# Quantum State Discrimination and Concept Learning

---

A Thesis
Presented to
The Division of Mathematics and Natural Sciences
Reed College

---

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Arts

---

Daniel R. Copeland

May 2011

Approved for the Division
(Mathematics)

_____

Jamie Pommersheim

# Acknowledgements

I would like to thank very much my advisor, Jamie, for insight and helpful guidance. And thanks for putting up with me! I also thank my brother Nicholas for lending a listening ear and a bright countenance.

# Table of Contents

# List of Figures

# Abstract

This thesis provides preliminaries to the field of quantum concept learning, including the formalism of quantum computing and general quantum measurement schemes. Once this is developed, we examine problems in quantum concept learning, particularly focused on Hamming distance oracles.

# Introduction

The theory of quantum mechanics was developed to its current state over the last century. In 1982, Richard Feynman, noting the computational difficulty in simulating quantum systems, hypothesized that quantum experiments could be used as computational devices, given tools to produce and measure quantum states. In 1992, David Deutsch and Richard Jozsa proved the theoretical existence of quantum computers by providing the first quantum algorithm. Since then, various quantum algorithms have been discovered that offer speed-up over known classical algorithms, such as Shor's quantum factoring algorithm.

The algorithms we consider are those that try to identify a bit string $a \in \mathbb{Z}_2^n$ from a "black-box", or oracle, that encodes the string somehow. The algorithm is allowed to query the oracle by passing information through it and recording the result. The efficiency of such algorithms is measured by how many queries are required to identify $a$, or the likelihood of determining $a$ with a small number of queries. It turns out that there is enormous speed-up in many such problems when they are converted from classical to quantum algorithms.

A classical bit is simply a binary distinction; an element of $\{0, 1\}$. Under the framework of quantum mechanics, we extend this notion to include any "superposition" of 0 or 1. A qubit (quantum bit), typically expressed in Dirac notation may be written as $|\Psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$. Physicists call this a quantum state; we will additionally refer to it as a qubit. Here $\alpha, \beta \in \mathbb{C}$ such that $|\alpha|^2 + |\beta|^2 = 1$. However, unlike a classical bit, as observers we are unable to extract the complete information contained in the state; namely the coefficients $\alpha$ and $\beta$. Instead, qubits must be "measured" in some manner. One simple example is measurement in the computational basis. In accordance to quantum mechanics, the real values $|\alpha|^2$ and $|\beta|^2$ respectively represent the probability of measuring the states $|0\rangle$ or $|1\rangle$, the possible results of measuring in the computational basis. However, measurement generally changes the state.

How does this notion of a bit fit into a computational framework if we cannot perfectly identify the qubits? The key is that although we are unable to access all the information "contained" in a state, quantum systems are capable of handling large amounts of information in novel and surprising ways so that an observer may deduce specific properties of the system with the small amount of information allotted to her.

At the particle level, matter is observed not to follow classical laws but operate according to solutions of the Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi$$

This produces a wave-function $\Psi$, which encodes a probability density function for legitimate observable quantities, such as position or momentum. The space of all wave functions is the space of $L^2$-integrable functions, which is a Hilbert space. For our purposes, we suppose quantum states to be elements of some finite-dimensional Hilbert space $\mathcal{H}_s$ dubbed the *state space*. Specifically, for quantum algorithms, we require an isomorphism $\mathcal{H}_s \cong \mathbb{C}^{2^n}$ for some integer $n$. Hence, from now on, we forget that states correspond to wave-functions and treat them simply as vectors in a suitable complex inner product space. For a discussion of the physical theory regarding the Schrödinger equation, the reader is referred to [Gri04].

Chapter 1 provides an introduction to the linear algebra description of quantum mechanics, quantum computing, and classical concept learning. Chapter 2 addresses conditions for optimal measurements. Chapter 3 examines various quantum concept learning algorithms. Chapter 4 examines Hamming distance concept learning and provides new numerical results for some problems.

# Chapter 1

# Postulates of Quantum Computing

We now describe the details of the linear algebra formalism of quantum mechanics. This thesis does not address the century-long experimental motivation for this framework, or the many steps the theory has undergone to reach the linear algebra framework (for example the relationship between vectors in state space and the wave function described by the Schrödinger equation). Instead, it describes mathematically the setting in which idealized quantum information processing takes place. A complete introduction to quantum computing is found in [NC00].

## 1.1 The State Space

We begin by describing the space of possible states that may describe a physical system.

**Postulate**: The *state space* $\mathcal{H}_s$ of any closed physical system is a complex Hilbert space. The state of the system is described by a unit vector of $\mathcal{H}_s$ (at any time $t$), $|\Psi_t\rangle$. (Two unit vectors $|\Psi\rangle, |\Psi'\rangle$ cannot be distinguished if $|\Psi\rangle = e^{i\theta} |\Psi'\rangle$ [1].)

The state space $\mathcal{H}_s$ is therefore equipped with a complex inner product.

**Definition.** A **complex inner product** $\langle \mid \rangle : \mathcal{H}_s \to \mathbb{C}$ satisfies the following conditions:

1. $\langle a|\lambda b\rangle = \lambda \langle a|b\rangle$ for $\lambda \in \mathbb{C}$
2. $\langle a|b\rangle = \langle b|a\rangle^*$
3. $\langle a|a\rangle \geq 0$ with equality if and only if $a = 0$.

Throughout this thesis we assume that $\mathcal{H}_s$ is finite-dimensional (unworried by whatever physical implications that carries), and isomorphic to $\mathbb{C}^N$ with the familiar

[1] A factor of $e^{i\theta}$ is called a *global phase factor*. The indistinguishability of two states differing by a global phase factor is due to the fact that any quantum measurement applied to the states admits the same statistical result. This phenomenon will be illuminated in Chapter 2.

dot product for an inner product. The inner-product supplies a norm $\|\Psi\| = \sqrt{\langle\Psi|\Psi\rangle}$. A unit vector is an element of $\mathcal{H}_s$ with unit norm.

## 1.2 Dirac Notation

An extremely useful notation when describing quantum states is *Dirac notation*. According to this, a column vector $\Psi$ is called a *ket* and written $|\Psi\rangle$. The conjugate transpose of $|\Psi\rangle$ is called a *bra* and written $|\Psi\rangle^\dagger = \langle\Psi|$. This notation is clearly useful when working in finite-dimensional state spaces, because the vector product $(\langle\Phi|)\,|\Psi\rangle = \sum_i \Phi_i^* \Psi_i = \langle\Phi|\Psi\rangle$ thus corresponds to the dot product, hence inner product of the state space. The notation may be used further to define an operator formed by the *outer product* of two vectors, i.e. $|\Phi\rangle\langle\Psi|$. As this is the product of a row vector by a column vector, we expect the formula to describe a linear transformation on the state space. Indeed,

**Definition.** The **outer product** of two vectors, written $|\Phi\rangle\langle\Psi|$ is the operator which acts by $(|\Phi\rangle\langle\Psi|)\,|x\rangle = \langle\Psi|x\rangle\,|\Phi\rangle$.

Clearly any such operator is rank-1, because the subspace of $\mathcal{H}_s$ that is orthogonal to a vector $|\Psi\rangle$ has dimension $dim(\mathcal{H}_s) - 1$. An important property of the outer product is its relation to the inner product:

**Proposition.** *Let $|\Phi\rangle$, $|\Psi\rangle \in \mathcal{H}_s$. Then*

$$Tr(|\Phi\rangle\langle\Psi|) = \langle\Psi|\Phi\rangle$$

*where $Tr(\ )$ denotes the trace of an operator* [2].

*Proof.* Suppose $\{|i\rangle\}$ is an orthonormal basis of $\mathcal{H}_s$. Then:

$$Tr(|\Phi\rangle\langle\Psi|) = \sum_i \langle i|\,(|\Phi\rangle\langle\Psi|)\,i\rangle = \sum_i \langle i|\Phi\rangle\langle i|\Psi\rangle^* = \langle\Psi|\Phi\rangle$$

□

**Example.** If $|\Psi\rangle$ is a state vector, the operator $P = |\Psi\rangle\langle\Psi|$ describes a rank-1 projection operator onto the vector $|\Psi\rangle$. Indeed,

$$\begin{aligned} P^2(|x\rangle) &= (|\Psi\rangle\langle\Psi|)\,|\Psi\rangle\langle\Psi|\,(|x\rangle) \\ &= \langle\Psi|x\rangle\,(|\Psi\rangle\langle\Psi|)\,|\Psi\rangle = \langle\Psi|x\rangle\,|\Psi\rangle = P(|x\rangle) \end{aligned}$$

[2] Typically, the trace is defined as a matrix function, i.e. $tr(A) = \sum_i A_{ii}$. However, it is noted (from matrix multiplication) that $tr(AB) = tr(BA)$. Then for any change of basis matrix $M$, we have $tr(MAM^{-1}) = tr(M^{-1}MA) = tr(A)$. Therefore $tr(A)$ is invariant under any matrix representation of the operator associated with $A$; hence the trace of an operator is well-defined.

