MPRI Module 2.34.2: Quantum information and cryptography **2022 - 2023**

Lecture 2: Quantum walk search and collision finding

Lecturer: Simon Apers

1 Motivation: Collision finding and Grover search

Assume that we are given an array of integers $x_1, x_2, \ldots, x_N$. A *collision* is a pair of distinct i, j such that $x_i = x_j$. How many elements do we have to query in order to find a collision (or decide that no collision exists)? Classically this essentially requires to query the full array, and so the classical query complexity is $\Omega(N)$. In contrast, using a quantum algorithm we can find a collision with a *sublinear* number of queries. We first describe a quantum algorithm based on Grover search that finds a collision with only $O(N^{3/4})$ queries.[1]

1.1 Grover search

First we consider Grover search, which we briefly recall. We are given a set V of n elements, and a subset $M \subseteq V$ of m elements are marked. The algorithm is as follows:

1. Set up the uniform superposition $|\pi\rangle = \frac{1}{\sqrt{n}} \sum_{x \in V} |x\rangle$.
2. Do $O(\sqrt{n/m})$ times:
   (a) Reflect around the subspace of marked elements (i.e., apply $2\Pi_M - I = 2 \sum_{x \in M} |x\rangle\langle x| - I$).
   (b) Reflect around the uniform superposition (i.e., apply $2|\pi\rangle\langle\pi| - I$).
3. Measure the state.

The algorithm corresponds to a rotation in the 2-dimensional subspace spanned by the states

$$|\pi_M\rangle = \frac{1}{\sqrt{m}} \sum_{x \in M} |x\rangle \quad \text{and} \quad |\pi_U\rangle = \frac{1}{\sqrt{n-m}} \sum_{x \notin M} |x\rangle .$$

Note that the initial state $|\pi\rangle = \sqrt{m/n}\,|\pi_M\rangle + \sqrt{1-m/n}\,|\pi_U\rangle$, and this is essentially $|\pi_U\rangle$ if $m \ll n$. Grover's algorithm rotates this state closer to $|\pi_M\rangle$: after $O(\sqrt{n/m})$ iterations the state has a constant overlap with the marked elements.

We wish to quantify the query complexity of Grover's algorithm. To this end, let the "setup cost" $\mathcal{S}$ denote the number of queries needed to create the initial state $|\pi\rangle$ in step 1. I.e., it is the number of queries required to implement a unitary U_π such that $U|0\rangle = |\pi\rangle$. Now note that we can use U_π (and its adjoint $U_\pi^\dagger$) to reflect around $|\pi\rangle$ as well, using that

$$2|\pi\rangle\langle\pi| - I = U_\pi(2|0\rangle\langle 0| - I)U_\pi^\dagger. \tag{1}$$

Hence, the query complexity of step 2.(b) is also essentially $\mathcal{S}$. Then, let the "checking cost" $\mathcal{C}$ be the number of queries required to check whether an element is marked in step 2.(a). The query complexity of a single Grover iteration is then $\mathcal{S} + \mathcal{C}$, and the total query complexity of Grover's algorithm scales as $\sqrt{\frac{n}{m}}(\mathcal{S} + \mathcal{C})$.

[1] While we focus on query complexity for ease of exposition, all algorithms can be implemented with a similar runtime.

1.2 Grover search for finding collisions

Naively we could apply Grover search to the set of all $O(N^2)$ pairs of distinct i, j and mark the pairs that form a collision. However, this would trivially require $\Omega(N)$ queries. A better algorithm was proposed by Buhrman, Dürr, Heiligman, Høyer, Magniez, Santha and de Wolf [BDH+01] by running Grover search over larger *subsets* of indices rather than just pairs.

