Search via Quantum Walk*

Frédéric Magniez† Ashwin Nayak‡ Jérémie Roland§ Miklos Santha¶

February 11, 2011

Abstract

We propose a new method for designing quantum search algorithms for finding a "marked" element in the state space of a classical Markov chain. The algorithm is based on a quantum walk *à la* Szegedy (2004) that is defined in terms of the Markov chain. The main new idea is to apply quantum phase estimation to the quantum walk in order to implement an approximate reflection operator. This operator is then used in an amplitude amplification scheme. As a result we considerably expand the scope of the previous approaches of Ambainis (2004) and Szegedy (2004). Our algorithm combines the benefits of these approaches in terms of being able to find marked elements, incurring the smaller cost of the two, and being applicable to a larger class of Markov chains. In addition, it is conceptually simple and avoids some technical difficulties in the previous analyses of several algorithms based on quantum walk.

1 Introduction

1.1 Background

At an abstract level, many search problems may be cast as the problem of finding a "marked" element from a set X with n elements. Let $M \subseteq X$ be the set of the so called marked elements. One approach to finding an element of M, if it is not empty, is to repeatedly sample from X uniformly until a marked element is picked. A more cost-effective approach reuses resources expended in generating the first sample (time, random bits, black-box queries, etc.) by simulating the steps of a Markov chain with state space X to generate the next sample. This approach often takes advantage of some structure present in the ground set X and the Markov chain, and leads to a more efficient algorithm. In this article, we study quantum analogues of this randomized scheme.

There are several ways of defining quantum analogues of Markov chains, including both discrete and continuous time versions (see, for example, Ref. [29] for a detailed introduction). We restrict our attention to discrete time analogues.

Discrete time quantum walks emerged gradually in the field of quantum algorithms. On the line they are related to the quantum cellular automaton model of Meyer [25]. Watrous [32] introduced quantum walks on regular graphs, and used them to show that randomized logarithmic space is included in quantum logarithmic space. Afterwards notions related to quantum walks, such as mixing time, and deviation from the starting state, were studied for restricted graphs by several researchers [26, 4, 2, 27], suggesting the possibility of speed-up of classical algorithms based on random walk.

*A preliminary version of this work appeared in *Proceedings of 39th ACM Symposium on Theory of Computing*, pages 575–584, 2007.

†LIAFA, Univ. Paris 7, CNRS, F-75205 Paris, France. magniez@liafa.jussieu.fr

‡Department of Combinatorics and Optimization, and Institute for Quantum Computing, University of Waterloo; and Perimeter Institute for Theoretical Physics; 200 University Ave. W., Waterloo, ON, N2L 3G1, Canada. Email: ashwin.nayak@uwaterloo.ca.

§NEC Laboratories America; Princeton, NJ 08540, USA. jroland@nec-labs.com

¶LIAFA, Univ. Paris 7, CNRS, F-75205 Paris, France, and Centre for Quantum Technologies, National University of Singapore, Singapore 117543. santha@lri.fr

Shenvi, Kempe, and Whaley [28] pointed out the algorithmic potential of quantum walks by designing a walk based algorithm to emulate Grover Search [16]. The first algorithm using quantum walks that goes beyond the capability of Grover Search is due to Ambainis [6] for Element Distinctness. In his seminal paper he resolved the quantum query complexity of the problem, settling a difficult question that had been open for several years [12, 1]. Finally Szegedy [29] developed a theory of quantum walk based algorithms. He designed a quantum search algorithm based on any symmetric, ergodic Markov chain that detects the presence of a marked element. He defined a notion of quantum hitting time that is quadratically smaller than the classical average hitting time. Since then, in the framework of Ambainis or Szegedy, many new algorithms with substantially better complexity emerged in a variety of contexts [5, 24, 13, 22, 15].

This work develops a new schema for quantum search algorithms, based on any ergodic Markov chain. We adapt the quantum analogue of classical Markov chains due to Szegedy to possibly non-symmetric Markov chains, but use it more in the style of the Ambainis algorithm. Departing from the two algorithms, however, we use quantum walks only indirectly. In conjunction with the well known phase estimation algorithm [19, 20, 14], the quantum walk helps us implement an approximate reflection operator. This operator may then be used within amplitude amplification algorithms [16, 10, 17] for search. As a result, our work generalizes previous ones by extending the class of possible Markov chains, and improving the complexity in terms of its relation with the eigenvalue or singular value gap of the related Markov chain. In addition, our approach is conceptually simple, avoids several technical difficulties in the analysis of the earlier approaches, and leads to improvements in various aspects of the algorithms.

1.2 Two subtly different search algorithms

We identify a Markov chain over state space X with its transition matrix $P = (p_{xy})_{x,y\in X}$, where p_{xy} is the probability of transition from x to y. A chain is *irreducible* if every state is reachable from every other state, and an irreducible chain is *ergodic* if it is also aperiodic (equivalently, its reachability graph is non-bipartite). The eigenvalues of a Markov chain are at most 1 in magnitude. By the Perron-Frobenius theorem, an irreducible chain has a unique stationary distribution $\pi = (\pi_x)$, that is, a unique left eigenvector π with eigenvalue 1 and positive coordinates summing up to 1. If the chain is ergodic, the eigenvalue 1 is the only eigenvalue of P with magnitude 1. We denote by $\delta = \delta(P)$ the *eigenvalue gap* of P, that is $1 - \lambda$, where $\lambda = \lambda(P) = \max_{\nu\in\Lambda} |\nu|$, where Λ is the set of eigenvalues of P different from 1. The *time-reversed Markov chain* $P^* = (p^*_{xy})$ of P is defined by the equations $\pi_x p_{xy} = \pi_y p^*_{yx}$. The chain P is said to be *reversible* if $P^* = P$. The Markov chain P is *symmetric* if $P = P^\mathsf{T}$ where P^T denotes the matrix transpose of P. The stationary distribution of any symmetric chain is the uniform distribution.

The optimal quantum algorithm for Element Distinctness discovered by Ambainis [6] recasts the problem in terms of search for a marked state in a Johnson graph defined by the problem instance. The algorithm may be viewed as a quantum analogue of the following search process, where P is a Markov chain defined on state space X.

Search Algorithm 1

1. Initialize x to a state sampled from a probability distribution s over X.
2. Repeat for t_2 steps
   (a) If the state y reached in the previous step is marked, then stop and output y.
   (b) Else, simulate t_1 steps of the Markov chain P starting with the current state y.
3. If the algorithm has not terminated, stop, and output 'no marked element exists'.

The parameters t_1 and t_2 in the algorithm are determined by the properties of the Markov chain and the marked subset M. The idea behind this algorithm is illustrated by considering an ergodic Markov chain P. When t_1 is large enough, the state y in step (2a) above is distributed (approximately) according to the stationary distribution of P. Then, the outer loop represents sampling from the stationary distribution until

a marked element is found. When t_2 is chosen to be inversely proportional to the probability that a state is marked according to the stationary distribution, the algorithm succeeds with high probability.

The analysis of the Ambainis quantum algorithm depends heavily on the form of marked states, and was presented for subsets M arising out of k-Collision, a generalization of Element Distinctness, with the assumption of a unique collision. Inspired by this algorithm, Szegedy [29] designed a quantum search algorithm with uniform initial distribution, based on any symmetric, ergodic Markov chain. The Szegedy algorithm may be viewed as a quantum analogue of a subtly different, but more natural, classical process.

Search Algorithm 2

1. Initialize x to a state sampled from a probability distribution s over X.
2. Repeat for t steps
   (a) If the state y reached in the previous step is marked, then stop and output y.
   (b) Else, simulate *one* step of the Markov chain P from the current state y.
3. If the algorithm has not terminated, stop, and output 'no marked element exists'.

The parameter t is also determined by the Markov chain P, and the set M of marked states. This algorithm is a greedy version of the first algorithm: a check is performed after every step of the Markov chain to determine if a marked state has been reached, irrespective of whether the Markov chain has mixed.

Let us formally derive the complexity of the two algorithms to clarify their differences. Assume that the search algorithms maintain a data structure d that associates some data $d(x)$ with every state $x \in X$. From $d(x)$, we would like to determine if $x \in M$. When operating with d, we distinguish three types of cost.

Set-up cost S**:** The cost of sampling $x \in X$ according to the initial distribution s and of constructing the data structure $d(x)$ for the state x.

Update cost U**:** The cost of simulating a transition from x to y for a state $x \in X$ according to the Markov chain P and of updating $d(x)$ to $d(y)$.

Checking cost C**:** The cost of checking if $x \in M$ using $d(x)$.

These costs may be thought of as vectors listing all the measures of complexity of interest, such as query and time complexity. We may now state generic bounds on the efficiency of the two search algorithms in terms of our cost parameters. Note that throughout this paper, we say that an event happens *with high probability* if it happens with probability at least some universal constant. All the search algorithms (classical and quantum) we discuss have one-sided error. The algorithms may fail with some probability to report any marked element even when they exist. This error probability may be driven down to the desired level in the standard manner by sequential repetition of the algorithms.

Proposition 1. *Let $\delta > 0$ be the eigenvalue gap of an ergodic, symmetric Markov chain P on a state space X of size n, and let $\frac{|M|}{|X|} \geq \varepsilon > 0$ whenever $M \subset X$ is non-empty. For the uniform initial distribution s,*

1. ***Search Algorithm 1*** *determines if a marked element exists and finds one such element with high probability if $t_1 \in \mathrm{O}(\frac{1}{\delta})$ and $t_2 \in \mathrm{O}(\frac{1}{\varepsilon})$ are chosen to be suitably large. The cost incurred is of order $\mathrm{S} + \frac{1}{\varepsilon}\left(\frac{1}{\delta}\mathrm{U} + \mathrm{C}\right)$.*
2. ***Search Algorithm 2*** *determines if a marked element exists and finds one such element with high probability if $t \in \mathrm{O}(\frac{1}{\delta\varepsilon})$ is chosen to be suitably large. The cost incurred is of order $\mathrm{S} + \frac{1}{\delta\varepsilon}\left(\mathrm{U} + \mathrm{C}\right)$.*

Proof. The stopping time of **Search Algorithm 2** is the average hitting time of the set M for the Markov chain P. We may therefore take t to be a constant factor more than this hitting time. As mentioned before, this time is bounded above by the stopping time for the first algorithm. Therefore part 2 of the proposition follows from part 1.

In the first algorithm, we may take t_2 to be proportional to the average hitting time of the set M for the Markov chain P^{t_1}. The quantity $\lambda(P)$ is bounded by $1-\delta$ by hypothesis. The analogous quantity $\lambda(P^{t_1})$ is therefore bounded by $(1-\delta)^{t_1} \leq \mathrm{e}^{-\delta t_1}$. Taking $t_1 = 1/\delta$, we get a spectral gap $\tilde{\delta}$ of at least $1/2$ for P^{t_1}. We may now bound the average hitting time of M for P^{t_1} by, for example, Equation (15) in Ref. [29] and Lemma 1 in Ref. [11] (also stated as Lemma 10 in Ref. [29]). This bound evaluates to $\frac{1}{\varepsilon\tilde{\delta}} \leq \frac{2}{\varepsilon}$. The expression for the cost of the algorithm now follows. □

For special classes of graphs, for example for the 2-d toroidal grid, the hitting time may be significantly smaller than the generic bound $t = \mathrm{O}(1/\delta\varepsilon)$ given in part 2 (see Ref. [3, Page 11, Chapter 5]).

1.3 Quantum analogues

As in the classical case, the quantum search algorithms look for a marked element in a finite set X, where a data structure d is maintained during the algorithm. Let X_d be the set of items along with their associated data, that is $X_d = \{(x, d(x)) : x \in X\}$. For convenience we suppose that $\bar{0} \in X$ and that $d(\bar{0}) = \bar{0}$.

