Algorithmica
https://doi.org/10.1007/s00453-018-0527-1

Approximate Span Programs

Tsuyoshi Ito[1] · **Stacey Jeffery**[2]

Received: 16 July 2017 / Accepted: 8 November 2018

Abstract
Span programs are a model of computation that have been used to design quantum algorithms, mainly in the query model. It is known that for any decision problem, there exists a span program that leads to an algorithm with optimal quantum query complexity, however finding such an algorithm is generally challenging. We consider new ways of designing quantum algorithms using span programs. We show how any span program that decides a function f can also be used to decide "threshold" versions of the function f, or more generally, approximate a quantity called the *span program witness size*, which is some property of the input related to f. We achieve these results by relaxing the requirement that 1-inputs hit some *target* exactly in the span program, which could potentially make design of span programs significantly easier. In addition, we give an exposition of span program structure, which increases the general understanding of this important model. One implication of this is alternative algorithms for estimating the witness size when the phase gap of a certain unitary can be lower bounded. We show how to lower bound this phase gap in certain cases. As an application, we give the first upper bounds in the adjacency query model on the quantum time complexity of estimating the effective resistance between s and t, $R_{s,t}(G)$. For this problem we obtain $\widetilde{O}(\frac{1}{\varepsilon^{3/2}} n\sqrt{R_{s,t}(G)})$, using $O(\log n)$ space. In addition, when μ is a lower bound on $\lambda_2(G)$, by our phase gap lower bound, we can obtain an upper bound of $\widetilde{O}\left(\frac{1}{\varepsilon} n\sqrt{R_{s,t}(G)/\mu}\right)$ for estimating effective resistance, also using $O(\log n)$ space.

Keywords Quantum algorithms · Span programs · Quantum query complexity · Effective resistance

SJ completed parts of this work while at the Institute for Quantum Information and Matter (IQIM), Caltech. This work was supported by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NFS Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028).

✉ Stacey Jeffery
jeffery@cwi.nl

[1] Tokyo, Japan

[2] QuSoft and CWI, Amsterdam, The Netherlands

Published online: 14 November 2018

1 Introduction

Span programs are a model of computation first used to study logspace complexity [15], and more recently, introduced to the study of quantum algorithms in [21]. They are of immense theoretical importance, having been used to show that the general adversary bound gives a tight lower bound on the quantum query complexity of any decision problem [19,20]. As a means of designing quantum algorithms, it is known that for any decision problem, there exists a span-program-based algorithm with asymptotically optimal quantum query complexity, but this fact alone gives no indication of how to find such an algorithm. Despite the relative difficulty in designing quantum algorithms this way, there are many applications, including formula evaluation [20,21], a number of algorithms based on the learning graph framework [5], st-connectivity [7] and k-distinctness [4]. Although generally quantum algorithms designed via span programs can only be analyzed in terms of their query complexity, in some cases their time complexity can also be analyzed, as is the case with the quantum algorithm for st-connectivity. In the case of the quantum algorithm for k-distinctness, the ideas used in designing the span program could be turned into a quantum algorithm for 3-distinctness with time complexity matching its query complexity up to logarithmic factors [3].

In this work, we consider new ways of designing quantum algorithms via span programs. Consider Grover's quantum search algorithm, which, on input $x \in \{0,1\}^n$, decides if there is some $i \in [n]$ such that $x_i = 1$ using only $O(\sqrt{n})$ quantum operations [11]. The ideas behind this algorithm have been used in innumerable contexts, but in particular, a careful analysis of the ideas behind Grover's algorithm led to algorithms for similar problems, including a class of *threshold functions*: given $x \in \{0,1\}^n$, decide if $|x| \geq t$ or $|x| < \varepsilon t$, where $|x|$ denotes the Hamming weight; and approximate counting: given $x \in \{0,1\}^n$, output an estimate of $|x|$ to some desired accuracy. The results in this paper offer the possibility of obtaining analogous results for any span program. That is, given a span program for some problem f, our results show that one can obtain, not only an algorithm for f, but algorithms for a related class of threshold functions, as well as an algorithm for estimating a quantity called the *span program witness size*, which is analogous to $|x|$ in the above example (and is in fact exactly $1/|x|$ in the span program for the OR function — see Sect. 3.3).

New Algorithms from Span Programs We give several new means of constructing quantum algorithms from span programs. Roughly speaking, a span program can be turned into a quantum algorithm that decides between two types of inputs: those that "hit" a certain "target vector", and those that don't. We show how to turn a span program into an algorithm that decides between inputs that get "close to" the target vector, and those that don't. Whereas traditionally a span program has been associated with some decision problem, we can now associate, with one span program, a whole class of threshold problems.

In addition, for any span program P, we can construct a quantum algorithm that estimates the *positive witness size*, $w_+(x)$, to accuracy ε in $\frac{1}{\varepsilon^{3/2}}\sqrt{w_+(x)\widetilde{W}_-}$ queries, where $\widetilde{W}_-$ is the *approximate negative witness complexity* of P. This construction is useful whenever we can construct a span program for which $w_+(x)$ corresponds to some function we care to estimate, as is the case with the span program for OR, in which

$w_+(x) = \frac{1}{|x|}$, or the span from for st-connectivity, in which $w_+(G) = \frac{1}{2}R_{s,t}(G)$, where G is a graph, and $R_{s,t}(G)$ is the *effective resistance* between s and t in G. We show similar results for estimating the negative witness size.

Structural Results Our analysis of the structure of span programs increases the theoretical understanding of this important model. One implication of this is alternative algorithms for estimating the witness size when the phase gap (or spectral gap) of a certain unitary associated with the span program can be lower bounded. This is in contrast to previous span program algorithms, including those mentioned in the previous paragraph, which have all relied on *effective spectral gap* analysis. We show how the phase gap can be lower bounded by $\frac{\sigma_{\max}(A)}{\sigma_{\min}(A(x))}$, where A and $A(x)$ are linear operators associated with the span program and some input x, and $\sigma_{\min}$ and $\sigma_{\max}$ are the smallest and largest nonzero singular values.

In addition, our exposition highlights the relationship between span programs and estimating the size of the smallest solution to a linear system, which is a problem solved by Harrow et al. [12]. It is not yet clear if this relationship can lead to new algorithms, but it is an interesting direction for future work, which we discuss in Sect. 6.

Application to Effective Resistance An immediate application of our results is a quantum algorithm for estimating the effective resistance between two vertices in a graph, $R_{s,t}(G)$. This example is immediate, because in [7], a span program for st-connectivity was presented, in which the positive witness size corresponds to $R_{s,t}(G)$. The results of Belovs and Reichardt [7], combined with our new span program algorithms, immediately yield an upper bound of $\widetilde{O}(\frac{1}{\varepsilon^{3/2}}n\sqrt{R_{s,t}(G)})$ for estimating the effective resistance to relative accuracy ε. This upper bound also holds for time complexity, due to the time complexity analysis of Belovs and Reichardt [7]. Using our new spectral analysis techniques, we are also able to get an often better upper bound of $\widetilde{O}\left(\frac{1}{\varepsilon}n\sqrt{R_{s,t}(G)/\mu}\right)$, on the time complexity of estimating effective resistance, where μ is a lower bound on $\lambda_2(G)$, the second smallest eigenvalue of the Laplacian of G. These upper bounds are incomparable. The second is preferable if it is promised that $\lambda_2(G) \geq \mu$ for some μ that is larger than the desired error bound ε, and otherwise the first upper bound is better. Both algorithms use $O(\log n)$ space. We also show that a linear dependence on n is necessary, so our results cannot be significantly improved.

These are the first quantum algorithms for this problem in the adjacency query model. Previous results have studied the problem in the edge-list model [23]. At the end of Sect. 5, we compare our results to Wang [23]. Classically, this quantity can be computed exactly by inverting the Laplacian, which costs $O(m) = O(n^2)$, where m is the number of edges in the input graph.

Outline In Sect. 2, we describe the algorithmic subroutines and standard linear algebra that will form the basis of our algorithms. In Sect. 3.1, we review the use of span programs in the context of quantum query algorithms, followed in Sect. 3.2 by our new paradigm of *approximate* span programs. At this point we will be able to formally state our results about how to use span programs to construct quantum algorithms. In Sect. 3.4, we describe the structure of span programs, giving several results that will help us develop algorithms. The new algorithms from span programs are developed in Sect. 4, and finally, in Sect. 5, we present our applications to estimating effective resistance. In Sect. 6, we discuss open problems.

2 Preliminaries

To begin, we fix notation. For vector spaces V and W, we let $\mathcal{L}(V, W)$ denote the set of linear operators from V to W. For any operator $A \in \mathcal{L}(V, W)$, we denote by col$A$ the columnspace, row$A$ the rowspace, and ker A the kernel of A. $\sigma_{\min}(A)$ and $\sigma_{\max}(A)$ denote the smallest and largest non-zero singular values, respectively. A^+ denotes the Moore-Penrose pseudo-inverse.

The algorithms in this paper solve either decision problems, or estimation problems. For $f : X \subseteq [q]^n \to \{0, 1\}$, we say that an algorithm *decides f with bounded error* if for any $x \in X$, with probability at least $2/3$, the algorithm outputs $f(x)$ on input x. For $f : X \subseteq [q]^n \to \mathbb{R}_{\geq 0}$, we say that an algorithm *estimates f to relative accuracy ε* with bounded error if for any $x \in X$, with probability at least $2/3$, on input x the algorithm outputs $\tilde{f}$ such that $|f(x) - \tilde{f}| \leq \varepsilon f(x)$. In both cases, using $2/3$ is without loss of generality: any algorithm with success probability bounded above $1/2$ by a constant can be amplified to success probability arbitrarily close to 1 by taking the median of the outputs of a constant number of repetitions of the algorithm. We generally omit the description "with bounded error", as all of our algorithms have bounded error.

All algorithms presented in this paper are based on the following structure. We have some initial state $|\phi_0\rangle$, and some unitary operator U, and we want to estimate $\|\Pi_0|\phi_0\rangle\|$, where Π_0 is the orthogonal projector onto the 1-eigenspace of U. The first step in this process is a quantum algorithm that estimates, in a new register, the phase of U applied to the input state.

Theorem 1 (Phase Estimation [8,14]) *Let U be a unitary with spectral decomposition $U = \sum_{j=1}^{m} e^{i\theta_j}|\psi_j\rangle\langle\psi_j|$, for $\theta_1, \ldots, \theta_m \in (-\pi, \pi]$. For any $\Theta \in (0, \pi)$ and $\varepsilon \in (0, 1)$, there exists a quantum algorithm that makes $O\left(\frac{1}{\Theta}\log\frac{1}{\varepsilon}\right)$ controlled calls to U and, on input $|\psi_j\rangle$, outputs a state $|\psi_j\rangle|\omega\rangle$ such that if $\theta_j = 0$, then $|\omega\rangle = |0\rangle$, and if $|\theta_j| \geq \Theta$, $|\langle 0|\omega\rangle|^2 \leq \varepsilon$. If U acts on s qubits, the algorithm uses $O(s + \log\frac{1}{\Theta})$ space.*

The precision needed to isolate $\Pi_0|\phi_0\rangle$ depends on the smallest nonzero phase of U, the *phase gap*.

Definition 1 (*Phase Gap*) Let $\{e^{i\theta_j}\}_{j\in S}$ be the eigenvalues of a unitary operator U, with $\{\theta_j\}_{j\in S} \subset (-\pi, \pi]$. Then the *phase gap* of U is $\Delta(U) := \min\{|\theta_j| : \theta_j \neq 0\}$.

In order to estimate $\|\Pi_0|\phi_0\rangle\|^2$, given a state $|0\rangle\Pi_0|\phi_0\rangle + |1\rangle(I - \Pi_0)|\phi_0\rangle$, we use the following.

Theorem 2 (Amplitude Estimation [6]) *Let $\mathcal{A}$ be a quantum algorithm that outputs $\sqrt{p(x)}|0\rangle|\Psi_x(0)\rangle + \sqrt{1 - p(x)}|1\rangle|\Psi_x(1)\rangle$ on input x. Then there exists a quantum algorithm that estimates $p(x)$ to precision ε using $O\left(\frac{1}{\varepsilon}\frac{1}{\sqrt{p(x)}}\right)$ calls to $\mathcal{A}$.*

If we know that the amplitude is either $\leq p_1$ or $\geq p_0$ for some $p_1 < p_0$, then we can use amplitude estimation to distinguish between these two cases.

Corollary 1 (Amplitude Gap) *Let $\mathcal{A}$ be a quantum algorithm that outputs $\sqrt{p(x)}$ $|0\rangle|\Psi_x(0)\rangle + \sqrt{1 - p(x)}|1\rangle|\Psi_x(1)\rangle$ on input x. For any $0 \leq p_1 < p_0 \leq 1$, we*

can distinguish between the cases $p(x) \geq p_0$ *and* $p(x) \leq p_1$ *with bounded error using* $O\left(\frac{\sqrt{p_0}}{p_0-p_1}\right)$ *calls to* $\mathcal{A}$.

Proof By [6, Thm. 12], using M calls to $\mathcal{A}$, we can obtain an estimate $\tilde{p}$ of $p(x)$ such that

$$|\tilde{p} - p(x)| \leq \frac{2\pi\sqrt{p(x)(1-p(x))}}{M} + \frac{\pi^2}{M^2}$$

with probability 3/4. Let $M = 4\pi\frac{\sqrt{p_0+p_1}}{p_0-p_1}$. Then note that for any x_1 and x_0 such that $p(x_1) \leq p_1$ and $p(x_0) \geq p_0$, using $\sqrt{p_0+p_1} \geq (\sqrt{p_0}+\sqrt{p_1})/\sqrt{2}$,

$$\begin{aligned} M &\geq 2\sqrt{2}\pi\frac{\sqrt{p_0}+\sqrt{p_1}}{p_0-p_1} = 2\sqrt{2}\pi\frac{1}{\sqrt{p_0}-\sqrt{p_1}} \geq 2\sqrt{2}\pi\frac{1}{\sqrt{p(x_0)}-\sqrt{p(x_1)}} \\ &= 2\sqrt{2}\pi\frac{\sqrt{p(x_0)}+\sqrt{p(x_1)}}{p(x_0)-p(x_1)}. \end{aligned}$$

If $\tilde{p}_1$ is the estimate obtained on input x_1, then we have, with probability 3/4:

$$\begin{aligned} \tilde{p}_1 &\leq p(x_1) + \frac{2\pi\sqrt{p(x_1)(1-p(x_1))}}{M} + \frac{\pi^2}{M^2} \\ &\leq p(x_1) + \frac{\sqrt{p(x_1)}(p(x_0)-p(x_1))}{\sqrt{2}(\sqrt{p(x_0)}+\sqrt{p(x_1)})} + \frac{(p_0-p_1)^2}{16(p_0+p_1)}. \end{aligned}$$

On the other hand, if $\tilde{p}_0$ is an estimate of $p(x_0)$, then with probability 3/4:

$$\begin{aligned} \tilde{p}_0 &\geq p(x_0) - \frac{2\pi\sqrt{p(x_0)(1-p(x_0))}}{M} - \frac{\pi^2}{M^2} \\ &\geq p(x_0) - \frac{\sqrt{p(x_0)}(p(x_0)-p(x_1))}{\sqrt{2}(\sqrt{p(x_0)}+\sqrt{p(x_1)})} - \frac{(p_0-p_1)^2}{16(p_0+p_1)}. \end{aligned}$$

We complete the proof by showing that $\tilde{p}_1 < \tilde{p}_0$, so we can distinguish these two events. We have:

$$\begin{aligned} \tilde{p}_0 - \tilde{p}_1 &\geq p(x_0) - p(x_1) - \frac{(p(x_0)-p(x_1))(\sqrt{p(x_0)}+\sqrt{p(x_1)})}{\sqrt{2}(\sqrt{p(x_0)}+\sqrt{p(x_1)})} - \frac{(p_0-p_1)^2}{8(p_0+p_1)} \\ &\geq \left(1-\frac{1}{\sqrt{2}}\right)(p_0-p_1) - \frac{1}{8}(p_0-p_1) \geq \frac{1}{6}(p_0-p_1) > 0. \end{aligned}$$

Thus, using $4\pi\frac{\sqrt{p_0+p_1}}{p_0-p_1} = O\left(\frac{\sqrt{p_0}}{p_0-p_1}\right)$ calls to $\mathcal{A}$, we can distinguish between $p(x) \leq p_1$ and $p(x) \geq p_0$ with success probability 3/4. □

In order to make use of phase estimation, we will need to analyze the spectrum of a particular unitary, which, in our case, consists of a pair of reflections. The following was first used in the context of quantum algorithms in [22]:

Theorem 3 *Let $U = (2\Pi_A - I)(2\Pi_B - I)$ be a product of two reflections on a space H containing $A = \mathrm{span}\{|\psi_1\rangle, \ldots, |\psi_a\rangle\}$ and $B = \mathrm{span}\{|\phi_1\rangle, \ldots, |\phi_b\rangle\}$, with $\Pi_A = \sum_{i=1}^{a} |\psi_i\rangle\langle\psi_i|$ and $\Pi_B = \sum_{i=1}^{b} |\phi_i\rangle\langle\phi_i|$. Let $D = \Pi_A\Pi_B$ be the discriminant of U with singular value decomposition $\sum_{j=1}^{r} \cos\theta_j |\alpha_j\rangle\langle\beta_j|$, with $\theta_j \in [0, \frac{\pi}{2}]$. Then the spectrum of U is $\{e^{\pm 2i\theta_j}\}_j$. The 1-eigenspace of U is $(A \cap B) \oplus (A^\perp \cap B^\perp)$ and the (-1)-eigenspace is $(A \cap B^\perp) \oplus (A^\perp \cap B)$.*

Let $\Lambda_A = \sum_{j=1}^{a} |\psi_j\rangle\langle j|$ and $\Lambda_B = \sum_{j=1}^{b} |\phi_j\rangle\langle j|$. We note that in the original statement of Theorem 3, the discriminant is defined $D' = \Lambda_A^\dagger \Lambda_B$. However it is easy to see that D' and D have the same singular values: if $D' = \sum_i \sigma_i |v_i\rangle\langle u_i|$ is a singular value decomposition of D', then $D = \sum_i \sigma_i \Lambda_A |v_i\rangle\langle u_i| \Lambda_B^\dagger$ is a singular value decomposition of D.

The following corollary to Theorem 3 will be useful in the analysis of several algorithms.

