Introduction to quantum computation: Grover's algorithm

Lecture Notes for Quantum Computing

Lecturer: Dorit Aharonov
Scribe: Uri Barkay and Omer Levy

April 23-May 1, 2001

Summary of the sixth week of semester 2 (23.4.2001 - 1.5.2001)

Abstract

We learn about Grover's algorithm with provable quadratic improvement: find an element in a database of N entries using $\sqrt{N}$ queries. We show the lower bound by BBBV, which shows that $\Omega(\sqrt(N))$ queries are required to solve the Grover problem, i.e. to distinguish between an oracle with one item for which $f(i) = 1$ and an oracle with no such items. We define the density matrix of a general quantum system. We will then give Ambainis's lower bound on quantum speed up, which is based on quantum intuition, and shows that quadratic improvement is the best possible for Grover's case. This lower bound uses density matrices.

1 Grover's Algorithm (1995)

1.1 Introduction

Assume we are given a database, and we would like to extract some information from the database, for example search for a phone number in a phone book which is sorted by names. This problem is classically solvable in $O(N)$ time, where N is the size of the database. We show here a quantum algorithm to solve the problem in $O(\sqrt{N})$ time.

1.2 Oracle database

Assume we are searching for ω in our database. Then we define the oracle $f_\omega(i)$ as follows: $f_\omega(i) = 1$ if $i = \omega$, otherwise $f_\omega(i) = 0$.

We define the oracle transformation U_ω as the transformation $|i\rangle \longmapsto (-1)^{f_\omega(i)}|i\rangle$. This transformation may be accomplished by using an ordinary oracle call for $|i\rangle|\alpha\rangle \longmapsto |i\rangle|\alpha \oplus f(\omega)\rangle$, where $\alpha = \frac{|0\rangle - |1\rangle}{\sqrt{2}}$

If we apply U_ω on $\sum |i\rangle$, the result is flipping amplitude of $|\omega\rangle$:

$$U_\omega : \sum_i |i\rangle \longmapsto (\sum_{i \neq \omega} |i\rangle) - |\omega\rangle$$

We would like to start with a uniform superposition of all possible i's,

$$|\alpha\rangle = \frac{1}{\sqrt{N}} \sum_i |i\rangle$$

and slowly increase the amplitude of the item for which the oracle is 1. Then, when we measure, we will get a large amplitude for ω.

The idea of the algorithm is as follows. After one call to the oracle U_ω, the amplitude of w is flipped. For a given state

$$|\phi\rangle = \sum a_i |i\rangle$$

we define the **average** of the state as

$$< a >= \frac{1}{N} \sum_i a_i.$$

Note that after applying the oracle on the state $|\alpha\rangle$ the average almost has not changed, it is very close to $1/\sqrt{N}$. The amplitudes of all the states except ω are very close to the average. The only state which is pretty far from the average is ω. If we could **invert** all amplitudes in $|\alpha\rangle$ about the average of $|\alpha\rangle$, then the amplitudes of most items will hardly change, except the amplitude of ω which would increase significantly: it would become about $3/\sqrt{N}$ instead of $1/\sqrt{N}$ as it was in the beginning. Now assume we could repeat this process again and again (call the oracle and apply inversion about the average). If at all times, the amplitude of ω grows by order of $1/\sqrt{N}$, then after $O(\sqrt{N})$ iterations the amplitude of ω would be constant, and if we measure the state we will have constant probability to see it. We proceed by first constructing an operator which inverts about the mean.

1.3 Inversion about average

Define

$$|\alpha\rangle = \sum_{i=0}^{N-1} |i\rangle$$

We claim that a transformation which transfers $|a\rangle \longmapsto |\alpha\rangle$, and $|\beta\rangle \longmapsto -|\beta\rangle \forall \beta \perp \alpha$ is indeed an inversion about the mean.