## 1.3 Qubits

The building blocks of quantum information, and the quantum analogues of classic bits are *qubits*. Although a valid state space may have arbitrary finite dimension, we associate only states of a 2 dimensional state space with the notion of a qubit in order to more readily allow comparison between classical and quantum algorithms.

**Definition.** A **qubit** is a quantum state (unit vector) in a 2-dimensional state space $\mathcal{H}_s$.

Equivalently, given an isomorphism $\mathcal{H}_s \cong \mathbb{C}^2$, a qubit is a unit vector in $\mathbb{C}^2$.

**Definition.** When $\mathcal{H}_s \cong \mathcal{C}^2$, the **computational basis** is the orthonormal basis $\{|0\rangle, |1\rangle\} = \{\binom{1}{0}, \binom{0}{1}\}$.

An important quantum phenomenon is *superposition*. Quantum superposition illuminates an essential difference between classical bits and qubits: a classical bit is equal to 0 or 1 while a qubit is any (unit) linear combination of the computational basis states $\{|0\rangle, |1\rangle\}$. Hence a general vector $|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is a qubit iff $|\alpha|^2 + |\beta|^2 = 1$.

**Definition.** When $\mathcal{H}_s \cong \mathcal{C}^2$ we define the states $\{|+\rangle, |-\rangle\}$ by

$$|+\rangle = \frac{1}{\sqrt{(2)}}(|0\rangle + |1\rangle)$$
$$|-\rangle = \frac{1}{\sqrt{(2)}}(|0\rangle - |1\rangle)$$

$|+\rangle$ is called the *equal superposition state*.

## 1.4 Dynamics of a Quantum System

Quantum states may undergo only linear transformations. Particularly, only unitary transformations are allowed.

**Postulate**: The time evolution of a closed system is described by unitary operators. That is, in a state space $\mathcal{H}_s$, $\forall\ t_1 \leq t_2$ if the state of the system at time $t_1$ is $|\Psi_{t_1}\rangle$ and the state of the system at time $t_2$ is $|\Psi_{t_2}\rangle$, there exists a unitary operator $U_{t_1,t_2} : \mathcal{H}_s \to \mathcal{H}_s$ such that $|\Psi_{t_2}\rangle = U_{t_1,t_2}|\Psi_{t_1}\rangle$. Any unitary operator describes a valid evolution of a quantum system.

Hence the set of unitary operators form the "toolbox" for changing quantum states. Intuitively, this makes sense because unitary operators preserves the inner product and hence the norm. The set of linear transformations $\mathcal{H}_s \to \mathcal{H}_s$ is a vector space with dimension $dim(\mathcal{H}_s)^2$ (again, supposing $dim(\mathcal{H}_s) = n$ is finite; then each linear transformation is associated with an $n \times n$ matrix). The second part of the postulate

states that any unitary transformation is available to us in the construction of quantum algorithms. In our computational model, unitary transformations correspond to classical logic gates; they provide the valid ways of manipulating the information stored in qubits. This comparison also highlights the necessary difference between classical and quantum computation. Unitary operators are necessarily reversible (because $U^{-1} = U^\dagger$), while many classical logic gates are irreverisble, such as OR or NAND.

**Example.** The following matrices correspond to unitary operators expressed in the computational basis:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

We now see, according to our previous definition,

$$|+\rangle = H\,|0\rangle$$
$$|-\rangle = H\,|1\rangle$$

Examining the action of $X$ we see that

$$X\,|0\rangle = |1\rangle$$
$$X\,|1\rangle = |0\rangle$$

For this reason, $X$ is sometimes referred to as the *quantom NOT gate*.

## 1.5 Composite Systems

We are often interested in quantum systems that are composed of multiple distinct component systems. A third postulate states that composite systems are formed from the *tensor product* of component systems. Rather than define the tensor product, we assume the following proposition is true:

**Proposition.** Suppose $V$ and $W$ are $n$ and $m$ dimensional complex Hilbert spaces, respectively. Then the **tensor product** of $V$ and $W$, written $V \otimes W$, is an $nm$ dimensional complex Hilbert space satisfying the properties:

1. For all $\lambda \in \mathbb{C}$ and $|v\rangle \in V$, $|w\rangle \in W$:

$$\lambda(|v\rangle \otimes |w\rangle) = (\lambda\,|v\rangle) \otimes |w\rangle = |v\rangle \otimes (\lambda\,|w\rangle)$$

2. For all $|v_1\rangle, |v_2\rangle \in V$ and $|w_1\rangle, |w_2\rangle \in W$:

$$(|v_1\rangle + |v_2\rangle) \otimes |w_1\rangle = |v_1\rangle \otimes |w_1\rangle + |v_2\rangle \otimes |w_1\rangle$$

$$|v_1\rangle \otimes (|w_1\rangle + |w_2\rangle) = |v_1\rangle \otimes |w_2\rangle + |v_1\rangle \otimes |w_2\rangle$$

3. If $\langle \,|\, \rangle_V$ and $\langle \,|\, \rangle_W$ are the inner products of $V$ and $W$, then the inner product on $V \otimes W$ is defined as

$$\langle |v_1\rangle \otimes |w_1\rangle \,|\, |v_2\rangle \otimes |w_2\rangle \rangle = \langle v_1|v_2\rangle_V \, \langle w_1|w_2\rangle_W$$

The tensor product is also useful for describing linear operators on the space $V \otimes W$:

**Definition.** Suppose $V$ and $W$ are $n$ and $m$ dimensional vector spaces and $A$ and $B$ are linear operators on $V$ and $W$, respectively. Then the operator $A \otimes B : V \otimes W \to V \otimes W$ is defined by:

$$(A \otimes B)(|v\rangle \otimes |w\rangle) = A\,|v\rangle \otimes B\,|w\rangle$$

Some obvious basic properties of this definition include that if $A$ and $B$ are unitary (invertible) operators, then $A \otimes B$ is unitary (invertible).

The definition of the tensor product of linear operators leads to a matrix formulation of this definition.

**Definition.** Suppose $A, B$ are $n \times n$ and $m \times m$ matrices, respectively. Then the matrix $A \otimes B$ is the $nm \times nm$ matrix (expressed in block form)

$$A \otimes B = \begin{pmatrix} A_{11}B & A_{12}B & \dots & A_{1n}B \\ A_{21}B & A_{22}B & \dots & A_{2n}B \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1}B & A_{n2}B & \dots & A_{nn}B \end{pmatrix}$$

Calculations on a basis state (not included) show that the two definitions for the tensor product coincide.

We are ready to state the third postulate:

**Postulate** The state space of a composite system is the tensor product of the component systems. That is, if $\mathcal{H}_1, \dots \mathcal{H}_k$ are the state spaces of component systems, then the full system $\mathcal{H}_s = \bigotimes_{i=1}^{k} \mathcal{H}_i$. If each component is in the state $|\Psi_i\rangle$ then the full system is in the state $\bigotimes_{i=1}^{k} |\Psi_i\rangle$.

**Remark** For ease of notation, we write $|v\rangle \otimes |w\rangle \equiv |vw\rangle$. Also, for either operators or vectors we write $A^{\otimes n} \equiv \bigotimes_{i=1}^{n} A$.