Corollary 2 (Phase gap and discriminant) *Let D be the discriminant of a unitary $U = (2\Pi_A - I)(2\Pi_B - I)$. Then $\Delta(-U) \geq 2\sigma_{\min}(D)$.*

Proof By Theorem 3, if $\{\sigma_0 = \cos\theta_0 < \sigma_1 = \cos\theta_1 < \cdots \sigma_m = \cos\theta_m\}$ are the non-zero singular values of D, for $\theta_j \in [0, \frac{\pi}{2})$, then the phases of U in $(-\pi, \pi)$ are $\{\pm 2\theta_j\}_{j=0}^{m}$ (U might also have $(-1) = e^{i\pi}$ as an eigenvalue), and so $-U$ has non-zero phases $\{\pm 2\theta_j \mp \pi\}_{j=0}^{m} = \{\pm(2\theta_j - \pi)\}_{j=0}^{m}$. Thus

$$\begin{aligned}\Delta(-U) &= \min\{|\pi - 2\theta_j| : \theta_j \neq \pi/2\} = |\pi - 2\cos^{-1}\min\{\sigma_j : \sigma_j \neq 0\}| \\ &= |\pi - 2\cos^{-1}\sigma_{\min}(D)|.\end{aligned}$$

We have $\theta \geq \sin\theta = \cos(\pi/2 - \theta)$, so $\sigma_{\min}(D) \geq \cos(\pi/2 - \sigma_{\min}(D))$. Then since cos is decreasing on the interval $[0, \pi/2]$, we have $\cos^{-1}(\sigma_{\min}(D)) \leq \pi/2 - \sigma_{\min}(D)$, and thus

$$\Delta(-U) \geq |\pi - 2(\pi/2 - \sigma_{\min}(D))| = 2\sigma_{\min}(D).$$

□

Finally, we will make use of the following lemma, which first appeared in this form in [16]:

Lemma 1 (Effective Spectral Gap Lemma) *Let $U = (2\Pi_A - I)(2\Pi_B - I)$ be the product of two reflections, and let Π_Θ be the orthogonal projector onto $\mathrm{span}\{|u\rangle : U|u\rangle = e^{i\theta}|u\rangle, |\theta| \leq \Theta\}$. Then if $\Pi_A|u\rangle = 0$, $\|\Pi_\Theta \Pi_B |u\rangle\| \leq \frac{\Theta}{2} \||u\rangle\|$.*

3 Approximate Span Programs

3.1 Span Programs and Decision Problems

In this section, we review the concept of span programs, and their use in quantum algorithms.

Definition 2 (*Span Program*) A span program $P = (H, V, \tau, A)$ on $[q]^n$ consists of

1. finite-dimensional inner product spaces $H = H_1 \oplus \cdots \oplus H_n \oplus H_{\text{true}} \oplus H_{\text{false}}$, and $\{H_{j,a} \subseteq H_j\}_{j\in[n],a\in[q]}$ such that $H_{j,1} + \cdots + H_{j,q} = H_j$,
2. a vector space V,
3. a *target vector* $\tau \in V$, and
4. a linear operator $A \in \mathcal{L}(H, V)$.

To each $x \in [q]^n$, we associate a subspace $H(x) := H_{1,x_1} \oplus \cdots \oplus H_{n,x_n} \oplus H_{\text{true}}$.

Although our notation in Definition 2 deviates from previous span program definitions, the only difference in the substance of the definition is that the spaces $H_{j,a}$ and $H_{j,b}$ for $a \neq b$ need not be orthogonal in our definition. This has the effect of removing $\log q$ factors in the equivalence between span programs and the dual adversary bound (for details see [13, Sec. 7.1]). The spaces H_{true} and H_{false} can be useful for designing a span program, but are never required, since we can always add an $(n+1)^{\text{th}}$ variable, set $x_{n+1} = 1$, and let $H_{n+1,0} = H_{\text{false}}$ and $H_{n+1,1} = H_{\text{true}}$.

A span program on $[q]^n$ partitions $[q]^n$ into two sets: *positive* inputs, which we call P_1, and *negative* inputs, which we call P_0. The importance of this partition stems from the fact that a span program may be converted into a quantum algorithm for deciding this partition in the quantum query model [19,20]. Thus, if one can construct a span program whose partition of $[q]^n$ corresponds to a problem one wants to solve, an algorithm follows. In order to describe how a span program partitions $[q]^n$ and the query complexity of the resulting algorithm, we need the concept of positive and negative witnesses and witness size.

Definition 3 (*Positive and Negative Witness*) Fix a span program P on $[q]^n$, and a string $x \in [q]^n$. We say that $|w\rangle$ is a *positive witness for x in P* if $|w\rangle \in H(x)$, and $A|w\rangle = \tau$. We define the *positive witness size of x* as:

$$w_+(x, P) = w_+(x) = \min\{\||w\rangle\|^2 : |w\rangle \in H(x) : A|w\rangle = \tau\},$$

if there exists a positive witness for x, and $w_+(x) = \infty$ else. We say that $\omega \in \mathcal{L}(V, \mathbb{R})$ is a *negative witness for x in P* if $\omega A \Pi_{H(x)} = 0$ and $\omega\tau = 1$. We define the *negative witness size of x* as:

$$w_-(x, P) = w_-(x) = \min\{\|\omega A\|^2 : \omega \in \mathcal{L}(V, \mathbb{R}) : \omega A \Pi_{H(x)} = 0, \omega\tau = 1\},$$

if there exists a negative witness, and $w_-(x) = \infty$ otherwise. If $w_+(x)$ is finite, we say that x is *positive* (wrt. P), and if $w_-(x)$ is finite, we say that x is *negative*. We let P_1 denote the set of positive inputs, and P_0 the set of negative inputs for P. Note that for every $x \in [q]^n$, exactly one of $w_-(x)$ and $w_+(x)$ is finite; that is, (P_0, P_1) partitions $[q]^n$.

For a decision problem $f : X \subseteq [q]^n \to \{0, 1\}$, we say that P *decides* f if $f^{-1}(0) \subseteq P_0$ and $f^{-1}(1) \subseteq P_1$. In that case, we can use P to construct a quantum algorithm that decides f.

Theorem 4 [19] *Fix $f : X \subseteq [q]^n \to \{0,1\}$, and let P be a span program on $[q]^n$ that decides f. Let $W_+(f,P) = \max_{x \in f^{-1}(1)} w_+(x,P)$ and $W_-(f,P) = \max_{x \in f^{-1}(0)} w_-(x,P)$. Then there exists a quantum algorithm that decides f using $O(\sqrt{W_+(f,P)W_-(f,P)})$ queries.*

We call $\sqrt{W_+(f,P)W_-(f,P)}$ the *complexity* of P. It is known that for any decision problem, there exists a span program whose complexity is equal, up to constants, to its query complexity [19,20] ([13, Sec. 7.1] removes log factors in this statement), however, it is generally a difficult task to find such an optimal span program.

3.2 Span Programs and Approximate Decision Problems

Consider a span program P and $x \in P_0$. Suppose there exists $|w\rangle \in H(x)$ such that $A|w\rangle$ is very close to τ. We might say that x is very close to being in P_1. If all vectors in $H(y)$ for $y \in P_0 \setminus \{x\}$ are mapped far from τ, it might be more natural to consider the partition $(P_0 \setminus \{x\}, P_1 \cup \{x\})$ rather than (P_0, P_1).

As further motivation, we mention a construction of Reichardt [19, Sec. 3 of full version] that takes any quantum query algorithm with one-sided error, and converts it into a span program whose complexity matches the query complexity of the algorithm. The target of the span program is the vector $|1,\bar{0}\rangle$, which corresponds to a quantum state with a 1 in the answer register and 0s elsewhere. If an algorithm has no error on 1-inputs, it can be modified so that it always ends in exactly this state, by uncomputing all but the answer register. An algorithm with two-sided error cannot be turned into a span program using this construction, because there is error in the final state. This is in intuitive opposition to the evidence that span programs characterize bounded (two-sided) error quantum query complexity.

This motivates us to consider the *positive error* of an input, or how close it comes to being positive. Since there is no meaningful notion of distance in V, we consider closeness in H.

Definition 4 (*Positive Error*) For any span program P on $[q]^n$, and $x \in [q]^n$, we define the *positive error of x in P* as:

$$e_+(x) = e_+(x,P) := \min\left\{ \left\| \Pi_{H(x)^\perp} |w\rangle \right\|^2 : A|w\rangle = \tau \right\}.$$

Note that $e_+(x,P) = 0$ if and only if $x \in P_1$. Any $|w\rangle$ such that $\left\| \Pi_{H(x)^\perp} |w\rangle \right\|^2 = e_+(x)$ is called a *min-error positive witness for x in P*. We define

$$\tilde{w}_+(x) = \tilde{w}_+(x,P) := \min\left\{ \||w\rangle\|^2 : A|w\rangle = \tau, \left\| \Pi_{H(x)^\perp} |w\rangle \right\|^2 = e_+(x) \right\}.$$

A min-error positive witness that also minimizes $\||w\rangle\|^2$ is called an *optimal min-error positive witness for x*.

Note that if $x \in P_1$, then $e_+(x) = 0$. In that case, a min-error positive witness for x is just a positive witness, and $\tilde{w}_+(x) = w_+(x)$.

We define a similar notion for positive inputs, to measure their closeness to being negative.

Definition 5 (*Negative Error*) For any span program P on $[q]^n$ and $x \in [q]^n$, we define the *negative error of x in P* as:

$$e_-(x) = e_-(x, P) := \min\left\{ \left\| \omega A \Pi_{H(x)} \right\|^2 : \omega(\tau) = 1 \right\}.$$

Again, $e_-(x, P) = 0$ if and only if $x \in P_0$. Any ω such that $\left\| \omega A \Pi_{H(x)} \right\|^2 = e_-(x, P)$ is called a *min-error negative witness for x in P*. We define

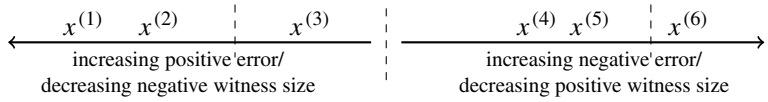
$$\tilde{w}_-(x) = \tilde{w}_-(x, P) := \min\left\{ \|\omega A\|^2 : \omega(\tau) = 1, \left\| \omega A \Pi_{H(x)} \right\|^2 = e_-(x, P) \right\}.$$

A min-error negative witness that also minimizes $\|\omega A\|^2$ is called an *optimal min-error negative witness for x*.

It turns out that the notion of span program error has a very nice characterization as exactly the reciprocal of the witness size:

$$\forall x \in P_0,\ w_-(x) = \frac{1}{e_+(x)}, \qquad \text{and} \qquad \forall x \in P_1,\ w_+(x) = \frac{1}{e_-(x)},$$

which we prove shortly in Theorems 8 and 9. This is a very nice state of affairs, for a number of reasons. It allows us two ways of thinking about approximate span programs: in terms of how small the error is, or how large the witness size is. That is, we can say that an input $x \in P_0$ is *almost positive* either because its positive error is small, or equivalently, because its negative witness size is large. In general, we can think of P as not only partitioning P into (P_0, P_1), but inducing an ordering on $[q]^n$ from most negative—smallest negative witness, or equivalently, largest positive error—to most positive—smallest positive witness, or equivalently, largest negative error. For example, on the domain $\{x^{(1)}, \ldots, x^{(6)}\} \subset [q]^n$, P might induce the following ordering:



The inputs $\{x^{(1)}, x^{(2)}, x^{(3)}\}$ are in P_0, and $w_-(x^{(1)}) < w_-(x^{(2)}) < w_-(x^{(3)})$ (although it is generally possible for two inputs to have the same witness size). The inputs $\{x^{(4)}, x^{(5)}, x^{(6)}\}$ are in P_1, and $w_+(x^{(4)}) > w_+(x^{(5)}) > w_+(x^{(6)})$. The span program exactly decides partition $(\{x^{(1)}, x^{(2)}, x^{(3)}\}, \{x^{(4)}, x^{(5)}, x^{(6)}\})$, but we say it *approximates* any partition that respects the ordering. If we obtain a partition by drawing a line somewhere on the left side, for example $(\{x^{(1)}, x^{(2)}\}, \{x^{(3)}, x^{(4)}, x^{(5)}, x^{(6)}\})$, we say P *negatively* approximates the function corresponding to that partition, whereas if we obtain a partition by drawing a line on the right side, for example $(\{x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x^{(5)}\}, \{x^{(6)}\})$, we say P *positively* approximates the function.

Definition 6 (*Functions Approximately Associated with P*) Let P be a span program on $[q]^n$, and $f : X \subseteq [q]^n \to \{0, 1\}$ a decision problem. For any $\lambda \in (0, 1)$, we say that P *positively* λ*-approximates* f if $f^{-1}(1) \subseteq P_1$, and for all $x \in f^{-1}(0)$, either $x \in P_0$, or $w_+(x, P) \geq \frac{1}{\lambda} W_+(f, P)$ (note that since $f^{-1}(1) \subseteq P_1$, $W_+(f, P) = \max_{x \in f^{-1}(1)} w_+(x, P)$ is finite). We say that P *negatively* λ*-approximates* f if $f^{-1}(0) \subseteq P_0$, and for all $x \in f^{-1}(1)$, either $x \in P_1$, or $w_-(x, P) \geq \frac{1}{\lambda} W_-(f, P)$ (note that since $f^{-1}(0) \subseteq P_0$, $W_-(f, P)$ is finite). If P decides f exactly, then both conditions hold for any value of λ, and so we can say that P 0-approximates f.

This allows us to consider a much broader class of functions associated with a particular span program. This association is useful, because as with the standard notion of association between a function f and a span program, if a function is approximated by a span program, we can convert the span program into a quantum algorithm that decides f using a number of queries related to the witness sizes. Specifically, we get the following, proven in Sect. 4.

Theorem 5 (Approximate Span Program Decision Algorithms) *Fix $f : X \subseteq [q]^n \to \{0, 1\}$, and let P be a span program that positively λ-approximates f. Define*

$$W_+ = W_+(f, P) := \max_{x \in f^{-1}(1)} w_+(x, P)$$
$$\text{and} \qquad \widetilde{W}_- = \widetilde{W}_-(f, P) := \max_{x \in f^{-1}(0)} \tilde{w}_-(x, P).$$

There is a quantum algorithm that decides f with bounded error in query complexity $O\left(\frac{\sqrt{W_+ \widetilde{W}_-}}{(1-\lambda)^{3/2}} \log \frac{1}{1-\lambda}\right)$. Similarly, let P be a span program that negatively λ-approximates f. Define

$$W_- = W_-(f, P) := \max_{x \in f^{-1}(0)} w_-(x, P)$$
$$\text{and} \qquad \widetilde{W}_+ = \widetilde{W}_+(f, P) := \max_{x \in f^{-1}(1)} \tilde{w}_+(x, P).$$

There is a quantum algorithm that decides f with bounded error in query complexity $O\left(\frac{\sqrt{W_- \widetilde{W}_+}}{(1-\lambda)^{3/2}} \log \frac{1}{1-\lambda}\right)$.

With the ability to distinguish between different witness sizes, we can obtain algorithms for estimating the witness size.

Theorem 6 (Witness Size Estimation Algorithm) *Fix $f : X \subseteq [q]^n \to \mathbb{R}_{\geq 0}$. Let P be a span program such that for all $x \in X$, $f(x) = w_+(x, P)$ and define $\widetilde{W}_- = \widetilde{W}_-(f, P) = \max_{x \in X} \tilde{w}_-(x, P)$. There exists a quantum algorithm that estimates f to accuracy ε in $\widetilde{O}\left(\frac{1}{\varepsilon^{3/2}} \sqrt{w_+(x) \widetilde{W}_-}\right)$ queries. Similarly, let P be a span program such that for all $x \in X$, $f(x) = w_-(x, P)$ and define $\widetilde{W}_+ = \widetilde{W}_+(f, P) =$*

$\max_{x \in X} \tilde{w}_+(x, P)$. *Then there exists a quantum algorithm that estimates* f *to accuracy* ε *in* $\widetilde{O}\left(\frac{1}{\varepsilon^{3/2}}\sqrt{w_-(x)\widetilde{W}_+}\right)$ *queries.*

The algorithms of Theorems 5 and 6 involve phase estimation of a particular unitary U, as with previous span program algorithms, in order to distinguish the 1-eigenspace of U from its other eigenspaces. In general, it may not be feasible to calculate the phase gap of U, so for the algorithms of Theorems 5 and 6, as with previous algorithms, we use the effective spectral gap lemma to bound the overlap of a particular initial state with eigenspaces of U corresponding to small phases. However, by relating the phase gap of U to the spectrum of A and $A(x) := A\Pi_{H(x)}$, we show how to lower bound the phase gap in some cases, which may give better results. In particular, in our application to effective resistance, it is not difficult to bound the phase gap in this way, which leads to an improved upper bound. In general we have the following theorem.

Theorem 7 (Witness Size Estimation Algorithm Using Real Phase Gap) *Fix* $f : X \subseteq [q]^n \to \mathbb{R}_{\geq 0}$ *and let* $P = (H, V, \tau, A)$ *be a normalized span program (see Definition 7) on* $[q]^n$ *such that* $\forall x \in X$, $f(x) = w_+(x, P)$ *(resp.* $f(x) = w_-(x)$*). If* $\kappa \geq \frac{\sigma_{\max}(A)}{\sigma_{\min}(A\Pi_{H(x)})}$, $\forall x \in X$, *then the quantum query complexity of estimating* $f(x)$ *to relative accuracy* ε *is at most* $\widetilde{O}\left(\frac{\sqrt{f(x)}\kappa}{\varepsilon}\right)$.

Theorem 5 is proven in Sect. 4.2, Theorem 6 is proven in Sect. 4.3, and Theorem 7 is proven in Sect. 4.4.

3.3 Example

To illustrate how these ideas might be useful, we give a brief example of how a span program that leads to an algorithm for the OR function can be combined with our results to additionally give algorithms for threshold functions and approximate counting. We define a span program P on $\{0, 1\}^n$ as follows:

$$V = \mathbb{R}, \quad \tau = 1, \quad H_i = H_{i,1} = \text{span}\{|i\rangle\}, \quad H_{i,0} = \{0\}, \quad A = \sum_{i=1}^{n} \langle i|.$$

So $H = \text{span}\{|i\rangle : i \in [n]\}$ and $H(x) = \text{span}\{|i\rangle : x_i = 1\}$. It is not difficult to see that P decides OR. In particular, we can see that the optimal positive witness for any x such that $|x| > 0$ is $|w_x\rangle = \sum_{i:x_i=1} \frac{1}{|x|}|i\rangle$. The only linear map $\omega : \mathbb{R} \to \mathbb{R}$ such that $\omega\tau = 1$ is the identity, and indeed, this is a negative witness for the all-zeros string $\bar{0}$, since $H(\bar{0}) = \{0\}$, and so $\omega A\Pi_{H(\bar{0})} = 0$.