Write this transformation in a basis of $|\alpha\rangle$ and orthonormal vectors to it. The transformation is:

$$\begin{pmatrix} 1 & & 0 \\ & -1 & \\ . & . & . \\ 0 & & -1 \end{pmatrix} = \begin{pmatrix} 2 & & \\ & 0 & \\ . & . & . \\ & & 0 \end{pmatrix} - I$$

We note that

$$\begin{pmatrix} 2 & & \\ & 0 & \\ . & . & . \\ & & 0 \end{pmatrix} = 2|\alpha\rangle\langle\alpha|$$

So we define

$$U_\alpha = 2|\alpha\rangle\langle\alpha| - I$$

We now check that U_α is indeed inversion about the mean for an arbitrary $|\psi\rangle$. Let

$$|\psi\rangle = \sum a_i|i\rangle, < a >= \frac{\sum a_i}{N}.$$

Then

$$U_\alpha|\psi\rangle = 2|\alpha\rangle\langle\alpha||\psi\rangle - |\psi\rangle$$

$$\langle\alpha|\psi\rangle = \frac{1}{\sqrt{N}}(\sum_i \langle i|)(\sum_j a_j|j\rangle) = \frac{1}{\sqrt{N}}\sum a_i = \sqrt{N} < a >$$

$$U_\alpha|\psi\rangle = 2|\alpha\rangle\sqrt{N} < a > -|\psi\rangle = 2\frac{\sum|i\rangle}{\sqrt{N}}\sqrt{N} < a > - \sum a_i|i\rangle = \sum_i (< a > +(< a > -a_i))|i\rangle$$

If we compare this to

$$|\psi\rangle = \sum_i a_i|i\rangle = \sum_i < a > +(a_i - < a >)|i\rangle$$

we see that the difference of each amplitude from the average indeed flipped sign, so U_α is the inversion we were looking for.

1.4 Probability proof

We claim that if we start with state $|\alpha\rangle$, and apply $U_\alpha U_\omega$ $O(\sqrt{N})$ times and then measure, then the probability of getting $|\omega\rangle$ is larger than $\frac{1}{2}$.

Note that U_ω is inversion about $|\omega^\perp\rangle$, and U_α is inversion about $|\alpha\rangle$. Denote the angle between $|\omega^\perp\rangle$ and $|\alpha\rangle$ as θ. Inverting twice in a two dimensional space, first around one vector and than around the other, is actually a rotation in the two dimensional space by twice the angle between the two vectors. Hence, applying $U_\alpha U_\omega$ is actually rotation by an angle of 2θ.

We approximate θ: $sin\theta = cos(\frac{\pi}{2} - \theta) = \langle\alpha|\omega\rangle = \frac{1}{\sqrt{N}}$. Assuming N is large enough, we can resolve $\theta \approx sin\theta = \frac{1}{\sqrt{N}}$.

We are starting with the vector $||\alpha\rangle\rangle$, which is almost orthogonal to $||w\rangle\rangle$; The actual angle between them is approximately $\pi/2 - 1/\sqrt{N}$. Therefore, in order to reach $|\omega\rangle$ we need to rotate by 2θ $\frac{\pi/2}{2\theta} = \frac{\pi}{4\theta} \approx \frac{\pi}{4}\sqrt{N}$ times. Since we made some approximations (eg. the number of times should be an integer...) we will not reach the actual state $||w\rangle\rangle$ but a state very close to it. When we measure, we will get w with very high probability.

1.5 Multiple Solutions

Suppose we are searching for one of r different solutions in our database. That is, $|\omega\rangle$ is a superposition of r different solutions. Then $\theta \approx \langle\omega|\alpha\rangle = \sqrt{\frac{r}{N}}$. By repeating the above process $\frac{\pi}{4}\sqrt{N}$ times we rotate the space by $2\theta\frac{\pi}{4}\sqrt{N} = \frac{\sqrt{r}\pi}{2}$. For $r = 4$, this is exactly π; This means that we reach a vector which is almost aligned with the initial vector $|\alpha\rangle\rangle$, except it is in opposite direction. Its projection on $||w\rangle\rangle$ would be exponentially small, and the probability to measure ω is close to 0. It seems that we need to know r in order to know how many iterations are required to reach ω...

Solution 1: View the algorithm as solving an NP problem: f(i)=1 if i is a solution. So we solved an NP problem in $\sqrt{N}polylog(N)$ time. We may now use a classical reduction to reduce the problem to an NP problem with a single solution, and then apply the quantum algorithm.

Solution 2: (sketch) We do not know r, but the probability to measure a solution (which is one of the components in ω) is periodic in the number of iterations, and the period is a function of r. We can find r by finding the period of the probability, in an approach similar to the one taken in Shor's algorithm. We will not describe this here.

