Hypothesis elimination on a quantum computer

Andrei N. Soklakov† and Rüdiger Schack‡

Department of Mathematics, Royal Holloway, University of London
Egham, Surrey, TW20 0EX, UK

3 December 2004

Abstract

Hypothesis elimination is a special case of Bayesian updating, where each piece of new data rules out a set of prior hypotheses. We describe how to use Grover’s algorithm to perform hypothesis elimination for a class of probability distributions encoded on a register of qubits, and establish a lower bound on the required computational resources.

1 Introduction

In the standard setting of Bayesian inference one starts from a set of hypotheses $\mathbb{H} = \{h\}$ and a set of possible data $\mathbb{D} = \{d\}$. Hypotheses and data are connected via conditional probabilities $p(d|h)$, known as the model. Given a prior probability distribution $p(h)$, the updated probability $p(h|d)$ of the hypothesis $h$ to be true given that the piece of data $d$ was observed is obtained from Bayes’s rule, \[ p(h|d) = \frac{p(d|h)p(h)}{\sum_{h \in \mathbb{H}} p(d|h)p(h)}. \] \[ (1) \]

In this paper, we consider the problem of hypothesis elimination, which is a special case of Bayesian updating where the model is of the form

$$p(d|h) = \begin{cases} 
0 & \text{if } h \text{ is ruled out by } d, \\
 c_h & \text{otherwise.} 
\end{cases}$$ \[ (2) \]

The positive constant $c_h$ does not depend on $d$ and is determined by normalization. We assume that there is a finite number, $N$, of hypotheses, which we label 0 to $N - 1$, i.e.,
$\mathbb{H} = \{0, 1, \ldots, N - 1\}$. Furthermore, we assume that the prior $p(h)$ has been obtained by hypothesis elimination from an initial uniform prior distribution on $\mathbb{H}$.

There are several possible ways of encoding a probability distribution $p(h)$ on a quantum register. Here we represent $p(h)$ by the state

$$|\psi\rangle = \sum_{h=0}^{N-1} \sqrt{p(h)} |h\rangle ,$$

where $|h\rangle$ are computational basis states [2] of a register formed of $\lceil \log_2 N \rceil$ qubits.

To be specific, we assume that the prior $p(h)$ has been obtained by $k - 1$ hypothesis elimination steps ($k > 1$). We thus assume that the prior is given in the form of a sequence of oracles $o_1, \ldots, o_{k-1}$, where

$$o_j(h) = \begin{cases} 0 & \text{if } h \text{ is ruled by the data in step } j, \\ 1 & \text{otherwise.} \end{cases}$$

Likewise, we assume that the model Eq. (2) is given as an oracle $o_k$, where

$$o_k(h) = c_h^{-1} p(d|h).$$

The sequence of oracles $o_j$ gives rise to a new sequence $O_1, \ldots, O_k$ defined by

$$O_j(h) = \begin{cases} 1 & \text{if } \forall i \leq j : o_i(h) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

For each $O_j$, we define the set of solutions, $\Omega_j = \{h : O_j(h) = 1\}$, and denote by $M_j$ the corresponding number of solutions,

$$M_j = \sum_{h=0}^{N-1} O_j(h).$$

For $j \in \{k - 1, k\}$, we now define the states

$$|\Psi_j\rangle = \frac{1}{\sqrt{M_j}} \sum_{h \in \Omega_j} |h\rangle .$$

The state $|\Psi_{k-1}\rangle$ encodes the prior $p(h)$, and $|\Psi_k\rangle$ encodes the posterior $p(h|d)$.

### 2 Quantum hypothesis elimination

The problem of hypothesis elimination now takes the following form: Given a quantum register in the prior state $|\Psi_{k-1}\rangle$ and given the oracles $O_1, \ldots, O_k$, transform the register state into the posterior state $|\Psi_k\rangle$. One can use Grover’s algorithm [3] to solve this problem as follows.