Let $\lambda \in (0, 1)$, $t \in [n]$, and let f be a threshold function defined by $f(x) = 1$ if $|x| \geq t$ and $f(x) = 0$ if $|x| \leq \lambda t$, with the promise that one of these conditions holds. If $f(x) = 1$, then $w_+(x) = \||w_x\rangle\|^2 = \frac{1}{|x|} \leq \frac{1}{t}$, so $W_+(f, P) = \frac{1}{t}$. On the other hand, if $f(x) = 0$, then $w_+(x) = \frac{1}{|x|} \geq \frac{1}{\lambda t} = \frac{1}{\lambda} W_+(f, P)$, so P positively λ-approximates f. The only approximate negative witness is ω the identity, so we have $\widetilde{W}_- = \|\omega A\|^2 = \|A\|^2 = n$. By Theorem 5, there is a quantum algorithm for f with query complexity $\frac{1}{(1-\lambda)^{3/2}}\sqrt{W_+\widetilde{W}_-} = \frac{1}{(1-\lambda)^{3/2}}\sqrt{n/t}$.

Furthermore, since $w_+(x) = \frac{1}{|x|}$, by Theorem 6, we can estimate $\frac{1}{|x|}$ to relative accuracy ε, and therefore we can estimate $|x|$ to relative accuracy 2ε, in quantum query complexity $\frac{1}{\varepsilon^{3/2}}\sqrt{n/|x|}$.

These upper bounds do not have optimal scaling in ε, as the actual quantum query complexities of these problems are $\frac{1}{1-\lambda}\sqrt{n/t}$ and $\frac{1}{\varepsilon}\sqrt{n/|x|}$ [1,2,6], however, using Theorem 7, the optimal query complexities can be recovered.

3.4 Span Program Structure and Scaling

In this section, we present some observations about the structure of span programs that will be useful in the design and analysis of our algorithms, and for general intuition. We begin by formally stating and proving Theorems 8 and 9, relating error to witness size.

Theorem 8 *Let P be a span program on $[q]^n$ and $x \in P_0$. If $|\tilde{w}\rangle$ is a min-error positive witness for x, and ω is an optimal negative witness for x,*

$$(\omega A)^\dagger = \frac{\Pi_{H(x)^\perp}|\tilde{w}\rangle}{\left\|\Pi_{H(x)^\perp}|\tilde{w}\rangle\right\|^2}, \quad \textit{and so} \quad w_-(x) = \frac{1}{e_+(x)}.$$

Proof Let $|\tilde{w}\rangle$ be a min-error positive witness for x, and ω an optimal negative witness for x. We have $(\omega A)|\tilde{w}\rangle = \omega\tau = 1$ and, since $\omega A \Pi_{H(x)} = 0$, we have $(\omega A)\Pi_{H(x)^\perp}|\tilde{w}\rangle = 1$. Thus, write $(\omega A)^\dagger = \frac{\Pi_{H(x)^\perp}|\tilde{w}\rangle}{\left\|\Pi_{H(x)^\perp}|\tilde{w}\rangle\right\|^2} + |u\rangle$ such that $\langle u|\Pi_{H(x)^\perp}|\tilde{w}\rangle = 0$. Define $|w_{\text{err}}\rangle = \Pi_{H(x)^\perp}|\tilde{w}\rangle$. We have $A(|\tilde{w}\rangle - \Pi_{\ker A}|w_{\text{err}}\rangle) = A|\tilde{w}\rangle = \tau$, so by assumption that $|\tilde{w}\rangle$ has minimal error,

$$\begin{aligned}\left\|\Pi_{H(x)^\perp}|\tilde{w}\rangle\right\| &\le \left\|\Pi_{H(x)^\perp}(|\tilde{w}\rangle - \Pi_{\ker A}|w_{\text{err}}\rangle)\right\| \le \left\|\Pi_{H(x)^\perp}|\tilde{w}\rangle - \Pi_{\ker A}|w_{\text{err}}\rangle\right\| \\ &= \left\|\Pi_{(\ker A)^\perp}|w_{\text{err}}\rangle\right\|,\end{aligned}$$

so $\||w_{\text{err}}\rangle\| \le \left\|\Pi_{(\ker A)^\perp}|w_{\text{err}}\rangle\right\|$, and so we must have $|w_{\text{err}}\rangle \in (\ker A)^\perp$. Thus, $\ker\langle w_{\text{err}}| \subseteq \ker A$, so by the fundamental homomorphism theorem, there exists a linear function $\bar{\omega} : \text{col}A \to \mathbb{R}$ such that $\bar{\omega}A = \langle w_{\text{err}}|$. Furthermore, we have $\bar{\omega}\tau = \bar{\omega}A|\tilde{w}\rangle = \langle\tilde{w}|\Pi_{H(x)^\perp}|\tilde{w}\rangle = \left\|\Pi_{H(x)^\perp}|\tilde{w}\rangle\right\|^2 = e_+(x)$, so $\omega' = \frac{\bar{\omega}}{e_+(x)}$ has $\omega'\tau = 1$. By the optimality of ω, we must have $\|\omega A\|^2 \le \left\|\omega' A\right\|^2$, so

$$\left\|\frac{\Pi_{H(x)^\perp}|\tilde{w}\rangle}{e_+(x)} + |u\rangle\right\|^2 \le \left\|\frac{\Pi_{H(x)^\perp}|\tilde{w}\rangle}{e_+(x)}\right\|^2$$

and so $|u\rangle = 0$. Thus $(\omega A)^\dagger = \frac{\Pi_{H(x)^\perp}|\tilde{w}\rangle}{e_+(x)}$ and

$$w_-(x) = \|\omega A\|^2 = \frac{\left\|\Pi_{H(x)^\perp}|\tilde{w}\rangle\right\|^2}{e_+(x)^2} = \frac{1}{e_+(x)}.$$

□

Theorem 9 *Let P be a span program on $[q]^n$ and $x \in P_1$. If $|w\rangle$ is an optimal positive witness for x, and $\tilde{\omega}$ is a min-error negative witness for x,*

$$|w\rangle = \frac{\Pi_{H(x)}(\tilde{\omega}A)^\dagger}{\left\|\tilde{\omega}A\Pi_{H(x)}\right\|^2} \quad \textit{and so} \quad w_+(x) = \frac{1}{e_-(x)}.$$

Proof Let $\tilde{\omega}$ be a min-error negative witness for x, and define $|w'\rangle = \frac{\Pi_{H(x)}(\tilde{\omega}A)^\dagger}{\left\|\tilde{\omega}A\Pi_{H(x)}\right\|^2}$. First note that $|w'\rangle \in H(x)$. We will show that $|w'\rangle$ is a positive witness for x by showing $A|w'\rangle = \tau$. Suppose τ and $A|w'\rangle$ are linearly independent, and let $\alpha \in \mathcal{L}(V, \mathbb{R})$ be such that $\alpha(A|w'\rangle) = 0$ and $\alpha(\tau) = 1$. Then for any $\varepsilon \in [0, 1]$, we have $(\varepsilon\tilde{\omega} + (1-\varepsilon)\alpha)\tau = 1$, so by optimality of $\tilde{\omega}$,

$$\begin{aligned}
\left\|\tilde{\omega}A\Pi_{H(x)}\right\|^2 &\le \left\|(\varepsilon\tilde{\omega} + (1-\varepsilon)\alpha)A\Pi_{H(x)}\right\|^2 \\
&= \varepsilon^2 \left\|\tilde{\omega}A\Pi_{H(x)}\right\|^2 + (1-\varepsilon)^2 \left\|\alpha A\Pi_{H(x)}\right\|^2 \\
(1-\varepsilon^2)\left\|\tilde{\omega}A\Pi_{H(x)}\right\|^2 &\le (1-\varepsilon)^2 \left\|\alpha A\Pi_{H(x)}\right\|^2,
\end{aligned}$$

where the equality follows from the fact that $\alpha(A\Pi_{H(x)}(\tilde{\omega}A)^\dagger) = 0$. This implies $\left\|\tilde{\omega}A\Pi_{H(x)}\right\| \le 0$, a contradiction, since $\left\|\tilde{\omega}A\Pi_{H(x)}\right\| > 0$. Thus, we must have $A|w'\rangle = r\tau$ for some scalar r, so $\tilde{\omega}(A|w'\rangle) = r\tilde{\omega}(\tau)$. We then have $\tilde{\omega}(A|w'\rangle) = \tilde{\omega}A\frac{\Pi_{H(x)}(\tilde{\omega}A)^\dagger}{\left\|\tilde{\omega}A\Pi_{H(x)}\right\|^2} = 1$, and so we have $r = 1$, and thus $A|w'\rangle = \tau$. So $|w'\rangle$ is a positive witness for x. Let $|w\rangle \in H(x)$ be an optimal positive witness for x, so $\||w\rangle\|^2 = w_+(x)$. We have

$$\langle w'|w\rangle = \frac{\tilde{\omega}A\Pi_{H(x)}|w\rangle}{\left\|\tilde{\omega}A\Pi_{H(x)}\right\|^2} = \frac{\tilde{\omega}\tau}{\left\|\tilde{\omega}A\Pi_{H(x)}\right\|^2} = \frac{1}{\left\|\tilde{\omega}A\Pi_{H(x)}\right\|^2} = \left\||w'\rangle\right\|^2.$$

Thus $\left\||w'\rangle\right\|^2 \le \left\||w'\rangle\right\| \||w\rangle\|$ by the Cauchy–Schwarz inequality, so since $|w\rangle$ is optimal, we must have $\||w\rangle\| = \left\||w'\rangle\right\|$. Since the the smallest $|w\rangle$ such that $A\Pi_{H(x)}|w\rangle = \tau$ is uniquely defined as $(A\Pi_{H(x)})^+\tau$, we have $|w\rangle = |w'\rangle$. Thus $w_+(x) = \||w\rangle\|^2 = \left\||w'\rangle\right\|^2 = \frac{1}{\left\|\tilde{\omega}A\Pi_{H(x)}\right\|^2} = \frac{1}{e_-(x)}$. □

Positive Witnesses Fix a span program $P = (H, V, \tau, A)$ on $[q]^n$. In general, a positive witness is any $|w\rangle \in H$ such that $A|w\rangle = \tau$. Assume the set of all such vectors is non-empty, and let $|w\rangle$ be any vector in H such that $A|w\rangle = \tau$. Then the set of positive witnesses is exactly

$$W := |w\rangle + \ker A = \{|w\rangle + |h\rangle : |h\rangle \in \ker A\}.$$

It is well known, and a simple exercise to prove, that the unique shortest vector in W is $A^+\tau$, and it is the unique vector in $W \cap (\ker A)^\perp$. We can therefore talk about the unique smallest positive witness, whenever W is non-empty.

Definition 7 Fix a span program P, and suppose $W = \{|h\rangle \in H : A|h\rangle = \tau\}$ is non-empty. We define the *minimal positive witness of* P to be $|w_0\rangle \in W$ with smallest norm—that is, $|w_0\rangle = A^+\tau$. We define $N_+(P) := \||w_0\rangle\|^2$.

Since $|w_0\rangle \in (\ker A)^\perp$, we can write any positive witness $|w\rangle$ as $|w_0\rangle + |w_0^\perp\rangle$ for some $|w_0^\perp\rangle \in \ker A$. If we let $T = A^{-1}(\text{span}\{\tau\})$, we can write $T = \text{span}\{|w_0\rangle\} \oplus \ker A$.

Negative Witnesses Just as we can talk about a minimal positive witness, we can also talk about a minimal negative witness of P: any $\omega_0 \in \mathcal{L}(V, \mathbb{R})$ such that $\omega_0\tau = 1$, that minimizes $\|\omega_0 A\|$. Define $N_-(P) = \min_{\omega_0:\omega_0(\tau)=1} \|\omega_0 A\|^2$. Note that unlike $|w_0\rangle$, ω_0 might not be unique. There may be distinct $\omega_0, \omega_0' \in \mathcal{L}(V, \mathbb{R})$ that map τ to 1 and have minimal complexity, however, one can easily show that in that case, $\omega_0 A = \omega_0' A$, and that the unique minimal negative witness in $\text{col}A$ is $\frac{\langle\tau|}{\|\tau\|^2}$.

For any minimal negative witness, ω_0, $\omega_0 A$ is related to the minimal positive witness $|w_0\rangle$ by $(\omega_0 A)^\dagger = \frac{|w_0\rangle}{N_+(P)}$, and $N_+(P) = \frac{1}{N_-(P)}$. This can be proven, for example, by replacing $H(x)$ with $\{0\}$ in the proof of Theorem 8.

Span Program Scaling and Normalization By scaling τ to get a new target $\tau' = B\tau$, we can scale a span program by an arbitrary positive real number B, so that all positive witnesses are scaled by B, and all negative witnesses are scaled by $\frac{1}{B}$. Note that this leaves W_+W_- unchanged, so we can in some sense consider the span program invariant under this scaling.

Definition 8 A span program P is *normalized* if $N_+(P) = N_-(P) = 1$.

Any span program can be converted to a normalized span program by replacing the target with $\tau' = \frac{\tau}{N_+}$. However, it will turn out to be desirable to normalize a span program, and also scale it, independently. We can accomplish this to some degree, as shown by the following theorem.

Theorem 10 (Span program scaling) *Let $P = (H, V, \tau, A)$ be any span program on $[q]^n$, and let $N = \||w_0\rangle\|^2$ for $|w_0\rangle$ the minimal positive witness of P. For $\beta \in \mathbb{R}_{>0}$, define $P^\beta = (H^\beta, V^\beta, \tau^\beta, A^\beta)$ as follows, for $|\hat{0}\rangle$ and $|\hat{1}\rangle$ two vectors orthogonal to H and V:*

$$\forall j \in [n], a \in [q],\ H^\beta_{j,a} := H_{j,a},$$
$$H^\beta_{\text{true}} = H_{\text{true}} \oplus \text{span}\{|\hat{1}\rangle\}, \quad H^\beta_{\text{false}} = H_{\text{false}} \oplus \text{span}\{|\hat{0}\rangle\}$$
$$V^\beta = V \oplus \text{span}\{|\hat{1}\rangle\}, \quad A^\beta = \beta A + |\tau\rangle\langle\hat{0}| + \frac{\sqrt{\beta^2 + N}}{\beta}|\hat{1}\rangle\langle\hat{1}|, \quad \tau^\beta = |\tau\rangle + |\hat{1}\rangle$$

Then we have the following:

- $\forall x \in P_1,\ w_+(x, P^\beta) = \frac{w_+(x,P)}{\beta^2} + \frac{\beta^2}{N+\beta^2}$ *and* $\tilde{w}_-(x, P^\beta) \leq \beta^2 \tilde{w}_-(x, P) + 2$;

– $\forall x \in P_0$, $w_-(x, P^\beta) = \beta^2 w_-(x, P) + 1$ *and* $\tilde{w}_+(x, P^\beta) \leq \frac{1}{\beta^2}\tilde{w}_+(x, P) + 2$;
– *the minimal witness of* P^β *is* $|w_0^\beta\rangle = \frac{\beta}{\beta^2+N}|w_0\rangle + \frac{N}{\beta^2+N}|\hat{0}\rangle + \frac{\beta}{\sqrt{\beta^2+N}}|\hat{1}\rangle$, *and* $\left\||w_0^\beta\rangle\right\|^2 = 1$.

Proof of Theorem 10 appears in "Appendix A".

4 Span Program Algorithms

In this section we describe several ways in which a span program can be turned into a quantum algorithm. As in the case of algorithms previously constructed from span programs, our algorithms will consist of many applications of a unitary on H, applied to some initial state. Unlike previous applications, we will use $|w_0\rangle$, the minimal positive witness of P, as the initial state, assuming P is normalized so that $\||w_0\rangle\| = 1$. This state is independent of the input, and so can be generated with 0 queries. For *negative span program algorithms*, where we want to decide a function negatively approximated by P, we will use a unitary $U(P, x)$, defined as follows:

$$U(P, x) := (2\Pi_{\ker A} - I)(2\Pi_{H(x)} - I) = (2\Pi_{(\ker A)^\perp} - I)(2\Pi_{H(x)^\perp} - I).$$

This is similar to the unitary used in previous span program algorithms. Note that $(2\Pi_{\ker A} - I)$ is input-independent, and so can be implemented in 0 queries. However, in order to analyze the time complexity of a span program algorithm, this reflection must be implemented (as we are able to do for our applications, following [7]). The reflection $(2\Pi_{H(x)} - I)$ depends on the input, but it is not difficult to see that it requires two queries to implement. Since our definition of span programs varies slightly from previous definitions, we provide a proof of this fact.

Lemma 2 *The reflection* $2\Pi_{H(x)} - I$ *can be implemented using 2 queries to* x.

Proof For every $i \in [n]$ and $a \in [q]$, let $R_{i,a} = (I - 2\Pi_{H_{i,a}^\perp \cap H_i})$, the operator that reflects every vector in H_i that is orthogonal to $H_{i,a}$. This operation is input independent, and so, can be implemented in 0 queries. For every $i \in [n]$, let $\{|\psi_{i,1}\rangle, \ldots, |\psi_{i,m_i}\rangle\}$ be an orthonormal basis for H_i. Recall that the spaces H_i are orthogonal, so we can map $|\psi_{i,j}\rangle \mapsto |i\rangle|\psi_{i,j}\rangle$. Then using one query, we can map $|i\rangle|\psi_{i,j}\rangle \mapsto |i\rangle|x_i\rangle|\psi_{i,j}\rangle$. We then perform R_{i,x_i} on the last register, conditioned on the first two registers, and then uncompute the first two registers, using one additional query. □

For *positive span program algorithms*, where we want to decide a function positively approximated by P, or estimate the positive witness size, we will use a slightly different unitary:

$$U'(P, x) = (2\Pi_{H(x)} - I)(2\Pi_T - I),$$

where $T = \ker A \oplus \text{span}\{|w_0\rangle\}$, the span of positive witnesses. We have $U' = U^\dagger(I - 2|w_0\rangle\langle w_0|)$.

We begin by analyzing the overlap of the initial state, $|w_0\rangle$, with the phase spaces of the unitaries U and U' in Sect. 4.1. In particular, we show that the projections of $|w_0\rangle$ onto the 0-phase spaces of U and U' are exactly related to the witness size. Using the effective spectral gap lemma (Lemma 1), we show that the overlap of $|w_0\rangle$ with small nonzero phase spaces is not too large. Using this analysis, in Sect. 4.2, we describe how to convert a span program into an algorithm for any decision problem that is approximated by the span program, proving Theorem 5, and in Sect. 4.3, we describe how to convert a span program into an algorithm that estimates the span program witness size, proving Theorem 6.