1.6 Calculating the transformation U_α

How do we calculate the transformation U_α we used above? (Inversion about $|\alpha\rangle$)

Inversion about $|0\rangle$: The transformation $U : |0\rangle \longmapsto |0\rangle, |1\rangle \longmapsto -|1\rangle$ is a simple transformation on a single qubit.

Inversion about $|00...0\rangle$: Define the transformation U_0 as follows: use a classical algorithm to achieve $|0\rangle|00...0\rangle \longmapsto |0\rangle|00...0\rangle, |0\rangle|i\rangle \longmapsto |1\rangle|i\rangle\forall|i\rangle \neq |00...0\rangle$ (examine the second register). Then apply the transformation U on the first qubit to get $|00...0\rangle$ in the second register if we started with $|00...0\rangle$, and $-|i\rangle$i if we started with $|i\rangle \neq |00...0\rangle$

Inversion about α: Define the transformation U_α as follows: first apply Fourier Transform to transform $|\alpha\rangle$ to $|00...0\rangle$. Then apply U_0 to inverse about $|00...0\rangle$. Apply Fourier transform again to achieve inversion about α.

1.7 Using Grover's Algorithm

Example: finding minimum. Let $f : 0,1^n \longmapsto 0,1^n$ be a binary function, and we search for an i such that f(i) is minimal. We use binary search to search the minimum: Define $g_1(i) = 1$ if $f(i) < 2^{n-1}$, $g(i) = 0$ otherwise. Apply Grover's algorithm to get an estimate for the first bit of the solution (determine if the minimum is above or under 2^{n-1}), and continue.

2 Lower bounds on oracle model

Grover's algorithm gives us a general method of solving NP-complete problems in $O(\sqrt{2^n})$ time: we could use our oracle for testing potential solutions for the problem (of course this can be done efficiently) and then use Grover's algorithm for searching all 2^n possible solutions. For this reason, it is only natural to ask whether the quadratic speedup to the database search problem achieved by this algorithm may be further improved, so as to help solving NP-complete problems. The answer is unfortunately no, as we will prove in this section. We will show a $\Omega(\sqrt{N})$ lower bound for the number of calls to the oracle that any algorithm must make for computing a closely related problem, $Or(Oracle)$ (e.g. for determining whether a certain formula is satisfiable or not).

The idea behind the proof is that if the number of calls to the oracle is too small, then there must be a certain input variable x, which is not queried enough to determine its value; we may change this value and still get the same output from the algorithm, so it cannot possibly compute $Or(Oracle)$ correctly.

Any algorithm that solves $Or(Oracle)$ can be divided into a chain of operators:

$$U_T O U_{T-1} ... O U_1 O U_0 |0\rangle$$

where each O is a call to the oracle and the U_is are unitary transformations not involving calls to the oracle. We denote by $|\phi_t\rangle$ the state before the t'th call to the oracle ($|\phi_t\rangle = U_{t-1} O ... U_0 |0\rangle$). We also denote by $q_x(\phi_t)$ the sum of square amplitudes of configurations which are querying the oracle on x during this call. We refer to $q_x(\phi_t)$ as the query magnitude of x in $|\phi_t\rangle$.

Since the sum of query magnitudes for each ϕ_t is 1, $\sum_{t=1}^{T} \sum_x q_x(\phi_t) = \sum_{t=1}^{T} 1 = T$, we have that there exists x such that

$$\sum_{t=1}^{T} q_x(\phi_t) \leq \frac{T}{N}$$

Let us consider two oracles: O such that $O(y) \equiv 0$ for all y, and O' such that $O'(y) = \delta_{xy}$. If our algorithm is to successfully distinguish between O and O', then we must be able to tell $|\phi_T\rangle$ from $|\phi'_T\rangle$ up to a constant probability, which means (as we have proved in exercise 1) that their difference must be greater than a constant, say $\frac{1}{2}$:

$$\frac{1}{2} < \| |\phi_T\rangle - |\phi'_T\rangle \| = \| U_T O ... O U_0 |0\rangle - U_T O' ... O' U_0 |0\rangle \|.$$

We want to bound the difference between the case in which we always apply O and the case in which we always apply O'. To do this, we write this difference as a telescopic sum of differences, between the following states. Define

$$|\phi_T\rangle_i = U_T O ... O U_i O' .. O' U_0 |0\rangle$$

to be the state in which we apply O' in the first i calls to the oracle, and we apply O from then on. Clearly, $|\phi_T\rangle_0 = |\phi_T\rangle$ and $|\phi_T\rangle_T = |\phi'_T\rangle$,

$$|||\phi_T\rangle - |\phi'_T\rangle|| = ||\sum_{t=1}^{T}(|\phi_T\rangle_t - |\phi'_T\rangle_{t-1})|| \leq \sum_{t=1}^{T} |||\phi_T\rangle_t - |\phi'_T\rangle_{t-1}||.$$