More specifically, we will search over elements that correspond to "words" $\mathcal{Y} = (Y, x_Y)$, consisting of (i) a size-k subset $Y \subseteq [N]$, and (ii) the list x_Y of integers x_j with index $j \in Y$. With $n = \binom{N}{k}$ the number of elements, the algorithm starts from the superposition

$$\frac{1}{\sqrt{n}} \sum_{Y \subseteq [N]:|Y|=k} |\mathcal{Y} = (Y, x_Y)\rangle .$$

A state $|\mathcal{Y}\rangle$ is marked if the corresponding subset Y contains an index of a collision. We now bound the query complexity of Grover search.

Exercise 1.
- *Let m denote the number of marked elements. Show that $m/n \geq k/N$ if $k \ll N$.*
- *Show that the checking cost $\mathcal{C}$ is $O(\sqrt{N})$ by (again) using Grover search to check if a state $|\mathcal{Y}\rangle$ is marked.*

It remains to bound the setup cost $\mathcal{S}$. Note that (i) we can create the state $\frac{1}{\sqrt{n}} \sum_{Y \subseteq [N]:|Y|=k} |(Y, 0)\rangle$ without making any queries, and (ii) we can map $|(Y, 0)\rangle \mapsto |(Y, x_Y)\rangle$ by only making k queries to the elements x_j with index $j \in Y$. By linearity, we can use this to map $\frac{1}{\sqrt{n}} \sum_{Y \subseteq [N]:|Y|=k} |(Y, 0)\rangle$ to $|\pi\rangle$ with only k queries, and hence the setup cost $\mathcal{S}$ is $O(k)$.

The total query complexity then scales as $\sqrt{\frac{n}{m}}(\mathcal{S}+\mathcal{C}) \approx \sqrt{\frac{N}{k}}(k+\sqrt{N})$. For $k = \sqrt{N}$ this yields an algorithm for collision finding with query complexity $\widetilde{O}(N^{3/4})$, which is an improvement over the classical $\Omega(N)$ query complexity.

In the following, we will use quantum walk search instead of Grover search to improve the query complexity to $\widetilde{O}(N^{2/3})$, which is essentially optimal.

2 Quantum walk search

Equation (1) in the previous section shows that the cost of reflecting around $|\pi\rangle$ is bounded by the cost of preparing $|\pi\rangle$. However, reflecting around $|\pi\rangle$ can be much cheaper than preparing $|\pi\rangle$. This is exploited in the quantum walk search algorithm, which generalizes Grover search to graphs.

Consider a regular graph $G = (V, E)$ with n nodes, and let $M \subseteq V$ denote a subset of marked nodes of size m. Rather than states $|x\rangle$ (corresponding to vertices of the graph), we consider basis states described by star states $|\psi_x\rangle$ of the graph. A state $|\psi_x\rangle$ is marked iff x is a marked element. The algorithm then starts from $|\pi\rangle = \frac{1}{\sqrt{n}} \sum_{x \in V} |\psi_x\rangle$ and iteratively applies $O(\sqrt{n/m})$ times the operator

$$(2 |\pi\rangle \langle\pi| - I)(2\Pi_M - I) = (2 |\pi\rangle \langle\pi| - I) \left(2 \sum_{x \in M} |\psi_x\rangle \langle\psi_x| - I \right).$$

Similarly to Grover search, the algorithm takes place in the 2-dimensional subspace spanned by $|\pi_M\rangle = \frac{1}{\sqrt{m}} \sum_{x \in M} |\psi_x\rangle$ and $|\psi_U\rangle = \frac{1}{\sqrt{n-m}} \sum_{x \notin M} |\psi_x\rangle$.