Define a quantum oracle corresponding to $O_k$ via

$$\hat{O}_k|h\rangle = (-1)^{O_k(h)} |h\rangle .$$

There are standard techniques [2] to implement $\hat{O}_k$ in the form of a quantum circuit. The Grover operator $\hat{G}(O_k)$ associated with the oracle is then defined as

$$\hat{G}(O_k) = (2|\Psi_0\rangle\langle\Psi_0| - \mathbb{I}) \hat{O}_k .$$
where $\mathbb{1}$ is the identity operator and

$$
|\Psi_0\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle 
$$

is the equal superposition state. The posterior state $|\Psi_k\rangle$ can now be prepared by repeated application of the Grover operator $\hat{G}(O_k)$ to the equal superposition state $|\Psi_0\rangle$. This requires $(\pi/4)\sqrt{N/M_k}$ calls of the oracle $\hat{O}_k$.

Notice that the hypothesis elimination algorithm outlined above makes no direct use of the prior state $|\Psi_{k-1}\rangle$. This raises the following question: Is it possible to reduce the number of Grover iterations (and therefore oracle calls) required to prepare $|\Psi_k\rangle$ by starting from the prior state $|\Psi_{k-1}\rangle$ instead of the equal superposition state $|\Psi_0\rangle$? In other words, can one make use of the computational effort that went into preparing the prior state $|\Psi_{k-1}\rangle$ in order to obtain the posterior state $|\Psi_k\rangle$ more efficiently? As already suggested by the results in Ref. [4], the answer to this question is negative. Here we prove the following result.

Consider the family of oracles $\mathcal{O}$ that consists of $O_k$, $O_{k-1}$ and all possible combinations of $O_k$ and $O_{k-1}$ (see Eq (29) for the precise definition of this family). Now consider all possible algorithms that consist of applying the corresponding Grover operators $\mathcal{A}_{abc...} = \ldots \hat{G}(O_c)\hat{G}(O_b)\hat{G}(O_a)$, $O_a, O_b, O_c, \cdots \in \mathcal{O}$. (11)

Then, in the limit of large $N$, any algorithm $\mathcal{A}_{abc...}$ requires at least $(\sqrt{2}/8)\sqrt{N/M_k}$ oracle calls from the above family to convert $|\Psi_{k-1}\rangle$ into $|\Psi_k\rangle$. In other words, making direct use of the prior state $|\Psi_{k-1}\rangle$ does not improve the asymptotic cost of $O(\sqrt{N/M_k})$ oracles calls required to prepare $|\Psi_k\rangle$.

### 3 Proof

Consider an oracle $O$ which accepts $M$ out of the total $N$ hypotheses $h$:

$$
\sum_{h=0}^{N-1} O(h) = M.
$$

We shall call such hypotheses good, as opposed to bad hypotheses that are rejected by the oracle. Using different notation [5] for the amplitudes of good and bad hypotheses, we have that after $t$ consecutive applications of the Grover operator $\hat{G}(O)$ an arbitrary quantum state

$$
|\Psi_{\text{ini}}\rangle = \sum_{\text{good } h} g^\text{ini}_h |h\rangle + \sum_{\text{bad } h} b^\text{ini}_h |h\rangle 
$$

is transformed into

$$
|\Psi_{\text{fin}}\rangle = \hat{G}^t(O) |\Psi_{\text{ini}}\rangle = \sum_{\text{good } h} g^\text{fin}_h |h\rangle + \sum_{\text{bad } h} b^\text{fin}_h |h\rangle. 
$$

(14)
Let $\bar{g}^{\text{ini}}$ and $\bar{b}^{\text{ini}}$ be the averages of the initial amplitudes corresponding to the good and the bad hypotheses respectively:

$$
\bar{g}^{\text{ini}} = \frac{1}{M} \sum_{\text{good } h} g_h^{\text{ini}}, \quad \bar{b}^{\text{ini}} = \frac{1}{N - M} \sum_{\text{bad } h} b_h^{\text{ini}},
$$

and similarly for the final amplitudes

$$
\bar{g}^{\text{fin}} = \frac{1}{M} \sum_{\text{good } h} g_h^{\text{fin}}, \quad \bar{b}^{\text{fin}} = \frac{1}{N - M} \sum_{\text{bad } h} b_h^{\text{fin}},
$$

Let us also define

$$
\Delta g^{\text{ini}}_h = g_h^{\text{ini}} - \bar{g}^{\text{ini}}, \quad \Delta b^{\text{ini}}_h = b_h^{\text{ini}} - \bar{b}^{\text{ini}}.
$$