Thus, we gradually move from applying O to applying O', and the difference between two subsequent states is only in one call to the oracle. If we let $E_t = |\phi_T\rangle_t - |\phi_T\rangle_{t-1}$ we get that the above sum is

$$\sum_i ||E_t||.$$

Let us try to estimate E_t. We have two algorithms, which are the same up to the t'th query. Both are in the state $|\phi'_t\rangle$ before the t'th query. Then in the t'th query, one calls O and the other calls O'. The call to O doesn't change the state, while the call to O' flips the sign of all configurations in $|\phi'_t\rangle$ which are querying x. The vector which is the difference between the states after these two queries, E_t, is thus exactly twice the projection of the state on all the configurations which query x. We know that the norm of this projection is $\sqrt{q_x(\phi_t)}$, so $||E_i|| = 2\sqrt{q_x(\phi_t)}$, or $||E_i||^2 = 4q_x(\phi_t)$. We have:

$$\sum_i ||E_i|| \cdot 1 \leq \sqrt{\sum_i ||E_i||^2 \cdot \sum_i 1^2} \leq \sqrt{4\sum_x q_x(\phi_t)} \cdot \sqrt{T} \leq \sqrt{\frac{4T^2}{N}} = \frac{2T}{\sqrt{N}}$$

where we have used Cauchy-Schwarz for the first inequality. Since we know that $\sum_t ||E_t||$ must be larger than half, we have $T = \Omega(\sqrt{N})$.

What does this tell us about NP and quantum computation? We have shown that one cannot expect to solve NP-complete problems through the general Oracle model. This does not mean that we will never be able to solve NP-complete problems quantumly, but that if anyone were ever to solve them, they would be using a certain “insight” for the specific problem at hand. Such an insight about the structure of NP complete problems seems very far from us right now.

Next we wish to prove this lower bound again with different arguments, but first we must introduce the important concept of *density matrices.*

3 Density matrices

The model we have been using so far for quantum computation restricted our attention to systems with pure states (states which can be described by a unit vector in a Hilbert space of dimension 2^n). In general, however, a quantum system is not in a pure state. This may be attributed to the fact that we have only partial knowledge about the system (e.g. after performing a measurement of some of the qubits) or that the system is not isolated from the rest of the universe. We say that the system is in a *mixed state*, and assign with the system a probability distribution, or *mixture* of pure states, denoted by $\{p_\alpha, |\alpha\rangle\}$. This means that the system is with probability p_α in the pure state $|\alpha\rangle$. As an alternative description, We now wish to introduce the notion of **density matrices**, invented by Von Neumann in 1927, which facilitates the treatment of such systems.

Given a pure state $|\alpha\rangle$, we define its density matrix as

$$\rho_\alpha = |\alpha\rangle\langle\alpha|$$

or equivalently, if $|\alpha\rangle = \sum_i c_i|i\rangle$,

$$(\rho_\alpha)_{ij} = C_i C_j^*.$$

Given a mixed state, with probability distribution $\{p_\alpha, |\alpha\rangle\}$, we define its density matrix as:

$$\rho_\alpha = \sum p_\alpha |\alpha\rangle\langle\alpha|$$

Is this representation of mixed states by density matrices unique? The answer is no, as we may notice from the next two single-qubit systems. First, consider the mixed state

$$Pr|0\rangle = \frac{1}{2}, Pr|1\rangle = \frac{1}{2}$$

Whose density matrix is:

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}.$$

Now consider the mixed state

$$Pr\frac{|0\rangle + |1\rangle}{\sqrt{2}} = \frac{1}{2}, Pr\frac{|0\rangle - |1\rangle}{\sqrt{2}} = \frac{1}{2}$$