The quantum walks due to Ambainis and Szegedy, as in our work, may be thought of as walks on *edges* of the original Markov chain, rather than its vertices. Thus, the associated state space is a linear subspace of the vector space $\mathcal{H} = \mathbb{C}^{X\times X}$, or $\mathcal{H}_d = \mathbb{C}^{X_d \times X_d}$ when we also include the data structure. For the sake of elegance in the mathematical analyses, our data structure keeps the data for both vertices of an edge, whereas in previous works the data was kept only for one of them.

There is a natural isomorphism $|\psi\rangle \mapsto |\psi\rangle_d$ between $\mathcal{H}$ and $\mathcal{H}_d$, where on basis states $|x\rangle_d = |x, d(x)\rangle$. This isomorphism maps a unitary operation U on $\mathcal{H}$ into U_d on $\mathcal{H}_d$ defined by $U_d|\psi\rangle_d = (U|\psi\rangle)_d$. Our walks are discussed in the space $\mathcal{H}_d$ when, for implementation and cost considerations, it is important to properly deal with the data structure. However, for convenience, we analyze the mathematical properties of the walks without the data structure, in the space $\mathcal{H}$. This is justified by the isomorphism between $\mathcal{H}_d$ and $\mathcal{H}$.

The initial state of the algorithm is explicitly related to the stationary distribution π of P. At each step, the right end-point of an edge (x, y) is "mixed" over the neighbors of x, and then the left end-point is mixed over the neighbors of the new right end-point. We again distinguish three types of cost generalizing those of the classical search. They are of the same order as the corresponding costs in the algorithms of Ambainis and Szegedy. Some operations of the algorithms not entering into these costs are not taken into account. This is justified by the fact that in all quantum search algorithms the overall complexity is of the order of the accounted part, which is expressed in terms of the costs below.

(Quantum) Set-up cost S**:** The cost of constructing the state $\sum_x \sqrt{\pi_x}|x\rangle_d|\bar{0}\rangle_d$ from $|\bar{0}\rangle_d|\bar{0}\rangle_d$.

(Quantum) Update cost U**:** The cost of realizing any of the unitary transformations

$$\begin{aligned} |x\rangle_d|\bar{0}\rangle_d &\mapsto |x\rangle_d \textstyle\sum_y \sqrt{p_{xy}}|y\rangle_d, \\ |\bar{0}\rangle_d|y\rangle_d &\mapsto \textstyle\sum_x \sqrt{p^*_{yx}}|x\rangle_d|y\rangle_d, \end{aligned}$$

and their inverses, where $P^* = (p^*_{xy})$ is the time-reversed Markov chain defined in Section 1.2.

(Quantum) Checking cost C**:** The cost of realizing the following conditional phase flip

$$|x\rangle_d|y\rangle_d \quad \mapsto \quad \begin{cases} -|x\rangle_d|y\rangle_d & \text{if } x \in M, \\ |x\rangle_d|y\rangle_d & \text{otherwise.} \end{cases}$$

The quantum search algorithms due to Ambainis and Szegedy give a quadratic speed up in the times t_1, t_2 and t, with respect to the classical algorithms. Let us recall that for integers $0 < r < m$ and $0 < l < r$ the vertices of the Johnson graph with parameters m, r, l are the subsets of size r of a universe of size m, and there is an edge between two vertices if the size of their intersection is l. The eigenvalue gap δ of the symmetric walk on the Johnson graph with $l = r-1$, and $r < m/2$ is in $\Theta(1/r)$. If the set of marked vertices consists of vertices that contain a fixed subset of constant size $k \leq r$, then their fraction ε is in $\Omega(\frac{r^k}{m^k})$.

Theorem 1 (Ambainis [6]). *Let P be the random walk on the Johnson graph on r-subsets of a universe of size m, where $r = o(m)$, and with intersection size $r-1$. Let M be either empty, or the class of all r-subsets that contain a fixed subset of constant size $k \leq r$. Then, there is a quantum algorithm that with high probability, determines if M is empty or finds the k-subset, with cost of order $\mathsf{S} + \frac{1}{\sqrt{\varepsilon}}(\frac{1}{\sqrt{\delta}}\mathsf{U} + \mathsf{C})$.*

Theorem 2 (Szegedy [29]). *Let $\delta > 0$ be the eigenvalue gap of an ergodic, symmetric Markov chain P, and let $\frac{|M|}{|X|} \geq \varepsilon > 0$ whenever M is non-empty. There exists a quantum algorithm that determines, with high probability, if M is non-empty with cost of order $\mathsf{S} + \frac{1}{\sqrt{\delta\varepsilon}}(\mathsf{U} + \mathsf{C})$.*

If the checking cost C is substantially greater than that of performing one step of the walk, an algorithm with the cost structure of the Ambainis algorithm would be more efficient. Moreover, the algorithm would *find* a marked element if one exists. These advantages are illustrated by the algorithm for Triangle Finding [24]. This algorithm uses two quantum walks *à la* Ambainis recursively; the Szegedy framework seems to give a less efficient algorithm. Nonetheless, the Szegedy approach has other advantages—it applies to a wider class of Markov chains and for arbitrary sets of marked states. Moreover, the quantity $1/\sqrt{\delta\epsilon}$ in Theorem 2 may be replaced by the square-root of the classical hitting time [29]. These features make it more suitable for applications such as the near-optimal algorithm for Group Commutativity [22] which has no equivalent using the Ambainis approach.

1.4 Contribution, relation with prior work, and organization

We present an algorithm that is a quantum analogue of **Search Algorithm 1** and works for any ergodic Markov chain. It is most easily described for *reversible* Markov chains.

Theorem 3. *Let $\delta > 0$ be the eigenvalue gap of a reversible, ergodic Markov chain P, and let $\varepsilon > 0$ be a lower bound on the probability that an element chosen from the stationary distribution of P is marked whenever M is non-empty. Then, there is a quantum algorithm that with high probability, determines if M is empty or finds an element of M, with cost of order $\mathsf{S} + \frac{1}{\sqrt{\varepsilon}}(\frac{1}{\sqrt{\delta}}\mathsf{U} + \mathsf{C})$.*

This algorithm considerably expands the scope of the approaches embodied in Theorems 1 and 2 above. It combines the benefits of the two approaches in terms of being able to find marked elements, incurring the smaller cost of the two, and being applicable to a larger class of Markov chains. In addition, it is conceptually simple, avoids several technical difficulties in the analysis of the earlier approaches, and leads to improvements in various aspects of algorithms for Element Distinctness, Matrix Product Verification, Triangle Finding, and Group Commutativity. Namely, we give a single-shot method for any algorithm *à la* Ambainis in presence of multiple solutions, without the need for a reduction to special cases such as that of a unique solution. This applies to Element Distinctness and Triangle Finding. For Element Distinctness, Matrix Product Verification, and Group Commutativity, where an algorithm *à la* Szegedy only detects the existence of a solution, we find one with the same time and query complexity. Finally, we improve the query complexity of the best previously known algorithm for Triangle Finding by a polylog(n) factor.

In Section 2, we describe a quantum analogue of a Markov chain based on the work of Szegedy [29] who defined such a quantum process $W(P,Q)$ for a classical bipartite walk (P,Q). By letting $Q = P$, he related the spectrum of the quantum walk $W(P)$ to that of P for symmetric Markov chains. Using an absorbing version of P as in **Search Algorithm 2**, he designed a quantum analogue of this classical scheme. Even when P is not symmetric, letting $Q = P^*$, the time-reversed Markov chain corresponding to P, leads to a natural connection between P and $W(P)$. If P is reversible, then the eigenvalues of $W(P)$ are closely related to those of P, as in the symmetric case. For an arbitrary, possibly non-reversible, ergodic Markov chain, this connection relates the eigenvalues of $W(P)$ to the singular values of a "discriminant" matrix $D(P)$ associated with P.

In Section 3, we use the quantum walk $W(P)$ associated with the unperturbed walk P in a completely different way, more in the style of the Ambainis approach. Ambainis directly uses a power of $W(P)$ to replace the "diffusion" operator in the Grover search algorithm. The beauty of this step, and the difficulty of proving its correctness, lies in the fact that even if no power of $W(P)$ closely approximates the diffusion

operator, some powers have sufficient properties to mimic its essential features (see Lemma 3 in Ref. [6]). While this lemma is sufficient to prove Theorem 1, it alone is not powerful enough to imply Theorem 3. The spectral gap of classical Markov chains and that of some special cases of quantum walks (such as the quantum walk on Johnson graphs proposed by Ambainis) may be amplified by sequential repetition. Nevertheless, this method and its obvious variants break down when we consider the walk $W(P)$ for arbitrary chains P, and arbitrary sets of marked elements. Instead, we introduce a novel way to approximate the diffusion operator. Our approach is both conceptually simpler, and more general. We observe that $W(P)$ amplifies the spectral gap of a reversible Markov chain quadratically. We translate this to an efficient approximation to the Grover diffusion operator (**Theorem 6**), using the well known phase estimation algorithm. We then begin an exposition of our algorithm by considering reversible Markov chains. To explain the basic idea of our approach, we first prove our main result with an additional logarithmic factor (**Theorem 7**).

In Section 4, using a technique developed by Høyer, Mosca, and de Wolf [17] we show how to eliminate the logarithmic factor in the previous theorem, thus proving **Theorem 3**.

In Section 5, we extend the algorithm to a possibly non-reversible Markov chain whose discriminant has non-zero singular value gap (**Theorem 8**). The complexity of the algorithm in the general case is similar to the one for reversible Markov chains. The sole difference is that the singular value gap of the discriminant matrix $D(P)$ takes the place of the spectral gap of P. While the eigenvalues of Markov chains are well studied, we are not aware of a similar theory for singular values of this matrix. Nonetheless, such a general result may prove useful for future applications.

1.5 Subsequent work

Since our work first appeared, much progress has been made on a question we had left unresolved. For any symmetric Markov chain, Szegedy [29] gave a procedure that detects the existence of marked elements in time of the order of the square-root of the classical hitting time. This result does not carry over to the potentially harder problem of *finding* a marked element. The latter (finding) problem has received particular attention in the case of the $\sqrt{N} \times \sqrt{N}$ grid. The classical hitting time for this graph is in $\mathrm{O}(N \log N)$ for any (non-zero) number of marked elements. Algorithms due to Ambainis, Kempe, and Rivosh [5] and Szegedy [29] find a unique marked state in time $\mathrm{O}(\sqrt{N} \log N)$, a $\sqrt{\log N}$ factor larger than the detection time. In a recent paper, Tulsi [30] finally shows how we may find a unique marked element in time $\mathrm{O}(\sqrt{N \log N})$.

Magniez, Nayak, Richter, and Santha [23] define new, Monte Carlo type classical and quantum hitting times that are potentially smaller than the existing notion of (Las Vegas type) hitting times. They also present new quantum algorithms for the detection and finding problems whose complexities are related to the Monte Carlo quantum hitting time. The detection algorithm is based on phase estimation, and the finding algorithm combines a similar phase estimation based procedure with an idea introduced by Tulsi. Extending Tulsi's result for the 2D grid, they show that for any state-transitive Markov chain with a unique marked state, the quantum hitting time is of the same order for both the detection and finding problems. Krovi, Magniez, Ozols, and Roland [21] make a significant improvement to this result by presenting a quantum algorithm for finding multiple marked elements in any reversible Markov chain. Taking a new, simpler, and more general approach, they introduce a notion of interpolation between any reversible chain and a perturbed version of this chain, in which the marked states are absorbing. The quantum analogue of the interpolated walk not only detects but also finds marked states with a quadratic speed-up over the classical hitting time.