Since the cost of reflecting around $|\pi\rangle$ will be different than that of preparing $|\pi\rangle$, we denote it by a separate "reflection cost" $\mathcal{R}$. The total cost is then

$$\mathcal{S} + \sqrt{\frac{n}{m}}(\mathcal{R} + \mathcal{C}).$$

2.1 Phase estimation and quantum walk reflection

We now demonstrate how to use quantum walks to efficiently reflect around $|\pi\rangle$. This procedure crucially relies on (i) $|\pi\rangle$ being a stationary state of the QW operator W, while (ii) its orthogonal complement $|\pi^\perp\rangle$ (in the 2-dimensional subspace $\mathrm{span}\{|\pi_U\rangle, |\pi_M\rangle\}$) being spanned by nonstationary eigenvectors of W (see last lecture). As a consequence, for any α, β we have a decomposition

$$\alpha |\pi\rangle + \beta |\pi^\perp\rangle = \alpha |\pi\rangle + \sum_{j>0} \beta_j |\phi_j\rangle ,$$

where $W |\phi_j\rangle = e^{i\theta_j} |\phi_j\rangle$ for some $\theta_j > 0$. Moreover, recall that the QW operator W has a phase gap $\Delta \in \Omega(\sqrt{\delta})$ with δ the random walk spectral gap, and hence $\theta_j \in \Omega(\sqrt{\delta})$.

Reflecting around $|\pi\rangle$ in this subspace amounts to transforming $\alpha |\pi\rangle + \beta |\pi^\perp\rangle$ into $\alpha |\pi\rangle - \beta |\pi^\perp\rangle$. Equivalently, we want to put a minus sign before any non-stationary eigenvector of W. To this end we can use *quantum phase estimation*. This corresponds to a unitary U_P such that

$$U_P \left(\alpha |\pi\rangle |0\rangle + \sum_{j>0} \beta_j |\phi_j\rangle |0\rangle \right) = \alpha |\pi\rangle |0\rangle + \sum_{j>0} \beta_j |\phi_j\rangle |\tilde{\theta}_j\rangle ,$$

where $\tilde{\theta}_j$ is an estimate of the phase θ_j to (additive) precision ε.[2] The cost of quantum phase estimation roughly corresponds to applying $\widetilde{O}(1/\varepsilon)$ times the QW operator W. If Δ is the smallest nonzero phase θ_j, then choosing $\varepsilon < \Delta$ correctly differentiates a stationary eigenvector from a nonstationary one (see Fig. 1).

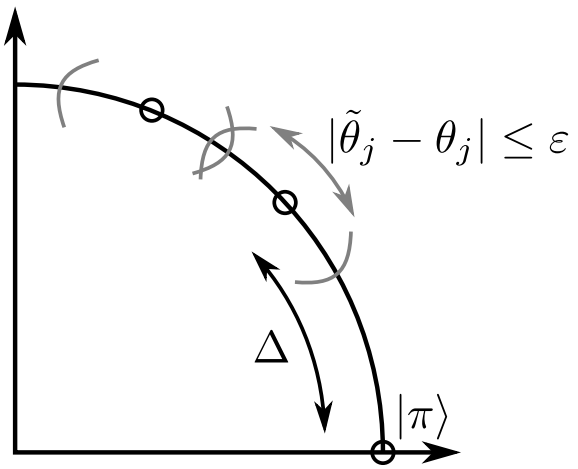

Figure 1: Quantum phase estimation to precision ε. We can differentiate the stationary state $|\pi\rangle$ from the non-stationary states $|\phi_j\rangle$ by choosing ε smaller than the phase gap Δ.

We can then apply a minus sign *conditioned* on the phase estimate being nonzero. Equivalently, we apply the reflection $R_0 = 2\Pi_0 - I$ with projector $\Pi_0 = I \otimes |0\rangle\langle 0|$. Finally, we apply the inverse operation $U_P^\dagger$ to "erase" the phase estimation registers.

Exercise 2. *Verify that $U_P^\dagger R_0 U_P$ implements a reflection around $|\pi\rangle$ in the relevant subspace. I.e.,*

$$U_P^\dagger R_0 U_P \left(\alpha |\pi\rangle + \beta |\pi^\perp\rangle \right) |0\rangle = \left(\alpha |\pi\rangle - \beta |\pi^\perp\rangle \right) |0\rangle .$$

[2] In fact, phase estimation returns a *superposition* over different estimates, but we will ignore this technicality.