Let us write down its density matrix:

$$\rho = \frac{1}{2}\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}.$$

At first glance, it seems as though we've lost information with this new notation: two different states have the same representation. However, it turns out that these states are not really that different. From the practical viewpoint of measurement, they are actually completely equivalent! We will now prove that generally, two states whose density matrices are the same, cannot be distinguished by measurement in any basis. Suppose we have a mixed state $\{p_\alpha, |\alpha\rangle\}$, and that we wish to measure it in the basis $\{|\beta_k\rangle\}$. The probability of measuring the basis state $|\beta_k\rangle$ is given by

$$Pr(\beta_k) = \sum P_\alpha |\langle\alpha|\beta_k\rangle|^2 = \sum P_\alpha \langle\beta_k|\alpha\rangle\langle\alpha|\beta_k\rangle = \langle\beta_k|(\sum_\alpha P_\alpha|\alpha\rangle\langle\alpha|)|\beta_k\rangle = \langle\beta_k|\rho|\beta_k\rangle$$

The probability is dependent only of ρ. This shows that there is no loss of information in the representation by density matrices.

3.1 Properties of the density matrix

1. $trace(\rho) = 1$.
   Proof: As the diagonal elements of ρ are simply the probabilities of measuring the corresponding basis states, their sum is obviously 1.

2. ρ is hermitian: $\rho = \rho^\dagger$.
   Proof: This follows directly from the definition of the density matrix of a pure state, and the fact that a sum of hermitian matrices is hermitian.

3. The eigenvalues of ρ are nonnegative.
   Proof: Let $|\beta\rangle$ be an eigenvector of ρ, then $Pr(\beta) = \langle\beta|\rho|\beta\rangle = \lambda_\beta |\beta|^2$. Therefore $\lambda_\beta \geq 0$.

3.2 Reduced density matrices

An important property of the density matrix model is that any subset of the qubits can be described within the model. Suppose we have two sets of qubits, A and B, whose state is described completely by the vector $|\psi\rangle$. Suppose also that we aren't interested in the state of any of the qubits in B, and we wish to describe the state of the bits in A alone. This process is referred to as "tracing out" the bits in B. The general idea behind the process is averaging over all possible states in B.

Before we start our calculations, let us note that operations on A's qubits and operations on B's qubits commute. We have already shown this in week 2. Now, denote

$$|\psi\rangle = \sum_{i,j} \alpha_{ij} |i\rangle_A |j\rangle_B$$

then the corresponding density matrix for $|\psi\rangle$ would be

$$\rho = |\psi\rangle\langle\psi| = \sum_{i,j,i',j'} \alpha_{ij}\alpha_{i'j'} |i\rangle|j\rangle\langle i'|\langle j'|$$

How can we describe the state of A's qubits? It is a mixed state, whose pure states correspond to the possible outcomes of measurement for B's qubits. For each basis state $|j\rangle$ of B, we use the "extended inner product" $\langle j|\psi\rangle = \sum_i \alpha_{ij}|i\rangle$ to describe the state of A conditioned that $|j\rangle$ was measured on B. The state of A is thus a mixture of these vectors (normalized), where the probability of each state is the probability of measuring the corresponding $|j\rangle_B$:

$$\{Pr(j), |\frac{\langle j|\psi\rangle}{\sqrt{Pr(j)}}\}$$

¿From which we get the **reduced density matrix** $\rho|_A$:

$$\rho|_A = \sum_j Pr(j) \cdot \frac{\langle j|\psi\rangle\langle\psi|j\rangle}{Pr(j)} = \sum_j \langle j|\psi\rangle\langle\psi|j\rangle = \sum_j \langle j|\rho|j\rangle$$

Note that $\langle j|\rho|j\rangle$ is simply the j'th block on the main diagonal of ρ, so $\rho|_A$ can be viewed as "folding" ρ to A's dimensions.

Example. Consider the EPR-pair $\frac{|00\rangle+|11\rangle}{\sqrt{2}}$. Its density matrix is given by:

$$\rho = \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

Tracing out the second qubit, we have:

$$\rho|_A = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

This agrees with our previous result, regarding the density matrix of a qubit which is $|0\rangle$ with 50% percent probability and $|1\rangle$ with 50% percent.