In other words, $\Delta g^{\text{ini}}_h$ and $\Delta b^{\text{ini}}_h$ define the features of the initial amplitude functions $g_h^{\text{ini}}$ and $b_h^{\text{ini}}$ relative to their averages $\bar{g}^{\text{ini}}$ and $\bar{b}^{\text{ini}}$. Biham et. al. [4] have shown that the change of the amplitudes is essentially determined by the change of the averages:

$$
g_h^{\text{fin}} = \bar{g}^{\text{fin}} + \Delta g^{\text{ini}}_h, \quad b_h^{\text{fin}} = \bar{b}^{\text{fin}} + (-1)^t \Delta b^{\text{ini}}_h,
$$

where the averages $\bar{g}^{\text{fin}}$ and $\bar{b}^{\text{fin}}$ are given as follows. Define

$$
\omega = \arccos \left( 1 - \frac{2M}{N} \right),
$$

$$
\alpha = \sqrt{ |\bar{b}^{\text{ini}}|^2 + |\bar{g}^{\text{ini}}|^2 \frac{M}{N - M} },
$$

$$
\phi = \arctan \left( \frac{\bar{g}^{\text{ini}}}{\bar{b}^{\text{ini}} \sqrt{M/(N - M)}} \right).
$$

The averages are given by

$$
\bar{g}^{\text{fin}} = \sqrt{\frac{N - M}{M}} \alpha \sin(\omega t + \phi),
$$

$$
\bar{b}^{\text{fin}} = \alpha \cos(\omega t + \phi).
$$

Let us also define the separation of the averages

$$
\Delta^{\text{ini}} = \bar{g}^{\text{ini}} - \bar{b}^{\text{ini}}, \quad \Delta^{\text{fin}} = \bar{g}^{\text{fin}} - \bar{b}^{\text{fin}}.
$$

Equations (18) imply that after $t$ applications of the Grover operator the values of individual amplitudes $g_h^{\text{fin}}$ of accepted hypotheses do not change relative to their average $\bar{g}^{\text{fin}}$

$$
g_h^{\text{fin}} - \bar{g}^{\text{fin}} = \Delta g_h^{\text{ini}} = g_h^{\text{ini}} - \bar{g}^{\text{ini}}.
$$

The same is true for the rejected hypotheses if $\Delta b^{\text{ini}} = 0$ or $t$ is even. This observation suggests that the action of any algorithm of the type

$$
A_{aabbcc...} = \ldots \hat{G}^2(O_c)\hat{G}^2(O_b)\hat{G}^2(O_a)
$$

(25)
can be analyzed by looking at the changes of the average amplitudes of the accepted and rejected hypotheses relative to various oracles $\hat{O}_a, \hat{O}_b, \ldots$. Before we proceed with this analysis let us first clarify what kind of oracles are relevant to this problem.

Let $f$ be a real-valued step function which takes only three values,

$$f(h) = \begin{cases} 
  f_1 & \text{if } O_k(h) = 1, \\
  f_2 & \text{if } O_{k-1}(h) = 1, \text{ but } O_k(h) = 0, \\
  f_3 & \text{if } O_{k-1}(h) = 0.
\end{cases} \quad (26)$$

Let us also require that $\sum_{h=0}^{N-1} f^2(h) = 1$, and introduce a quantum state $|\Psi(f)\rangle$ that is defined by $f$ in a natural way,

$$|\Psi(f)\rangle = \sum_{h=0}^{N-1} f(h)|h\rangle. \quad (27)$$

Evidently, both $|\Psi_k\rangle$ and $|\Psi_{k-1}\rangle$ can be written in this way. It follows from the above discussion that the action of the operator $\hat{G}^2(O_k)$ on the state (27) can be completely described by the changes of $f_1, f_2$ and $f_3$. Moreover $\hat{G}^2(O_k)$ preserves the value of $\delta_{k-1} = f_2 - f_3$. Similarly, the action of $\hat{G}^2(O_{k-1})$ on (27) can be completely described by the changes of $f_1, f_2$ and $f_3$, and it preserves the value of $\delta_k = f_1 - f_2$. In general, for any oracle $O$, the corresponding operator $\hat{G}^2(O)$ preserves the amplitude differences between any two hypotheses for as long as either both hypotheses are accepted or both are rejected by $O$. However, if $f_{\text{good}}$ and $f_{\text{bad}}$ denote the amplitudes of an accepted and a rejected hypothesis, respectively, then the difference $\delta = f_{\text{good}} - f_{\text{bad}}$ is changed by an amount $\Delta$ which satisfies the inequality