2 Quantum analogue of a classical Markov chain

Let $P = (p_{xy})$ be the transition matrix of any irreducible Markov chain on a finite space X with $|X| = n$. We define a quantum analogue of P, based on and extending the notion of quantum Markov chain due to Szegedy [29]. The latter was inspired by an earlier notion of quantum walk due to Ambainis [6]. We also point out that a similar process on regular graphs was studied by Watrous [32]. Recall that P^* denotes the time-reversed Markov chain of P.

For a state $|\psi\rangle \in \mathcal{H}$, let $\Pi_\psi = |\psi\rangle\langle\psi|$ denote the orthogonal projector onto $\mathsf{Span}(|\psi\rangle)$, and let $\mathrm{ref}(\psi) = 2\Pi_\psi - \mathrm{Id}$ denote the reflection through the line generated by $|\psi\rangle$, where Id is the identity operator on $\mathcal{H}$. For a subspace $\mathcal{K}$ of $\mathcal{H}$ spanned by a set of mutually orthogonal states $\{|\psi_i\rangle : i \in I\}$, let $\Pi_\mathcal{K} = \sum_{i\in I} \Pi_{\psi_i}$ be the orthogonal projector onto $\mathcal{K}$, and let $\mathrm{ref}(\mathcal{K}) = 2\Pi_\mathcal{K} - \mathrm{Id}$ be the reflection through $\mathcal{K}$.

Let $\mathcal{A} = \mathsf{Span}(|x\rangle|p_x\rangle : x \in X)$ and $\mathcal{B} = \mathsf{Span}(|p_y^*\rangle|y\rangle : y \in X)$ be vector subspaces of $\mathcal{H} = \mathbb{C}^{X\times X}$, where

$$|p_x\rangle \quad = \quad \sum_{y\in X} \sqrt{p_{xy}}\,|y\rangle \qquad \text{and} \qquad |p_y^*\rangle \quad = \quad \sum_{x\in X} \sqrt{p_{yx}^*}\,|x\rangle.$$

Definition 1 (Quantum walk)**.** *The unitary operation $W(P)$ defined on $\mathcal{H}$ by $W(P) = \mathrm{ref}(\mathcal{B}) \cdot \mathrm{ref}(\mathcal{A})$ is called the* quantum walk *based on the classical chain P.*

This quantum walk extends to a walk $W(P)_d$ on the space $\mathcal{H}$ augmented with data structures, as explained in Section 1.3. Recall that U is the quantum update cost as defined in the same section.

Proposition 2. *The quantum walk with data, $W(P)_d$, can be implemented at cost $4\mathsf{U}$.*

Proof. Recall that $W(P)_d = \mathrm{ref}(\mathcal{B})_d \cdot \mathrm{ref}(\mathcal{A})_d$. The reflection $\mathrm{ref}(\mathcal{A})_d$ is implemented by mapping states $|x\rangle_d|p_x\rangle_d$ to $|x\rangle_d|\bar{0}\rangle_d$, applying $\mathrm{ref}(|\bar{0}\rangle_d)$ on the second register, and inverting the first transformation. While the first and last steps each have cost U, we only charge unit cost for the second step since it does not depend on the data structure ($|\bar{0}\rangle_d = |\bar{0},\bar{0}\rangle$ by definition). Therefore the implementation of $\mathrm{ref}(\mathcal{A})_d$ is of cost $2\mathsf{U}$. The reflection $\mathrm{ref}(\mathcal{B})_d$ may be implemented similarly. □

The eigen-spectrum of the transition matrix P plays an important role in the analysis of a classical Markov chain. Similarly, the behavior of the quantum process $W(P)$ may be inferred from its spectral decomposition. We consider the *discriminant* matrix $D(P) = (\sqrt{p_{xy}p_{yx}^*})$. Since $\sqrt{p_{xy}p_{yx}^*} = \sqrt{\pi_x}p_{xy}/\sqrt{\pi_y}$, the discriminant matrix is equal to

$$D(P) \quad = \quad \mathrm{diag}(\pi)^{1/2} \cdot P \cdot \mathrm{diag}(\pi)^{-1/2},$$

where $\mathrm{diag}(\pi)$ is the invertible diagonal matrix with the coordinates of the distribution π in its diagonal. Since the singular values of $D(P)$ all lie in the range $[0,1]$, we may express them as $\cos\theta$, for some angles $\theta \in [0, \frac{\pi}{2}]$. (Note that this is a second type of use of the Greek letter 'π' in this article, and it denotes the usual Mathematical constant. A third type of use occurs later in the article. The meaning of the letter can be inferred from the context in which it is used.) For later reference, we rewrite Theorem 1 due to Szegedy [29] which relates the singular value decomposition of $D(P)$ to the spectral decomposition of $W(P)$. This theorem is a variant of a result due to Jordan [18] (see also Ref. [8, Section VII.1, page 201]), and may be derived from it.

Theorem 4 (Szegedy [29])**.** *Let P be an irreducible Markov chain, and let $\cos\theta_1, \ldots, \cos\theta_l$ be an enumeration of those singular values (possibly repeated) of $D(P)$ that lie in the open interval $(0,1)$. Then:*

1. *On $\mathcal{A}+\mathcal{B}$ those eigenvalues of $W(P)$ that have non-zero imaginary part are exactly $\mathrm{e}^{\pm 2\mathrm{i}\theta_1}, \ldots, \mathrm{e}^{\pm 2\mathrm{i}\theta_l}$, with the same multiplicity.*

2. *On $\mathcal{A}\cap\mathcal{B}$ the operator $W(P)$ acts as the identity* Id*. The linear subspace $\mathcal{A}\cap\mathcal{B}$ is spanned by the left (and right) singular vectors of $D(P)$ with singular value* 1*.*

3. *On $\mathcal{A}\cap\mathcal{B}^\perp$ and $\mathcal{A}^\perp\cap\mathcal{B}$ the operator $W(P)$ acts as $-$*Id*. The linear subspace $\mathcal{A}\cap\mathcal{B}^\perp$ (respectively, $\mathcal{A}^\perp\cap\mathcal{B}$) is spanned by the set of left (respectively, right) singular vectors of $D(P)$ with singular value* 0*.*

4. *$W(P)$ has no other eigenvalues on $\mathcal{A}+\mathcal{B}$; on $\mathcal{A}^\perp\cap\mathcal{B}^\perp$ the operator $W(P)$ acts as* Id*.*

We define $\Delta(P)$, the *phase gap* of $W(P)$ as 2θ, where θ is the smallest angle in $(0, \frac{\pi}{2})$ such that $\cos\theta$ is a singular value of $D(P)$. This definition is motivated by the previous theorem: in the complex plane, the angular distance of 1 from any other eigenvalue is at least $\Delta(P)$.

3 From quantum walk to search

3.1 Outline of search algorithm

We now describe a search algorithm that may be viewed as a quantum analogue of **Search Algorithm 1** of Section 1.2. Consider the following quantum state in the Hilbert space $\mathcal{H}$:

$$|\pi\rangle \quad = \quad \sum_{x\in X} \sqrt{\pi_x}\,|x\rangle|p_x\rangle \quad = \quad \sum_{y\in X} \sqrt{\pi_y}\,|p^*_y\rangle|y\rangle.$$

(Note that its use as a label for the quantum state above is the third type of use of the letter 'π' in this article.) This state serves as the initial state for our algorithm, and corresponds to starting in the stationary distribution π in the classical search algorithms. Taking into account the data structure, preparing $|\pi\rangle_d$ from $|\bar{0}\rangle_d|\bar{0}\rangle_d$ has cost $\mathsf{S}+\mathsf{U}$ as it requires one set-up operation to prepare $\sum_{x\in X}\sqrt{\pi_x}\,|x\rangle_d|\bar{0}\rangle_d$, followed by one update operation to map this state to $|\pi\rangle_d$. Assume that $M \neq \emptyset$. Let $\mathcal{M} = \mathbb{C}^{M\times X}$ denote the subspace with marked items in the first register. We would like to transform the initial state $|\pi\rangle$ to the target state $|\mu\rangle$, which is the (normalized) projection of $|\pi\rangle$ onto the "marked subspace" $\mathcal{M}$:

$$|\mu\rangle \quad = \quad \frac{\Pi_{\mathcal{M}}|\pi\rangle}{\|\Pi_{\mathcal{M}}|\pi\rangle\|} \quad = \quad \frac{1}{\sqrt{p_M}}\sum_{x\in M}\sqrt{\pi_x}\,|x\rangle|p_x\rangle,$$

where $p_M = \|\Pi_{\mathcal{M}}|\pi\rangle\|^2 = \sum_{x\in M}\pi_x$ is the probability of a set M of marked states under the stationary distribution π. Roughly speaking, we effect this transformation by implementing a rotation *à la* Grover [16] in the two-dimensional real subspace $\mathcal{S} = \mathsf{Span}(|\pi\rangle,\ |\mu\rangle)$ generated by the states.

Ideally, we would like to effect the rotation $\mathrm{ref}(\pi)_d\cdot\mathrm{ref}(\mu^\perp)_d$ in $\mathcal{S}_d$, where $|\mu^\perp\rangle$ is the state in $\mathcal{S}$ orthogonal to $|\mu\rangle$ which makes an acute angle with $|\pi\rangle$. The angle φ between $|\pi\rangle$ and $|\mu^\perp\rangle$ is given by $\sin\varphi = \langle\mu|\pi\rangle = \sqrt{p_M}$. The product of the two reflections above is a rotation by an angle of 2φ within the space $\mathcal{S}$. Therefore, after $\mathrm{O}(1/\varphi) = \mathrm{O}(1/\sqrt{p_M})$ iterations of this rotation starting with the state $|\pi\rangle$, we would have approximated the target state $|\mu\rangle$.

Restricted to the subspace $\mathcal{S}$, the operators $\mathrm{ref}(\mu^\perp)$ and $-\mathrm{ref}(\mathcal{M})$ are identical. Therefore, if we ensure that the state of the algorithm remain close to the subspace $\mathcal{S}$ throughout, we would be able to implement $\mathrm{ref}(\mu^\perp)_d$. This involves checking at cost C whether an item in the first register is marked.

The reflection $\mathrm{ref}(\pi)_d$ is computationally harder to perform. The straightforward strategy would be to rotate $|\pi\rangle_d$ to the state $|\bar{0}\rangle_d|\bar{0}\rangle_d$, use $\mathrm{ref}(|\bar{0}\rangle_d|\bar{0}\rangle_d)$, and then undo the first rotation. However, rotating $|\pi\rangle_d$ to $|\bar{0}\rangle_d|\bar{0}\rangle_d$ is exactly the inverse operation of the preparation of the initial state $|\pi\rangle_d$ from $|\bar{0}\rangle_d|\bar{0}\rangle_d$, and therefore requires the same cost $\mathsf{S}+\mathsf{U}$. This may be much more expensive than the update cost $4\mathsf{U}$ incurred by the walk $W(P)_d$. To use $W(P)_d$ instead, our idea is to apply phase estimation to it, and exploit this procedure to approximate the required diffusion operator on $\mathcal{A}_d + \mathcal{B}_d$ which contains the subspace $\mathcal{S}_d$.

The above approach is only valid when the probability p_M is known in advance. This assumption may be removed using standard techniques, without increasing the asymptotic complexity of the algorithms [9]. Indeed, if only a lower bound $\varepsilon > 0$ on p_M is known for non-empty M, then the above argument can be modified in order to determine if M is empty or find an element of M. We first sample from the stationary distribution a few times to accommodate the case that $p_M > 1/4$. If no marked element is found, we proceed as if $p_M \leq 1/4$. Following [9, Lemma 2], we iterate the rotation $\mathrm{ref}(\pi)_d\cdot\mathrm{ref}(\mu^\perp)_d$ a total of T times on the initial state, where T is chosen uniformly at random in $[0, 1/\sqrt{\varepsilon}\,]$. If M is not empty, a marked element is found with probability at least $1/4$, and otherwise no marked element is found. We refer to this version of the Grover algorithm as the *randomized Grover* algorithm.