Finally, since the quantum walk operator W has a phase gap $\Delta \in \Omega(\sqrt{\delta})$ it suffices to run phase estimation to precision $\varepsilon \in O(\sqrt{\delta})$. This requires $O(\frac{1}{\sqrt{\delta}})$ calls to the QW operator. If we let $\mathcal{U}$ denote the cost of implementing the QW operator, then this bounds the reflection cost $\mathcal{R}$ by roughly $\frac{1}{\sqrt{\delta}}\mathcal{U}$.

2.2 Quantum walk search

The quantum walk search algorithm now runs in parallel to Grover's algorithm, except that we use quantum walks and phase estimation to implement the reflection around $|\pi\rangle$. Recalling that the total cost of the search algorithm is $\mathcal{S} + \sqrt{\frac{n}{m}}(\mathcal{R}+\mathcal{C})$, we can bound the cost of quantum walk search by

$$\mathcal{S} + \sqrt{\frac{n}{m}}\left(\frac{1}{\sqrt{\delta}}\mathcal{U} + \mathcal{C}\right).$$

Hence, if we are given the state $|\pi\rangle$ we can find a marked element with only $O(\frac{1}{\sqrt{\delta}}\sqrt{\frac{n}{m}})$ steps of a quantum walk.

We can compare this with the cost of using a *random walk* to find a marked element. In the last lecture we saw that a random walk, starting from the stationary distribution π, hits a marked element after an expected number of steps given by $HT(M) \in O(\frac{1}{\delta}\frac{n}{m})$. Quantum walk search quadratically improves the number of steps as compared to this upper bound.

3 Collision finding with quantum walk search

Ambainis [Amb07] used quantum walk search to describe a quantum algorithm for collision finding that is essentially optimal.

Similar to the Grover algorithm for collision finding, the basis states are indexed by elements $\mathcal{Y} = (Y, x_Y)$, with $Y \subseteq [N]$ a size-k subset of indices and x_Y the list of integers x_j with index $j \in Y$. Slightly differently, we will now call an element $\mathcal{Y}$ marked only if Y contains *both* indices of a collision (equivalently, x_Y must contain a collision).

Exercise 3. *Let $n = \binom{N}{k}$ denote the number of elements and m the number of marked elements. Show that $m/n \in \Omega(k^2/N^2)$.*

To use quantum walk search, we consider a graph G with vertex set V indexed by the elements $\mathcal{Y}$. There is an edge between $\mathcal{Y} = (Y, x_Y)$ and $\mathcal{Y}' = (Y', x_{Y'})$ if the subsets Y and Y' differ in exactly one element (i.e., we can obtain Y' from Y by replacing one index). The resulting graph G has $n = \binom{N}{k}$ vertices and is $k(n-k)$-regular. It corresponds to a so-called *Johnson graph*, and one can show that its spectral gap is $\delta \in \Omega(1/k)$ when $k \ll n$.

A star state $|\psi_{\mathcal{Y}}\rangle$ centered on a vertex $\mathcal{Y}$ of G is given by the state

$$|\psi_{\mathcal{Y}}\rangle = \frac{1}{\sqrt{k(n-k)}} \sum_{\mathcal{Y}' \sim \mathcal{Y}} |\mathcal{Y}, \mathcal{Y}'\rangle,$$

where the sum runs over neighboring elements $\mathcal{Y}'$ of $\mathcal{Y}$. The quantum walk search algorithm then starts from the uniform superposition

$$|\pi\rangle = \frac{1}{\sqrt{n}} \sum_{\mathcal{Y}} |\psi_{\mathcal{Y}}\rangle,$$

and the algorithm has cost

$$\mathcal{S} + \frac{N}{k}(\sqrt{k}\mathcal{U} + \mathcal{C}),$$

where $\mathcal{U}$ is the *update cost* or cost of implementing a single quantum walk step on the Johnson graph G. We now bound the different costs.