$$|\Delta| \leq \frac{4\sqrt{2}}{\sqrt{N - M}}, \quad (28)$$

where $M$ is the number of accepted hypotheses with respect to $O$ (see the Appendix). Using this inequality the action of any algorithm of the type (25) can be analyzed by calculating how individual changes of $\delta_k$ and $\delta_{k-1}$ accumulate during the action of the algorithm. In order to convert $|\Psi_{k-1}\rangle$ into $|\Psi_k\rangle$, the net result of such changes must be sufficient to increase $\delta_k$ from 0 to $1/\sqrt{M_k}$ and decrease $\delta_{k-1}$ from $1/\sqrt{M_{k-1}}$ to 0.

It follows that all oracles that are relevant for this task can be obtained from $O_k$ and $O_{k-1}$. Since each oracle is completely characterized by the set of acceptable hypotheses, the relevant family of oracles generated by $O_k$ and $O_{k-1}$ can be written out as oracles that correspond to the sets

$$\Omega_k, \bar{\Omega}_k, \Omega_{k-1}, \bar{\Omega}_{k-1}, \Omega_k \cap \bar{\Omega}_{k-1}, \text{ and the complementary } \bar{\Omega}_k, \Omega_{k-1}, \Omega_k \cup \bar{\Omega}_{k-1}. \quad (29)$$

Let us consider the first three oracles from the family defined by the sets (29), namely the oracles that accept hypotheses from the sets

$$\Omega_k, \Omega_{k-1}, \Omega_{k-1} \cap \bar{\Omega}_k. \quad (30)$$
Oracles that correspond to the complementary sets in (29) can be analyzed in a completely analogous manner. The oracles corresponding to the sets (30) are \( O_k, O_{k-1} \) and \( O_{k-1}^c \), where

\[
O_{k-1}^c(h) = \begin{cases} 1 & \text{if } h \in \Omega_{k-1} \cap \tilde{\Omega}_k \\ 0 & \text{otherwise} \end{cases}.
\] (31)

Using the inequality (28) we see that, regardless of its position in the algorithm, the operator \( \hat{G}^2(O_k) \) changes \( \delta_k \) by at most \( d_1 = 4\sqrt{2}/\sqrt{N - M_k} \) without changing \( \delta_{k-1} \). Similarly, the operator \( \hat{G}^2(O_{k-1}) \) changes \( \delta_{k-1} \) by at most \( d_2 = 4\sqrt{2}/\sqrt{N - M_{k-1}} \) without affecting the value of \( \delta_k \). The Grover operator with the combined oracle \( \hat{G}(O_{k-1}^c) \) changes both \( \delta_k \) and \( \delta_{k-1} \) by an equal amount that does not exceed \( d_3 = 4\sqrt{2}/\sqrt{N - (M_{k-1} - M_k)} \).

Let \( N(O) \) be the number of times that the oracle \( O \) is called within the algorithm. We would like to find a lower bound on the total number of oracle calls, \( N_{\text{total}} = N(O_k) + N(O_{k-1}) + N(O_{k-1}^c) \), that is needed by the algorithm to convert \( |\Psi_{k-1}\rangle \) into \( |\Psi_k\rangle \). Let \( x, y \) and \( z \) be the number of times that the operators \( \hat{G}^2(O_k), \hat{G}^2(O_{k-1}) \) and \( \hat{G}^2(O_{k-1}^c) \) appear in the algorithm. Then \( N_{\text{total}} \) is bounded from below by the minimal value of \( 2(x + y + z) \) subject to the constraints

\[
\begin{align*}
xd_1 + zd_3 & \geq U, \\
yd_2 + zd_3 & \geq V,
\end{align*}
\] (32)

where \( U = 1/\sqrt{M_k} \) and \( V = 1/\sqrt{M_{k-1}} \) are the required changes of \( \delta_k \) and \( \delta_{k-1} \) respectively. This is a simple linear optimization problem that should be considered in the nonnegative \( x, y, z \) octant. Keeping \( M_k \) and \( M_{k-1} \) constant we obtain that in the limit of large \( N \) the value of \( N_{\text{total}} \) is approaching \((\sqrt{2}/4)\sqrt{N/M_k}\).