3.2 Diffusion operator from quantum walk

To explain our approach, in the rest of this section, and in the next one, we assume that the classical Markov chain P is ergodic and reversible. For a reversible chain the corresponding discriminant $D(P)$ is symmetric. Symmetry implies that the singular values of $D(P)$ equal the absolute values of its eigenvalues.

Since $D(P) = \mathrm{diag}(\pi)^{1/2} \cdot P \cdot \mathrm{diag}(\pi)^{-1/2}$ is similar to the matrix P, their spectra are the same. Therefore, we only study the spectrum of P. The Perron-Frobenius theorem and the ergodicity of P imply that the eigenvalue 1 has multiplicity 1, and is the only eigenvalue of P with absolute value 1. The corresponding eigenvector of $D(P)$ is $(\sqrt{\pi_x})$, and every singular (or eigen-) vector of $D(P)$ orthogonal to this has singular value strictly less than 1. Transferring this property to the quantum walk $W(P)$ via Theorem 4, $|\pi\rangle$ is the unique eigenvector of the unitary operator $W(P)$ in $\mathcal{A}+\mathcal{B}$ with eigenvalue 1, and the remaining eigenvalues in $\mathcal{A}+\mathcal{B}$ are bounded away from 1. We use this observation to identify the component of any state $|\psi\rangle \in \mathcal{S}$ perpendicular to $|\pi\rangle$.

The main idea in our implementation of the above approach is to use phase estimation [19, 20, 14].

Theorem 5 (Phase estimation; Cleve, Ekert, Macchiavello, and Mosca [14])**.** *For every pair of integers $m, s \geq 1$, and a unitary operator U of dimension $2^m \times 2^m$, there exists a quantum circuit $C(U)$ that acts on $m+s$ qubits and satisfies the following properties:*

1. *The circuit $C(U)$ uses $2s$ Hadamard gates, $\mathrm{O}(s^2)$ controlled phase rotations, and makes 2^{s+1} calls to the controlled unitary operator c-U.*
2. *For any eigenvector $|\psi\rangle$ of U with eigenvalue 1, i.e., if $U|\psi\rangle = |\psi\rangle$, then $C(U)|\psi\rangle|0^s\rangle = |\psi\rangle|0^s\rangle$.*
3. *If $U|\psi\rangle = \mathrm{e}^{2\mathrm{i}\theta}|\psi\rangle$, where $\theta \in (0,\pi)$, then $C(U)|\psi\rangle|0^s\rangle = |\psi\rangle|\omega\rangle$, where $|\omega\rangle$ is an s-qubit state such that $|\langle 0^s|\omega\rangle| = \sin(2^s\theta)/(2^s \sin\theta)$.*

Moreover the family of circuits C parametrized by m and s is uniform.

This circuit is called phase estimation because measuring the state $|\omega\rangle$ in the computational basis yields an approximation to θ/π. In our case we only need to discriminate between the eigenvalue 1 and the remaining eigenvalues. In the following theorem we show how phase estimation is used to design a quantum circuit $R(P)$ which implements an operation that is close to the reflection $\mathrm{ref}(\pi)$.

Theorem 6. *Let P be an ergodic Markov chain on a state space of size $n \geq 2$, such that the phase gap of the quantum walk $W(P)$ based on P is $\Delta(P)$. Then for any integer k there exists a quantum circuit $R(P)$ that acts on $2\lceil \log_2 n \rceil + ks$ qubits, where $s \in \log_2(\frac{1}{\Delta(P)}) + \mathrm{O}(1)$, and satisfies the following properties:*

1. *The circuit $R(P)$ uses $2ks$ Hadamard gates, $\mathrm{O}(ks^2)$ controlled phase rotations, and makes at most $k\,2^{s+1}$ calls to the controlled quantum walk c-$W(P)$ and its inverse c-$W(P)^\dagger$.*
2. *If $|\pi\rangle$ is the unique 1-eigenvector of $W(P)$ as defined above, then $R(P)|\pi\rangle|0^{ks}\rangle = |\pi\rangle|0^{ks}\rangle$.*
3. *If $|\psi\rangle$ lies in the subspace of $\mathcal{A}+\mathcal{B}$ orthogonal to $|\pi\rangle$, then $\|(R(P)+\mathrm{Id})|\psi\rangle|0^{ks}\rangle\| \leq 2^{1-k}$.*

Moreover the family of circuits $R(P)$ parametrized by n and k is uniform.

Proof. We describe the circuit $R(P)$. Let $m = n^2$ and $s = \left\lceil \log_2(\frac{2\pi}{\Delta(P)}) \right\rceil$. We start by applying the phase estimation circuit $C(U)$ to the quantum walk $W(P)$, a unitary operator of dimension $m \times m$. To increase the accuracy of the phase estimation, we repeat the circuit k times, creating k identical copies of the s-qubit state $|\omega\rangle$ holding estimates of the phase. Observe that only the number of ancillary qubits increases from s to ks in this process. Since $C(U)$ leaves the eigenvectors of $W(P)$ in the first register unchanged, we do not need additional copies of the state $|\psi\rangle$.

The above operations approximately resolve any state $|\psi\rangle$ in $\mathcal{A}+\mathcal{B}$ along the eigenvectors of $W(P)$ by labeling them with estimates of the corresponding eigenvalue phases. We now flip the phase (i.e., multiply it by -1) of all computational basis states with a non-zero estimate of the phase in any of the k copies. Our intention is to flip the phase of all eigenvectors other than $|\pi\rangle$. Finally, we reverse the phase estimation. All these operations together constitute $R(P)$.

The state $|\pi\rangle|0^{ks}\rangle$ stays unchanged under the action of $R(P)$. When $|\psi\rangle$ is orthogonal to $|\pi\rangle$ it is a linear combination of eigenvectors of $W(P)$ whose eigenvalues are of the form $\mathrm{e}^{\pm 2\mathrm{i}\theta}$, where $\Delta(P)/2 \leq \theta < \pi/2$. By

definition of s, the state $|\omega\rangle$ holding the estimate for any phase $\theta \neq 0$ then satisfies $|\langle 0^s|\omega\rangle| \leq 1/2$. With k repetitions of the phase estimation, we can therefore decompose $|\psi\rangle|0^{ks}\rangle$ into a sum $|\psi_0\rangle + |\psi_1\rangle$, such that the phase estimate is zero in each of the k copies of $|\omega\rangle$ on the state $|\psi_0\rangle$, is non-zero in at least one copy on the state $|\psi_1\rangle$, and $\|\psi_0\| \leq 2^{-k}$. Then $R(P)|\psi\rangle|0^{ks}\rangle = |\psi_0\rangle - |\psi_1\rangle$, and $(R(P) + \mathrm{Id})|\psi\rangle|0^{ks}\rangle = 2|\psi_0\rangle$, whose norm is at most 2^{1-k}. □

3.3 The search algorithm for reversible Markov chains

Let us consider the following quantum procedure.

Quantum Search(P, ε)

1. Repeat 5 times:
   (a) Sample a state x from the stationary distribution π of P.
   (b) If $x \in M$, output x and stop.
2. Choose T uniformly at random in $[0, 1/\sqrt{\varepsilon}\,]$, let $k \in \log_2(T) + \mathrm{O}(1)$, and let s be as given by Theorem 6.
3. Prepare the initial state $|\pi\rangle_d|0^{Tks}\rangle$.
4. Repeat T times:
   (a) For any basis vector $|x\rangle_d|y\rangle_d|z\rangle$ of $\mathcal{H}_d$ and the ancillary (Tks)-qubit space, flip the phase if $x \in M$:
   $$|x\rangle_d|y\rangle_d|z\rangle \mapsto \begin{cases} -|x\rangle_d|y\rangle_d|z\rangle, & \text{if } x \in M \\ |x\rangle_d|y\rangle_d|z\rangle, & \text{otherwise.} \end{cases}$$
   (b) Apply circuit $R(P)_d$ of Theorem 6 with k as above, using a fresh set of ancilla qubits $|0^{ks}\rangle$ in each iteration.
5. Observe the first register.
6. Output x if $x \in M$, otherwise output 'no marked element exists'.

Theorem 7. *Let $\delta > 0$ be the eigenvalue gap of a reversible, ergodic Markov chain P, and let $\varepsilon > 0$ be a lower bound on the probability that an element chosen from the stationary distribution of P is marked whenever M is non-empty. Then, with high probability, the procedure* ***Quantum Search****(P, ε) determines if M is empty or else finds an element of M with cost of order* $\mathsf{S} + \frac{1}{\sqrt{\varepsilon}}\left[\left(\frac{1}{\sqrt{\delta}}\log\frac{1}{\sqrt{\varepsilon}}\right)\mathsf{U} + \mathsf{C}\right]$.

Proof. For convenience, we reason in the Hilbert space $\mathcal{H}$, without the data structures, and also omit the ancilla qubits used by the circuit $R(P)$. Between applications of $R(P)$ the ancilla qubits remain in a state close to $|0^{Tks}\rangle$.

First observe that if M is empty then no marked element is found by **Quantum Search**(P, ε). We assume now that M is non-empty. When $p_M > 1/4$, we detect a marked element in Step 1 with probability at least $1 - (3/4)^5 > 3/4$. In analyzing the correctness of the remaining steps, we may therefore assume that $p_M \leq 1/4$. Let $\mathcal{S}$ be the two-dimensional subspace $\mathcal{S} = \mathsf{Span}(|\pi\rangle,\ |\mu\rangle)$. Recall that the randomized Grover algorithm consists in T iterations of $\mathrm{ref}(\pi) \cdot \mathrm{ref}(\mu^\perp)$, where T is chosen uniformly at random from $[0, 1/\sqrt{\varepsilon}\,]$. Since $\varepsilon \leq p_M \leq 1/4$, with constant probability the randomized Grover algorithm rotates the vector $|\pi\rangle$ in the space $\mathcal{S}$ into a state whose inner product with $|\mu\rangle$ is a constant. Using a hybrid argument as in Refs. [7, 31], we prove that the algorithm **Quantum Search**(P, ε) simulates, with an arbitrarily small constant probability of error, the randomized Grover algorithm, and therefore finds a marked element with high probability, whenever such an element exists.

For $i \geq 0$, we define $|\phi_i\rangle$ as the result of i Grover iterations applied to $|\pi\rangle$, and $|\psi_i\rangle$ as the result of i iterations of step (4) in **Quantum Search**(P, ε) applied to $|\pi\rangle$. We show by induction on i, that $\||\psi_i\rangle - |\phi_i\rangle\| \leq i2^{1-k}$. Indeed, we can write $|\psi_i\rangle$ as $|\phi_i\rangle + (|\psi_i\rangle - |\phi_i\rangle)$. The actions of $\mathrm{ref}(\mu^\perp)$ and $-\mathrm{ref}(\mathcal{M})$ are identical on $|\phi_i\rangle$ since the state is in $\mathcal{S}$. Set $|\tau\rangle = |\phi_{i+1}\rangle - R(P)\cdot\mathrm{ref}(\mathcal{M})|\phi_i\rangle$. Since $\mathrm{ref}(\mathcal{M})|\phi_i\rangle$ is in $\mathcal{S}$, and $\mathcal{S}$ is a subspace of $\mathcal{A}+\mathcal{B}$, conclusion (3) of Theorem 6 can be applied, which implies that $\|\tau\| \leq 2^{1-k}$. Using $\||\psi_{i+1}\rangle - |\phi_{i+1}\rangle\| \leq \||\tau\rangle\| + \||\psi_i\rangle - |\phi_i\rangle\|$, the statement follows. For $k \in \log_2(T) + c$, where c is a constant, this implies that $\||\psi_T\rangle - |\phi_T\rangle\| \leq 2^{1-c}$, which can be made arbitrarily small by choosing c sufficiently large.