Finally, in Sect. 4.4, we give a lower bound on the phase gap of U in terms of the spectra of A and $A(x) = A\Pi_{H(x)}$, giving an alternative analysis to the effective spectral gap analysis of Sect. 4.1 that may be better in some cases, and proving Theorem 7.

4.1 Analysis

Negative Span Programs In this section we analyze the overlap of $|w_0\rangle$ with the eigenspaces of $U(P, x)$. For any angle $\Theta \in [0, \pi)$, we define Π_Θ^x as the orthogonal projector onto the $e^{i\theta}$-eigenspaces of $U(P, x)$ for which $|\theta| \leq \Theta$.

Lemma 3 *Let P be a normalized span program on $[q]^n$. For any $x \in [q]^n$,*

$$\left\|\Pi_\Theta^x |w_0\rangle\right\|^2 \leq \frac{\Theta^2}{4}\tilde{w}_+(x) + \frac{1}{w_-(x)}.$$

In particular, for any $x \in P_1$, $\left\|\Pi_\Theta^x |w_0\rangle\right\|^2 \leq \frac{\Theta^2}{4} w_+(x)$.

Proof Suppose $x \in P_1$, and let $|w_x\rangle$ be an optimal positive witness for x, so $\Pi_{(\ker A)^\perp}|w_x\rangle = |w_0\rangle$. Then since $\Pi_{H(x)^\perp}|w_x\rangle = 0$, by the effective spectral gap lemma (Lemma 1):

$$\left\|\Pi_\Theta^x |w_0\rangle\right\|^2 = \left\|\Pi_\Theta^x \Pi_{(\ker A)^\perp}|w_x\rangle\right\|^2 \leq \frac{\Theta^2}{4}\||w_x\rangle\|^2 = \frac{\Theta^2}{4} w_+(x).$$

Suppose $x \in P_0$ and let ω_x be an optimal zero-error negative witness for x and $|\tilde{w}_x\rangle$ an optimal min-error positive witness for x. First note that $\Pi_{(\ker A)^\perp}|\tilde{w}_x\rangle = |w_0\rangle$, so $\Pi_{(\ker A)^\perp}\Pi_{H(x)}|\tilde{w}_x\rangle + \Pi_{(\ker A)^\perp}\Pi_{H(x)^\perp}|\tilde{w}_x\rangle = |w_0\rangle$. Since $\Pi_{H(x)^\perp}\left(\Pi_{H(x)}|\tilde{w}_x\rangle\right) = 0$, we have, by Lemma 1,

$$\begin{aligned}
\left\|\Pi_\Theta \Pi_{(\ker A)^\perp}\Pi_{H(x)}|\tilde{w}_x\rangle\right\|^2 &\leq \frac{\Theta^2}{4}\left\|\Pi_{H(x)}|\tilde{w}_x\rangle\right\|^2 \\
\left\|\Pi_\Theta\left(|w_0\rangle - \Pi_{(\ker A)^\perp}\Pi_{H(x)^\perp}|\tilde{w}_x\rangle\right)\right\|^2 &\leq \frac{\Theta^2}{4}\||\tilde{w}_x\rangle\|^2 \\
\left\|\Pi_\Theta\left(|w_0\rangle - \Pi_{(\ker A)^\perp}\frac{(\omega_x A)^\dagger}{w_-(x)}\right)\right\|^2 &\leq \frac{\Theta^2}{4}\||\tilde{w}_x\rangle\|^2.
\end{aligned}$$

In the last step, we used the fact that $\frac{(\omega_x A)^\dagger}{w_-(x)} = \Pi_{H(x)^\perp}|\tilde{w}_x\rangle$, by Theorem 8. Next note that $\Pi_{(\ker A)^\perp}(\omega_x A)^\dagger = (\omega_x A)^\dagger$ and $\Pi_{H(x)^\perp}(\omega_x A)^\dagger = (\omega_x A)^\dagger$, so $U(\omega_x A)^\dagger = (\omega_x A)^\dagger$, and therefore, $\Pi_\Theta(\omega_x A)^\dagger = (\omega_x A)^\dagger$. Thus:

$$\begin{aligned}
\left\| \Pi_\Theta|w_0\rangle - \frac{(\omega_x A)^\dagger}{w_-(x)} \right\|^2 &\leq \frac{\Theta^2}{4} \||\tilde{w}_x\rangle\|^2 \\
\|\Pi_\Theta|w_0\rangle\|^2 + \frac{1}{w_-(x)} - 2\frac{1}{w_-(x)}\langle w_0|\Pi_\Theta(\omega_x A)^\dagger &\leq \frac{\Theta^2}{4}\tilde{w}_+(x) \\
\|\Pi_\Theta|w_0\rangle\|^2 + \frac{1}{w_-(x)} - 2\frac{1}{w_-(x)}(\omega_x A|w_0\rangle)^\dagger &\leq \frac{\Theta^2}{4}\tilde{w}_+(x) \\
\|\Pi_\Theta|w_0\rangle\|^2 + \frac{1}{w_-(x)} - 2\frac{1}{w_-(x)}(\omega_x \tau)^\dagger &\leq \frac{\Theta^2}{4}\tilde{w}_+(x) \\
\|\Pi_\Theta|w_0\rangle\|^2 &\leq \frac{\Theta^2}{4}\tilde{w}_+(x) + \frac{1}{w_-(x)},
\end{aligned}$$

where in the last line we used the fact that $\omega_x \tau = 1$. □

Lemma 4 *Let P be a normalized span program on $[q]^n$. For any $x \in [q]^n$,*

$$\left\| \Pi_0^x|w_0\rangle \right\|^2 = \frac{1}{w_-(x)}.$$

In particular, for any $x \in P_1$, $\left\| \Pi_0^x|w_0\rangle \right\| = 0$.

Proof By Lemma 3, we have $\left\| \Pi_0^x|w_0\rangle \right\|^2 \leq \frac{1}{w_-(x)}$. To see the other direction, let ω_x be an optimal zero-error negative witness for x (if none exists, then $w_-(x) = \infty$ and the statement is vacuously true). Define $|u\rangle = (\omega_x A)^\dagger$. By the proof of Lemma 3, $U|u\rangle = |u\rangle$. We have $\langle u|w_0\rangle = \omega_x A|w_0\rangle = \omega_x \tau = 1$ and $\||u\rangle\|^2 = \|\omega_x A\|^2 = w_-(x)$, so we have: $\left\| \Pi_0^x|w_0\rangle \right\|^2 \geq \left\| \frac{|u\rangle\langle u|}{\||u\rangle\|^2}|w_0\rangle \right\|^2 = \frac{1}{w_-(x)}$.

Positive Span Programs We now prove results analogous to Lemmas 3 and 4 for the unitary $U'(P, x)$. For any angle $\Theta \in [0, \pi)$, we define $\overline{\Pi}_\Theta^x$ as the projector onto the θ-phase spaces of $U'(P, x)$ for which $|\theta| \leq \Theta$.

Lemma 5 *Let P be a normalized span program on $[q]^n$. For any $x \in [q]^n$,*

$$\left\| \overline{\Pi}_\Theta^x|w_0\rangle \right\|^2 \leq \frac{\Theta^2}{4}\tilde{w}_-(x) + \frac{1}{w_+(x)}.$$

In particular, if $x \in P_0$, then $\left\| \overline{\Pi}_\Theta^x|w_0\rangle \right\|^2 \leq \frac{\Theta^2}{4} w_-(x)$.

Proof If $x \in P_0$, then let ω_x be an optimal exact negative witness for x, so $\omega_x A \Pi_{H(x)} = 0$, and thus, by the effective spectral gap lemma (Lemma 1),

$$\left\| \overline{\Pi}^x_\Theta \Pi_T (\omega_x A)^\dagger \right\|^2 \leq \frac{\Theta^2}{4} \|\omega_x A\|^2 = \frac{\Theta^2}{4} w_-(x).$$

We have $\omega_x A \Pi_T = \omega_x A(\Pi_{\ker A} + |w_0\rangle\langle w_0|) = \omega_x A |w_0\rangle\langle w_0| = \omega_x \tau \langle w_0| = \langle w_0|$, so $\left\| \overline{\Pi}^x_\Theta |w_0\rangle \right\|^2 \leq \frac{\Theta^2}{4} w_-(x)$.

Suppose $x \in P_1$, and let $|w_x\rangle$ be an optimal zero-error positive witness for x, and $\tilde{\omega}_x$ an optimal min-error negative witness for x. By Theorem 9, we have $\frac{|w_x\rangle}{w_+(x)} = \Pi_{H(x)} (\tilde{\omega}_x A)^\dagger$. Since $\Pi_{H(x)} (\tilde{\omega}_x A \Pi_{H(x)^\perp})^\dagger = 0$, we have, by Lemma 1,

$$\begin{aligned}
\left\| \overline{\Pi}^x_\Theta \Pi_T (\tilde{\omega}_x A \Pi_{H(x)^\perp})^\dagger \right\|^2 &\leq \frac{\Theta^2}{4} \left\| \tilde{\omega}_x A \Pi_{H(x)^\perp} \right\|^2 \\
\left\| \overline{\Pi}^x_\Theta \Pi_T \left((\tilde{\omega}_x A)^\dagger - \frac{|w_x\rangle}{w_+(x)} \right) \right\|^2 &\leq \frac{\Theta^2}{4} \|\tilde{\omega}_x A\|^2 \\
\left\| \overline{\Pi}^x_\Theta \Pi_T (\tilde{\omega}_x A)^\dagger - \frac{|w_x\rangle}{w_+(x)} \right\|^2 &\leq \frac{\Theta^2}{4} \tilde{w}_-(x).
\end{aligned}$$

In the last line we used the fact that $\Pi_T |w_x\rangle = \Pi_{H(x)} |w_x\rangle = |w_x\rangle$, so $U' |w_x\rangle = |w_x\rangle$, and thus $\overline{\Pi}^x_\Theta |w_x\rangle = |w_x\rangle$.

Note that $\tilde{\omega}_x A \Pi_T = \tilde{\omega}_x A(\Pi_{\ker A} + |w_0\rangle\langle w_0|) = \tilde{\omega}_x A |w_0\rangle\langle w_0| = \tilde{\omega}_x \tau \langle w_0| = \langle w_0|$. Thus, we can continue from above as:

$$\begin{aligned}
\left\| \overline{\Pi}^x_\Theta |w_0\rangle - \frac{|w_x\rangle}{w_+(x)} \right\|^2 &\leq \frac{\Theta^2}{4} \tilde{w}_-(x) \\
\left\| \overline{\Pi}^x_\Theta |w_0\rangle \right\|^2 + \left\| \frac{|w_x\rangle}{w_+(x)} \right\|^2 - \frac{2}{w_+(x)} \langle w_0 | \overline{\Pi}^x_\Theta | w_x \rangle &\leq \frac{\Theta^2}{4} \tilde{w}_-(x) \\
\left\| \overline{\Pi}^x_\Theta |w_0\rangle \right\|^2 + \frac{1}{w_+(x)} - \frac{2}{w_+(x)} \langle w_0 | w_x \rangle &\leq \frac{\Theta^2}{4} \tilde{w}_-(x) \\
\left\| \overline{\Pi}^x_\Theta |w_0\rangle \right\|^2 &\leq \frac{\Theta^2}{4} \tilde{w}_-(x) + \frac{1}{w_+(x)},
\end{aligned}$$

where in the last line we used the fact that $\langle w_0 | w_x \rangle = 1$. □

Lemma 6 *Let P be a normalized span program on $[q]^n$. For any $x \in [q]^n$,*

$$\left\| \overline{\Pi}^x_0 |w_0\rangle \right\|^2 = \frac{1}{w_+(x)}.$$

In particular, if $x \in P_0$, then $\left\| \overline{\Pi}^x_0 |w_0\rangle \right\| = 0$.

Proof By Lemma 5, $\left\|\overline{\Pi}_0^x|w_0\rangle\right\|^2 \leq \frac{1}{w_+(x)}$. Let $|w_x\rangle = |w_0\rangle + |w_0^\perp\rangle$ be an optimal zero-error positive witness for x. Since $|w_x\rangle \in H(x) \cap T$, $U'|w_x\rangle = |w_x\rangle$, so $\left\|\overline{\Pi}_0^x|w_0\rangle\right\|^2 \geq \frac{\langle w_x|w_0\rangle}{\||w_x\rangle\|^2} \geq \frac{1}{w_+(x)}$.

4.2 Algorithms for Approximate Span Programs

Using the spectral analysis from Sect. 4.1, we can design an algorithm that decides a function that is approximated by a span program. We will give details for the negative case, using Lemmas 3 and 4. A nearly identical argument proves the analogous statement for the positive case, using Lemmas 5 and 6 instead.

Throughout this section, fix a decision problem f on $[q]^n$, and let P be a normalized span program that negatively λ-approximates f. By Lemmas 4 and 3, it is possible to distinguish between the cases $f(x) = 0$, in which $\frac{1}{w_-(x)} \geq \frac{1}{W_-}$, and $f(x) = 1$, in which $\frac{1}{w_-(x)} \leq \frac{\lambda}{W_-}$ using phase estimation to sufficient precision, and amplitude estimation on a 0 in the phase register. We give details in the following theorem.

Lemma 7 *Let P be a normalized λ-negative approximate span program for f. Then the quantum query complexity of f is $O\left(\frac{1}{(1-\lambda)^{3/2}} W_- \sqrt{\widetilde{W}_+} \log \frac{W_-}{1-\lambda}\right)$.*

Proof Let $U(P,x) = \sum_{j=1}^m e^{i\theta_j}|\psi_j\rangle\langle\psi_j|$, and let $|w_0\rangle = \sum_{j=1}^m \alpha_j|\psi_j\rangle$. Then applying phase estimation (Theorem 1) to precision $\Theta = \sqrt{\frac{4(1-\lambda)}{3W_-\widetilde{W}_+}}$ and error $\varepsilon = \frac{1}{6}\frac{1-\lambda}{W_-}$ produces a state $|w_0'\rangle = \sum_{j=1}^m \alpha_j|\psi_j\rangle|\omega_j\rangle$ such that if $\theta_j = 0$, then $|\omega_j\rangle = |0\rangle$, and if $|\theta_j| > \Theta$ then $|\langle\omega_j|0\rangle|^2 \leq \varepsilon$. Let Λ_0 be the projector onto states with 0 in the phase register. We have: $\left\|\Lambda_0|w_0'\rangle\right\|^2 = \sum_{j=1}^m |\alpha_j|^2|\langle 0|\omega_j\rangle|^2$. By Lemma 4, we have $\left\|\Pi_0^x|w_0\rangle\right\|^2 = \frac{1}{w_-(x)}$, so if $x \in f^{-1}(0)$, we have:

$$\left\|\Lambda_0|w_0'\rangle\right\|^2 \geq \sum_{j:\theta_j=0} |\alpha_j|^2|\langle 0|0\rangle|^2 = \left\|\Pi_0^x|w_0\rangle\right\|^2 = \frac{1}{w_-(x)} \geq \frac{1}{W_-} =: p_0.$$

On the other hand, suppose $x \in f^{-1}(1)$. Since P negatively λ-approximates f and $x \in f^{-1}(1)$, $w_-(x,P) \geq \frac{1}{\lambda}W_-(x,P)$. By Lemma 3, we have

$$\left\|\Pi_\Theta^x|w_0\rangle\right\|^2 \leq \frac{1}{w_-(x,P)} + \frac{\Theta^2}{4}\tilde{w}_+(x,P) \leq \frac{\lambda}{W_-} + \frac{1-\lambda}{3W_-\widetilde{W}_+}\widetilde{W}_+ = \frac{1}{3}\frac{1+2\lambda}{W_-}$$

and thus

$$\begin{aligned}\left\|\Lambda_0|w_0'\rangle\right\|^2 &\leq \sum_{j:|\theta_j|\leq\Theta} |\alpha_j|^2 + \sum_{j:|\theta_j|>\Theta} |\alpha_j|^2|\langle\omega_j|0\rangle|^2 \\ &= \left\|\Pi_\Theta^x|w_0\rangle\right\|^2 + \varepsilon \sum_{j:|\theta_j|>\Theta} |\alpha_j|^2 \leq \frac{1+2\lambda}{3W_-} + \frac{1-\lambda}{6W_-} =: p_1.\end{aligned}$$

By Corollary 1, we can distinguish between these cases using $O\left(\frac{\sqrt{p_0}}{p_0-p_1}\right)$ calls to phase estimation, which costs $\frac{1}{\Theta}\log\frac{1}{\varepsilon}$ calls to U. In this case, we have

$$p_0 - p_1 = \frac{1 - \frac{1}{3} - \frac{2}{3}\lambda - \frac{1}{6} + \frac{1}{6}\lambda}{W_-} = \frac{1}{2}\frac{1-\lambda}{W_-}.$$

The total number of calls to U is:

$$\frac{\sqrt{p_0}}{p_0 - p_1}\frac{1}{\Theta}\log\frac{1}{\varepsilon} = \frac{W_-}{\sqrt{W_-}(1-\lambda)}\sqrt{\frac{W_-\widetilde{W}_+}{1-\lambda}}\log\frac{W_-}{1-\lambda} = \frac{W_-\sqrt{\widetilde{W}_+}}{(1-\lambda)^{3/2}}\log\frac{W_-}{1-\lambda}.$$

□

In addition to wanting to extend this to non-normalized span programs, we note that this expression is not symmetric in the positive and negative error. Using Theorem 10, we can normalize any span program, while also scaling the positive and negative witnesses. This gives us the following.