Consider now the vector $\sum_i |\psi_i\rangle_A|i\rangle_B$. Suppose we wish to trace out A. We have:

$$\rho = \sum_{i,j} |\psi_i\rangle_A|i\rangle_B\langle j|_B\langle\psi_j|_B$$

Suppose $\{|k\rangle\}$ is a basis for A, then from our previous results, we can compute the reduced density matrix $\rho|_B$:

$$\rho|_B = \sum_k \langle k| \left(\sum_{i,j} |\psi_i\rangle_A|i\rangle_B\langle j|_B\langle\psi_j|_B \right) |k\rangle =$$

$$= \sum_{i,j,k} \langle k|\psi_i\rangle|i\rangle_B\langle j|_B\langle\psi_j|k\rangle = \sum_{i,j,k} |i\rangle_B\langle j|_B\langle\psi_j|k\rangle\langle k|\psi_i\rangle$$

$$= \sum_{i,j} |i\rangle_B\langle j|_B\langle\psi_j| \left(\sum_k |k\rangle\langle k| \right) |\psi_i\rangle = \sum_{i,j} |i\rangle\langle j|\langle\psi_j|\psi_i\rangle$$

in the last equality, $(\sum_k |k\rangle\langle k|)$ is the identity matrix.

The probability of measuring $|i\rangle$ is $(\rho|_B)_{ii}$. If we measure ρ in the standard basis, the off-diagonal elements will vanish as we will have a superposition of classical outcomes. These elements determine the amount of “quantumness” of the system; the higher their amplitudes, the more the qubits are entangled. We will use this property in our next proof.

4 Ambainis’s Lower Bound for Grover’s Algorithm

4.1 Introduction

We now apply Grover’s algorithm on a superposition of oracles. We begin with a full density matrix (all elements are non-zero), and we show that during the algorithm the entanglement of the different oracles exceeds. By the end of the algorithm, the density matrix is diagonal, thus we reach full entanglement. We use this property to show that Grover’s algorithm requires at least $O(\sqrt{N})$ oracle calls.

4.2 Calculating the density matrices

We define the *i'th oracle* as follows: $f(i) = 1$, $f(j) = 0$ $for\ i \neq j$. We define the unitary matrix O as the transformation matrix $|x\rangle \longmapsto (-1)^{f(x)}|x\rangle$. O' is the transformation which receives as input $|x\rangle \otimes |i\rangle$ and applies the *i'th oracle* on $|x\rangle$.

Now assume that Grover's algorithm was achieved by a series of unitary transformations, $U_T...U_1OU_0$, starting with a $|0...0\rangle$ input (The algorithm workspace). We replace each U_i with $U_i \otimes I$ and each occurrence of O with O', now starting with a $|0...0\rangle \otimes |i\rangle$ input (We add an Oracle workspace). The result will be some state in the form $|\psi_i\rangle \otimes |i\rangle \otimes |i\rangle$. ¿From linearity, if we now start with a superposition $\sum |i\rangle$ of all possible $i's$ in the Oracle workspace, the algorithm will produce

$$\sum_i |\psi_i\rangle \otimes |i\rangle \otimes |i\rangle$$

Assume our database's size is N. Our initial state is $|0...0\rangle \otimes |i\rangle$, so the density matrix, reduced to the oracle workspace, is

$$\rho_0 = \begin{pmatrix} \frac{1}{N} & \cdots & \frac{1}{N} \\ . & \cdots & . \\ . & \cdots & . \\ \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}$$

At the end of the algorithm, if the algorithm 'works properly', the density matrix reduced to the oracle workspace is

$$\rho_T = \begin{pmatrix} \frac{1}{N} & \cdots & 0 \\ . & \cdots & . \\ . & \cdots & . \\ 0 & \cdots & \frac{1}{N} \end{pmatrix} = \frac{1}{N} I$$

4.3 Error estimate

Let f(i) be the function which is computed by the algorithm on the i'th oracle, in our case f(i)=i.

Lemma: Let the starting state be a superposition of the oracles, $\sum \alpha_i |i\rangle$. Assume that the algorithm computes f with error probability $\leq \epsilon$. If $f(i) \neq f(j)$, then $|\rho_T(i,j)| \leq 2\sqrt{\epsilon}\sqrt{1-\epsilon}|\alpha_i||\alpha_j|$

Proof: At the end of the algorithm, the general state is $|\psi\rangle = \sum \alpha_i |\psi_i\rangle |i\rangle$.
$(\rho|_O)(i,j) = \alpha_i \alpha_j^* \langle \psi_j | \psi_i \rangle$
$|\rho_T(i,j)| = |\alpha_i||\alpha_j||\langle \psi_j | \psi_i \rangle|$

Since $f(i) \neq f(j)$, there is a differentiating bit b, assume the bit b to be $|0\rangle$ in $f(i)$, $|1\rangle$ in $f(j)$. Then there exist some $\epsilon_1, \epsilon_2 \leq \epsilon$ such that:

$|\psi_i\rangle = \sqrt{1-\epsilon_1}|\alpha\rangle|0\rangle + \sqrt{\epsilon_1}|\beta\rangle|1\rangle$ (the probability to measure 0 in the bit b in f(i)).