For the checking cost $\mathcal{C}$, note that we can check whether a given state $|\psi_{\mathcal{Y}}\rangle$ is marked simply by checking whether the list x_Y contains a collision. Since this list is given explicitly in the description of $|\psi_{\mathcal{Y}}\rangle$, this requires no queries and so $\mathcal{C} = 0$.

The setup cost $\mathcal{S}$ amounts to creating the state $|\pi\rangle = \frac{1}{\sqrt{n}} \sum_{\mathcal{Y}} |\psi_{\mathcal{Y}}\rangle$. We do this in a few steps. First, exactly as in the Grover algorithm for collision finding, we prepare the state $\frac{1}{\sqrt{n}} \sum_{\mathcal{Y}} |\mathcal{Y}\rangle |0\rangle$. This takes k queries. Then, we construct the mapping U_ψ defined by $U_\psi |\mathcal{Y}\rangle |0\rangle = |\psi_{\mathcal{Y}}\rangle$. We do this in two steps:

$$\begin{aligned} |\mathcal{Y}\rangle |0\rangle = |Y, x_Y\rangle |0\rangle &\overset{\text{(i)}}{\mapsto} \frac{1}{\sqrt{k(n-k)}} \sum_{\mathcal{Y}' \sim \mathcal{Y}} |Y, x_Y\rangle |Y', 0\rangle \\ &\overset{\text{(ii)}}{\mapsto} \frac{1}{\sqrt{k(n-k)}} \sum_{\mathcal{Y}' \sim \mathcal{Y}} |Y, x_Y\rangle |Y', x_{Y'}\rangle = |\psi_{\mathcal{Y}}\rangle . \end{aligned}$$

Step (i) requires no queries. Step (ii) amounts to gathering the elements $x_{Y'}$ with index in Y'. Since $x_{Y'}$ contains exactly one element not in x_Y, this requires exactly one query. The setup cost $\mathcal{S}$ is hence roughly k.

Finally, we bound the cost $\mathcal{U}$ of a single call to the quantum walk operator W. Recall from last lecture that $W = S \cdot C$ where S is a simple swap (i.e., $S |\mathcal{Y}, \mathcal{Y}'\rangle = |\mathcal{Y}', \mathcal{Y}\rangle$), requiring no queries, and $C = 2\left(\sum_{\mathcal{Y}} |\psi_{\mathcal{Y}}\rangle \langle\psi_{\mathcal{Y}}|\right) - I$ is a reflection around the star subspace. Using a similar trick as before, we can implement the reflection C by making two calls to the preparation operator U_ψ, which requires a single query. This proves that the update cost $\mathcal{U}$ is $O(1)$.

Exercise 4. *Verify that* $C = U_\psi R_0 U_\psi^\dagger$.

Combining these different arguments, we can bound the total cost by

$$\mathcal{S} + \sqrt{\frac{n}{m}} \left(\frac{1}{\sqrt{\delta}} \mathcal{U} + \mathcal{C} \right) \approx k + \frac{N}{\sqrt{k}}.$$

If we set $k = N^{2/3}$ then this yields a quantum algorithm for collision finding with complexity $\widetilde{O}(N^{2/3})$. This is essentially optimal by the $\Omega(N^{2/3})$ lower bound of Aaronson and Shi [AS04].

References

[Amb07] Andris Ambainis. Quantum walk algorithm for element distinctness. *SIAM Journal on Computing*, 37(1):210–239, 2007.

[AS04] Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *Journal of the ACM (JACM)*, 51(4):595–605, 2004.

[BDH+01] Harry Buhrman, Christoph Dürr, Mark Heiligman, Peter Høyer, Frédéric Magniez, Miklos Santha, and Ronald de Wolf. Quantum algorithms for element distinctness. In *Proceedings 16th Annual IEEE Conference on Computational Complexity*, pages 131–137. IEEE, 2001.