Corollary 3 *Let P be any span program that negatively λ-approximates f. Then the quantum query complexity of f is at most*

$$O\left(\frac{1}{(1-\lambda)^{3/2}}\sqrt{W_-(f,P)\widetilde{W}_+(f,P)}\log\frac{1}{1-\lambda}\right).$$

Proof We will use the scaled span program described in Theorem 10. Let $\beta = \frac{1}{\sqrt{W_-(f,P)}}$. Then P^β is a normalized span program with

$$W_-(f,P^\beta) = \max_{x\in f^{-1}(0)} w_-(x,P^\beta) = \beta^2 \max_{x\in f^{-1}(0)} w_-(x,P) + 1 = \frac{1}{W_-}W_- + 1 = 2,$$

and

$$\begin{aligned}\widetilde{W}_+(f,P^\beta) &= \max_{x\in f^{-1}(1)} \tilde{w}_+(x,P^\beta) \le \frac{1}{\beta^2}\max_{x\in f^{-1}(1)} \tilde{w}_+(x,P) + 2\\ &= W_-(f,P)\widetilde{W}_+(f,P) + 2.\end{aligned}$$

If we define $\lambda^{(\beta)} := \frac{\max_{x\in f^{-1}(0)} w_-(x,P^\beta)}{\min_{x\in f^{-1}(1)} w_-(x,P^\beta)} = \frac{\beta^2 W_-(f,P)+1}{\beta^2\frac{1}{\lambda}W_-(f,P)+1} = \frac{2}{\frac{1}{\lambda}+1}$, then clearly P^β negatively $\lambda^{(\beta)}$-approximates f, so we can apply Lemma 7. We have $\frac{1}{1-\lambda^{(\beta)}} = \frac{1}{1-\frac{2\lambda}{1+\lambda}} = \frac{1+\lambda}{1-\lambda}$ so we can decide f in query complexity (neglecting constants):

$$\left(\frac{1+\lambda}{1-\lambda}\right)^{\frac{3}{2}}\sqrt{2\left(W_-(f,P)\widetilde{W}_+(f,P)+2\right)}\log 2\frac{1+\lambda}{1-\lambda}$$

$$= \frac{1}{(1-\lambda)^{\frac{3}{2}}} \sqrt{W_-(f,P)\widetilde{W}_+(f,P)} \log \frac{1}{1-\lambda}.$$

□

By computations analogous to Lemma 7 and Corollary 3 (using $\beta = \sqrt{W_+}$), we can show that if P positively λ-approximates f, then f has quantum query complexity $O\left(\frac{1}{(1-\lambda)^{3/2}}\sqrt{W_+\widetilde{W}_-}\log\frac{1}{1-\lambda}\right)$. This and Corollary 3 imply Theorem 5.

4.3 Estimating the Witness Size

Using the algorithms for deciding approximate span programs (Theorem 5) as a black box, we can construct a quantum algorithm that estimates the positive or negative witness size of an input using standard algorithmic techniques. We give the full proof for the case of positive witness size, as negative witness size is virtually identical. This proves Theorem 6.

Theorem 11 (Estimating the Witness Size) *Fix $f : X \subseteq [q]^n \to \mathbb{R}_{>0}$. Let P be a span program on $[q]^n$ such that for all $x \in X$, $f(x) = w_+(x, P)$. The quantum query complexity of estimating f to accuracy ε is $\widetilde{O}\left(\frac{\sqrt{w_+(x)\widetilde{W}_-(P)}}{\varepsilon^{3/2}}\right)$.*

Proof We will estimate $e(x) = \frac{1}{w_+(x)}$. The basic idea is to use the algorithm from Theorem 5 to narrow down the interval in which the value of $e(x)$ may lie. Assuming that the span program is normalized (which is without loss of generality, since normalizing by scaling τ does not impact relative accuracy) we can begin with the interval $[0, 1]$. We stop when we reach an interval $[e_{\min}, e_{\max}]$ such that the midpoint $\tilde{e} = \frac{e_{\max}+e_{\min}}{2}$ satisfies $(1-\varepsilon)e_{\max} \le \tilde{e} \le (1+\varepsilon)e_{\min}$.

Let `Decide`(P, w, λ) be the quantum algorithm from Theorem 5 that decides the (partial) function $g : P_1 \to \{0, 1\}$ defined by $g(x) = 1$ if $w_+(x) \le w$ and $g(x) = 0$ if $w_+(x) \ge \frac{w}{\lambda}$. We will amplify the success probability so that with high probability, `Decide` returns $g(x)$ correctly every time it is called by the algorithm, and we will assume that this is the case. The full witness estimation algorithm consists of repeated calls to `Decide` as follows:

`WitnessEstimate`(P, ε):

1. $e_{\max}^{(1)} = 1, e_{\min}^{(1)} = 0, e_1^{(1)} = \frac{2}{3}, e_0^{(1)} = \frac{1}{3}$
2. For $i = 1, 2, \ldots$ repeat:
   (a) Run `Decide`(P, w, λ) with $w = 1/e_1^{(i)}$ and $\lambda = e_0^{(i)}/e_1^{(i)}$.
   (b) If `Decide` outputs 1, indicating $w_+(x) \le w$, set $e_{\max}^{(i+1)} = e_{\max}^{(i)}$ and $e_{\min}^{(i+1)} = e_0^{(i)}$.
   (c) Else, set $e_{\min}^{(i+1)} = e_{\min}^{(i)}$ and $e_{\max}^{(i+1)} = e_1^{(i)}$.
   (d) If $e_{\max}^{(i+1)} \le (1+\varepsilon)e_{\min}^{(i+1)}$, return $\tilde{e} = \frac{e_{\max}^{(i+1)}+e_{\min}^{(i+1)}}{2}$.
   (e) Else, set $e_1^{(i+1)} = \frac{2}{3}e_{\max}^{(i+1)} + \frac{1}{3}e_{\min}^{(i+1)}$ and $e_0^{(i+1)} = \frac{1}{3}e_{\max}^{(i+1)} + \frac{2}{3}e_{\min}^{(i+1)}$.

We can see by induction that for every i, $e^{(i)}_{\min} \le \frac{1}{w_+(x)} \le e^{(i)}_{\max}$. This is certainly true for $i = 1$, since $w_+(x) \ge \||w_0\rangle\|^2 = 1$. Suppose it is true at step i. At step i we run $\texttt{Decide}(P, w_i, \lambda_i)$ with $w_i = 1/e^{(i)}_1$ and $\frac{w_i}{\lambda_i} = 1/e^{(i)}_0$. If $\frac{1}{w_+(x)} \ge e^{(1)}_1$, then $\texttt{Decide}$ returns 1, so we have $\frac{1}{w_+(x)} \in [e^{(i)}_0, e^{(i)}_{\max}] = [e^{(i+1)}_{\min}, e^{(i+1)}_{\max}]$. If $\frac{1}{w_+(x)} \le e^{(i)}_0$, then $\texttt{Decide}$ returns 0, so we have $\frac{1}{w_+(x)} \in [e^{(i)}_{\min}, e^{(i)}_1] = [e^{(i+1)}_{\min}, e^{(i+1)}_{\max}]$. Otherwise, $\frac{1}{w_+(x)} \in [e^{(i)}_0, e^{(i)}_1]$, which is a subset of both $[e^{(i)}_0, e^{(i)}_{\max}]$ and $[e^{(i)}_{\min}, e^{(i)}_1]$, so in any case, $\frac{1}{w_+(x)} \in [e^{(i+1)}_{\min}, e^{(i+1)}_{\max}]$.

To see that the algorithm terminates, let $\Delta_i = e^{(i)}_{\max} - e^{(i)}_{\min}$ denote the length of the remaining interval at round i. We either have $\Delta_{i+1} = e^{(i)}_{\max} - e^{(i)}_0 = e^{(i)}_{\max} - \frac{1}{3}e^{(i)}_{\max} - \frac{2}{3}e^{(i)}_{\min} = \frac{2}{3}\Delta_i$, or $\Delta_{i+1} = e^{(i)}_1 - e^{(i)}_{\min} = \frac{2}{3}e^{(i)}_{\max} + \frac{1}{3}e^{(i)}_{\min} - e^{(i)}_{\min} = \frac{2}{3}\Delta_i$, so $\Delta_i = (2/3)^{i-1}$. We terminate at the smallest T such that $(2/3)^{T-1} = \Delta_T = e^{(T)}_{\max} - e^{(T)}_{\min} \le (1+\varepsilon-1)e^{(T)}_{\min} \le \frac{\varepsilon}{w_+(x)}$. Thus we terminate before $T = \lceil \log_{3/2} \frac{w_+(x)}{\varepsilon} + 1 \rceil$.

Next, we show that, assuming $\texttt{Decide}$ does not err, the estimate is correct to within ε. Let $\tilde{e} = \frac{1}{2}(e^{(T)}_{\max} + e^{(T)}_{\min})$ be the returned estimate. Recall that we only terminate when $e^{(T)}_{\max} \le (1+\varepsilon)e^{(T)}_{\min}$. We have

$$\frac{1}{\tilde{e}} = \frac{2}{e^{(T)}_{\max} + e^{(T)}_{\min}} \le \frac{2}{e^{(T)}_{\max}\left(1 + \frac{1}{1+\varepsilon}\right)} \le \frac{2}{\frac{1}{w_+(x)}\left(\frac{2+\varepsilon}{1+\varepsilon}\right)} \le (1+\varepsilon)\, w_+(x),$$

and

$$\begin{aligned}\frac{1}{\tilde{e}} &\ge \frac{2}{e_{\min}(1+1+\varepsilon)} \ge \frac{1}{\frac{1}{w_+(x)}(1+\varepsilon/2)} \\ &= \left(1 - \frac{\varepsilon/2}{1+\varepsilon/2}\right) w_+(x) \ge \left(1 - \frac{\varepsilon}{2}\right) w_+(x).\end{aligned}$$

Thus, $|1/\tilde{e} - w_+(x)| \le \varepsilon w_+(x)$.

By Theorem 5, $\texttt{Decide}(P, w, \lambda)$ runs in cost $O\left(\frac{\sqrt{w\widetilde{W}_-}}{(1-\lambda)^{3/2}} \log \frac{1}{1-\lambda}\right)$. Let $w_i = 1/e^{(i)}_1$ and $\lambda_i = e^{(i)}_0/e^{(i)}_1$ be the values used at the i^{th} iteration. Since $e^{(i)}_1 \le e^{(i)}_{\max} \le \frac{1}{w_+(x)} + \Delta_i$, we have

$$\begin{aligned}\frac{1}{1-\lambda_i} &= \frac{e^{(i)}_1}{e^{(i)}_1 - e^{(i)}_0} \le \frac{\frac{1}{w_+(x)} + \Delta_i}{\frac{2}{3}e^{(i)}_{\max} + \frac{1}{3}e^{(i)}_{\min} - \frac{1}{3}e^{(i)}_{\max} - \frac{2}{3}e^{(i)}_{\min}} \\ &= \frac{3}{w_+(x)\Delta_i} + 3 = O\left(\frac{1}{\varepsilon}\right),\end{aligned}$$

since $\Delta_i = (2/3)^{i-1} \ge (2/3)^{T-1} = \Omega\left(\frac{\varepsilon}{w_+(x)}\right)$. Next, observe that $\frac{\sqrt{w_i}}{(1-\lambda_i)^{3/2}} = \frac{e^{(i)}_1}{(e^{(i)}_1 - e^{(i)}_0)^{3/2}} \le \left(\frac{1}{w_+(x)} + \Delta_i\right)\frac{3^{3/2}}{\Delta_i^{3/2}}$, so, ignoring the $\log \frac{1}{1-\lambda_i} = O(\log \frac{1}{\varepsilon})$ factor, the

cost of the i^{th} iteration can be computed as:

$$C_i = \frac{\sqrt{w_i \widetilde{W}_-}}{(1-\lambda_i)^{3/2}} \leq \sqrt{\widetilde{W}_-}\left(\frac{1}{w_+(x)} + \Delta_i\right)\frac{3^{3/2}}{\Delta_i^{3/2}}$$
$$= 3^{3/2}\frac{\sqrt{\widetilde{W}_-}}{w_+(x)}\left(\frac{3}{2}\right)^{\frac{3}{2}(i-1)} + 3^{3/2}\sqrt{\widetilde{W}_-}\left(\frac{3}{2}\right)^{\frac{1}{2}(i-1)}.$$

We can thus compute the total cost (neglecting logarithmic factors):

$$\sum_{i=1}^{T} C_i \leq \frac{\sqrt{\widetilde{W}_-}}{w_+(x)}\sum_{i=1}^{T}\left(\frac{3}{2}\right)^{\frac{3}{2}(i-1)} + \sqrt{\widetilde{W}_-}\sum_{i=1}^{T}\left(\frac{3}{2}\right)^{\frac{1}{2}(i-1)}$$
$$\leq \frac{\sqrt{\widetilde{W}_-}}{w_+(x)}\frac{\left(\frac{3}{2}\right)^{\frac{3}{2}T}-1}{\left(\frac{3}{2}\right)^{3/2}-1} + \sqrt{\widetilde{W}_-}\frac{\left(\frac{3}{2}\right)^{\frac{1}{2}T}-1}{\left(\frac{3}{2}\right)^{1/2}-1}$$
$$\leq O\left(\frac{\sqrt{\widetilde{W}_-}}{w_+(x)}\left(\frac{w_+(x)}{\varepsilon}\right)^{3/2} + \sqrt{\widetilde{W}_-}\left(\frac{w_+(x)}{\varepsilon}\right)^{1/2}\right)$$
$$= O\left(\frac{\sqrt{\widetilde{W}_- w_+(x)}}{\varepsilon^{3/2}}\right),$$

using the fact that $(2/3)^T = \Theta\left(\frac{\varepsilon}{w_+(x)}\right)$.

Finally, we have been assuming that `Decide` returns the correct bit on every call. We now justify this assumption. At round i, we will amplify the success probability of `Decide` to $1 - \frac{1}{9}(2/3)^{i-1}$, incurring a factor of $\log(9(3/2)^{i-1}) = O(\log \frac{w_+(x)}{\varepsilon})$ in the complexity. Then the total error is at most:

$$\sum_{i=1}^{T}\frac{1}{9}(2/3)^{i-1} = \frac{1}{9}\frac{1-(2/3)^{T-1}}{1-\frac{2}{3}} = \frac{1}{3}\left(1 - \frac{\varepsilon}{w_+(x)}\right) \leq \frac{1}{3}.$$

Thus, with probability $\geq 2/3$, `Decide` never errs, and the algorithm is correct. □

4.4 Span Program Phase Gap

The scaling in the error from Theorem 11, $1/\varepsilon^{3/2}$, is not ideal. For instance, we showed in Sect. 3.3 how to construct a quantum algorithm for approximate counting based on a simple span program for the OR function with complexity that scales like $1/\varepsilon^{3/2}$ in the error, whereas the best quantum algorithm for this task has complexity scaling as $1/\varepsilon$ in the error. However, the following theorem, which is a corollary to Lemmas 4 and 6, gives an alternative analysis of the complexity of the algorithm in Theorem 11 that

may be better in some cases, and in particular, has the more natural error dependence $1/\varepsilon$.

Theorem 12 *Fix $f : X \subseteq [q]^n \to \mathbb{R}_{>0}$. Let P be a normalized span program on $[q]^n$ such that $X \subseteq P_0$, and $\forall x \in X$, $w_-(x,P) = f(x)$. Define $\Delta(f) = \min_{x\in X} \Delta(U(P,x))$. Then there is a quantum algorithm that estimates f to relative accuracy ε using $\widetilde{O}\left(\frac{1}{\varepsilon}\frac{\sqrt{w_-(x,P)}}{\Delta(f)}\right)$ queries. Similarly, let P be a normalized span program such that $X \subseteq P_1$, and $\forall x \in X$, $w_+(x,P) = f(x)$. Define $\Delta'(f) = \min_{x\in X} \Delta(U'(P,x))$. Then there is a quantum algorithm that estimates f with relative accuracy ε using $\widetilde{O}\left(\frac{1}{\varepsilon}\frac{\sqrt{w_+(x,P)}}{\Delta'(f)}\right)$ queries.*

Proof To estimate $w_-(x)$, we can use phase estimation of $U(P,x)$ applied to $|w_0\rangle$, with precision $\Delta = \Delta(f)$ and accuracy $\eta = \frac{\varepsilon}{8}\frac{1}{W_-(P,f)}$, however, this results in $\log W_-$ factors, and W_- may be significantly larger than $w_-(x)$. Instead, we will start with $\eta = \frac{1}{2}$, and decrease it by $1/2$ until $\eta \approx \frac{\varepsilon}{w_-(x,P)}$.

Let $|w_0'\rangle$ be the result of applying phase estimation to precision $\Delta = \Delta(f)$ and accuracy η, and let Λ_0 be the projector onto states with 0 in the phase register. We will then estimate $\left\|\Lambda_0|w_0'\rangle\right\|^2$ to relative accuracy $\varepsilon/4$ using amplitude estimation. Since $\Delta \le \Delta(U(P,x))$, we have $\left\|\Pi_0^x|w_0\rangle\right\|^2 \le \left\|\Lambda_0|w_0'\rangle\right\|^2 \le \left\|\Pi_\Delta^x|w_0\rangle\right\|^2 + \eta = \left\|\Pi_0^x|w_0\rangle\right\|^2 + \eta$. By Lemma 4, we have $\left\|\Pi_0^x|w_0\rangle\right\|^2 = \frac{1}{w_-(x)}$, so we will obtain an estimate $\tilde{p}$ of $\frac{1}{w_-(x)}$ such that

$$\left(1-\frac{\varepsilon}{4}\right)\frac{1}{w_-(x)} \le \tilde{p} \le \left(1+\frac{\varepsilon}{4}\right)\left(\frac{1}{w_-(x)}+\eta\right).$$

If $\tilde{p} > 2(1+\frac{\varepsilon}{4})\eta$, then we know that $\frac{1}{w_-(x)} \ge \eta$, so we perform one more estimate with accuracy $\eta' = \frac{\varepsilon}{8}\eta \le \frac{\varepsilon}{8}\frac{1}{w-(x)}$ and return the resulting estimate. Otherwise, we let $\eta' = \eta/2$ and repeat.

To see that we will eventually terminate, suppose $\eta \le \frac{1}{4w_-(x)}$. Then

$$\tilde{p} \ge (1-\varepsilon/4)\frac{1}{w_-(x)} \ge (3/4)4\eta \ge (3/4)(4/5)(1+\varepsilon/4)4\eta \ge 2(1+\varepsilon/4)\eta,$$

so the algorithm terminates. Upon termination, we have

$$\tilde{p} \le \left(1+\frac{\varepsilon}{4}\right)\left(\frac{1}{w_-(x)}+\eta\right) \le \left(1+\frac{\varepsilon}{4}\right)\left(\frac{1}{w_-(x)}+\frac{\varepsilon}{8}\frac{1}{w_-(x)}\right) \le \left(1+\frac{\varepsilon}{2}\right)\frac{1}{w_-(x)},$$

so $|1/\tilde{p} - w_-(x)| \le \varepsilon w_-(x)$. By Theorem 1 and 2, the number of calls to U is:

$$\sum_{i=0}^{\log 4w_-(x)} \frac{1}{\Delta}\frac{\sqrt{w_-(x)}}{\varepsilon}\log 2^i + \frac{\sqrt{w_-(x)}}{\Delta\varepsilon}\log\frac{w_-(x)}{\varepsilon}$$

$$= \frac{1}{\Delta}\frac{\sqrt{w_-(x)}}{\varepsilon}\left(\sum_{i=0}^{\log 6w_-(x)} i + \log\frac{w_-(x)}{\varepsilon}\right),$$

which is at most $\frac{\sqrt{w_-(x)}}{\Delta}\varepsilon\log^2\frac{w_-(x)}{\varepsilon} = \widetilde{O}\left(\frac{\sqrt{w_-(x)}}{\Delta\varepsilon}\right)$. Similarly, we can estimate $w_+(x)$ to relative accuracy ε using $\widetilde{O}\left(\frac{\sqrt{w_+(x)}}{\Delta'\varepsilon}\right)$ calls to U'.