It is clear to see that if $\{|v_i\rangle\}$ is a basis of $V$ and $\{|v_j\rangle\}$ is a basis of $W$, then $\{|v_i \otimes w_j\rangle\}$ form a basis of $V \otimes W$. Notice that $\mathbb{C}^{2^n}$ is trivially isomorphic to $(\mathbb{C}^2)^{\otimes n}$. With this in mind, we extend the definition of the computational basis to larger spaces:

**Definition.** When $dim(\mathcal{H}_s) = 2^n$, that is, $\mathcal{H}_s \cong \mathbb{C}^{2^n}$ the **computational basis** of $\mathcal{H}_s$ is the orthonormal basis

$$\{|0\ldots0\rangle, |0\ldots1\rangle \ldots |1\ldots1\rangle\} = \{|i\rangle \, | i \in \mathbb{Z}_2^n\}$$

**Example.** Suppose $\mathcal{H}_s \cong \mathbb{C}^4$. Then, from the previous example, we may examine the result of subjecting the computational basis states to $H^{\otimes 2}$:

$$\begin{aligned}
H^{\otimes 2}|00\rangle = |{+}{+}\rangle &= \left(\frac{1}{\sqrt{2}}|0\rangle + |1\rangle\right) \otimes \left(\frac{1}{\sqrt{2}}|0\rangle + |1\rangle\right) \\
&= \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) \\
H^{\otimes 2}|01\rangle = |{+}{-}\rangle &= \frac{1}{2}(|00\rangle - |01\rangle + |10\rangle - |11\rangle) \\
H^{\otimes 2}|10\rangle = |{-}{+}\rangle &= \frac{1}{2}(|00\rangle + |01\rangle - |10\rangle + |11\rangle) \\
H^{\otimes 2}|11\rangle = |{-}{-}\rangle &= \frac{1}{2}(|00\rangle - |01\rangle - |10\rangle + |11\rangle
\end{aligned}$$

Notice that $H^{\otimes 2}$ once again produces an equal sum of the basis states of $\mathbb{C}^2$. Therefore we also extend the definition of an equal superposition state:

**Definition.** If $\mathcal{H}_s \cong \mathbb{C}^{2^n}$, the **equal superposition state** is defined:

$$|\eta_n\rangle = \frac{1}{\sqrt{2^n}} \sum_{i \in \mathbb{Z}_2^n} |i\rangle$$

Clearly $|\eta_n\rangle = |+\rangle^{\otimes n} = H^{\otimes n}|0\ldots0\rangle$.

### 1.5.1 The Discrete Fourier Transform

The discrete Fourier transform is an often used tool in quantum algorithms.

**Definition.** The discrete Fourier transform is a linear operator $\mathcal{F} : \mathbb{C}^N \to \mathbb{C}^N$ defined on an orthonormal basis $\{|0\rangle, \ldots |N-1\rangle\}$ by:

$$\mathcal{F}^N |j\rangle = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{-2\pi i j k/N} |k\rangle$$

with inverse

$$(\mathcal{F}^N)^{-1} |k\rangle = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{2\pi i j k/N} |j\rangle$$

## 1.6 Hadamard Matrices

The matrix $H^{\otimes n} = \left(\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\right)^{\otimes n}$ is extremely useful and is referred to as a *Hadamard matrix*. It has many useful properties and occurs frequently in quantum algorithms.

**Proposition 1.1.**

$$H^{\otimes n} = \frac{1}{\sqrt{2}} \sum_{x,y\in\mathbb{Z}_2^n} (-1)^{x\cdot y} |x\rangle \langle y|$$

*where* $x \cdot y = \sum_{i=1}^{n} x_i y_i$.

Abusing terminology, we call the operation $x \cdot y$ the *dot product* of n-bit strings $x$ and $y$ (this is not the same as $\langle x|y\rangle = \delta_{x,y}$ for computational basis states). It is the number of shared 1's between $x$ and $y$. The proposition implies $H^{\otimes n}_{i,j} = (-1)^{i\cdot j}$ where $i, j$ are indexed by the computational basis states, i.e. elements of $\mathbb{Z}_2^n$.

*Proof.* We prove by induction. Recall that

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$$

which satisfies the claim. We note that $H$ can be written:

$$H = \frac{1}{\sqrt{2}} \left((|0\rangle + |1\rangle) \langle 0| + (|0\rangle - |1\rangle) \langle 1|\right)$$

and for the induction hypothesis we assume

$$H^{\otimes n-1} = \frac{1}{\sqrt{2^{n-1}}} \sum_{x,y\in\mathbb{Z}_2^{n-1}} (-1)^{x\cdot y} |x\rangle \langle y|$$

Now examine how $H^{\otimes n} = H^{\otimes n-1} \otimes H$ acts on an arbitrary computational basis element $|e\rangle = |w\rangle \otimes |a\rangle$ where $w \in \mathbb{Z}_2^{n-1}$ and $a \in \mathbb{Z}_2$:

$$\left(\frac{1}{\sqrt{2^{n-1}}} \sum_{x,y\in\mathbb{Z}_2^{n-1}} (-1)^{x\cdot y} |x\rangle \langle y| \otimes \frac{1}{\sqrt{2}} \left((|0\rangle + |1\rangle) \langle 0| + (|0\rangle - |1\rangle) \langle 1|\right)\right) (|w\rangle \otimes |a\rangle)$$

$$= \frac{1}{\sqrt{2^n}} \left(\sum_{x\in\mathbb{Z}_2^{n-1}} (-1)^{x\cdot w} |x\rangle\right) \otimes \left((|0\rangle + |1\rangle) \langle 0|a\rangle + (|0\rangle - |1\rangle) \langle 1|a\rangle\right)$$

Now suppose $|a\rangle = |0\rangle$. Then we have:

$$H^{\otimes n}(|w\rangle \otimes |0\rangle) = \frac{1}{\sqrt{2^n}} \sum_{x\in\mathbb{Z}_2^{n-1}} (-1)^{x\cdot w} (|x0\rangle + |x1\rangle)$$

$$= \frac{1}{\sqrt{2^n}} \sum_{z\in\mathbb{Z}_2^n} (-1)^{z\cdot e} |z\rangle$$

since $z \cdot e = x \cdot w$ for all $z \in \mathbb{Z}_2^n$. Next consider $|a\rangle = |1\rangle$. Then:

$$\begin{aligned} H^{\otimes n}(|w\rangle \otimes |1\rangle) &= \frac{1}{\sqrt{2^n}} \sum_{x \in \mathbb{Z}_2^{n-1}} (-1)^{x \cdot w}(|x0\rangle - |x1\rangle) \\ &= \frac{1}{\sqrt{2^n}} \sum_{z \in \mathbb{Z}_2^n} (-1)^{z \cdot e} |z\rangle \end{aligned}$$

because $z \cdot e = x \cdot w$ if the last bit of $z$ is 0, while $z \cdot e = x \cdot w + 1$ if the last bit of $z$ is 1 (i.e. $|z\rangle = |x1\rangle$). We have completed the proof, as we have shown that

$$H^{\otimes n} |e\rangle = \left( \frac{1}{\sqrt{2^n}} \sum_{x,y \in \mathbb{Z}_2^n} (-1)^{x \cdot y} |x\rangle \langle y| \right) |e\rangle$$

for all computational basis elements $|e\rangle$ of $\mathbb{C}^{2^n}$. □

It is clear that $H^{\otimes n}$ is symmetric. Since it is unitary, we have:

**Corollary.**

$$H^{\otimes n} = (H^{\otimes n})^\dagger = (H^{\otimes n})^{-1}$$

## 1.7 The Density Operator

We now discuss the *density operator* formalism, which is a way of describing a quantum system whose state is not known, that is, when a system that could be in one of a number of different states with certain probabilities. This formalism is not used in the discussion of quantum algorithms and is used solely in the description of general quantum measurement schemes discussed in Chapter 2.

**Definition.** Suppose a quantum system with (finite dimensional) state space $\mathcal{H}_s$ is in one of $n$ possible states $\{|\Psi_i\rangle\}$ with respective probabilities $\{p_i\}$. Then the system is an *ensemble of pure states*. The *density operator* associated with an ensemble is a linear operator $\rho : \mathcal{H}_s \to \mathcal{H}_s$:

$$\rho = \sum_{i=1}^{n} p_i |\Psi_i\rangle \langle \Psi_i|$$

Density operators are characterized by a useful theorem:

**Theorem 1.2.** *A linear operator $\rho$ acting on an $N$ dimensional complex Hilbert space $\mathcal{H}_s$ is a density operator if and only if:*

1. $Tr(\rho) = 1$
2. $\rho \geq 0$

*where the notation $\rho \geq 0$ asserts that $\rho$ is a positive operator.*

*Proof.* $\Longrightarrow$ Suppose $\rho = \sum_{i=1}^{N} p_i |\Psi_i\rangle$ for $|\Psi_i\rangle \in \mathcal{H}_s$. Let $\{j\}$ be an orthonormal basis in $\mathcal{H}_s$. Then

$$\begin{aligned} Tr(\rho) &= \sum_{i=1}^{N} p_i Tr(\langle\Psi_i| |\Psi_i\rangle) \\ &= \sum_{i=1}^{N} p_i \langle\Psi_i|\Psi_i\rangle = 1 \end{aligned}$$

Now take an arbitrary $|\Phi\rangle \in \mathcal{H}_s$. Then

$$\begin{aligned} \langle\Phi|\rho|\Phi\rangle &= \sum_{i=1}^{N} p_i \langle\Phi|\Psi_i\rangle \langle\Phi|\Psi_i\rangle \\ &= \sum_{i=1}^{N} p_i \| \langle\Phi|\Psi_i\rangle \|^2 \geq 1 \end{aligned}$$

$\Longleftarrow$ Suppose $\rho$ is a linear operator on $\mathcal{H}_s$ satisfying the two conditions. Since it is positive, it has a spectral decomposition [ref]:

$$\rho = \sum_{i=1}^{N} \lambda_i |\Psi_i\rangle \langle\Psi_i|$$

where each $\lambda_i \geq 0$. The trace condition requires $\sum_{i=1}^{N} \lambda_i = 1$. Therefore the ensemble of pure states $\{|\Psi_i\rangle\}$ with probabilities $\{p_i\}$ produce the density operator $\rho$. □

Let us examine the rules that the postulates discussed enforce on density operators. The previous theorem accounted for the first postulate, as density operators are formed only from unit vectors of a complex Hilbert space.