Let us now turn to the cost of the procedure. Since measuring $|\pi\rangle$ gives us a sample from the stationary distribution π, the cost of Step 1 is of the order of S. Preparing $|\pi\rangle_d$ costs $\mathsf{S} + \mathsf{U}$, and in each iteration the single phase flip costs C. In the circuit $R(P)_d$, the controlled quantum walk and its inverse can be implemented with four update operations, each of cost U. Indeed, the implementation of $W(P)$, described in the proof of Proposition 2 works also for the controlled quantum walk if we replace $\mathrm{ref}(|\bar{0}\rangle_d)$ by the controlled operator $\text{c-ref}(|\bar{0}\rangle_d)$. Since the controlled reflection is also of unit cost, this change does not alter the cost of the implementation.

In $R(P)_d$ the number of controlled quantum walks and its inverse is in $\mathrm{O}((1/\Delta(P))\log(1/\sqrt{\epsilon}))$. We claim that $\Delta(P) = \Omega(\sqrt{\delta})$. Let $\lambda_0, ..., \lambda_{n-1}$ be the eigenvalues of P, possibly with repetitions, such that $1 = \lambda_0 > |\lambda_1| \geq \ldots \geq |\lambda_{n-1}|$. Since the discriminant $D(P)$ is similar to P, their spectra are the same, and therefore the singular values of $D(P)$ are $|\lambda_0|, |\lambda_1|, \ldots, |\lambda_{n-1}|$. By definition, $\Delta(P) = 2\theta_1$, where $\cos\theta_1 = |\lambda_1|$. The following straightforward (in)equalities relate $\Delta(P)$ to $\delta(P)$: $\Delta(P) \geq \left|1 - \mathrm{e}^{2\mathrm{i}\theta_1}\right| = 2\sqrt{1 - |\lambda_1|^2} \geq 2\sqrt{\delta}$. This finishes the cost analysis. □

Let us observe that the origin of the quadratic speed-up due to quantum walks may be traced to the quadratic relationship between the phase gap $\Delta(P)$ of the quantum walk $W(P)$ and the eigenvalue gap δ of the classical Markov chain P, observed at the end of the above proof.

4 Search with approximate reflection operators

In this section, we describe how our approximate reflection operator may be incorporated into a search algorithm without incurring additional cost for reducing its error. The basic idea is to adapt the recursive amplitude amplification (RAA) algorithm due to Høyer, Mosca, and de Wolf [17] to our setting. To describe it, we use the notation from Section 3.1 where we discussed how the Grover algorithm works to rotate a starting state $|\pi\rangle$ into a target state $|\mu\rangle$, where $\langle\mu|\pi\rangle = \sin\varphi = \sqrt{p_M}$. We define procedures A_i recursively, for $i \geq 0$. Let the procedure A_0 be the identity map Id, and for $i > 0$, let

$$A_i \quad = \quad A_{i-1} \cdot \mathrm{ref}(\pi) \cdot A_{i-1}^\dagger \cdot \mathrm{ref}(\mu^\perp) \cdot A_{i-1}.$$

We define the states $|\pi_i\rangle$ as $A_i|\pi\rangle$. Then $|\pi_i\rangle$ forms an angle $3^i\varphi$ with $|\mu^\perp\rangle$, and therefore the state $|\pi_t\rangle$ is close to $|\mu\rangle$ when $t = \log_3 \frac{1}{\varphi} + \mathrm{O}(1)$. The final recursive algorithm is thus A_t.

We may estimate the cost $\mathsf{Cost}(t)$ of this search algorithm in terms of the cost c of implementing the two original reflections, $\mathrm{ref}(\pi)$ and $\mathrm{ref}(\mu^\perp)$. We have $\mathsf{Cost}(0) = 0$, and for $i \geq 1$, $\mathsf{Cost}(i) = 3 \cdot \mathsf{Cost}(i-1) + c$, and therefore the cost of A_t is $\mathrm{O}(c/\sqrt{\varepsilon})$.

The RAA algorithm is more suitable for situations where we have imperfect procedures that implement the basic reflections $\mathrm{ref}(\pi), \mathrm{ref}(\mu^\perp)$. Høyer *et al.* [17] demonstrated this when there is an ideal (error-free) procedure for $\mathrm{ref}(\pi)$, and a procedure for $\mathrm{ref}(\mu^\perp)$ that has ideal behavior only with high probability. Here, we adapt their approach to the case where it is the first reflection $\mathrm{ref}(\pi)$ which may only be approximated (it is probably possible to deal with the case where both reflections are imperfect, but for the sake of simplicity, we only deal with the case when the implementation of $\mathrm{ref}(\mu^\perp)$ is ideal since this is sufficient for our purpose). In the context of quantum walk based search, an imperfection appears in the form given by Theorem 6. The basic idea is to create an analogue of the recursive algorithms A_i when $\mathrm{ref}(\pi)$ is replaced by increasingly fine approximations based on Theorem 6.

We now state this precisely in full generality for potential further applications. Assume that for any $\beta > 0$, we have a quantum circuit $R(\beta)$ acting on $\mathcal{H} \otimes \mathcal{K}$, where $\mathcal{K}$ is an extra register of $s(\beta)$ qubits. For

a given integer t, and a precision parameter γ, the quantum circuit consists of t induction steps and acts on $\mathcal{H} \otimes \left[\bigotimes_{i=1}^{t} \mathcal{K}_i\right]$, where $\mathcal{K}_i$ is an extra register used at step i. Let $s_i = s(\beta_i)$ be the size of register $\mathcal{K}_i$, Let $S = \sum_{i=1}^{t} s_i$. We use $|\pi\rangle_d|0^S\rangle$ as the initial state of the algorithm.

The quantum circuit follows exactly the RAA algorithm explained above. We essentially replace $\mathrm{ref}(\pi)$ at step i by an approximation $R(\beta_i)$, acting on $\mathcal{H} \otimes \mathcal{K}_i$, and Id on the rest. Here is now one explicit step of the induction, where the basis case **Approximate RAA**$(0, \gamma)$ is simply the identity map :

Approximate RAA(i, γ)

1. Apply **Approximate RAA**$(i-1, \gamma)$.
2. For any basis vector $(|x\rangle_d|y\rangle_d) \otimes |z\rangle$, where $|x\rangle|y\rangle \in \mathcal{H}$, flip the phase if $x \in M$.
3. Undo **Approximate RAA**$(i-1, \gamma)$.
4. If any of the registers $\mathcal{K}_j$, with $j < i$, are not in state $|0^{s_j}\rangle$, respectively, then flip the phase of the state. Otherwise, apply $R(\beta_i)$ on $\mathcal{H} \otimes \mathcal{K}_i$, where $\beta_i = \frac{18}{4\pi^3}\gamma/i^2$.
5. Apply **Approximate RAA**$(i-1, \gamma)$.

We now prove that this algorithm can be used to find a marked element when p_M is known. We will later show how to modify the algorithm when only a lower bound on p_M is known.

Lemma 1. *Assume that for any $\beta > 0$, we have a quantum circuit $R(\beta)$ acting on $\mathcal{H} \otimes \mathcal{K}$, where $\mathcal{K}$ is an extra register of s qubits ($s = s(\beta)$ may depend on β), with the following properties:*

1. *The circuit $R(\beta)$ has cost $c_1 \log \frac{1}{\beta}$.*
2. *$R(\beta)|\pi\rangle|0^s\rangle = |\pi\rangle|0^s\rangle$.*
3. *$\|(R(\beta) + \mathrm{Id})|\psi\rangle|0^s\rangle\| \leq \beta$ when $|\psi\rangle$ is orthogonal to $|\pi\rangle$.*

*Further, assume that we are able to apply $-\mathrm{ref}(\mathcal{M})$ with cost c_2, and let t be the smallest non-negative integer such that $3^t \sin^{-1}\sqrt{p_M} \in [\pi/4, 3\pi/4]$. Then, for every real $\gamma > 0$, **Approximate RAA**(t, γ) maps $|\pi\rangle|0^S\rangle$ to a state that has projection of length at least $(\frac{1}{\sqrt{2}} - \gamma)$ in $\mathcal{M} \otimes \left[\bigotimes_{i=1}^{t} \mathcal{K}_i\right]$, and incurs a cost of order $3^t \cdot (c_1 \log \frac{1}{\gamma} + c_2)$.*

Proof. For simplicity, we omit the data structure in our error analysis, but take it into account in bounding the complexity of the algorithm. Let $s_i = s(\beta_i)$ be the size of register $\mathcal{K}_i$. Let $S = \sum_{i=1}^{t} s_i$. Recall that we use $|\phi_0\rangle = |\pi\rangle|0^S\rangle$ as the initial state. We also denote by $|\phi_i\rangle$ the output state of **Approximate RAA**(i, γ) on input $|\phi_0\rangle$. Note that the component of $|\phi_i\rangle$ on $\mathcal{K}_j$ is $|0^{s_j}\rangle$, for all $j > i$. Define the reflection operator R_i as the product of the recursive steps 3-5 of **Approximate RAA**(i, γ).

In order to understand the behavior of R_i, let us examine the action of $R(\beta_i)$ in step 4. At the beginning of that step, the algorithm state still has component $|0^{s_j}\rangle$ on $\mathcal{K}_j$, for all $j \geq i$. Therefore the conditioning, and the fact that $R(\beta_i)$ is an approximation to $\mathrm{ref}(\pi)$, directly gives that R_i behaves on the current state as an approximation to $\mathrm{ref}(\phi_{i-1})$. To be more precise, let $E_i = R_i - \mathrm{ref}(\phi_{i-1})$ be the error made in our implementation of $\mathrm{ref}(\phi_{i-1})$. We state the following fact without proof since it directly derives from the hypothesis on $R(\beta_i)$.

Fact 1. *E_i satisfies the following properties:*

1. *$E_i|\phi_{i-1}\rangle = 0$, and*
2. *$\|E_i|\psi\rangle|0^{S_i}\rangle\| \leq \beta_i$, for all $|\psi\rangle \in \mathcal{H} \otimes \left[\bigotimes_{j=1}^{i-1} \mathcal{K}_j\right]$ such that $|\psi\rangle|0^{S_i}\rangle \perp |\phi_{i-1}\rangle$, where $S_i = \sum_{j=i}^{t} s_j$.*

To analyze this algorithm, we keep track of the projection of $|\phi_i\rangle$ on the marked subspace. The marked subspace corresponds to $\mathcal{M}\otimes\left[\bigotimes_j \mathcal{K}_j\right]$; it consists of states in which the first register of the $\mathcal{H}$-part is marked. We denote this space by $\tilde{\mathcal{M}}$. Define the normalized projections of $|\phi_i\rangle$ on the marked subspace $\tilde{\mathcal{M}}$ and on its orthogonal complement as:

$$\begin{aligned} |\mu_i\rangle &= \frac{\Pi_{\tilde{\mathcal{M}}}|\phi_i\rangle}{\|\Pi_{\tilde{\mathcal{M}}}|\phi_i\rangle\|} \\ |\mu_i^\perp\rangle &= \frac{(\mathrm{Id}-\Pi_{\tilde{\mathcal{M}}})|\phi_i\rangle}{\|(\mathrm{Id}-\Pi_{\tilde{\mathcal{M}}})|\phi_i\rangle\|}. \end{aligned}$$

We thus have

$$|\phi_i\rangle = \sin\varphi_i\,|\mu_i\rangle + \cos\varphi_i\,|\mu_i^\perp\rangle. \quad (1)$$

where $\sin^2\varphi_i = \|\Pi_{\tilde{\mathcal{M}}}|\phi_i\rangle\|^2$ is the probability of finding a marked item by measuring the first register according to $\{\Pi_{\tilde{\mathcal{M}}}, \mathrm{Id}-\Pi_{\tilde{\mathcal{M}}}\}$. For later use, let us also define $|\phi_i^\perp\rangle$ as the state in the 2-dimensional subspace spanned by $|\mu_i\rangle$ and $|\mu_i^\perp\rangle$ that is orthogonal to $|\phi_i\rangle$:

$$|\phi_i^\perp\rangle = \cos\varphi_i\,|\mu_i\rangle - \sin\varphi_i\,|\mu_i^\perp\rangle.$$

For the initial state $|\phi_0\rangle$, we have $\sin^2\varphi_0 = p_M$. If all the errors β_i were zero, **Approximate RAA** would implement the RAA algorithm in the subspace spanned by $|\mu_i\rangle = |\mu_0\rangle$ and $|\mu_i^\perp\rangle = |\mu_0^\perp\rangle$, with the angles $\varphi_{i+1} = 3\varphi_i$, that is $\varphi_i = 3^i\varphi_0$. Therefore by recursively iterating our procedure for a total number of t steps, we would end up with a state whose inner product with $|\mu_0\rangle$ is at least $\frac{1}{\sqrt{2}}$.