Theorem 12 is only useful if a lower bound on the phase gap of $U(P,x)$ or $U'(P,x)$ can be computed. This may not always be feasible, but the following two theorems show it is sufficient to compute the spectral norm of A, and the spectral gap, or specifically, smallest nonzero singular value, of the matrix $A(x) = A\Pi_{H(x)}$. This may still not be an easy task, but in Sect. 5, we show that we can get a better algorithm for estimating the effective resistance by this analysis, which, in the case of effective resistance, is very simple.

Theorem 13 *Let P be any span program on $[q]^n$. For any $x \in [q]^n$, we have $\Delta(U(P,x)) \geq 2\frac{\sigma_{\min}(A(x))}{\sigma_{\max}(A)}$.*

Proof Let $U = U(P,x)$. Consider $-U = (2\Pi_{(\ker A)^\perp} - I)(2\Pi_{H(x)} - I)$. By Corollary 2, if D is the discriminant of $-U$, then $\Delta(U) \geq 2\sigma_{\min}(D)$, so we will lower bound $\sigma_{\min}(D)$. Since the orthogonal projector onto $(\ker A)^\perp = \mathrm{row}A$ is A^+A, we have $D = A^+A\Pi_{H(x)} = A^+A(x)$.

We have $\sigma_{\min}(D) = \min_{|u\rangle\in\mathrm{row}D}\frac{\|D|u\rangle\|}{\||u\rangle\|}$, so let $|u\rangle \in \mathrm{row}D$ be a unit vector that minimizes $\|D|u\rangle\|$. Since $|u\rangle \in \mathrm{row}D \subseteq \mathrm{row}A(x)$, we have $\|A(x)|u\rangle\| \geq \sigma_{\min}(A(x))$. Since $A(x)|u\rangle \in \mathrm{col}A(x) \subseteq \mathrm{col}A = \mathrm{row}A^+$, we have

$$\begin{aligned}\sigma_{\min}(D) &= \left\|A^+A(x)|u\rangle\right\| \geq \sigma_{\min}(A^+)\,\|A(x)|u\rangle\| \\ &\geq \sigma_{\min}(A^+)\sigma_{\min}(A(x)) = \frac{\sigma_{\min}(A(x))}{\sigma_{\max}(A)},\end{aligned}$$

since $\sigma_{\min}(A^+) = \frac{1}{\sigma_{\max}(A)}$. Thus $\Delta(U) \geq 2\frac{\sigma_{\min}(A(x))}{\sigma_{\max}(A)}$. □

Theorem 14 *Let P be any span program. $\forall x \in P_1$, $\Delta(U'(P,x)) \geq 2\frac{\sigma_{\min}(A(x))}{\sigma_{\max}(A)}$.*

Proof We have

$$\begin{aligned}-U'(P,x)^\dagger &= (2(I - \Pi_{\ker A\oplus\mathrm{span}\{|w_0\rangle\}}) - I)(2\Pi_{H(x)} - I) \\ &= (2(I - \Pi_{\ker A} - \Pi_{|w_0\rangle}) - I)(2\Pi_{H(x)} - I),\end{aligned}$$

since $|w_0\rangle \in (\ker A)^\perp$, so $-U'(P,x)^\dagger$ has discriminant:

$$\begin{aligned}D' &= (\Pi_{(\ker A)^\perp} - \Pi_{|w_0\rangle})\Pi_{H(x)} = \Pi_{(\ker A)^\perp}\Pi_{H(x)} - \Pi_{|w_0\rangle}\Pi_{(\ker A)^\perp}\Pi_{H(x)} \\ &= \Pi_{|w_0\rangle^\perp}D.\end{aligned}$$

Since $x \in P_1$, let $|w_x\rangle = A(x)^+|\tau\rangle$. Then $D|w_x\rangle = A^+A(x)|w_x\rangle = A^+|\tau\rangle = |w_0\rangle$, so $|w_0\rangle \in \mathrm{col}D$. Let $\{|\phi_0\rangle = |w_0\rangle, |\phi_1\rangle, \ldots, |\phi_{r-1}\rangle\}$ be an orthogonal basis for $\mathrm{col}D$. Then we can write $D = \sum_{i=0}^{r-1} |\phi_i\rangle\langle v_i|$ for $|v_i\rangle = D^\dagger|\phi_i\rangle \neq 0$ (not necessarily orthogonal). Then $D' = \sum_{i=0}^{r-1} \Pi_{|w_0\rangle^\perp}|\phi_i\rangle\langle v_i| = \sum_{i=1}^{r-1} |\phi_i\rangle\langle v_i|$, so $\mathrm{col}D' = \mathrm{span}\{|\phi_1\rangle, \ldots, |\phi_{r-1}\rangle\} = \{|\phi\rangle \in \mathrm{col}D : \langle\phi|w_0\rangle = 0\}$. Thus:

$$\sigma_{\min}(D') = \min_{|u\rangle\in\mathrm{col}D'} \frac{\|\langle u|D'\|}{\||u\rangle\|} = \min_{|u\rangle\in\mathrm{col}D:\langle w_0|u\rangle=0} \frac{\|\langle u|\Pi_{|w_0\rangle^\perp}D\|}{\||u\rangle\|}$$
$$= \min_{|u\rangle\in\mathrm{col}D:\langle w_0|u\rangle=0} \frac{\|\langle u|D\|}{\||u\rangle\|} \geq \min_{|u\rangle\in\mathrm{col}D} \frac{\|\langle u|D\|}{\||u\rangle\|} = \sigma_{\min}(D).$$

By the proof of Theorem 13, we have $\sigma_{\min}(D) \geq \frac{\sigma_{\min}(A(x))}{\sigma_{\max}(A)}$ and by Corollary 2, we have $\Delta(U'(P,x)^\dagger) = \Delta(U'(P,x)) \geq 2\sigma_{\min}(D') \geq 2\sigma_{\min}(D) \geq 2\frac{\sigma_{\min}(A(x))}{\sigma_{\max}(A)}$. □

Combining the last three theorems, we get the following, which has Theorem 7 as a special case:

Theorem 15 *Fix $f : X \subseteq [q]^n \to \mathbb{R}_{>0}$, and define $\kappa(f) = \max_{x\in X} \frac{\sigma_{\max}(A)}{\sigma_{\min}(A(x))}$. Let P be any span program on $[q]^n$ such that $X \subseteq P_0$ (resp. $X \subseteq P_1$), and for all $x \in X$, $f(x) = w_-(x,P)$ (resp. $f(x) = w_+(x,P)$). Let $N = \||w_0\rangle\|^2$. Then there is a quantum algorithm that estimates f to relative accuracy ε using $\widetilde{O}\left(\frac{\kappa(f)}{\varepsilon}\sqrt{Nf(x)}\right)$ (resp. $\widetilde{O}\left(\frac{\kappa(f)}{\varepsilon}\sqrt{\frac{f(x)}{N}}\right)$) queries.*

Proof Let P' be the span program that is the same as P, but with target $\tau' = \frac{\tau}{\sqrt{N}}$. Then it is clear that $\frac{|w_0\rangle}{\sqrt{N}}$ is the minimal positive witness of P', and furthermore, it has norm 1, so P' is normalized. We can similarly see that for any $x \in P_1$, if $|w_x\rangle$ is an optimal positive witness for x in P, then $\frac{1}{\sqrt{N}}|w_x\rangle$ is an optimal positive witness for x in P', so $w_+(x,P') = \frac{w_+(x,P)}{N}$. Similarly, for any $x \in P_0$, if ω_x is an optimal negative witness for x in P, then $\sqrt{N}\omega_x$ is an optimal negative witness for x in P', so $w_-(x,P') = Nw_-(x,P)$. By Theorems 13 and 14, for all $x \in X$, $\frac{1}{\Delta(U(P',x))} \leq \kappa(f)$ (resp. $\frac{1}{\Delta(U'(P',x))} \leq \kappa(f)$). The result then follows from Theorem 12. □

5 Applications

In this section, we will demonstrate how to apply the ideas from Sect. 4 to get new quantum algorithms. Specifically, we will give upper bounds of $\widetilde{O}(n\sqrt{R_{s,t}}/\varepsilon^{3/2})$ and $\widetilde{O}(n\sqrt{R_{s,t}/\lambda_2}/\varepsilon)$ on the time complexity of estimating the effective resistance, $R_{s,t}$, between two vertices, s and t, in a graph. Unlike previous upper bounds, we study this problem in the adjacency model, however, there are similarities between the ideas of this upper bound and a previous quantum upper bound in the edge-list model due to Wang [23], which we discuss further at the end of this section.

A *unit flow* from s to t in G is a real-valued function θ on the directed edges $\vec{E}(G) = \{(u, v) : \{u, v\} \in E(G)\}$ such that:

1. $\forall (u, v) \in \vec{E}, \theta(u, v) = -\theta(v, u)$;
2. $\forall u \in [n] \setminus \{s, t\}, \sum_{v \in \Gamma(u)} \theta(u, v) = 0$, where $\Gamma(u) = \{v \in [n] : \{u, v\} \in E\}$;
3. $\sum_{u \in \Gamma(s)} \theta(s, u) = \sum_{u \in \Gamma(t)} \theta(u, t) = 1$.

Let $\mathcal{F}$ be the set of unit flows from s to t in G. The *effective resistance* from s to t in G is defined:

$$R_{s,t}(G) = \min_{\theta \in \mathcal{F}} \sum_{\{u,v\} \in E(G)} \theta(u, v)^2.$$

This quantity gives the resistance of a network of unit resistors described by G, but is also an interesting quantity for graph theoretic reasons. For instance, the commute time between s and t, which is the expected number of steps in a random walk starting from s to reach t, and then return to s, is exactly the product of the number of edges in G, and $R_{s,t}(G)$ [9].

In the adjacency model, we are given, as input, a string $x \in \{0, 1\}^{n \times n}$, representing a graph $G_x = ([n], \{\{i, j\} : x_{i,j} = 1\})$ (we assume that $x_{i,i} = 0$ for all i, and $x_{i,j} = x_{j,i}$ for all i, j). The problem of st-connectivity is the following. Given as input $x \in \{0, 1\}^{n \times n}$ and $s, t \in [n]$, decide if there exists a path from s to t in G_x; that is, whether or not s and t are in the same component of G_x. A span-program-based algorithm for this problem was given in [7], with time complexity $\widetilde{O}(n\sqrt{p})$, under the promise that, if s and t are connected in G_x, they are connected by a path of length $\leq p$. They use the following span program, defined on $\{0, 1\}^{n \times n}$:

$$H_{(u,v),0} = \{0\}, \ H_{(u,v),1} = \text{span}\{|u, v\rangle\}, \ V = \mathbb{R}^n,$$
$$A = \sum_{u,v \in [n]} (|u\rangle - |v\rangle)\langle u, v|, \ |\tau\rangle = |s\rangle - |t\rangle.$$

We have $H = \text{span}\{|u, v\rangle : u, v \in [n]\}$, and $H(x) = \text{span}\{|u, v\rangle : \{u, v\} \in E(G_x)\}$. Throughout this section, P will denote the above span program. We will use this span program to define algorithms for estimating the effective resistance. Ref. [7] are even able to show how to efficiently implement a unitary similar to $U(P, x)$, giving a time efficient algorithm. In "Appendix B", we adapt their proof to our setting, showing how to efficiently implement $U'(P^\beta, x)$ for any $n^{-O(1)} \leq \beta \leq n^{O(1)}$ and efficiently construct the initial state $|w_0\rangle$, making our algorithms time efficient as well.

The effective resistance between s and t is related to st-connectivity by the fact that if s and t are not connected, then $R_{s,t}$ is infinite (there is no flow from s to t) and if s and t are connected then $R_{s,t}$ is related to the number and length of paths from s to t. In particular, if s and t are connected by a path of length p, then $R_{s,t}(G) \leq p$ (take the unit flow that simply travels along this path). In general, if s and t are connected in G, then $\frac{2}{n} \leq R_{s,t}(G) \leq n - 1$. The span program for st-connectivity is amenable to the task of estimating the effective resistance due to the following.

Lemma 8 [7] *For any graph G_x on $[n]$, $x \in P_1$ if and only if s and t are connected, and in that case, $w_+(x, P) = \frac{1}{2} R_{s,t}(G_x)$.*

The following is a near immediate consequence of Lemma 8 and Theorem 6.

Theorem 16 *There exists a quantum algorithm for estimating $R_{s,t}(G_x)$ to accuracy ε with time complexity $\widetilde{O}\left(\frac{n\sqrt{R_{s,t}(G_x)}}{\varepsilon^{3/2}}\right)$ and space complexity $O(\log n)$.*

Proof We merely observe that if G is a connected graph, an approximate negative witness is $\omega : [n] \to \mathbb{R}$ that minimizes $\left\|\omega A \Pi_{H(x)}\right\|^2 = \sum_{\{u,v\} \in E} (\omega(u) - \omega(v))^2$ and satisfies $\omega(s) - \omega(t) = 1$. That is, ω is the voltage induced by a unit potential difference between s and t (see [10] for details). This is not unique, but if we fix $\omega(s) = 1$ and $\omega(t) = 0$, then the ω that minimizes $\left\|\omega A \Pi_{H(x)}\right\|^2$ is unique, and this is without loss of generality. In that case, for all $u \in [n]$, $0 \leq \omega(u) \leq 1$, so $\tilde{w}_-(x) = \|\omega A\|^2 = \sum_{u,v \in [n]} (\omega(u) - \omega(v)) \leq 2n^2 \quad$ and thus $\quad \widetilde{W}_- \leq 2n^2$.

By Theorem 6, we can estimate $R_{s,t}$ to precision ε using $\widetilde{O}\left(\frac{\sqrt{\widetilde{W}_- w_+(x)}}{\varepsilon^{3/2}}\right) = \widetilde{O}\left(\frac{n\sqrt{R_{s,t}(G_x)}}{\varepsilon^{3/2}}\right)$ calls to $U'(P^\beta, x)$ for some β, which, by Theorem 19, costs $O(\log n)$ time and space. □

By analyzing the spectra of A and $A(x)$, and applying Theorem 7, we can get an often better algorithm (Theorem 17). The *spectral gap* of a graph G, denoted $\lambda_2(G)$, is the second smallest eigenvalue (including multiplicity) of the Laplacian of G, which is defined $L_G = \sum_{u \in [n]} d_u |u\rangle\langle u| - \sum_{u \in [n]} \sum_{v \in \Gamma(u)} |u\rangle\langle v|$, where d_u is the degree of u, and $\Gamma(u)$ is the set of neighbours of u. The smallest eigenvalue of L_G is 0 for any graph G. A graph G is connected if and only if $\lambda_2(G) > 0$. A connected graph G has $\frac{2}{n^2} \leq \lambda_2(G) \leq n$.

The following theorem is an improvement over Theorem 16 when $\lambda_2(G) > \varepsilon$. In particular, it is an improvement for all ε when we know that $\lambda_2(G) > 1$.

Theorem 17 *Let $\mathcal{G}$ be a family of graphs such that for all $x \in \mathcal{G}$, $\lambda_2(G_x) \geq \mu$. Let $f : \mathcal{G} \times [n] \times [n] \to \mathbb{R}_{>0}$ be defined by $f(x, s, t) = R_{s,t}(G_x)$. There exists a quantum algorithm for estimating f to relative accuracy ε that has time complexity $\widetilde{O}\left(\frac{1}{\varepsilon} n\sqrt{R_{s,t}(G_x)/\mu}\right)$ and space complexity $O(\log n)$.*

Proof Before applying Theorem 7, we compute $\||w_0\rangle\|^2$, to normalize P.

Lemma 9 $N = \||w_0\rangle\|^2 = \frac{1}{n}$.

Proof Since $H(x) = H$ when G_x is the complete graph, by Lemma 8, we need only compute $R_{s,t}$ in the complete graph. It is simple to verify that the optimal unit st-flow in the complete graph has $\frac{1}{n}$ units of flow on every path of the form (s, u, t) for $u \in [n] \setminus \{s, t\}$, and $\frac{2}{n}$ units of flow on the edge (s, t). Thus, $R_{s,t}(K_n) = \sum_{u \in [n] \setminus \{s,t\}} 2(1/n)^2 + (2/n)^2 = 2/n$. So $\||w_0\rangle\|^2 = \frac{1}{2} R_{s,t}(K_n) = \frac{1}{n}$.

Next, we compute the following:

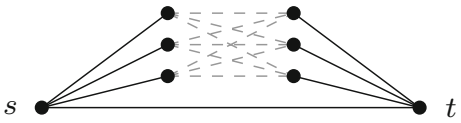


Fig. 1 The graphs in $\mathcal{G}_0$ contain only the solid edges. The graphs in $\mathcal{G}_1$ contain the solid edges and one of the dashed edges. We can embed an instance of OR in the dashed edges. If a dashed edge is included, the number of st-paths increases, decreasing the effective resistance

Lemma 10 *For any* $x \in \mathcal{G}$, $\frac{\sigma_{\max}(A)}{\sigma_{\min}(A(x))} = \sqrt{\frac{n}{\lambda_2(G_x)}} \leq \sqrt{\frac{n}{\mu}}$, *so* $\kappa(f) \leq \sqrt{\frac{n}{\mu}}$.