**Proposition 1.3.** *Suppose a closed quantum system is an ensemble of pure states in a n N dimensional state space $\mathcal{H}_s$ with $\rho_{t_1} = \sum_{i=1}^{N} p_i |\Psi_i\rangle \langle\Psi_i|$. Then for all $t_2 \geq t_1$, if the system is an ensemble of pure states with density operator $\rho_{t_2}$, then there exists some unitary operator $U_{t_1,t_2} : \mathcal{H}_s \to \mathcal{H}_s$ such that*

$$\rho_{t_2} = U_{t_1,t_2} \rho_{t_1} U_{t_1,t_2}^{\dagger}$$

*Proof.* From the second postulate, if a quantum system is in state $|\Psi_{t_1}\rangle$ at time $t_1$, then for all $t_2 \geq t_1$, if the system is in state $|\Psi_{t_2}\rangle$ at time $t_2$, $|\Psi_{t_2}\rangle = U_{t_1,t_2} |\Psi_{t_1}\rangle$. Therefore, each pure state in an ensemble must evolve according to the same unitary operator. Hence, for all $t_2 \geq t_1$,

$$\begin{aligned} \rho_{t_2} &= \sum_{i=1}^{N} |\Psi_{t_2}\rangle \langle\Psi_{t_2}| \\ &= \sum_{i=1}^{N} U_{t_1,t_2} |\Psi_{t_1}\rangle \langle\Psi_{t_1}| U_{t_1,t_2}^{\dagger} \\ &= U_{t_1,t_2} \left( \sum_{i=1}^{N} |\Psi_{t_1}\rangle \langle\Psi_{t_1}| \right) U_{t_1,t_2}^{\dagger} = U_{t_1,t_2} \rho_{t_1} U_{t_1,t_2}^{\dagger} \end{aligned}$$

□

## 1.8 The Measurement Postulate

The process of extracting information from a quantum state requires further assumptions. Measurement is determined by a set of linear operators, $\{M_m\}$, acting on a finite dimensional state space $\mathcal{H}_s$. The index $m$ refers to the possible measurement outcomes, which we generally take to be finite. The results of measurement are provided by the Measurement Postulate:

**Postulate:** Given a finite set of linear operators $\{M_m\}$ acting on a finite dimensional state space $\mathcal{H}_s$ such that $\sum_m M_m = I$ and a quantum system in the state $|\Psi\rangle$, the probability of measuring outcome $m$ is

$$P(m) = \langle\Psi|M_m^\dagger M_m|\Psi\rangle \tag{1.1}$$

and the post-measurement state of the system is

$$\frac{M_m|\Psi\rangle}{\|M_m|\Psi\rangle\|} \tag{1.2}$$

We remark that the condition $\sum_m M_m = I$ (called the **completeness relation**) is required so that measurements are consistent with the notions of probability.

Clearly $1 = \sum_{m=1}^N P(m)$. Therefore, from the Measurement Postulate,

$$\begin{aligned} 1 &= \sum_{m=1}^N \langle\Psi|M_m^\dagger M_m|\Psi\rangle \\ &= \langle\Psi|\sum_{m=1}^N M_m^\dagger M_m|\Psi\rangle \end{aligned}$$

due to the sesquilinearity of the inner product. Therefore

$$\begin{aligned} &(\sum_{m=1}^N M_m^\dagger M_m)\,|\Psi\rangle = |\Psi\rangle \\ \implies &\sum_{m=1}^N M_m^\dagger M_m = I \end{aligned} \tag{1.3}$$

A frequently used measurement scheme is a *von Neumann measurement*, in which case each the set of measurement operators forms an orthogonal set of projection operators (recall that the space of linear operators acting on $\mathcal{H}_s$ is itself a Hilbert space, equipped with the trace inner product).

**Definition.** A measurement $\{M_m\}$ is a **von Neumann measurement** if the measurement operators form a set of mutually orthogonal projection operators, i.e.

$$\begin{aligned} M_m^2 &= M_m && \forall M_m \\ Tr(M_i^* M_j) &= 0 && i \neq j \end{aligned}$$

A measurement $\{M_m\}$ is a **complete von Neumann measurement** if measurement operators form a set of mutually orthogonal rank-1 projection operators; i.e. $M_m = |\mu_m\rangle \langle\mu_m|$ for some $|\mu_m\rangle \in \mathcal{H}_s$ and $\langle\mu_i|\mu_j\rangle = 0$ for $i \neq j$.

The orthogonality condition $\langle\mu_i|\mu_j\rangle = 0$ for $i \neq j$ is a direct consequence of the Hilbert-Schmidt orthogonality of the projection operators, assuming they are 1-dimensional.

**Remark.** Note that this definition, along with the completeness relation, requires each $|\mu_i\rangle$ to be normalized.

$$\begin{aligned} |\mu_i\rangle &= \left(\sum_m E_m\right) |\mu_i\rangle = \left(\sum_m |\mu_m\rangle \langle\mu_m|\right) |\mu_i\rangle \\ &= \langle\mu_i|\mu_i\rangle |\mu_i\rangle \implies \langle\mu_i|\mu_i\rangle = 1 \end{aligned}$$

The following complete von Neumann measurement is the most common measurement employed in quantum algorithms.

**Definition.** In a state space of dimension $2^n$ supplied with the orthonormal basis $\{|j\rangle \, | j \in \mathbb{Z}_2^n\}$, we define **measurement in the computational basis** to be a quantum measurement using the operators

$$M_{|j\rangle} = |j\rangle \langle j| \tag{1.4}$$

**Example.** Let us examine the result of measurement in the computational basis on the general single-qubit state $|\Psi\rangle = \alpha |0\rangle + \beta |1\rangle$. From the above definition, this measurement consists of two operators

$$\begin{aligned} M_{|0\rangle} &= |0\rangle \langle 0| \\ M_{|1\rangle} &= |1\rangle \langle 1| \end{aligned}$$

Therefore, by the Measurement Postulate, the probability of receiving outcome $|0\rangle$ is

$$\begin{aligned} P(|0\rangle) &= \langle\Psi|M_{|0\rangle}^\dagger M_{|0\rangle}|\Psi\rangle \\ &= \langle\Psi|0\rangle \langle 0|0\rangle \langle 0|\Psi\rangle \\ &= \alpha^*\alpha = |\alpha|^2 \end{aligned}$$

By the Measurement Postulate we can also compute the post-measurement state if the output is $|0\rangle$:

$$\frac{M_{|0\rangle} |0\rangle}{\| M_{|0\rangle} |0\rangle \|} = \frac{|0\rangle \langle 0|0\rangle}{\| |0\rangle \langle 0|0\rangle \|} = |0\rangle$$

Similarly, $P(|1\rangle) = |\beta|^2$ and the post-measurement state after measuring $|1\rangle$ is $|1\rangle$.

For many applications, including those in this thesis, only the measurement outcome is significant and so the post-measurement state is irrelevant. This means the The Measurement Postulate may be simplified by defining, for any measurement, positive semi-definite linear operators $E_m = M_m^\dagger M_m$. These operators are positive semi-definite and Hermitian by definition (and are usually referred to as Positive Operator-Valued Measures, or POVMs).

**Definition.** A POVM, or Positive Operator-Valued Measure, is a finite set $\{E_1, \dots E_k\}$ of positive semi-definite operators acting on a finite dimensional state space $\mathcal{H}_s$ such that $\sum_m E_m = I$.

The probability of measuring outcome $m$ may now be written:

$$P(m) = \langle \Psi | E_m | \Psi \rangle \tag{1.5}$$

Hence, from now on, a quantum measurement is specified by a set of positive semidefinite linear operators $\{E_m\}$.