Analysis of the errors — We show that **Approximate RAA** still works when the errors β_i are sufficiently small. In that case, the 2-dimensional subspace $\mathsf{Span}(|\mu_i\rangle, |\mu_i^\perp\rangle)$ may drift away from the initial subspace $\mathsf{Span}(|\mu_0\rangle, |\mu_0^\perp\rangle)$, and the angles φ_i may be different from the ideal value $\bar{\varphi}_i = 3^i\varphi_0$. We derive bounds on the error e_i:

$$e_i = |\sin\varphi_i - \sin\bar{\varphi}_i|\,, \quad (2)$$

the difference between the amplitude $\sin\varphi_i$ of the marked part of the state $|\phi_i\rangle$ and the ideal amplitude, $\sin\bar{\varphi}_i$.

We assume without loss of generality that $0 < \gamma < \frac{1}{\sqrt{2}}$ since the case $\gamma \geq \frac{1}{\sqrt{2}}$ is vacuous. We prove that after t steps $e_t \leq \gamma$. This will conclude the error analysis since $\frac{1}{\sqrt{2}} \leq \sin\bar{\varphi}_t \leq 1$.

We have

$$\begin{aligned} |\phi_{i+1}\rangle &= R_{i+1}\cdot \mathrm{ref}(\tilde{\mathcal{M}}^\perp)\,|\phi_i\rangle \\ &= \mathrm{ref}(\phi_i)\cdot\mathrm{ref}(\tilde{\mathcal{M}}^\perp)\,|\phi_i\rangle + E_{i+1}\cdot\mathrm{ref}(\tilde{\mathcal{M}}^\perp)\,|\phi_i\rangle \\ &= \sin 3\varphi_i\,|\mu_i\rangle + \cos 3\varphi_i\,|\mu_i^\perp\rangle + |\omega_{i+1}\rangle, \end{aligned} \quad (3)$$

where we used the fact that $\mathrm{ref}(\phi_i)\cdot\mathrm{ref}(\tilde{\mathcal{M}}^\perp)$ implements a perfect amplitude amplification step, and we introduced an error state $|\omega_i\rangle$, defined as

$$\begin{aligned} |\omega_{i+1}\rangle &= E_{i+1}\cdot\mathrm{ref}(\tilde{\mathcal{M}}^\perp)\,|\phi_i\rangle \\ &= E_{i+1}\cdot\mathrm{ref}(\tilde{\mathcal{M}}^\perp)\left(\sin\varphi_i\,|\mu_i\rangle + \cos\varphi_i\,|\mu_i^\perp\rangle\right) \\ &= E_{i+1}\left(-\sin\varphi_i\,|\mu_i\rangle + \cos\varphi_i\,|\mu_i^\perp\rangle\right) \\ &= E_{i+1}\left(\cos 2\varphi_i\,|\phi_i\rangle - \sin 2\varphi_i\,|\phi_i^\perp\rangle\right) \\ &= -\sin 2\varphi_i\;E_{i+1}\,|\phi_i^\perp\rangle, \end{aligned}$$

where we used Fact 1, property 1. Moreover, $|\phi_i^\perp\rangle \perp |\phi_i\rangle$, so $\|\omega_{i+1}\| \leq \beta_{i+1}\,|\sin 2\varphi_i|$ by Fact 1, property 2. Finally, comparing Eq. (1) and Eq. (3), we get

$$|\sin\varphi_{i+1} - \sin 3\varphi_i| \leq \beta_{i+1}\,|\sin 2\varphi_i|\,.$$

We may now bound the error defined in Eq. (2) as:

$$\begin{aligned} e_{i+1} &\leq |\sin\varphi_{i+1} - \sin 3\varphi_i| + |\sin 3\varphi_i - \sin\bar{\varphi}_{i+1}| \\ &\leq \beta_{i+1}\,|\sin 2\varphi_i| + |\sin 3\varphi_i - \sin 3\bar{\varphi}_i| \\ &\leq \beta_{i+1}(\sin 2\bar{\varphi}_i + |\sin 2\varphi_i - \sin 2\bar{\varphi}_i|) \\ &\quad + |\sin 3\varphi_i - \sin 3\bar{\varphi}_i| \\ &\leq \beta_{i+1}(\sin 2\bar{\varphi}_i + 2e_i) + 3e_i \\ &\leq 2\beta_{i+1}(\bar{\varphi}_i + e_i) + 3e_i, \end{aligned} \tag{4}$$

where we have used the triangle inequality and the following trigonometric inequalities

$$\begin{aligned} |\sin 2A - \sin 2B| &\leq 2|\sin A - \sin B| \\ |\sin 3A - \sin 3B| &\leq 3|\sin A - \sin B| \\ \sin A &\leq A \end{aligned}$$

that hold for any angles $A, B \in [0, \pi/4]$.

We define a quantity $\tilde{e}_i$, intended to be an upper bound on e_i (it would be if $\tilde{e}_i \leq \bar{\varphi}_i$). Let

$$\begin{aligned} \tilde{e}_0 &= 0 \\ \tilde{e}_{i+1} &= 4\beta_{i+1}\bar{\varphi}_i + 3\tilde{e}_i. \end{aligned}$$

We show that $\tilde{e}_i \leq \gamma$ for every $i \leq t$. Indeed, let us define u_i as

$$\tilde{e}_i = \gamma\,\bar{\varphi}_i\,u_i.$$

We therefore have the following recursion for u_i

$$\begin{aligned} u_0 &= 0 \\ u_{i+1} &= u_i + \frac{4}{3\gamma}\beta_{i+1}, \qquad (\forall i \geq 0) \end{aligned}$$

so that

$$u_i = \frac{4}{3\gamma}\sum_{j=1}^{i}\beta_j.$$

Recall that we have chosen $\beta_i = \frac{18}{4\pi^3}\gamma/i^2$, so that $\{\beta_i\}$ define a convergent series and the non-decreasing sequence (u_i) tends to $1/\pi$ when $i \to \infty$. We therefore have $\tilde{e}_i \leq \gamma\bar{\varphi}_i/\pi \leq \gamma$ since $0 \leq \bar{\varphi}_t \leq \pi$ for $i \leq t$.

Since $0 < \gamma \leq 1$, we have $\tilde{e}_i \leq \bar{\varphi}_i$, and we can show by induction that $e_i \leq \tilde{e}_i$ for all $i \leq t$. This finishes the error analysis.

Complexity — We now evaluate the complexity of our algorithm. We know from the hypotheses of the theorem that applying $R(\beta_i)$ costs $c_1 \log\frac{1}{\beta_i}$, while applying $\mathrm{ref}(\tilde{\mathcal{M}}^\perp) = -\mathrm{ref}(\mathcal{M}) \otimes \mathrm{Id}_{\bigotimes_j \mathcal{K}_j}$ costs c_2. Moreover, by definition of **Approximate RAA**, applying **Approximate RAA**(i, γ) requires 3 calls to **Approximate RAA**$(i-1, \gamma)$, one call to $R(\beta_i)$ and one call to $\mathrm{ref}(\mathcal{M})$. Hence, if we denote by $\mathsf{Cost}(i)$ the cost of applying **Approximate RAA**(i, γ), we have

$$\begin{aligned} \mathsf{Cost}(0) &= 0 \\ \mathsf{Cost}(i) &= 3\,\mathsf{Cost}(i-1) + c_1 \log\frac{1}{\beta_i} + c_2. \end{aligned}$$

Since we have fixed $\beta_i = \frac{18}{4\pi^2}\gamma/i^2$, we find that $\mathsf{Cost}(i)$ equals

$$\begin{aligned} & c_1 \sum_{j=1}^{i} 3^{i-j}\left(2\log j + \log\frac{1}{\gamma} + \mathrm{O}(1)\right) + c_2\sum_{j=1}^{i} 3^{i-j} \\ &= 3^i\left[\left(c_1\log\frac{1}{\gamma} + c_2 + \mathrm{O}(1)\right)\sum_{j=1}^{i}\frac{1}{3^j} + 2c_1\sum_{j=1}^{i}\frac{\log j}{3^j}\right] \end{aligned}$$

where both sums converge as $i \to \infty$. After t steps we have $\mathsf{Cost}(t) \in \mathrm{O}\left(3^t \cdot (c_1 \log \frac{1}{\gamma} + c_2)\right)$.

If the cost refers to time complexity, then there is an additional term pertaining to the reflection R_i. This arises from the check to see if the ancilla are in state $|0^S\rangle$. This does not change the asymptotic complexity of the algorithm. □

Note that Lemma 1 requires knowledge of p_M to infer the necessary number of iterations t. When only a lower bound ε on p_M is known, we can use the algorithm **Tolerant RAA**(t, γ), which only adds a constant factor overhead with respect to **Approximate RAA**(t, γ).

Tolerant RAA$(t_{\max}, \gamma)$

1. Sample a state x from the stationary distribution π of P.
2. if $x \in M$, output x, and stop.
3. Prepare the initial state $|\pi\rangle_d|0^S\rangle$ and set $i = 0$.
4. Increment i. Apply **Approximate RAA**(i, γ).
5. Measure the first register according to $\Pi_{\mathcal{M}}$.
   If successful, observe and output the first register, and stop.
6. If $i < t_{\max}$ go back to Step 4, otherwise output "No marked element".

Lemma 2. *Assume that for any $\beta > 0$, we have a quantum circuit $R(\beta)$ acting on $\mathcal{H} \otimes \mathcal{K}$, where $\mathcal{K}$ is an extra register of s qubits ($s = s(\beta)$ may depend on β), with the following properties:*

1. *The circuit $R(\beta)$ has cost $c_1 \log \frac{1}{\beta}$.*
2. *$R(\beta)|\pi\rangle|0^s\rangle = |\pi\rangle|0^s\rangle$.*
3. *$\|(R(\beta) + \mathrm{Id})|\psi\rangle|0^s\rangle\| \leq \beta$ when $|\psi\rangle$ is orthogonal to $|\pi\rangle$.*

*Further, assume that we are able to apply $-\mathrm{ref}(\mathcal{M})$ with cost c_2, and let $t_{\max}$ be the smallest non-negative integer such that $3^{t_{\max}} \sin^{-1} \sqrt{\epsilon} \in [\pi/4, 3\pi/4]$, where $p_M \geq \epsilon > 0$ whenever $p_M > 0$. Then, for every real γ such that $0 < \gamma \leq \frac{1}{40}$, **Tolerant RAA**$(t_{\max}, \gamma)$ always outputs "No marked element" if M is empty, otherwise it ends with a marked element with probability at least $1/12 - 3\gamma$, and incurs a cost of order $3^{t_{\max}} \cdot (c_1 \log \frac{1}{\gamma} + c_2)$.*

Proof. First, if M is empty then clearly the algorithm always outputs "No marked element". We now assume that M is non-empty and $p_M \geq \varepsilon$. If $p_M \geq 1/2$, the first two steps of the algorithm succeed with probability at least $1/2$. So in the analysis of the remaining steps, we additionally assume that $p_M < 1/2$.