For any algorithm \( \mathcal{A}_{abc...} \) the action of the corresponding sequence of Grover operators can be rewritten in the form similar to that in Eq. (25). Using the definition of the Grover operator, we have

\[
\hat{G}(O_{a_1}) \ldots \hat{G}(O_{a_3}) \hat{G}(O_{a_2}) \hat{G}(O_{a_1}) = (\hat{A} \hat{O}_{a_1}) \ldots (\hat{A} \hat{O}_{a_3}) (\hat{A} \hat{O}_{a_2}) (\hat{A} \hat{O}_{a_1}),
\] (33)

where \( \hat{A} = 2|\Psi_0\rangle\langle\Psi_0| - \mathbf{1} \). Using the fact that \( \hat{G}(O_k)^2 = \mathbf{1} \) and \( \hat{A}^2 = \mathbf{1} \) we obtain

\[
(\hat{A} \hat{O}_{a_1}) \ldots (\hat{A} \hat{O}_{a_3}) (\hat{A} \hat{O}_{a_2}) (\hat{A} \hat{O}_{a_1}) = \hat{A} \hat{O}_{a_1} \ldots \hat{A} \hat{O}_{a_3} \hat{A} \hat{O}_{a_2} (\hat{A} \hat{O}_{a_1})^{-1} (\hat{A} \hat{O}_{a_1})^2
= \hat{A} \hat{O}_{a_1} \ldots \hat{A} \hat{O}_{a_3} \hat{A}(\hat{O}_{a_1} a_2 \hat{A}) (\hat{A} \hat{O}_{a_1})^2,
\] (34)

where \( \hat{O}_{a_1 a_2} = \hat{O}_{a_1} \hat{O}_{a_2} \). Denoting \( \hat{O}_{a_1 a_2 \ldots a_j} = \prod_{p=1}^{j} \hat{O}_{a_p} \) we proceed

\[
\begin{align*}
\hat{A} \hat{O}_{a_1} \ldots \hat{A} \hat{O}_{a_3} \hat{A}(\hat{O}_{a_1 a_2} \hat{A})(\hat{A} \hat{O}_{a_1})^2 &= \ldots \hat{A} \hat{O}_{a_1} \hat{A}(\hat{O}_{a_1 a_2} \hat{A})^{-1} (\hat{O}_{a_1 a_2} \hat{A})^2 (\hat{A} \hat{O}_{a_1})^2 \\
&= \ldots \hat{A} \hat{O}_{a_1} \hat{A}(\hat{O}_{a_1 a_2 a_3} \hat{A})^2 (\hat{A} \hat{O}_{a_1})^2 \\
&= \hat{R} \hat{G}^{\pm 2}(O_{a_1 \ldots a_n}) \ldots \hat{G}^{2}(O_{a_1 a_2 a_3}) \hat{G}^{-2}(O_{a_1 a_2}) \hat{G}^{2}(O_{a_1})
\end{align*}
\] (35)

where the + and − signs are chosen for odd and even values of \( n \), respectively, \( O_{a_1 \ldots a_j} \) denote classical oracles that correspond to the quantum oracle \( \hat{O}_{a_1 \ldots a_j} \), and \( \hat{R} \) is a residual operator