**Theorem 1.4.** *Suppose $|\Psi_1\rangle, |\Psi_2\rangle$ are quantum states in an $n$ dimensional state space $\mathcal{H}_s$. Then there exists a measurement to distinguish $|\Psi_1\rangle, |\Psi_2\rangle$ with certainty iff $|\Psi_1\rangle, |\Psi_2\rangle$ are orthonormal.*

*Proof.* First assume $|\Psi_1\rangle, |\Psi_2\rangle$ are orthonormal. Then the measurement

$$\{|\Psi_1\rangle\langle\Psi_1|,\ |\Psi_2\rangle\langle\Psi_2|,\ I - (|\Psi_1\rangle\langle\Psi_1| + |\Psi_2\rangle\langle\Psi_2|)\}$$

is a valid measurement that will measure outcome 1 iff the state to be measured is $|\Psi_1\rangle$ and outcome 2 iff the state to be measured is $|\Psi_2\rangle$.

We prove the converse by contradiction. Assume $|\Psi_1\rangle, |\Psi_2\rangle$ is not orthogonal, but there is some measurement $\{E_1, E_2, \dots E_k\}$ that distinguishes them with certainty. Then for some outcomes $j \in \{1 \dots k\}$ we deduce the state is $|\Psi_1\rangle$ and for some outcomes $i \in \{1 \dots k\}$ we deduce the state is $|\Psi_2\rangle$. Therefore we introduce the operators

$$E^1 = \sum_j E_j$$
$$E^2 = \sum_i E_i$$

where the indices $i, j$ are defined as above. Clearly $E^1, E^2 \geq 0$, and by the hypothesis, we must guess that the state is either $|\Psi_1\rangle$ or $|\Psi_2\rangle$; therefore $E^1 + E^2 = I$. Therefore $\{E^1, E^2\}$ supplies a valid measurement. We know therefore that

$$1 = \langle \Psi_1 | E^1 \Psi_1 \rangle = \langle \Psi_2 | E^2 \Psi_2 \rangle = \langle \Psi_1 | (E^1 + E^2) \Psi_1 \rangle$$

From which we deduce $\langle \Psi_1 | E^2 \Psi_1 \rangle = 0$. Since every positive semi-definite operator admits a unique positive semi-definite square root, $\sqrt{E^2}$ is well-defined and we have shown that $\sqrt{E^2}|\Psi_1\rangle = 0$. Now write

$$|\Psi_2\rangle = \alpha |\Psi_1\rangle + \beta |\phi\rangle$$

where $|\phi\rangle$ is orthonormal to $|\Psi_1\rangle$. We assume that $|\Psi_1\rangle, |\Psi_2\rangle$ are not orthogonal, so $|\beta|^2 < 1$. Therefore $\sqrt{E^2}|\Psi_2\rangle = \beta\sqrt{E^2}|\phi\rangle$.. However:

$$1 = \langle\Psi_2|E^2\Psi_2\rangle = \langle\sqrt{E^2}\Psi_2|\sqrt{E^2}\Psi_2\rangle = |\beta|^2\langle\sqrt{E^2}\phi|\sqrt{E^2}\phi\rangle = |\beta|^2 < 1$$

A contradiction. Therefore there is no such measurement that distinguishes non-orthogonal $|\Psi_1\rangle, |\Psi_2\rangle$. □

Although not included here, this result easily generalizes to the corollary:

**Corollary.** *Suppose $|\Psi_1\rangle, \ldots |\Psi_n\rangle$ are quantum states in an n dimensional state space $\mathcal{H}_s$. Then there is a measurement $\{E_1, \ldots E_n\}$ which distinguishes those states with certainty iff $\{|\Psi_1\rangle \ldots |\Psi_n\rangle\}$ form an orthonormal set. If that is the case, then the measurement $\{E_1, \ldots E_n\} = \{|\Psi_1\rangle\langle\Psi_1|, \ldots |\Psi_n\rangle\langle\Psi_n|\}$ is a complete von Neumann measurement.*

The Measurement Postulate is generalized further to describe its action on general quantum ensembles, specified by a density operator, in the following Proposition.

**Proposition 1.5.** *Given a quantum system described by a density operator $\rho$ and a set of measurement operators $\{E_m\}$, the probability of measuring outcome m is $P(m) = Tr(E_m\rho)$.*

*Proof.* Let $\rho = \sum_{i=0}^{k} p_i |\Psi_i\rangle\langle\Psi_i|$. Hence the quantum system exists in state $|\Psi_i\rangle$ with probability $p_i$. Now consider applying the measurement operators to this system. By the Measurement Postulate, the conditional probability $p(m|i) = \langle\Psi_i|E_m|\Psi_i\rangle$. Therefore:

$$\begin{aligned}
P(m) &= \sum_{i=0}^{k} p_i p(m|i) \\
&= \sum_{i=0}^{k} p_i \langle\Psi_i|E_m|\Psi_i\rangle \\
&= \sum_{i=0}^{k} p_i Tr(E_m|\Psi_i\rangle\langle\Psi_i|) \qquad (1.6) \\
&= Tr(\sum_{i=0}^{k} E_m p_i |\Psi_i\rangle\langle\Psi_i|) \qquad (1.7) \\
&= Tr(E_m\rho) \qquad (1.8)
\end{aligned}$$

where (3.8) and (3.9) follow from the linearity of the trace and $\{E_m\}$ operators. □

# Chapter 2

# Optimal Quantum Measurements

## 2.1 Necessary and Sufficient Condition for Optimal Measurement of Multiple States

We are now equipped to consider the problem of determining how effective a set of measurement operators are in distinguishing between a set of known quantum systems. This problem seems to have been originally solved by Kennedy and Yuen [YKL75], but Eldar and Forney [EMV08] have since condensed the literature to provide a direct proof of the main theorem describing necessary and sufficient conditions for optimal measurement of $m$ general quantum ensembles.

To describe what is meant by "distinguishing" between a set of known quantum systems, we consider a communicational analogy to the problem. Suppose Alice and Bob share knowledge of a fixed set of quantum systems $\{\rho_i\}$. Then Alice "picks" a state by applying a set of prior probabilities $\{p_i\}$ (i.e. she sends the system $\rho_i$ with probability $p_i$). Bob is equipped with a set of measurement operators $\{E_m\}$ and hopes to determine which system Alice transmitted such that measuring outcome $m$ corresponds to deciding that Alice transmitted $\{\rho_m\}$. An optimal measurement is a set of positive linear operators $\{E_m\}$ that maximize the probability of Bob correctly guessing Alice's transmission. If $P(i|i)$ is the probability of choosing outcome $i$ assuming the system is described by $\rho_i$, then the optimal measurement optimizes:

$$\max \sum_{i=1}^{N} p_i P(i|i)$$

The theorem relies on the partial ordering of operators supplied by the notion of positive semi-definiteness: for linear operators $\lambda_1, \lambda_2$ acting on $\mathcal{H}_s$,

$$\lambda_1 \geq \lambda_2 \iff \lambda_1 - \lambda_2 \geq 0$$

that is, $\lambda_1 - \lambda_2$ is positive semi-definite. The necessary and sufficient conditions for an optimal measurement $\{\hat{E}_m\}$ are provided by the following theorem, compiled from [EMV08], [YKL75], and [Hel76].

**Theorem 2.3.** *Let $\{\rho_i\}$ be a fixed set of $N$ quantum ensembles in a an $N$ dimensional state space $\mathcal{H}_s$ equipped with a set of prior probabilities $\{p_i | p_i \geq 0, \sum_{i=1}^{N} p_i = 1\}$. Then a measurement $\{\hat{E}_m\}$ is optimal iff:*

$$(\hat{\lambda} - \rho_i')E_i = E_i(\hat{\lambda} - \rho_i') = 0, \qquad i = 1, 2, ...N \tag{2.1}$$

$$\hat{\lambda} - \rho_i' \geq 0, \qquad i = 1, 2, ...N \tag{2.2}$$

*where*

$$\rho_i' = p_i \rho_i \tag{2.3}$$

$$\hat{\lambda} \equiv \sum_{j=1}^{N} E_j \rho_j' = \sum_{j=1}^{N} \rho_j' E_j \tag{2.4}$$

The theorem makes use of the following lemmas, the proofs of which are found in Appendix 1.