Proof Let L_x denote the Laplacian of G_x. We have:

$$\begin{aligned} A(x)A(x)^T &= \sum_{u\in[n]} \sum_{v\in\Gamma(u)} (|u\rangle - |v\rangle)(\langle u| - \langle v|) \\ &= 2\sum_{u\in[n]} d_u |u\rangle\langle u| - 2\sum_{u\in[n]} \sum_{v\in\Gamma(u)} |u\rangle\langle v| = 2L_x. \end{aligned}$$

Thus, if L denotes the Laplacian of the complete graph, we also have $AA^T = 2L$. Letting J denote the all ones matrix, we have $L = (n-1)I - (J-I) = nI - J$, and since $J = n|u\rangle\langle u|$ where $|u\rangle = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} |i\rangle$, if $|u_1\rangle, \ldots, |u_{n-1}\rangle, |u\rangle$ is any orthonormal basis of $\mathbb{R}^n$, then $L = n\sum_{i=1}^{n-1} |u_i\rangle\langle u_i| + n|u\rangle\langle u| - n|u\rangle\langle u| = \sum_{i=1}^{n-1} n|u_i\rangle\langle u_i|$, so the spectrum of L is 0, with multiplicity 1, and n with multiplicity $n-1$. Thus, the only nonzero singular value of A is $\sqrt{2n} = \sigma_{\max}(A)$. Furthermore, since $\lambda_2(G_x)$ is the smallest nonzero eigenvalue of L_x, and $A(x)A(x)^T = 2L_x$, $\sigma_{\min}(A(x)) = \sqrt{2\lambda_2(G_x)}$. The result follows. □

Finally, by Lemma 8, we have $w_+(x, P) = \frac{1}{2}R_{s,t}(G_x)$, so, applying Theorem 15, we get an algorithm that makes $\widetilde{O}\left(\frac{\kappa(f)}{\varepsilon}\sqrt{\frac{w_+(x,P)}{N}}\right) = \widetilde{O}\left(\frac{1}{\varepsilon}\sqrt{n/\mu}\sqrt{R_{s,t}n}\right)$ calls to $U'(P, x)$. By Theorem 19, the time complexity of this algorithm is $\widetilde{O}\left(\frac{1}{\varepsilon}n\sqrt{R_{s,t}/\mu}\right)$ and the space complexity is $O(\log n)$. □

Both of our upper bounds have linear dependence on n, and the following theorem shows that this is optimal.

Theorem 18 (Lower Bound) *There exists a family of graphs* $\mathcal{G}$ *such that estimating effective resistance on* $\mathcal{G}$ *costs at least* $\Omega(n)$ *queries.*

Proof Let $\mathcal{G}_0$ be the set of graphs consisting of two stars $K_{1,n/2-1}$, centered at s and t, with an edge connecting s and t (see Fig. 1). Let $\mathcal{G}_1$ be the set of graphs consisting of graphs from $\mathcal{G}_0$ with a single edge added between two degree one vertices from different stars. Let $\mathcal{G} = \mathcal{G}_0 \cup \mathcal{G}_1$. We first note that we can distinguish between $\mathcal{G}_0$ and $\mathcal{G}_1$ by estimating effective resistance on $\mathcal{G}$ to accuracy $\frac{1}{10}$: If $G \in \mathcal{G}_0$, then there is a single st-path, consisting of one edge, so the effective resistance is 1. If $G \in \mathcal{G}_1$, then there are two st-paths, one of length 1 and one of length 3. We put a flow of $\frac{1}{4}$ on the length-3 path and $\frac{3}{4}$ on the length-1 path to get effective resistance at most $(3/4)^2 + 3(1/4)^2 = \frac{3}{4}$.

We now describe how to embed an instance $y \in \{0,1\}^{(n/2-1)^2}$ of $\mathrm{OR}_{(n/2-1)^2}$ in a graph. We let $s = 1$ be connected to every vertex in $\{2,\ldots,n/2\}$, and $t = n$ be connected to every vertex in $\{n/2+1,\ldots,n-1\}$. Let the values of $\{G_{i,j} : i \in \{2,\ldots,n/2\}, j \in \{n/2,\ldots,n-1\}\}$ be determined by y. Let all other values $G_{i,j}$ be 0. Then clearly $R_{s,t}(G) \geq 1$ if and only if $y = 0\ldots0$ (in that case $G \in \mathcal{G}_0$) and otherwise, $R_{s,t}(G) \leq 3/4$, since there is at least one extra path from s to t (in that case $G \in \mathcal{G}_1$). The result follows from the lower bound of $\Omega(\sqrt{(n/2-1)^2}) = \Omega(n)$ on $\mathrm{OR}_{(n/2-1)^2}$. □

Discussion The algorithms from Theorems 16 and 17 are the first quantum algorithms for estimating the effective resistance in the adjacency model, however, the problem has been studied previously in the edge-list model [23], where Wang obtains a quantum algorithm with complexity $\widetilde{O}\left(\frac{d^{3/2}\log n}{\Phi(G)^2\varepsilon}\right)$, where $\Phi(G) \leq 1$ is the conductance (or edge-expansion) of G. In the edge-list model, the input $x \in [n]^{[n]\times[d]}$ models a d-regular graph (or d-bounded degree graph) G_x by $x_{u,i} = v$ for some $i \in [d]$ whenever $\{u,v\} \in E(G_x)$. Wang requires edge-list queries to simulate walking on the graph, which requires constructing a superposition over all neighbours of a given vertex. This type of edge-list query can be simulated by $\sqrt{n/d}$ adjacency queries to a d-regular graph, using quantum search, so Wang's algorithm can be converted to an algorithm in the adjacency query model with cost $\widetilde{O}\left(\frac{d^{3/2}}{\Phi(G)^2\varepsilon}\sqrt{\frac{n}{d}}\right)$. We can compare our results to this by noticing that $R_{s,t} \leq \frac{1}{\lambda_2(G)}$ [9], implying that our algorithm always runs in time at most $\widetilde{O}\left(\frac{1}{\varepsilon}\frac{n}{\mu}\right)$. If G is a connected d-regular graph, then $\lambda_2(G) = d\delta(G)$, where $\delta(G)$ is the spectral gap of a random walk on G. By Cheeger inequalities, we have $\frac{\Phi^2}{2} \leq \delta$ [17], so the complexity of the algorithm from Theorem 17 is at most $\widetilde{O}\left(\frac{1}{\varepsilon}\frac{n}{d\delta}\right) = \widetilde{O}\left(\frac{1}{\varepsilon}\frac{n}{d\Phi^2}\right)$, which is an improvement over the bound of $\widetilde{O}\left(\frac{1}{\varepsilon}\frac{d^{3/2}}{\Phi^2}\sqrt{\frac{n}{d}}\right) = \widetilde{O}\left(\frac{1}{\varepsilon}\frac{d}{\Phi^2}\sqrt{n}\right)$ given by naively adapting Wang's algorithm to the adjacency model whenever $d > \sqrt[4]{n}$. In general our upper bound may be much better than $\frac{1}{\varepsilon}\frac{n}{d\Phi^2}$, since the Cheeger inequality is not tight, and $R_{s,t}$ can be much smaller than $\frac{1}{\lambda_2}$.

It is worth further discussing Wang's algorithms for estimating effective resistance, due to their relationship with the ideas presented here. In order to get a time-efficient algorithm for st-connectivity, Belovs and Reichardt show how to efficiently reflect about the kernel of A (see also "Appendix B"), A being related to the Laplacian of a complete graph, L, by $AA^T = 2L$. This implementation consists, in part, of a quantum walk on the complete graph. Wang's algorithm directly implements a reflection about the kernel of $A(x)$ by instead using a quantum walk on the graph G, which can be done efficiently in the edge-list model. For general span programs, when a reflection about the kernel of $A(x)$ can be implemented efficiently in such a direct way, this can lead to an efficient quantum algorithm for estimating the witness size.

We also remark on another quantum algorithm for estimating effective resistance, also from Wang [23] with worse complexity $\widetilde{O}\left(\frac{d^8\mathrm{polylog}n}{\Phi(G)^{10}\varepsilon^2}\right)$, obtained by using the HHL algorithm [12] to estimate $\left\|A(x)^+|\tau\rangle\right\|^2$, which is the positive witness size of x, or in this case, the effective resistance. We remark that, for any span program, $w_+(x) =$

$\||w_x\rangle\|^2 = \left\|A(x)^+|\tau\rangle\right\|^2$, so HHL may be another means of estimating the positive witness size. There are several caveats: $A(x)$ must be efficiently row-computable, and the complexity additionally depends on $\frac{\sigma_{\max}(A(x))}{\sigma_{\min}(A(x))}$, the *condition number* of $A(x)$ (We remark that this is upper bounded by $\frac{\sigma_{\max}(A)}{\sigma_{\min}(A(x))}$, upon which the complexity of some of our algorithms depends as well). However, if this approach yields an efficient algorithm, it is efficient in time complexity, not only query complexity. We leave further exploration of this idea for future research.

6 Conclusion and Open Problems

Summary We have presented several new techniques for turning span programs into quantum algorithms, which we hope will have future applications. Specifically, given a span program P, in addition to algorithms for deciding any function f such that $f^{-1}(0) \subseteq P_0$ and $f^{-1}(1) \subseteq P_1$, we also show how to get several different algorithms for deciding a number of related threshold problems, as well as estimating the witness size. In addition to algorithms based on the standard effective spectral gap lemma, we also show how to get algorithms by analyzing the real phase gap.

We hope that the importance of this work lies not only in its potential for applications, but in the improved understanding of the structure and power of span programs. A number of very important quantum algorithms rely on a similar structure, using phase estimation of a unitary that depends on the input to distinguish between different types of inputs. Span-program-based algorithms represent a very general class of such algorithms, making them not only important to the study of the quantum query model, but to quantum algorithms in general.

Further Applications The main avenue for future work is in applications of our techniques to obtain new quantum algorithms. We stress that *any* span program for a decision problem can now be turned into an algorithm for estimating the positive or negative witness size, if these correspond to some meaningful function, or deciding threshold functions related to the witness size. A natural source of potential future applications is in the rich area of property testing problems (for a survey, see [18]).

Span Programs and HHL One final open problem, briefly discussed at the end of the previous section, is the relationship between estimating the witness size and the HHL algorithm [12]. The HHL algorithm can be used to estimate $\left\|M^+|u\rangle\right\|^2$, given the state $|u\rangle$ and access to a row-computable linear operator M. When $M = A(x)$, this quantity is exactly $w_+(x)$, so if $A(x)$ is row-computable — that is, there is an efficient procedure for computing the i^{th} nonzero entry of the j^{th} row of $A(x)$, then HHL gives us yet another means of estimating the witness size, whose time complexity is known, rather than only its query complexity. It may be interesting to explore this connection further.

Acknowledgements The authors would like to thank David Gosset, Shelby Kimmel, Ben Reichardt, and Guoming Wang for useful discussions about span programs. We would especially like to thank Shelby Kimmel for valuable feedback and suggestions on an earlier draft of this paper. Finally, S.J. would like to thank Moritz Ernst for acting as a sounding board throughout the writing of this paper.

Appendix A: Span Program Scaling

In this section we prove Theorem 10. Let $P = (H, V, \tau, A)$ be any span program on $[q]^n$, and let $N = \||w_0\rangle\|^2$ for $|w_0\rangle$ the optimal positive witness of P. We define $P^\beta = (H^\beta, A^\beta, \tau^\beta, V^\beta)$ as follows. Let $|\hat{0}\rangle$ and $|\hat{1}\rangle$ be two vectors orthogonal to H and V. We define:

$$\forall j \in [n], a \in [q],\ H^\beta_{j,a} = H_{j,a},\ H^\beta_{\text{true}} = H_{\text{true}} \oplus \text{span}\{|\hat{1}\rangle\},\ H^\beta_{\text{false}} = H_{\text{false}} \oplus \text{span}\{|\hat{0}\rangle\}$$

$$V^\beta = V \oplus \text{span}\{|\hat{1}\rangle\}, \quad A^\beta = \beta A + \tau|\hat{0}\rangle + \frac{\sqrt{\beta^2+N}}{\beta}|\hat{1}\rangle\langle\hat{1}|, \quad \tau^\beta = \tau + |\hat{1}\rangle$$

We then have and $H^\beta = H \oplus \text{span}\{|\hat{0}\rangle, |\hat{1}\rangle\}$ and $H^\beta(x) = H(x) \oplus \text{span}\{|\hat{1}\rangle\}$. In order to prove Theorem 10, we show that:

- For all $x \in P_1$, $w_+(x, P^\beta) = \frac{1}{\beta^2} w_+(x, P) + \frac{\beta^2}{N+\beta^2}$ and $\tilde{w}_-(x, P^\beta) \leq \beta^2 \tilde{w}_-(x, P) + 2$;
- for all $x \in P_0$, $w_-(x, P^\beta) = \beta^2 w_-(x, P) + 1$ and $\tilde{w}_+(x, P^\beta) \leq \frac{1}{\beta^2}\tilde{w}_+(x, P) + 2$;
- P^β has optimal positive witness $|w_0^\beta\rangle = \frac{\beta}{\beta^2+N}|w_0\rangle + \frac{N}{\beta^2+N}|\hat{0}\rangle + \frac{\beta}{\sqrt{\beta^2+N}}|\hat{1}\rangle$, and $\left\||w_0^\beta\rangle\right\|^2 = 1$.

Lemma 11 *The optimal positive witness in P^β is $|w_0^\beta\rangle = \frac{\beta}{\beta^2+N}|w_0\rangle + \frac{N}{\beta^2+N}|\hat{0}\rangle + \frac{\beta}{\sqrt{\beta^2+N}}|\hat{1}\rangle$. It is easily verified that $\left\||w_0^\beta\rangle\right\|^2 = 1$.*

Proof Let $|w_0'\rangle = |h\rangle + b|\hat{0}\rangle + c|\hat{1}\rangle$ be the smallest witness in P^β, for some $|h\rangle \in H$. Since $A^\beta|w_0'\rangle = \beta A|h\rangle + b\tau + c\frac{\sqrt{\beta^2+N}}{\beta}|\hat{1}\rangle = \tau + |\hat{1}\rangle$, we must have $c = \frac{\beta}{\sqrt{\beta^2+N}}$ and $A|h\rangle = \frac{1-b}{\beta}\tau$, so $|h\rangle = \frac{1-b}{\beta}|w\rangle$ for some positive witness $|w\rangle$ of P. We have:

$$\left\||w_0'\rangle\right\|^2 = \frac{(1-b)^2}{\beta^2}\||w\rangle\|^2 + b^2 + \frac{\beta^2}{\beta^2+N}.$$

This is minimized by taking $|w\rangle = |w_0\rangle$, the smallest witness of P, and setting $b = \frac{N}{\beta^2+N}$, giving:

$$|w_0^\beta\rangle = \frac{\beta}{\beta^2+N}|w_0\rangle + \frac{N}{\beta^2+N}|\hat{0}\rangle + \frac{\beta}{\sqrt{\beta^2+N}}|\hat{1}\rangle.$$

Lemma 12 *$\forall x \in P_1$, $w_+(x, P^\beta) = \frac{1}{\beta^2} w_+(x, P) + \frac{\beta^2}{N+\beta^2}$ and $\tilde{w}_-(x, P^\beta) \leq \beta^2 \tilde{w}_-(x, P) + 2$.*

Proof The proof is similar to that of Lemma 11, however, we have $H^\beta(x) = H(x) \oplus \text{span}\{|\hat{1}\rangle\}$, so a positive witness for x has the form $|w_x'\rangle = |h\rangle + \frac{\beta}{\sqrt{\beta^2+N}}|\hat{1}\rangle$ with $\beta|h\rangle$

some witness for x in P. Clearly $\| |w'_x\rangle \|$ is minimized by setting $|h\rangle = \frac{1}{\beta}|w_x\rangle$ for $|w_x\rangle$ the minimal positive witness for x in P, so we have $w_+(x, P^\beta) = \frac{1}{\beta^2} w_+(x, P) + \frac{\beta^2}{\beta^2+N}$, as required.

Let $\tilde{\omega}$ be an optimal min-error witness for x in P, and define

$$\tilde{\omega}' = \frac{(\beta^2+N)w_+(x,P)}{\beta^4+(\beta^2+N)w_+(x,P)}\tilde{\omega} + \frac{\beta^4}{\beta^4+(\beta^2+N)w_+(x,P)}\langle\hat{1}|.$$

We have $\tilde{\omega}'(\tau+|\hat{1}\rangle) = \frac{(\beta^2+N)w_+(x,P)}{\beta^4+(\beta^2+N)w_+(x,P)}\tilde{\omega}(\tau) + \frac{\beta^4}{\beta^4+(\beta^2+N)w_+(x,P)} = 1$, and:

$$\begin{aligned}
&\left\|\tilde{\omega}' A^\beta \Pi_{H^\beta(x)}\right\|^2 \\
&= \left\| \frac{(\beta^2+N)w_+(x,P)}{\beta^4+(\beta^2+N)w_+(x,P)}\tilde{\omega}\beta A \Pi_{H(x)} \right\|^2 + \left\| \frac{\beta^4}{\beta^4+(\beta^2+N)w_+(x,P)} \frac{\sqrt{\beta^2+N}}{\beta}\langle\hat{1}| \right\|^2 \\
&= \frac{(\beta^2+N)^2 w_+(x,P)^2\beta^2}{(\beta^4+(\beta^2+N)w_+(x,P))^2}\frac{1}{w_+(x,P)} + \frac{\beta^8}{(\beta^4+(\beta^2+N)w_+(x,P))^2}\frac{\beta^2+N}{\beta^2} \\
&= \frac{(\beta^2+N)^2 w_+(x,P)\beta^2 + \beta^6(\beta^2+N)}{(\beta^4+(\beta^2+N)w_+(x,P))^2} = \frac{\beta^2(\beta^2+N)}{\beta^4+(\beta^2+N)w_+(x,P)} = \frac{1}{w_+(x,P^\beta)}
\end{aligned}$$

so $\tilde{\omega}'$ is a min-error witness for x in P^β. Thus, letting $\varepsilon = \frac{(\beta^2+N)w_+(x,P)}{\beta^4+(\beta^2+N)w_+(x,P)}$, we have

$$\begin{aligned}
\tilde{w}_-(x,P^\beta) \le \left\|\tilde{\omega}' A^\beta\right\|^2 &= \left\| \varepsilon\tilde{\omega}\beta A + \varepsilon\tilde{\omega}(\tau)\langle\hat{0}| + \frac{\sqrt{\beta^2+N}}{\beta}\tilde{\omega}'(\hat{1})\langle\hat{1}| \right\|^2 \\
&\le \beta^2 \|\tilde{\omega}A\|^2 + 1 + \frac{\beta^2+N}{\beta^2}\frac{\beta^8}{(\beta^4+(\beta^2+N)w_+(x,P))^2} \\
&\le \beta^2\tilde{w}_-(x,P) + 1 + \frac{\beta^6(\beta^2+N)}{(\beta^4+\beta^2 w_+(x,P))^2} \;\le\; \beta^2\tilde{w}_+(x,P) + 2,
\end{aligned}$$

where in the last line, we use the fact that $w_+(x, P) \ge N$. □

Lemma 13 *$\forall x \in P_0$, $w_-(x, P^\beta) = \beta^2 w_-(x, P) + 1$, and $\tilde{w}_+(x, P^\beta) \le \frac{1}{\beta^2}\tilde{w}_+(x, P) + 2$.*

Proof Let ω'_x be an optimal negative witness for x in P^β. Since $\omega'_x \Pi_{H^\beta(x)} = 0$, $\omega'_x|\hat{1}\rangle = 0$, so $\omega'_x(\tau^\beta) = \omega'_x(\tau) + \omega'_x(|\hat{1}\rangle) = \omega'_x(\tau) = 1$. Furthermore, ω'_x minimizes

$$\left\|\omega'_x A^\beta\right\|^2 = \left\|\beta\omega'_x A + \omega'_x(\tau)|\hat{0}\rangle\right\|^2 = \beta^2\left\|\omega'_x A\right\|^2 + 1.$$

This is minimized by taking $\omega'_x|_V$ to be the minimal negative witness of x in P, so $\left\|\omega'_x A\right\|^2 = w_-(x, P)$, and thus $w_-(x, P^\beta) = \beta^2 w_-(x, P) + 1$.