$|\psi_j\rangle = \sqrt{1-\epsilon_2}|\gamma\rangle|1\rangle + \sqrt{\epsilon_2}|\delta\rangle|0\rangle$ (the probability to measure 0 in the bit b in f(j)).

Therefore $\langle \psi_j | \psi_i \rangle = \sqrt{1-\epsilon_1}\sqrt{\epsilon_2} + \sqrt{1-\epsilon_2}\sqrt{\epsilon_1} \leq 2\sqrt{1-\epsilon}\sqrt{\epsilon}$

4.4 Proof of the lower bound

Let

$$S_k = \sum_{i \neq j} |\rho_{i,j}^{(k)}|$$

Then $S_0 = N-1$, $S_T \leq 2\sqrt{\epsilon}\sqrt{1-\epsilon}(N-1)$. We will show that for every k, $S_{k-1} - S_k \leq 2\sqrt{N-1}$, which will prove the lower bound of $T = O(\sqrt{N})$.

By the triangular inequality we get:

$$S_{k-1} - S_k = \sum_{i \neq j} |\rho_{k-1}^{(ij)}| - \sum_{i \neq j} |\rho_k^{(ij)}| \leq \sum_{i \neq j} |\rho_{k-1}^{(ij)} - \rho_k^{(ij)}|$$

We now write the state as a combination of the algorithm's base states, rather than the oracle's:

$$\psi_{k-1} = \sum_{i,z} \sqrt{P_{i,z}} |i,z\rangle_A \otimes |\psi_{i,z}\rangle_O$$

And then

$$\rho_{k-1} = \sum_{i,z} P_{i,z} |\psi_{i,z}\rangle\langle\psi_{i,z}|$$

$$S_{k-1} - S_k \leq \sum_{i \neq j} \sum_{l,z} P_{lz} |\rho_{k-1}^{lz}(ij) - \rho_k^{lz}(ij)| = \sum_{l,z} P_{lz} \sum_{i \neq j} |\rho_{k-1}^{lz}(ij) - \rho_k^{lz}(ij)|$$

We notice that the probability factor P_{lz} is independent on k, since the oracle only changes the bit's phase when $i = j$

We conclude that it is enough to bound the last expression for fixed l, z.

We claim that:

$$\sum_{i \neq j} |\rho^{lz}_{k-1}(ij) - \rho^{lz}_{k}(ij)| \leq 4\sqrt{N-1}.$$

We write:

$$|\phi_{l,z}\rangle = \sum_j \alpha_{l,z,j} |j\rangle.$$

We know that the density matrix of the oracle does not change by the unitary operator on the algorithm's register. hence, the density matric $\rho^k_{l,z}$ is achieved by applying the oracle on $|\phi_{l,z}$, and so

$$\rho^k_{l,z} = |\phi'_{l,z}\rangle\langle\phi'_{l,z}|$$

where

$$|\phi_{l,z}\rangle = \sum_j \alpha'_{l,z,j} |j\rangle$$

Once l, z are fixed, $\alpha'_{l,z,j}$ is different from $\alpha_{l,z,j}$ only if $l = j$ since only the lth oracle applies a minus sign on l. In the density matrix language, the density matrices are different only on the l'th column and row. The change is always a phase flip, therefore the distance between the matrices is the sum of the distance on the l'th row and the l'th column. For the l'th column, the sum of the absolute values of the differences is

$$\sum_j 2|\alpha_{lzj}\alpha^*_{lzj}| = 2|\alpha^*_{lzj}| \sum_j |\alpha_{lzj}|$$

but $|\alpha^*_{lzj}| \leq 1$, and $\sum_j |\alpha_{lzj}| \leq \sqrt{N-1}\sqrt{\sum_j |\alpha_{l,z,j}|^2}$ by the Cauchy-Schwartz inequality. Since $\sum_j |\alpha_{l,z,j}|^2 \leq 1$ we have proved that the sum over the column is less than $2\sqrt{N-1}$. The same argument works for the row, which proves the above inequality.