**Lemma 1.** *Suppose $E_1, E_2$ are positive semidefinite linear operators. Then:*

$$Tr(E_1 E_2) \geq 0$$
$$Tr(E_1 E_2) = 0 \iff E_1 E_2 = E_2 E_1 = 0$$

*Proof.* See [YKL75]. □

**Lemma 2.** *Let $K$ denote the space of linear operators over $\mathcal{H}_s$. Then the space $K^{\otimes N}$ is a Hilbert space equipped with the inner product*

$$\langle U_1 \otimes \cdots \otimes U_N | V_1 \otimes \cdots \otimes V_n \rangle = \prod_{i=1}^{N} Tr(U_i V_i)$$

*Then the set of possible measurements, $\mathcal{M} = \{\{E_m\}_{m=1}^{N} | E_m \geq 0, \sum_{m=1}^{N} E_m = I\}$ is compact subset of $K^{\otimes n}$.*

*Proof.* See [CDS08]. □

**Lemma 3.** *(Separating Hyperplane Theorem) Any two disjoint convex sets can be separated by a hyperplane. That is, if $V$ is a real inner product space with $C, D \subset V$ disjoint convex sets, then $\exists x \in V, \exists \alpha \in \mathbb{R} :$* $\begin{cases} \langle x, c \rangle \geq \alpha & \forall c \in C \\ \langle x, d \rangle \leq \alpha & \forall d \in D \end{cases}$

*Proof of Theorem 2.3.* We wish to maximize the probability of correctly determining the quantum system. This translates to the optimization problem:

$$\max \sum_{i=1}^{N} p_i P(i|i)$$

where $P(i|i)$ is the probability of measuring outcome $i$ assuming the system is in state $\rho_i$. Therefore, by Prop. 3.1, our optimization problem is:

$$\max_{\substack{E_i \geq 0 \\ \sum_{m=1}^{N} E_m = I}} \sum_{i=1}^{N} p_i Tr(E_i \rho_i)$$
$$= \max_{\substack{E_i \geq 0 \\ \sum_{m=1}^{N} E_m = I}} \sum_{i=1}^{N} Tr(E_i \rho_i')$$

Let us denote $\mathcal{M}$ to be the set of all possible measurements with N operators, i.e. $\mathcal{M} = \{\{E_m\}_{m=1}^N | E_m \geq 0, \sum_{m=1}^N E_m = I\}$. Also define $M(\{E_m\}) = \sum_{i=1}^N Tr(E_i \rho_i')$ so that our optimization problem may again be rewritten:

$$\max_{\{E_m\} \in \mathcal{M}} M$$

By Lemma 2, $\mathcal{M}$ is compact, and M is a continuous linear functional. Therefore M attains a maximum over $\mathcal{M}$, that is $\exists \{\hat{E}_m\} \in \mathcal{M} : \hat{M} = M(\{\hat{E}_m\}) \geq M(\{E_m\})\ \forall \{E_m\} \in \mathcal{M}$.

The optimization problem is now a problem in the general field of convex programming. A common method for determining necessary and sufficient conditions for $\{\hat{E}_m\}$ is to formulate a *dual problem*. Denote by $\mathcal{L}$ the space of Hermitian operators on $\mathcal{H}_s$. The purpose of the dual problem is to produce a linear functional $T$ such that

$$\min_{\lambda \in \mathcal{L}} T(\lambda) = \hat{M}$$

This equality will allow us to find the necessary and sufficient conditions on $\{E_m\}$. Following [ElFo 2008], $T$ is constructed by producing a convex set and applying the separating hyperplane theorem to this set and $\{0\}$; we will see that the parameters of the hyperplane will be parameters for the dual problem.

The setting for the dual problem is the inner product space $\mathcal{L} \otimes \mathbb{R}$ equipped with the inner product:

$$\langle (E_1, r_1), (E_2, r_2) \rangle = Tr(E_1 E_2) + r_1 r_2$$

Note that as $E_1, E_2$ are Hermitian operators:

$$Tr(E_1 E_2) = Tr(E_2 E_1) = Tr((E_1 E_2)^\dagger) = (Tr(E_1 E_2))^\star$$

Therefore the trace of the product of two Hermitian operators is real; thus $\mathcal{L} \otimes \mathbb{R}$ is a real inner product space.

We now construct a convex set in the space $\mathcal{L} \otimes \mathbb{R}$. The constraints on the original (often referred to as the *primal*) problem are $\sum_{m=1}^N E_m = I$ and $\hat{M} \geq M(\{E_m\})$ $(\forall \{E_m\} \in \mathcal{M})$. First consider the set $\mathcal{C}_1 = \{-I + \sum_{m=1}^N E_m | E_m \in \mathcal{L}\}$. We

contend this set is convex; consider a convex combination of $X, Y \in \mathcal{C}_1$: for $0 \leq t \leq 1$,

$$\begin{aligned} tX + (1-t)Y &= -tI + t\sum_{m=1}^{N} X_m + (1-t)\sum_{m=1}^{N} Y_m \\ &= -I + \sum_{m=1}^{N} (tX_m + (1-t)Y_m) \end{aligned} \tag{2.5}$$

The summand in (2.5) is clearly positive semidefinite; therefore $\mathcal{C}_1$ is convex. Now consider $\mathcal{C}_2 = \{c - M(\{E_m\}) |\ c > \hat{M}, E_m \in \mathcal{L}\}$. By similar reasoning, $\mathcal{C}_2$ is convex. Hence $\mathcal{C}_1 \otimes \mathcal{C}_2 \subset \mathcal{L} \otimes \mathbb{R}$ is convex, and by definition $(0,0) \notin \mathcal{C}_1 \otimes \mathcal{C}_2$.

By applying the separating hyperplanes theorem to the sets $\{(0,0)\}$ and $\mathcal{C}_1 \otimes \mathcal{C}_2$, there exists a non-zero vector $(Z, a) \in \mathcal{L} \otimes \mathbb{R}$ such that

$$\begin{aligned} &\langle (Z,a), (-I + \sum_{m=1}^{N} E_m,\ c - \sum_{m=1}^{N} Tr(E_m \rho'_m)) \rangle \geq 0 \\ \implies &Tr\left(Z\left(-I + \sum_{m=1}^{N} E_m\right)\right) + a\left(c - \sum_{m=1}^{N} Tr(E_m \rho'_m)\right) \geq 0 \end{aligned} \tag{2.6}$$

for all possible measurements $\{E_m\}$ and any $c > \hat{M}$. (Note that, in applying the separating hyperplanes theorem, since $\langle (Z,a), (0,0) \rangle = 0$ we may take $\alpha = 0$ without loss of generality). We now examine critical points of the set $\mathcal{C}_1 \otimes \mathcal{C}_2$ using (2.6) to determine constraints on the vector $(Z, a)$.

First, consider $E_m = 0,\ m = 1, 2, \ldots N$ and $c \to \hat{M}$. Then (2.6) implies

$$a\hat{M} \geq Tr(Z) \tag{2.7}$$

Next, suppose $c = \hat{M} + 1$, $E_m = 0$ if $m \neq 1$, $E_1 = t\,|x\rangle\langle x|$ for some $|x\rangle \in \mathbb{R}^N$. Note that $E_1$ is positive semidefinite $\forall t \geq 0$. Now (2.6) implies

$$\begin{aligned} &tTr\left(Z\,|x\rangle\langle x| - a\rho'_i\,|x\rangle\langle x|\right) \geq Tr(Z) - a(\hat{M}+1) \\ \implies &tTr\left((Z - a\rho'_i)\,|x\rangle\langle x|\right) \geq Tr(Z) - a(\hat{M}+1) \\ \implies &t\,\langle x|(Z - a\rho'_i)x\rangle \geq Tr(Z) - a(\hat{M}+1) \end{aligned}$$

As $t \to \infty$, this implies $\langle x|(Z - a\rho'_i)x\rangle \geq 0\ \forall x \in \mathbb{R}^N$. Therefore

$$Z \geq a\rho'_i \qquad\qquad i = 1, 2, \ldots N \tag{2.8}$$

We now claim $a \neq 0$; indeed, setting $a = 0$ and $E_m = 0$, (2.7) states $Tr(Z) \leq 0$, while (2.8) means $Z \geq 0 \implies Z = 0$, contradicting the assumption that $(Z, a) \neq (0, 0)$. Hence we may define $\hat{\lambda} = Z/a$.