Next, let $|\tilde{w}\rangle$ be an optimal min-error positive witness for x in P. Define:

$$|\tilde{w}'\rangle := \frac{\beta w_-(x,P)}{1+\beta^2 w_-(x,P)}|\tilde{w}\rangle + \frac{1}{1+\beta^2 w_-(x,P)}|\hat{0}\rangle + \frac{\beta}{\sqrt{\beta^2+N}}|\hat{1}\rangle.$$

We have:

$$A|\tilde{w}'\rangle = \frac{\beta^2 w_-(x,P)}{1+\beta^2 w_-(x,P)}\tau + \frac{1}{1+\beta^2 w_-(x,P)}\tau + |\hat{1}\rangle = \tau + |\hat{1}\rangle = \tau^\beta,$$

and since $H^\beta(x)^\perp = H(x)^\perp \oplus \mathrm{span}\{|\hat{0}\rangle\}$:

$$\begin{aligned}
\left\|\Pi_{H^\beta(x)^\perp}|\tilde{w}'\rangle\right\|^2 &= \left\|\Pi_{H(x)^\perp}|\tilde{w}'\rangle\right\|^2 + \left\|\Pi_{|\hat{0}\rangle}|\tilde{w}'\rangle\right\|^2 \\
&= \frac{\beta^2 w_-(x,P)^2}{(1+\beta^2 w_-(x,P))^2}\left\|\Pi_{H(x)^\perp}|\tilde{w}\rangle\right\|^2 + \frac{1}{(1+\beta^2 w_-(x,P))^2} \\
&= \frac{\beta^2 w_-(x,P)^2}{(1+\beta^2 w_-(x,P))^2}\frac{1}{w_-(x,P)} + \frac{1}{(1+\beta^2 w_-(x,P))^2} \\
&= \frac{1}{1+\beta^2 w_-(x,P)} \quad = \quad \frac{1}{w_-(x,P^\beta)},
\end{aligned}$$

so $|\tilde{w}'\rangle$ has minimal error. Thus:

$$\begin{aligned}
\tilde{w}_+(x,P^\beta) \le \left\||\tilde{w}'\rangle\right\|^2 &= \frac{\beta^2 w_-(x,P)^2}{(1+\beta^2 w_-(x,P))^2}\||\tilde{w}\rangle\|^2 + \frac{1}{(1+\beta^2 w_-(x,P))^2} + \frac{\beta^2}{\beta^2+N} \\
&\le \frac{\beta^2 w_-(x,P)^2 \tilde{w}_+(x,P)}{(1+\beta^2 w_-(x,P))^2} + 2 \le \frac{\beta^2 w_-(x,P)^2 \tilde{w}_+(x,P)}{\beta^4 w_-(x,P)^2} + 2 = \frac{\tilde{w}_+(x,P)}{\beta^2} + 2.
\end{aligned}$$

□

Appendix B: Time Complexity Analysis

In [7], the authors analyze the time complexity of the reflections needed to implement their span program to give a time upper bound on st-connectivity. Since our algorithms look superficially different from theirs, we reproduce their analysis here to show an upper bound on the quantum time complexity of estimating effective resistance.

Theorem 19 *Let P be the span program for st-connectivity given in Sect. 5. Then for any β such that $1/n^{O(1)} \le \beta \le n^{O(1)}$, $U'(P^\beta, x)$ can be implemented in quantum time complexity $O(\log n)$ and space $O(\log n)$, and $|w_0^\beta\rangle$ can be constructed in quantum time complexity $O(\log n)$.*

Proof In order to implement $U'(P^\beta, x)$, we implement the reflections $R_x(\beta) = 2\Pi_{H^\beta(x)} - I$ and $R'_P(\beta) = 2\Pi_{\ker A^\beta \oplus \mathrm{span}\{|w_0^\beta\rangle\}} - I$. We remark that $R_x(\beta)$ is easily implemented in a single query and constant overhead. This proof deals with the

implementation of $R'_P(\beta)$, which can be easily implemented given an implementation of $R_P = 2\Pi_{\ker A} - I$.

In order to implement R_P, we describe a unitary $W = (2\Pi_Z - I)(2\Pi_Y - I)$ that can be efficiently implemented, and such that W can be used to implement R_P. In order to show that W implements R_P, we need to show that some isometry $M_Y : H \to Y$ maps $\ker A$ to the (-1)-eigenspace of W, and $(\ker A)^\perp$ to the 1-eigenspace of W. This allows us to implement R_P by first implementing the isometry M_Y, applying W, and then uncomputing M_Y.

Define the spaces Z and Y as follows:

$$Z = \operatorname{span}\left\{ |z_u\rangle := \frac{1}{\sqrt{2(n-1)}} \sum_{v \neq u} |0, u, u, v\rangle + \frac{1}{\sqrt{2(n-1)}} \sum_{v \neq u} |1, u, v, u\rangle : u \in [n] \right\}; \text{ and}$$

$$Y = \operatorname{span}\left\{ |y_{u,v}\rangle := (|0, u, u, v\rangle - |1, v, u, v\rangle)/\sqrt{2} : u, v \in [n], u \neq v \right\}.$$

Define isometries

$$M_Z = \sum_{u \in [n]} |z_u\rangle\langle u| \quad \text{and} \quad M_Y = \sum_{(u,v) \in [n]^2 : u \neq v} |y_{u,v}\rangle\langle u, v|.$$

Lemma 14 *Let $S = \{M_Y|\psi\rangle : |\psi\rangle \in \ker A\}$ and $S' = \{M_Y|\psi\rangle : |\psi\rangle \in (\ker A)^\perp\}$ be the images of $\ker A$ and $(\ker A)^\perp$ respectively under the isometry M_Y. Then $S = Y \cap Z^\perp$, which is exactly the intersection of Y and the (-1)-eigenspace of W, and $S' = Y \cap Z$, which is exactly the intersection of Y and the 1-eigenspace of W.*

Proof We have:

$$M_Z^\dagger M_Y = \frac{1}{2\sqrt{n-1}} \sum_{u \in [n]} \sum_{v \neq u} |u\rangle \left(\langle 0, u, u, v| + \langle 1, u, v, u|\right) \sum_{\substack{a,b \in [n]: \\ a \neq b}} (|0, a, a, b\rangle - |1, b, a, b\rangle)\langle a, b|$$

$$= \frac{1}{2\sqrt{n-1}} \sum_{u \in [n]} \sum_{v \neq u} |u\rangle\langle u, v| - \frac{1}{2\sqrt{n-1}} \sum_{u \in [n]} \sum_{v \neq u} |v\rangle\langle u, v| = \frac{1}{2\sqrt{n-1}} A.$$

Thus, for all $|\psi\rangle \in \ker A$, $M_Y|\psi\rangle \in Y \cap \ker M_Z^\dagger = Y \cap Z^\perp$, so $S \subseteq Y \cap Z^\perp$. On the other hand, if $|\psi\rangle \in (\ker A)^\perp$, then $M_Y|\psi\rangle \in Y \cap (\ker M_Z^\dagger)^\perp = Y \cap Z$. By Theorem 3, the (-1)-eigenspace of W is exactly $(Y \cap Z^\perp) \oplus (Y^\perp \cap Z)$ and the 1-eigenspace of W is exactly $(Y \cap Z) \oplus (Y^\perp \cap Z^\perp)$. □

Lemma 15 *M_Y, $R_Z = 2\Pi_Z - I$ and $R_Y = 2\Pi_Y - I$ can be implemented in time $O(\log n)$.*

Proof To implement R_Z and R_Y, we need only show how to implement the unitary versions of M_Z and M_Y. We begin with M_Z. For any $u \in [n]$, we can map $|u\rangle \mapsto |0, u, u, 0\rangle$ by initializing three new registers and copying u into one of them. Then we map:

$$|0, u, u, 0\rangle \mapsto |0, u, u\rangle \frac{1}{\sqrt{n-1}} \sum_{v \neq u} |v\rangle$$

$$\overset{H\otimes I^{\otimes 3}}{\mapsto} \frac{1}{\sqrt{2(n-1)}}\left(|0,u,u\rangle\sum_{v\neq u}|v\rangle+|1,u,u\rangle\sum_{v\neq u}|v\rangle\right)\mapsto|x_u\rangle,$$

where the last transformation is achieved by swapping the last two registers conditioned on the first. This can be implemented in $O(\log n)$ elementary gates.

For M_Y, we start by mapping any edge $|u,v\rangle$ to $|1,0,u,v\rangle$, followed by:

$$|1,0,u,v\rangle \overset{H\otimes I^{\otimes 3}}{\mapsto} \frac{1}{\sqrt{2}}(|0,0,u,v\rangle-|1,0,u,v\rangle)\mapsto\frac{1}{\sqrt{2}}(|0,u,u,v\rangle-|1,v,u,v\rangle)=|y_{u,v}\rangle,$$

where in the last step we copy either u or v into the second register depending on the value of the first register. This can be implemented in $O(1)$ elementary gates.

Then in order to implement R_Z, we simply apply $M_Z^\dagger$, reflect about $\mathrm{span}\{|0,u,u,0\rangle : u\in[n]\}$, and then apply M_Z again. To implement R_Y, we apply $M_Y^\dagger$, reflect about $\mathrm{span}\{|1,0,u,v\rangle : u,v\in[n], u\neq v\}$, and then apply M_Y. □

We now show how to efficiently implement the span program P^β when $1/n^{O(1)}\leq\beta\leq n^{O(1)}$. First, consider $|w_0\rangle$, the minimal positive witness for P. Since $|w_0\rangle$ corresponds to an optimal st-flow in the complete graph, it is easy to compute that

$$|w_0\rangle=\frac{1}{n}|s,t\rangle+\frac{1}{2n}\sum_{u\in[n]\setminus\{s,t\}}(|s,u\rangle+|u,t\rangle)-\frac{1}{n}|t,s\rangle-\frac{1}{2n}\sum_{u\in[n]}(|t,u\rangle+|u,s\rangle),$$

and $\||w_0\rangle\|^2=\frac{1}{n}$ (see also Lemma 9). We can construct this state by mapping $|s,0\rangle+|0,t\rangle\mapsto\sum_{u\neq s}|s,u\rangle+\sum_{u\neq t}|u,t\rangle$ and then performing a swap controlled on an additional register in the state $\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$. The initial state of the scaled span program P^β is (see Theorem 10):

$$|w_0^\beta\rangle=\frac{\beta}{\beta^2+\frac{1}{n}}|w_0\rangle+\frac{\frac{1}{n}}{\beta^2+\frac{1}{n}}|\hat{0}\rangle+\frac{\beta}{\sqrt{\beta^2+\frac{1}{n}}}|\hat{1}\rangle,$$

which we can also construct efficiently, as follows:

$$|\hat{0}\rangle\mapsto\frac{\beta\sqrt{n}}{\beta^2+\frac{1}{n}}|\hat{2}\rangle+\frac{1}{n\beta^2+1}|\hat{0}\rangle+\frac{\beta}{\sqrt{\beta^2+\frac{1}{n}}}|\hat{1}\rangle\mapsto\frac{\beta}{\beta^2+\frac{1}{n}}|w_0\rangle+\frac{\frac{1}{n}}{\beta^2+\frac{1}{n}}|\hat{0}\rangle+\frac{\beta}{\sqrt{\beta^2+\frac{1}{n}}}|\hat{1}\rangle.$$

The first step is accomplished by a pair of rotations using $O(\log\frac{n}{\beta})$ elementary gates, and the second is accomplished by mapping $|\hat{2}\rangle$ to $\frac{|w_0\rangle}{\||w_0\rangle\|}=\sqrt{n}|w_0\rangle$, which can be accomplished in $O(\log n)$ elementary gates.

Next, we have $A^\beta=\beta A+(|s\rangle-|t\rangle)\langle\hat{0}|+\frac{\sqrt{\beta^2+\frac{n}{2}}}{\beta}|\hat{1}\rangle\langle\hat{1}|$, so

$$\ker A^\beta\oplus\mathrm{span}\{|w_0^\beta\rangle\}=\ker A\oplus\mathrm{span}\{|\hat{0}\rangle-\frac{1}{\beta}|w_0\rangle\}\oplus\mathrm{span}\{|w_0^\beta\rangle\}.$$

We know how to reflect about $\ker A$, and since we can efficiently construct $|w_0^\beta\rangle$, we can reflect about it, so we need only consider how to reflect about $\operatorname{span}\{|\hat{0}\rangle - \frac{1}{\beta}|w_0\rangle\}$. Since we can compute $|w_0\rangle$ efficiently, we can compute:

$$|\hat{0}\rangle \mapsto \frac{\beta}{\sqrt{\beta^2+1}}|\hat{0}\rangle + \frac{1}{\sqrt{\beta^2+1}}|\hat{1}\rangle \mapsto \frac{\beta}{\sqrt{\beta^2+1}}|\hat{0}\rangle + \frac{1}{\sqrt{\beta^2+1}}|\bar{w}_0\rangle.$$

The first step is a rotation, which can be performed in $O(\log \frac{1}{\beta})$ elementary gates, and the second step is some mapping that maps $|\hat{1}\rangle$ to $|w_0\rangle$, which we know can be done in $O(\log n)$ elementary gates. Thus, the total cost to reflect about $\ker A^\beta$ is $O(\log n)$. □

References

1. Bennett, C.H., Bernstein, E., Brassard, G., Vazirani, U.: Strengths and weaknesses of quantum computing. SIAM J. Comput. (special issue on quantum computing) **26**, 1510–1523 (1997). arXiv:quant-ph/9701001v1
2. Beals, R., Buhrman, H., Cleve, R., Mosca, M., de Wolf, R.: Quantum lower bounds by polynomials. J. ACM **48**, 778–797 (2001)
3. Belovs, A., Childs, A.M., Jeffery, S., Kothari, R., Magniez, F.: Time efficient quantum walks for 3-distinctness. In: Proceedings of the 40th International Colloquium on Automata, Languages and Programming (ICALP 2013), pp. 105–122 (2013)
4. Belovs, A.: Learning-graph-based quantum algorithm for k-distinctness. In: Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2012), pp. 207–216 (2012)
5. Belovs, A.: Span programs for functions with constant-sized 1-certificates. In: Proceedings of the 44th Symposium on Theory of Computing (STOC 2012), pp. 77–84 (2012)
6. Brassard, G., Høyer, P., Mosca, M., Tapp, A.: Quantum amplitude amplification and estimation. In: Lomonaca, S.J., Brandt, H.E. (eds.) Quantum Computation and Quantum Information: A Millennium Volume, volume 305 of AMS Contemporary Mathematics Series Millennium Volume, pp. 53–74. AMS (2002). arXiv:quant-ph/0005055v1
7. Belovs, A., Reichardt, B.: Span programs and quantum algorithms for st-connectivity and claw detection. In: Proceedings of the 20th European Symposium on Algorithms (ESA 2012), pp. 193–204 (2012)
8. Cleve, R., Ekert, A., Macchiavello, C., Mosca, M.: Quantum algorithms revisited. Proc. Royal Soc. A Math. Phys. Eng. Sci. **454**(1969), 339–354 (1998)
9. Chandra, A.K., Raghavan, P., Ruzzo, W.L., Smolensky, R., Tiwari, P.: The electrical resistance of a graph captures its commute and cover times. Comput. Complex. **6**(4), 312–340 (1996)
10. Doyle, P.G., Snell, J.L.: Random Walks and Electrical Networks, volume 22 of the Carus Mathematical Monographs. The Mathematical Association of America, Washington (1984)
11. Grover, L.K.: A fast quantum mechanical algorithm for database search. In: Proceedings of the 28th ACM Symposium on Theory of Computing (STOC 1996), pp. 212–219 (1996)
12. Harrow, A.W., Hassidim, A., Lloyd, S.: Quantum algorithm for linear systems of equations. Phys. Rev. Lett. **103**, 150502 (2009)
13. Jeffery, S.: Frameworks for Quantum Algorithms. Ph.D. thesis, University of Waterloo (2014). http://uwspace.uwaterloo.ca/handle/10012/8710. Accessed 01 Jan 2015
14. Kitaev, A.: Quantum measurements and the Abelian stabilizer problem (1995). arXiv:quant-ph/9511026
15. Karchmer, M., Wigderson, A.: On span programs. In: Proceedings of the IEEE 8th Annual Conference on Structure in Complexity Theory, pp. 102–111 (1993)
16. Lee, T., Mittal, R., Reichardt, B., Špalek, R., Szegedy, M.: Quantum query complexity of state conversion. In: Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2011), pp. 344–353 (2011)

17. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov Chains and Mixing Times. American Mathematical Society, Providence (2009)
18. Montanaro, A., de Wolf, R.: A survey of quantum property testing (2013). arXiv:1310.2035
19. Reichardt, B.: Span programs and quantum query complexity: the general adversary bound is nearly tight for every Boolean function. In: Proceedings of the 50th IEEE Symposium on Foundations of Computer Science (FOCS 2009), pp. 544–551 (2009). arXiv:0904.2759 [quant-ph]
20. Reichardt, B.: Reflections for quantum query algorithms. In: Proceedings of the 22nd ACM-SIAM Symposium on Discrete Algorithms (SODA 2011), pp. 560–569 (2011)
21. Reichardt, B., Špalek, R.: Span-program-based quantum algorithm for evaluating formulas. Theory Comput. **8**(13), 291–319 (2012)
22. Szegedy, M.: Quantum speed-up of Markov chain based algorithms. In: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2004), pp. 32–41 (2004)
23. Wang, G.: Quantum algorithms for approximating the effective resistances in electrical networks (2013). arXiv:1311.1851