We will use the notations of Lemma 1, together with the following ones. For $i \geq 1$, define $|\psi_i\rangle$ as the state after Step 4, $\sin^2 \theta_i = \|\Pi_{\tilde{\mathcal{M}}}|\psi_i\rangle\|^2$ the probability to project $|\psi_i\rangle$ onto the marked subspace $\mathcal{M}$, and the normalized projections $|\nu_i\rangle = \Pi_{\tilde{\mathcal{M}}}|\psi_i\rangle / \sin\theta_i$ and $|\nu_i^\perp\rangle = \Pi_{\tilde{\mathcal{M}}^\perp}|\psi_i\rangle / \cos\theta_i$, where $\tilde{\mathcal{M}}^\perp$ is the orthogonal complement of $\tilde{\mathcal{M}}$. Initially, we set $|\nu_0^\perp\rangle = |\phi_0\rangle = |\pi\rangle|0^S\rangle$.

Let us denote by A_i the unitary operator corresponding to circuit **Approximate RAA**(i, γ), and let t be the smallest positive integer such that $3^t \sin^{-1} \sqrt{p_M} \in [\pi/4, 3\pi/4]$. By Lemma 1, Applying A_t on $|\phi_0\rangle = |\pi\rangle|0^S\rangle$ prepares a state $|\phi_t\rangle$ that has projection at least $1/\sqrt{2} - \gamma$ on $\mathcal{M}$.

Since we do not know t, we will apply A_i for all possible values $i \in [1, t_{\max}]$. To avoid having to prepare a fresh copy of $|\phi_0\rangle$ for each attempt, which would incur an additional cost, for $i > 1$ we apply A_i on the state $|\nu_{i-1}^\perp\rangle$ left over from the previous attempt, which produces the state $|\psi_i\rangle = A_i|\nu_{i-1}^\perp\rangle$ instead of $|\phi_i\rangle = A_i|\phi_0\rangle$.

Analysis of the errors — Let $\delta_i = \||\psi_i\rangle - |\phi_i\rangle\| = \||\nu_{i-1}^\perp\rangle - |\phi_0\rangle\|$ denote the error at step i. By construction, we have $\delta_1 = 0$ and, for $i \geq 1$,

$$\delta_{i+1} = \||\nu_i^\perp\rangle - |\phi_0\rangle\| \leq \||\nu_i^\perp\rangle - |\mu_i^\perp\rangle\| + \sum_{k=0}^{i-1} \||\mu_{k+1}^\perp\rangle - |\mu_k^\perp\rangle\| + \||\mu_0^\perp\rangle - |\phi_0\rangle\|. \tag{5}$$

Let us evaluate the first term. By definition we have

$$\begin{aligned} |\psi_i\rangle &= \sin\theta_i|\nu_i\rangle + \cos\theta_i|\nu_i^\perp\rangle, \\ |\phi_i\rangle &= \sin\varphi_i|\mu_i\rangle + \cos\varphi_i|\mu_i^\perp\rangle. \end{aligned} \tag{6}$$

Since $\||\psi_i\rangle - |\phi_i\rangle\| = \delta_i$ we also have

$$|\psi_i\rangle = \sin\varphi_i|\mu_i\rangle + \cos\varphi_i|\mu_i^\perp\rangle + |\xi_i\rangle, \tag{7}$$

where $\|\xi_i\| \leq \delta_i$. Projecting Equations (6) and (7) onto $\tilde{\mathcal{M}}^\perp$, we obtain

$$\cos\theta_i|\nu_i^\perp\rangle = \cos\varphi_i|\mu_i^\perp\rangle + \Pi_{\tilde{\mathcal{M}}^\perp}|\xi_i\rangle,$$

which implies that $|\cos\theta_i - \cos\varphi_i| \leq \delta_i$ and in turn

$$\||\nu_i^\perp\rangle - |\mu_i^\perp\rangle\| \leq \frac{2\delta_i}{\cos\varphi_i} \leq 3\delta_i,$$

for any $i < t$. For the last inequality, we have used the fact that $\bar{\varphi}_i < \frac{\pi}{4}$, and therefore $\cos\varphi_i \geq \cos\bar{\varphi}_i - e_i \geq \frac{\sqrt{2}}{2} - \gamma \geq \frac{2}{3}$, since $\gamma \leq \frac{1}{40}$.

Let us now evaluate the second term in Equation (5). Recall that

$$\begin{aligned} |\phi_{k+1}\rangle &= \sin\varphi_{k+1}\ |\mu_{k+1}\rangle + \cos\varphi_{k+1}\ |\mu_{k+1}^\perp\rangle \\ &= \sin 3\varphi_k\ |\mu_k\rangle + \cos 3\varphi_k\ |\mu_k^\perp\rangle + |\omega_{k+1}\rangle, \end{aligned}$$

where $\|\omega_{k+1}\| \leq \beta_{k+1}\sin 2\bar{\varphi}_k \leq 4\beta_{k+1}\bar{\varphi}_k$, by the calculations leading to Eq. (4), and the bound $e_k \leq \tilde{e}_k \leq \bar{\varphi}_k$. Projecting this equation onto $\tilde{\mathcal{M}}^\perp$, we obtain $|\cos 3\varphi_k - \cos\varphi_{k+1}| \leq \|\omega_{k+1}\|$ and in turn

$$\||\mu_{k+1}^\perp\rangle - |\mu_k^\perp\rangle\| \leq \frac{2\|\omega_{k+1}\|}{\cos\varphi_{k+1}} \leq 12\beta_{k+1}\bar{\varphi}_k = 12\beta_{k+1}3^k\varphi_0,$$

for any $k < t-1$, where we have used the fact that $\cos\varphi_{k+1} \geq \frac{2}{3}$.

For the last term of Equation (5), since $\langle\mu_0^\perp|\phi_0\rangle = \cos\varphi_0$, we have

$$\||\mu_0^\perp\rangle - |\phi_0\rangle\| = \sqrt{2 - 2\cos\varphi_0} = 2\sin(\varphi_0/2) \leq \varphi_0.$$

Putting everything back together, we have

$$\delta_{i+1} \leq 3\delta_i + 12\varphi_0 \sum_{k=0}^{i-1} 3^k\beta_{k+1} + \varphi_0,$$

which, from $\delta_1 = 0$, implies

$$\begin{aligned} \delta_t &\leq \varphi_0 \sum_{k=0}^{t-2} 3^k + 12\varphi_0 \sum_{k=0}^{t-2} \left(\sum_{j=0}^{t-k-2} 3^j \right) 3^k\beta_{k+1} \\ &\leq \frac{1}{2}\ 3^{t-1}\varphi_0 + 12\varphi_0 \sum_{k=0}^{t-2} \left(\frac{1}{2}\ 3^{t-k-1} \right) 3^k\beta_{k+1} \\ &\leq \frac{1}{2}\ 3^{t-1}\varphi_0 + 2\cdot 3^t\varphi_0 \sum_{k=0}^{t-2} \beta_{k+1} \\ &\leq \frac{\pi}{8} + \frac{9\gamma}{8}. \end{aligned}$$

Since the projection of $|\phi_t\rangle$ onto $\tilde{\mathcal{M}}$ has length at least $\frac{1}{\sqrt{2}} - \gamma$, the projection of $|\psi_t\rangle$ onto $\tilde{\mathcal{M}}$ has length at least $\frac{1}{\sqrt{2}} - \gamma - \delta_t \geq \frac{1}{\sqrt{12}} - 3\gamma$, which means that the next measurement projects this state onto $\tilde{\mathcal{M}}$ with probability at least $\frac{1}{12} - 3\gamma$.

Complexity — As for the complexity analysis, note that we apply A_i for all $i \in [1, t_{\max}]$. From Lemma 1, the cost of A_i is of order $\mathsf{Cost}(i) \in \mathrm{O}(3^i \cdot (c_1 \log \frac{1}{\gamma} + c_2))$, therefore the cost of **Tolerant RAA**$(t_{\max}, \gamma)$ is dominated by

$$\sum_{i=1}^{t_{\max}} \mathsf{Cost}(i) \in \mathrm{O}(3^{t_{\max}} \cdot (c_1 \log \tfrac{1}{\gamma} + c_2)), \tag{8}$$

since this defines a geometric sum. □

We now have all the elements to prove Theorem 3 (stated in Section 1.4).

Proof of Theorem 3. The algorithm consists in **Tolerant RAA**$(t_{\max}, \frac{1}{72})$ from Lemma 2, using for the approximate reflections $R(\beta)$ the quantum phase estimation circuit $R(P)$ from Theorem 6.

First, no marked element is found if M is empty. Assume for now that M is non empty. We will prove that the assumptions of Lemma 2 are satisfied. Therefore the probability of finding an element for **Tolerant RAA**$(t_{\max}, \frac{1}{72})$ is at least $1/24$.

Setting $k = \left\lceil \log_2(\frac{1}{\beta}) + 1 \right\rceil$ in Theorem 6, $R(P)$ simulates a reflection with an error upper bounded by $2^{1-k} \leq \beta$. Implementing $R(P)_d$ then requires $k\, 2^{s+1}$ calls to the controlled quantum walk c-$W(P)_d$ or its inverse, where $s \in \log_2(\frac{1}{\sqrt{\delta}}) + \mathrm{O}(1)$. Since implementing c-$W(P)_d$ or its inverse has a cost $4\mathsf{U}$, the cost of implementing the circuit $R(P)_d$ for a given error β is $c_1 \log \frac{1}{\beta}$, with c_1 of order $\frac{1}{\sqrt{\delta}}\mathsf{U}$. Furthermore, preparing the initial state $|\pi\rangle_d$ has a cost $\mathsf{S} + \mathsf{U}$, and implementing $-\mathrm{ref}(\mathcal{M})_d$ has a cost $c_2 = \mathsf{C}$. Finally, since $t_{\max} \in \log_3 \frac{1}{\sqrt{\varepsilon}} + \mathrm{O}(1)$, the total cost of **Tolerant RAA**$(t_{\max}, \frac{1}{72})$ is of order $\mathsf{S} + \frac{1}{\sqrt{\varepsilon}}(\frac{1}{\sqrt{\delta}}\mathsf{U} + \mathsf{C})$. □

5 Non-reversible Markov chains

In this section, we discuss the performance of the search algorithm presented earlier for any ergodic, but possibly non-reversible Markov chain P. For the analysis of the quantum walk $W(P)$ we directly examine the singular value decomposition of the discriminant matrix $D(P) = \mathrm{diag}(\pi)^{1/2} \cdot P \cdot \mathrm{diag}(\pi)^{-1/2}$. This matrix has the same eigenvalues as P, but the singular values of $D(P)$ may be different from the eigenvalues of P. The singular values of $D(P)$ lie in the interval $[0, 1]$. The vector $v = (\sqrt{\pi_x})$ is both a left and a right eigenvector of $D(P)$ with eigenvalue 1. Therefore, $\mathsf{Span}(v)$ and $\mathsf{Span}(v)^\perp$ are invariant subspaces of $D(P)$, and we may choose v to be a left and right singular vector. If every singular vector orthogonal to v has a singular value strictly smaller than 1, that is $D(P)$ has a non-zero singular value gap, then Theorem 3 and its proof stay valid when the eigenvalue gap of P is replaced by the singular value gap of $D(P)$.