Now (2.7), (2.8) describe the following conditions:

$$\begin{aligned} &\hat{M} \geq Tr(\hat{\lambda}) \\ &\hat{\lambda} \geq \rho'_i \qquad\qquad i = 1, 2, \ldots N \end{aligned}$$

Let $\Gamma = \{\lambda \in \mathcal{L} : \ \lambda \geq \rho'_i \mid i = 1, 2, \ldots N\}$. Then $\forall \lambda \in \Gamma, \{E_m\} \in \mathcal{M}$, from the completeness relation we see that:

$$Tr(\lambda) - M(\{E_m\}) = \sum_{m=1}^{N} E_m(\lambda - \rho'_m) \geq 0$$

Hence we have $Tr(\hat{\lambda}) = \hat{M}$; furthermore, we have proven that our original maximization problem is equivalent to the dual problem:

$$\max_{\substack{E_m \geq 0 \\ \sum_{i=1}^{N} E_i = I}} \sum_{i=1}^{N} Tr(E_i \rho'_i) = \min_{\lambda \geq \rho'_m} Tr(\lambda) \qquad\qquad m = 1, 2, \ldots N \tag{2.9}$$

Now suppose $\{\hat{E}_m\}$ and $\hat{\lambda}$ provide optimal solutions to (2.9). Then, using the completeness relation,

$$\sum_{m=1}^{N} Tr\left(\hat{E}_m(\hat{\lambda} - \rho'_m)\right) = 0$$

The operators $\hat{E}_m$ and $\hat{\lambda} - \rho'_m$ are both positive semidefinite. Therefore we may apply Lemma 1: the first statement of the lemma shows that each summand is equal to 0; the second part of the lemma requires:

$$\hat{E}_m(\hat{\lambda} - \rho'_m) = (\hat{\lambda} - \rho'_m)\hat{E}_m = 0 \qquad\qquad m = 1, 2, \ldots N \tag{2.10}$$

Summing (2.10) over all $m$ and using the completeness relation gives us

$$\hat{\lambda} = \sum_{m=1}^{N} \hat{E}_m \rho'_m = \sum_{m=1}^{N} \rho'_m \hat{E}_m \tag{2.11}$$

Now the condition $\hat{\lambda} \geq \rho'_m \ \ \forall m$ (note that this requires $\hat{\lambda}$ to be positive semidefinite, hence Hermitian), (2.10), and (2.11) supply the desired conditions stated in the theorem.

To prove sufficiency, assume $\{\hat{E}_m\}$ and $\hat{\lambda}$ satisfy the conditions of the theorem. It suffices to show that they satisfy (2.9); that is

$$\sum_{m=1}^{N} Tr(\hat{E}_m \rho'_m) = Tr(\hat{\lambda})$$

$$\iff \sum_{m=1}^{N} Tr\left(\hat{E}_m(\rho'_m - \hat{\lambda})\right) = 0$$

Indeed, from (2.1), $Tr\left(E_m(\hat{\lambda} - \rho'_m)\right) = 0$. Therefore summing over $m$ maintains a vanishing trace; hence $\{E_m\}$ is an optimal measurement.

□

We have seen the necessary and sufficient conditions for the optimal measurement. However, this unfortunately provides little clues to determining the optimal measurement; in fact, no analytic solution for the general problem is yet known. [EMV08] and [Hel76] provide some computational methods for approximating the desired measurement.

To find the optimal measurement for a given problem, it is necessary to consider the form of the quantum states $\rho_i$. For many applications, it is sufficient to examine the condition when the quantum states to be distinguished are linearly independent pure states, examined in the next section.

## 2.2 Optimum Testing of Linearly Independent Pure States

A frequently recurring situation, and one important for later investigations, occurs when the quantum systems to be determined are linearly independent pure states. That is, each density operator is a rank-1 projection operator associated with a single state, and these states are linearly independent in a finite dimensional state space $\mathcal{H}_s$. The main theorem concerning this situation is that a complete von Neumann measurement provides the optimal measurement. This result follows directly from the previous theorem, and is discussed in [Hel76], [Ken74], [EMV08].

**Theorem 2.4.** *Suppose $\{\rho_i = |\phi_i\rangle \langle\phi_i|\}$ are a set of N fixed pure quantum states in an N dimensional state space $\mathcal{H}_s$[1]. equipped with a set of prior probabilities $\{p_i\}$. Then there exists a unique complete von Neumann measurement that supplies the optimal measurement.*

*Proof.* We prove existence. Suppose that $\{E_m\}$ satisfy the conditions for optimality and, as above, $\lambda = \sum_{i=1}^{N} p_i E_i |\phi_i\rangle \langle\phi_i|$.

We claim that $\mathcal{H}_s$ contains $N$ linearly independent vectors $\{|F_j\rangle\}$ such that $\langle\phi_i|F_j\rangle = \alpha_i\delta_{ij}$ for some set of positive real numbers $\{\alpha_i\}$. Consider the $n-1$ dimensional subspace spanned by $\{|\phi_i\rangle \ : \ i \neq j\}$. There exists non-zero $|F_j'\rangle$ orthogonal to this subspace. Let c $= \langle\phi_j|F_j'\rangle$. Then $|F_j\rangle = c^* |F_j'\rangle$ is orthogonal to each $|\phi_i\rangle \ : i \neq j$ and $\langle\phi_j|F_j\rangle = c^*c > 0$.

We now calculate $\lambda |F_j\rangle$:

$$\lambda |F_j\rangle = \left(\sum_{i=1}^{N} p_i E_i |\phi_i\rangle \langle\phi_i|\right) |F_j\rangle = p_j\alpha_j E_j |\phi_j\rangle$$

[1]For most applications, the $\{\phi_i\}$ span $\mathcal{H}_s$, i.e. $dim(\mathcal{H}_s) = N$. However, when $dim(\mathcal{H}_s > N$ but still finite, we use $N+1$ measurement operators with $E_{N+1} = I - \mathcal{P}_{\mathcal{H}_N}$ where $\mathcal{P}_{\mathcal{H}_N}$ is projection onto $\mathcal{H}_N$. However, the probability of achieving the "outcome" $N+1$ is 0, as each possible outcome is in the null space of $E_{N+1}$. Now the completeness relation on the operators $\{E_m\}_{m=1}^N$ requires $\sum_{m=1}^N E_m = I_{\mathcal{H}_N}$. Examining the optimality conditions in Theorem 3.2.1 show that, due to the impossibility of measuring outcome $n+1$, we need consider only $\{E_m\}_{m=1}^N$ with $\lambda = \sum_{m=1}^N E_m\rho_m'$.

The conditions of Theorem 2.3 require

$$E_j(p_j \left|\phi_j\right\rangle \left\langle\phi_j\right| - \lambda) = 0 \tag{2.12}$$

Therefore

$$\begin{aligned} 0 &= E_j(p_j \left|\phi_j\right\rangle \left\langle\phi_j\right| - \lambda) \left|F_k\right\rangle \\ &= E_j p_j \left|\phi_j\right\rangle \left\langle\phi_j|F_k\right\rangle - p_k\alpha_k E_j E_k \left|\phi_k\right\rangle \\ \iff p_k\alpha_k E_j E_k \left|\phi_k\right\rangle &= \delta_{jk} p_k \alpha_k E_k \left|\phi_k\right\rangle \\ E_j E_k \left|\phi_k\right\rangle &= \delta_{jk} E_k \left|\phi_k\right\rangle \end{aligned}$$

This last fact invites us to consider a new set of vectors that will serve as a basis: define $\left|\mu_k\right\rangle = E_k \left|\phi_k\right\rangle$. Then $E_i \left|\mu_k\right\rangle = \delta_{ik} \left|\mu_k\right\rangle$. We claim these vectors are linearly independent: suppose they are not. Then $\exists \left|\beta\right\rangle \in \mathcal{H}_N$ such that $\left\langle\beta|\mu_i\right\rangle = 0$ for all $i$. Then the condition $\lambda \geq \rho_i'$ requires:

$$\begin{aligned} \left\langle\beta|\lambda|\beta\right\rangle &\geq p_i \left\langle\beta|\phi_i\right\rangle \left\langle\phi_i|\beta\right\rangle \\ \left\langle\beta\right| \sum_{m=1}^{N} E_m \left|\phi_m\right\rangle \left\langle\phi_m|\beta\right\rangle\right\rangle &= \sum_{m=1}^{N} \left\langle\phi_m|\beta\right\rangle \left\langle\beta|\mu_m\right\rangle \geq p_i \left|\left\langle\phi_i|\beta\right\rangle\right|^2 \end{aligned}$$

Each summand on the left vanishes by assumption, but the quantity on the right is strictly greater than 0 for some $\left|\phi_i\right\rangle$ since $span\left(\{\left|\phi_i\right\rangle\}\right) = \mathcal{H}_s$. Hence no such $\left|\beta\right\rangle$ may exist, so the vectors $\{\left|\mu_i\right\rangle\}$ are linearly independent.

Now, the operators are rank-1 projectors, for if $\left|\Psi\right\rangle = \sum_{k=1}^{N} a_k \left|\mu_k\right\rangle$ then

$$E_j E_i \left|\Psi\right\rangle = E_j(a_i \left|\mu_i\right\rangle) = a_i \delta_{ij} \left|\mu_i\right\rangle = \delta_{ij} E_i \left|\Psi\right\rangle$$

Therefore $\{E_m\}$ composes a set of projection operators of the form $\{E_m\} = \left|\mu_m\right\rangle \left\langle\mu_m\right|$.

