mathcal{PT}}(\mathcal{V}', \mathcal{B})$. Then,

$$d_{TV}(\Omega_{\mathcal{PT}}(\mathcal{V}, \mathcal{B}), \Omega_{\mathcal{PT}}(\mathcal{V}', \mathcal{B})) = d_{TV}(v, v').$$

By Lemma 7.1,

$$\begin{aligned}
d_{TV}(v,v') &\leq 2\sum_{i=1}^{T} d_{TV}((v_{i-1},v_i),(v'_{i-1},v'_i)) \\
&= 2\sum_{i=1}^{T} \left\| P_{\mathcal{PT}_{B_i}}(U_{i-1}\cdots U_1 |0\rangle^{\otimes \ell}, U_i) - P_{\mathcal{PT}_{B_i}}(U'_{i-1}\cdots U'_1 |0\rangle^{\otimes \ell}, U'_i) \right\|_1 \\
&\leq 2^{2\ell+1}\sum_{i=1}^{T} \left(\left\| U_{i-1}\cdots U_1 |0\rangle^{\otimes \ell} - U'_{i-1}\cdots U'_1 |0\rangle^{\otimes \ell} \right\|_2 + \|U_i - U'_i\| \right) \\
&\leq 2^{2\ell+1}\sum_{i=1}^{T} \left(\left\| U_{i-1}\cdots U_1 - U'_{i-1}\cdots U'_1 \right\| + \epsilon \right) \\
&\leq 2^{2\ell+1}\sum_{i=1}^{T} \left(\sum_{k=0}^{i-1} \|U_i - U'_i\| + \epsilon \right) \\
&\leq 2^{2\ell+1}\sum_{i=1}^{T} (T\epsilon + \epsilon) \\
&\leq \frac{1}{8},
\end{aligned}$$

as desired. □

B The DQP search time lower bound

As mentioned earlier, the proof in [2] that any algorithm for the search problem in DQP takes at least $\Omega(N^{1/3})$ time is flawed. The proof is based on the hybrid argument: it shows that changing the marked item from x to x^* does not affect any particular entry v_i of the hidden-variable history by very much (in the total variation distance).

The error is that there is no union bound for total variation distance. That is, the result of Lemma 7.1 cannot be strengthened to

$$d_{TV}(v,w) \leq \sum_{i=0}^{T} d_{TV}(v_i,w_i).$$

A specific counterexample is $T = 1$, where v is $(0,0)$ with probability $\frac{1}{2}$ and $(1,1)$ with probabilty $\frac{1}{2}$, and w is $(0,1)$ with probability $\frac{1}{2}$ and $(1,0)$ with probabilty $\frac{1}{2}$. Then,

$$d_{TV}(v,w) = 0$$

whereas

$$\sum_{i=0}^{T} d_{TV}(v_i,w_i) = 1$$

When Aaronson bounds "the probability of noticing the $x \to x^*$ change" by the union bound, he implicitly uses this strengthening of Lemma 7.1.

Acknowledgements

I thank my mentor Adam Bouland for teaching me all I know about quantum computation, as well as being extremely helpful throughout the entire process of writing this paper. I thank Pavel Etingof, Jacob Fox, and Scott Aaronson for valuable correspondence. I thank Slava Gerovitch for directing the SPUR program.

References

[1] Scott Aaronson. Quantum lower bound for the collision problem. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, STOC '02, pages 635–642, New York, NY, USA, 2002. ACM.

[2] Scott Aaronson. Quantum computing and hidden variables. *Phys. Rev. A*, 71:032325, Mar 2005.

[3] Scott Aaronson. Quantum computing, postselection, and probabilistic polynomial-time. In *Proceedings of the Royal Society A*, page 0412187, 2005.

[4] Leonard M. Adleman, Jonathan Demarrais, Ming-deh, and A. Huang. Quantum computability. *SIAM Journal of Computation*, pages 1524–1540, 1997.

[5] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.

[6] T. Baker, J. Gill, and R. Solovay. Relativizations of the $\mathcal{P} =? \mathcal{NP}$ question. *SIAM Journal on Computing*, 4(4):431–442, 1975.

[7] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and weaknesses of quantum computing. *SIAM J. Comput.*, 26(5):1510–1523, October 1997.

[8] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. In *in Proc. 25th Annual ACM Symposium on Theory of Computing, ACM*, pages 11–20, 1993.

[9] Christopher M. Dawson and Michael A. Nielsen. The Solovay-Kitaev algorithm. *Quantum Info. Comput.*, 6(1):81–95, January 2006.

[10] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, STOC '96, pages 212–219, New York, NY, USA, 1996. ACM.

[11] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information (Cambridge Series on Information and the Natural Sciences)*. Cambridge University Press, 1 edition, January 2004.

[12] A Peres. Two simple proofs of the Kochen-Specker theorem. *Journal of Physics A: Mathematical and General*, 24(4):L175, 1991.

[13] A. Sahai and S.P. Vadhan. A complete promise problem for statistical zero-knowledge. In *Foundations of Computer Science, 1997. Proceedings., 38th Annual Symposium on*, pages 448–457, 1997.

[14] Yaoyun Shi. Both Toffoli and controlled-NOT need little help to do universal quantum computing. *Quantum Info. Comput.*, 3(1):84–92, January 2003.