The discriminant of an irreducible walk does not necessarily have non-zero singular value gap, even if it is ergodic. Ergodicity implies a non-zero eigenvalue gap for P, but there are examples of ergodic Markov chains whose discriminants have 0 singular value gap. In the next proposition we show that if every state in the Markov chain has a transition to itself with non-zero probability, then its discriminant has non-zero singular value gap (the proof is given in the appendix). There is a standard and simple modification to any Markov chain P such that the resulting chain has this property: with some probability $\alpha \in (0, 1)$, stay at the current state, and with probability $1 - \alpha$, make a transition according to P.

Proposition 3. *Let $P = (p_{xy})$ be an irreducible Markov chain on a finite state space X, such that $p_{xx} > 0$, for every $x \in X$. Then, the discriminant matrix $D(P)$ has exactly one singular value equal to 1.*

Finally, we state the theorem on the performance of the quantum search algorithm presented in Section 4 when the underlying Markov chain is not necessarily reversible.

Theorem 8. *Let $P = (p_{xy})$ be an irreducible Markov chain on a finite state space X, such that $D(P)$ has exactly one singular value equal to 1. Let $\delta > 0$ be the singular value gap of $D(P)$, and let $\varepsilon > 0$ be a lower bound on the probability that an element chosen from the stationary distribution of P is marked whenever M is non-empty. Then, there is a quantum algorithm that with high probability, determines if M is empty or finds an element of M, with cost of order $\mathsf{S} + \frac{1}{\sqrt{\varepsilon}}(\frac{1}{\sqrt{\delta}}\mathsf{U} + \mathsf{C})$.*

6 Acknowledgments

A part of this work was done while the authors were visiting Institut Henri Poincaré, Paris, France, during the Programme on Quantum Information, Computation, and Complexity, January–April 2006.

This research was partially supported by the European Commission IST projects QAP 015848 and QCS 25596, and by the French ANR projects AlgoQP and QRAC 08-EMER-012. A. N. was supported in part by NSERC Canada, CIFAR, an ERA (Ontario), QuantumWorks, CFI, OIT, MITACS, and ARO/NSA (USA). Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through Industry Canada and by the Province of Ontario through MRI. Research at the Centre for Quantum Technologies is funded by the Singapore Ministry of Education and the National Research Foundation. J. R. acknowledges support from the Belgian FNRS, NSF Grant CCF-0524837 and ARO Grant DAAD 19-03-1-0082, and during this work he was affiliated with LRI, Université Paris-Sud; QuIC, Université Libre de Bruxelles; and Computer Science Division, U.C. Berkeley.

We thank the anonymous referees for their careful reading of the earlier drafts of this article, and for suggestions that vastly improved the quality of presentation.

References

[1] S. Aaronson and Y. Shi. Quantum lower bounds for the collision and the element distinctness problems. *Journal of the ACM*, pages 595–605, 2004.

[2] D. Aharonov, A. Ambainis, J. Kempe, and U. Vazirani. Quantum walks on graphs. In *Proceedings of the 33rd ACM Symposium on Theory of Computing*, pages 50–59, 2001.

[3] D. Aldous and J. A. Fill. *Reversible Markov Chains and Random Walks on Graphs.* http://www.stat.berkeley.edu/users/aldous/RWG/book.html. Monograph in preparation, August 2006 version.

[4] A. Ambainis, E. Bach, A. Nayak, A. Vishwanath, and J. Watrous. One-dimensional quantum walks. In *Proceedings of the 33rd ACM Symposium on Theory of Computing*, pages 37–49, 2001.

[5] A. Ambainis, J. Kempe, and A. Rivosh. Coins make quantum walks faster. In *Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms*, pages 1099–1108, 2005.

[6] A. Ambainis. Quantum walk algorithm for Element Distinctness. In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science*, pages 22–31. IEEE Computer Society Press, 2004.

[7] C. H. Bennett, E. Bernstein, G. Brassard, and U. Vazirani. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing*, 26(5):1510–1523, 1997.

[8] R. Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, 1993.

[9] M. Boyer, G. Brassard, P. Høyer, and A. Tapp. Tight bounds on quantum searching. *Fortschritte Der Physik*, 46(4-5):493–505, 1998.

[10] G. Brassard, P. Høyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. In J. S. J. Lomonaco and H. E. Brandt, editors, *Quantum Computation and Quantum Information: A Millennium Volume*, volume 305 of *Contemporary Mathematics Series*. American Mathematical Society, 2002.

[11] A. Z. Broder and A. R. Karlin. Bounds on the cover time. *Journal of Theoretical Probability*, 2(1):101–120, 1989.

[12] H. Buhrman, C. Dürr, M. Heiligman, P. Høyer, M. Santha, F. Magniez, and R. de Wolf. Quantum algorithms for Element Distinctness. *SIAM Journal of Computing*, 34(6):1324–1330, 2005.

[13] H. Buhrman and R. Špalek. Quantum verification of matrix products. In *Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms*, pages 880–889, 2006.

[14] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca. Quantum algorithms revisited. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 454(1969):339–354, 1998.

[15] S. Dörn and T. Thierauf. The quantum query complexity of algebraic properties. In *Proceedings of the 16th International Symposium on the Fundamentals of Computation Theory*, volume 4639 of *Lecture Notes in Computer Science*, pages 250–260, Berlin/Heidelberg, 2007. Springer.

[16] L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th ACM Symposium on Theory of Computing*, pages 212–219, 1996.

[17] P. Høyer, M. Mosca, and R. de Wolf. Quantum search on bounded-error inputs. In *Proceedings of the 30th International Colloquium on Automata, Languages and Programming*, volume 2719 of *Lecture Notes in Computer Science*, pages 291–299. Verlag, 2003.

[18] C. Jordan. Essai sur la géométrie à n dimensions. *Bulletin de la Société Mathématique de France*, 3:103–174, 1875.

[19] A. Kitaev. Quantum measurements and the Abelian stabilizer problem. Technical Report quant-ph/9511026, arXiv.org, 1995.

[20] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi. *Classical and Quantum Computation*, volume 47 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002.

[21] H. Krovi, F. Magniez, M. Ozols, and J. Roland. Finding is as easy as detecting for quantum walks. In *Proceedings of 37st International Colloquium on Automata, Languages and Programming*, volume 6198 of *Lecture Notes in Computer Science*, pages 540–551, Berlin/Heidelberg, 2010. Springer.

[22] F. Magniez and A. Nayak. Quantum complexity of testing group commutativity. *Algorithmica*, 48(3):221–232, July 2007.

[23] F. Magniez, A. Nayak, P. C. Richter, and M. Santha. On the hitting times of quantum versus random walks. In *Proceedings of the 20th ACM-SIAM Symposium on Discrete Algorithms*, pages 86–95, 2009.

[24] F. Magniez, M. Santha, and M. Szegedy. Quantum algorithms for the triangle problem. *SIAM Journal on Computing*, 37(2):413–424, 2007.

[25] D. Meyer. From quantum cellular automata to quantum lattice gases. *Journal of Statistical Physics*, 85(5-6):551–574, 1996.

[26] A. Nayak and A. Vishwanath. Quantum walk on the line. Technical Report quant-ph/0010117, arXiv.org, 2000.

[27] P. C. Richter. Almost uniform sampling via quantum walks. *New Journal of Physics*, 9(3):72, 2007.

[28] N. Shenvi, J. Kempe, and K. B. Whaley. Quantum random-walk search algorithm. *Physical Review A*, 67, 2003. Article no. 052307.

[29] M. Szegedy. Quantum speed-up of Markov chain based algorithms. In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science*, pages 32–41. IEEE Computer Society Press, 2004.

[30] A. Tulsi. Faster quantum walk algorithm for the two dimensional spatial search. *Physical Review A*, 78, 2008. Article no. 012310.

[31] U. Vazirani. On the power of quantum computation. *Philosophical Transactions of the Royal Society of London, Series A*, 356:1759–1768, 1998.

[32] J. Watrous. Quantum simulations of classical random walks and undirected graph connectivity. *Journal of Computer and System Sciences*, 62(2):376–391, 2001.

Proof of Proposition 3

We first state and prove that all the singular values of $D(P)$ lie in $[0, 1]$.

Lemma 3. *Let $P = (p_{xy})_{x,y\in X}$ be an irreducible Markov chain with stationary distribution $\pi = (\pi_x)_{x\in X}$. Then the singular values of the matrix $D(P)$ given by*

$$D(P) \quad = \quad \mathrm{diag}(\pi)^{1/2} \cdot P \cdot \mathrm{diag}(\pi)^{-1/2}$$

all lie in the interval $[0, 1]$.

Proof. Singular values are by convention taken to be non-negative real. To verify that $\|D(P)\|$, the largest singular value of $D(P)$ is at most 1, consider the inner product $u^\dagger D(P)v$, for some unit vectors u, v. The maximum absolute value that this inner product achieves is the norm of $D(P)$. By the Cauchy-Schwarz inequality, the inner product may be bounded as

$$\begin{aligned} & \left|u^\dagger D(P)v\right| \\ = \quad & \left|\sum_{xy} \bar{u}_x v_y \sqrt{\frac{\pi_x}{\pi_y}}\, p_{xy}\right| \\ \leq \quad & \left(\sum_{xy} |u_x|^2 p_{xy}\right)^{1/2} \left(\sum_{xy} |v_y|^2 \frac{\pi_x}{\pi_y} p_{xy}\right)^{1/2} \qquad (9) \\ \leq \quad & 1, \end{aligned}$$

since $\sum_x \pi_x p_{xy} = \pi_y$. □

Proof of Proposition 3. From Lemma 3, we know that the singular values of $D(P)$ all lie in $[0, 1]$. Further $v = (\sqrt{\pi_x})$ is a left (and right) singular vector with singular value 1. We show below that for any left and right singular vectors $u, w \in \mathbb{C}^X$, if $u^\dagger D(P) w = 1$, then $u = w = v$ (modulo an overall phase). This establishes the uniqueness of the singular value 1 and a non-zero singular value gap in $D(P)$.

Suppose $u^\dagger D(P) w = 1$. This implies that the Cauchy-Schwarz inequality in Equation (9) in the proof of Lemma 3 is tight. Then necessarily, the two unit vectors $u', w' \in \mathbb{C}^{X\times X}$ given by $u' = (u_x\sqrt{p_{xy}})_{x,y\in X}$ and $w' = (w_y\sqrt{\pi_x p_{xy}/\pi_y})_{x,y\in X}$ are parallel. Ignoring an overall phase, we may assume that they are in fact equal. This means that for every pair $x, y \in X$ such that $p_{xy} > 0$, $u_x = w_y\sqrt{\pi_x/\pi_y}$. In particular, since $p_{xx} > 0$, $u_x = w_x$ for every x, and so $u_x = u_y\sqrt{\pi_x/\pi_y}$ for every neighbor y of x in the graph underlying the Markov chain P.

Furthermore, for any path $x_1, x_2, \ldots, x_k$ in the graph, chaining together the equations

$$u_{x_{i+1}} \quad = \quad u_{x_i}\sqrt{\frac{\pi_{x_{i+1}}}{\pi_{x_i}}},$$

for $i = 1, \ldots, k-1$, we get that

$$u_{x_i} \quad = \quad u_{x_1}\sqrt{\frac{\pi_{x_i}}{\pi_{x_1}}},$$

for every i. Since the chain P is irreducible, i.e., the underlying graph is strongly connected, there is a path from x_1 to y for every $y \in X$. Thus,

$$u_y \quad = \quad u_{x_1} \sqrt{\frac{\pi_y}{\pi_{x_1}}},$$

for every y. Since the vector u is a unit vector, this implies that $u = w = (\sqrt{\pi_x})_{x \in X} = v$ (up to an unimportant global phase). □