\[
\hat{R} = \begin{cases} \hat{G}(O_{a_1 \ldots a_n}) & \text{if } n \text{ is even}, \\
\hat{G}^{-1}(O_{a_1 \ldots a_n}) & \text{otherwise}.
\end{cases}
\] (36)
Since all oracles, \( \hat{O}_{a_1...a_j} \), belong to the family associated with the sets (29), the above arguments allow us to derive a bound on the minimum number of oracle calls from this family that are required to convert \(|\Psi_{k-1}\rangle\) to \(|\Psi_k\rangle\). Indeed, transformations between Eqs. (33) and (36) at most double the number of oracle calls that are used by the original algorithm. To be more precise, if \( n \) is the number of oracle calls used by the original algorithm (see the left-hand side of Eq. (33)), then the equivalent modified algorithm, defined by the right-hand side of Eq. (35), requires at most \( 2n + 1 \) oracle calls. It remains to note that after the application of the residual operator that concludes the modified algorithm one has to arrive at the target state \(|\Psi_k\rangle\), or, which is equivalent, the algorithm

\[
\hat{G}^{\pm 2}(O_{a_1...a_n}) \cdots \hat{G}^2(O_{a_1a_2a_3})\hat{G}^{-2}(O_{a_1a_2})\hat{G}^2(O_{a_1})
\]

must prepare the state \(|\psi_k\rangle = G^{\pm 1}(O_{a_1...a_n})|\Psi_k\rangle\). In the limit of large \( N \) the state \(|\psi_k\rangle\) coincides with \(|\Psi_k\rangle\). Using an analysis analogous to that of algorithm (25) we therefore conclude that, in the limit of large \( N \), the algorithm (37) requires at least \( (\sqrt{2}/4)\sqrt{N/M} \) oracle calls to convert \( \Psi_{k-1} \) into \( |\psi_k\rangle \). The original algorithm, therefore, will need, asymptotically, at least \( (\sqrt{2}/8)\sqrt{N/M} \) oracle calls to convert \( |\Psi_{k-1}\rangle \) into \( |\Psi_k\rangle \).

Appendix

Using the notation of Eqs. (12–24), the inequality (28) can be written as

\[
|\Delta_{\text{fin}} - \Delta_{\text{ini}}| \leq \frac{4\sqrt{2}}{\sqrt{N-M}},
\]

where the number of iterations is \( t = 2 \). By definition, we have

\[
|\Delta_{\text{fin}} - \Delta_{\text{ini}}| = \alpha \sqrt{\frac{N}{M}} |\sin(2\omega + \xi) - \sin \xi|
= \alpha \sqrt{\frac{N}{M}} |2 \sin \omega \cos(\omega + \xi)|
\leq 4\alpha \sqrt{\frac{N}{M}} \left| \sin \frac{\omega}{2} \right|.
\]

Since \( 0 \leq \omega \leq \pi \)

\[
\sin \frac{\omega}{2} = \sqrt{\frac{1 - \cos \omega}{2}} = \sqrt{\frac{M}{N}},
\]

and therefore

\[
|\Delta_{\text{fin}} - \Delta_{\text{ini}}| \leq 4\alpha.
\]

A bound on \( \alpha \) can be obtained using the fact that, for any \( a_1, a_2, \ldots, a_M \in \mathbb{R} \) such that \( \sum_{k=1}^{M} (a_k)^2 = 1 \), we have

\[
\max \left( \frac{1}{M} \sum_{k=1}^{M} a_k \right) = \frac{1}{\sqrt{M}}.
\]
This can be easily shown using the method of Lagrange multipliers. Using (42) we obtain
\[
\max \bar{g} = \frac{1}{\sqrt{M}}, \quad \text{and} \quad \max \bar{b} = \frac{1}{\sqrt{N - M}}.
\]  
(43)
It then follows that
\[
\alpha < \sqrt{|\max \bar{b}|^2 + |\max \bar{g}|^2 \frac{M}{N - M}} = \frac{\sqrt{2}}{\sqrt{N - M}}.
\]  
(44)
Combining this bound with Eq. (41) we obtain Eq. (38) as intended.

References

[1] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory* (Wiley, Chichester, England, 1994).

[2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[3] L. K. Grover, “Quantum mechanics helps in searching for a needle in a haystack,” Phys. Rev. Lett. 79, 325 (1997).

[4] E. Biham, O. Biham, D. Biron, M. Grassl, and D. A. Lidar, “Grover’s Quantum Search Algorithm for an Arbitrary Initial Amplitude Distribution,” Phys. Rev. A 60, 2742 (1999).

[5] A. N. Soklakov and R. Schack, “Efficient state preparation for a register of quantum bits,” quant-ph/0408045