Finally, note that the completeness relation implies the orthonormality of the set $\{\left|\mu_i\right\rangle\}$:

$$\begin{aligned} \left|\mu_i\right\rangle &= \left( \sum_{j=1}^{N} \left|\mu_j\right\rangle \left\langle\mu_j\right| \right) \left|\mu_i\right\rangle \\ \iff 0 &= \left(\left\langle\mu_i|\mu_i\right\rangle - 1\right) \left|\mu_i\right\rangle + \sum_{j\neq i}^{N} \left\langle\mu_j|\mu_i\right\rangle \left|\mu_j\right\rangle \end{aligned}$$

The $\{\left|\mu_i\right\rangle\}$ are linearly independent, so $\left\langle\mu_i|\mu_j\right\rangle = \delta_{ij}$.

Proof of uniqueness is found in [Ken74] □

We conclude this section with an equivalent optimality condition for pure states that span the state space (actually a specification of a general condition derived from Theorem 2.3 that is outside the scope of this thesis). It will be extremely useful in formulating the optimal measurement problem as a matrix problem.

**Proposition 2.5.** *A complete von Neumann measurement* $\{E_m\} = \{|\mu_m\rangle \langle\mu_m|\}$ *is optimal in distinguishing a set of* $N$ *uniformly distributed pure quantum states* $\{\rho_i\} = \{|\phi_i\rangle \langle\phi_i|\}$ *if and only if*

$$\langle\mu_j|\phi_j\rangle \langle\mu_k|\phi_j\rangle^* = \langle\mu_j|\phi_k\rangle \langle\mu_k|\phi_k\rangle^* \tag{2.13}$$

$$\sum_{m=1}^{N} \langle\mu_m|\phi_m\rangle |\mu_m\rangle \langle\phi_m| \geq |\phi_i\rangle \langle\phi_i| \qquad \forall i = 1, 2, \ldots N \tag{2.14}$$

*for all* $j, k \in \{1, 2, \ldots N\}$.

*Proof.* We begin with the forward implication. Assume the orthonormal set $\{|\mu_m\rangle\}$ specify an optimal measurement. Then the condition $(\lambda - \rho'_k)E_k = 0$ implies

$$\begin{aligned} \lambda E_k &= \frac{1}{N}\rho_k E_k \\ E_j \left( \sum_{m=1}^{N} E_m \frac{1}{N} \rho_m \right) E_k &= E_j \frac{1}{N} \rho_k E_k \\ E_j \rho_j E_k &= E_j \rho_k E_k \end{aligned} \tag{2.15}$$

Expanding the final equality gives (2.13). The condition $\lambda \geq \rho'_i$ is clearly equivalent to (2.14)

To prove the converse, assume (2.15) is true. Then

$$E_j(\rho_j - \rho_k)E_k = 0$$

Summing over $j$ yields $(\lambda - \rho'_k)E_k = 0$. As noted, the condition (2.14) says $\lambda - \rho'_i$ is positive semidefinite. Finally, $E_i(\lambda - \rho'_i) = E_i\rho'_i - E_i\rho'_i = 0$. □

### 2.2.1 Distinguishing Linearly Dependent Pure States

We present an upper bound on the success probability for distinguishing between a finite number of linearly dependent pure states.

**Theorem 2.6.** *Suppose* $\{\rho_i = |\phi_i\rangle \langle\phi_i|\}$ *are a set of* $N$ *pure states in a finite dimensional state space* $\mathcal{H}_s$ *with a uniform distribution (the system is in state* $\rho_i$ *with probability* $1/N$*.. Suppose* $\{|\phi_i\rangle\}$ *span a* $k$ *dimensional subspace* $W$*. Then any measurement will correctly determine that a system is in a state* $|\phi_i\rangle$ *with probability*

$$\sum_{i=1}^{N} p_i P(i|i) \leq \frac{k}{N}$$

*Proof.* From [MP09]. Let $P_W$ denote projection onto $W$. Then $P_W - \rho_i \geq 0$ for all $\rho_i$. Let $\{E_j\}$ be an arbitrary measurement consisting of $N$ measurement operators. Then we may bound the success probability as follows:

$$\frac{1}{N}\sum_{i=1} NTr(E_i\rho_i) \leq \frac{1}{N}\sum_{i=1} NTr(E_i P_W) = \frac{1}{N} Tr(P_W) = \frac{k}{N}$$

□

## 2.3 Matrix Formulation

For our later purposes, due to the binary nature of qubits, we consider the problem of distinguishing states in a finite-dimensional state space with $\mathcal{H}_s \cong \mathbb{C}^{2^n}$. The previous corollary states that the optimal measurement for distinguishing $N = 2^n$ linearly independent pure states $\{\psi_i\}$ is some complete von Neumann measurement, specifying an orthonormal set $\{\mu_i\}$. Suppose they are distributed uniformly, so $p_i = \frac{1}{N}$. Then the probability of correctly determining the state is

$$\begin{aligned}\sum_{i=1}^{N} P(i|i) &= \sum_{i=1}^{N} Tr\left(|\mu_i\rangle \langle\mu_i|\psi_i\rangle \langle\psi_i|\right) \\ &= \sum_{i=1}^{N} |\langle\mu_i|\psi_i\rangle|^2 \end{aligned} \tag{2.16}$$

Recall that unitary transformations are equivalent to orthonormal basis transformations. Therefore measurement by an arbitrary complete von Neumann measurement is equivalent to subjecting the system to an arbitrary unitary transformation and measuring in the computational basis. Let $\{|i\rangle\}$ denote the computational basis. Consider the matrix $A$ whose $i^{th}$ column is $|\psi_i\rangle$ in the computational basis, that is

$$A = \sum_{k=1}^{N} |\psi_k\rangle \langle k|$$

Let

$$U = \sum_{j=1}^{N} |j\rangle \langle e_j|$$

be the unitary matrix mapping an arbitrary orthonormal basis $\{|e_j\rangle\}$ to the computational basis. Let $B = UA$. Then

$$B_{ij} = \langle i|Bj\rangle = \langle i|UAj\rangle = \langle e_i|\psi_j\rangle$$

Furthermore, the probability of successful measurement using the measurement vectors $\{|e_i\rangle\}$ is, from (2.16)

$$\sum_{i=1}^{N} |\langle e_i|\psi_i\rangle|^2 = \|d(B)\|^2$$

where $d(B)$ denotes projection onto diagonal matrices and $\|\ \|$ denotes the $L^2$ norm. Now Proposition 3.3.1 may be applied to describe conditions on $B$ when the $\{|e_i\rangle\}$ describes an optimal measurement:

**Corollary.** *For A, B, U as described above, $\{|e_i\rangle\}$ describes an optimal measurement of the states $\{|\phi_i\rangle\}$ if and only if*

$$B_{jj}B_{kj}^* = B_{jk}B_{kk}^*$$

$$\sum_{m=1}^{N} B_{mm} |e_m\rangle \langle\psi_m| \geq |\psi_i\rangle \langle\psi_i| \qquad \forall i = 1, 2, \ldots N$$

This is refined by noticing that the first condition describe the coefficients of the matrices $d(B)B^\dagger$ and $Bd(B)^\dagger$ respectively. Furthermore, as noted in [GQL], $\|d(B)\|^2$ is constant under left multiplication by diagonal unitary matrices, so we may assume $d(B)$ contains non-negative real entries.

**Corollary.** *For $A$, $B$, $U$ as described above, $\{|e_i\rangle\}$ describes an optimal measurement of the states $\{|\phi_i\rangle\}$ if and only if*

$$d(B)B^\dagger = Bd(B)$$
$$\sum_{m=1}^{N} B_{mm} |e_m\rangle \langle\psi_m| \geq |\psi_i\rangle \langle\psi_i| \qquad \forall i = 1, 2, \ldots N$$

The first condition requires $B$ to have constant diagonal (if there are enough terms in the off-diagonal part of $B$). An extremely important result grew from considering the Gram matrix of $A$:

**Definition.** The **Gram matrix** of an $n \times k$ matrix $A$ is the $k \times k$ matrix $G = A^\dagger A$.

We see from the definition that elements of the Gram matrix are the inner products of columns of $A$, that is

$$G_{i,j} = \langle A_i | A_j \rangle$$

where $A_i$ denotes the $i^{th}$ column of $A$. We know from the polar decomposition [linAlg] that there exists a unitary operator such that $A = U\sqrt{G}$. Therefore we consider the $\sqrt{G}$ matrix as a measurement (i.e., in the above formulation, we take $B = \sqrt{G}$. It turns out that if $\sqrt{G}$ has constant diagonal, then it supplies the optimal measurement!

This result is discussed frequently in the literature; its proof is given in [HMP$^+$03], [SKIH97] and discussed in [EJ01], [EMV08].

**Theorem 2.7.** *Suppose $\{\rho_i = |\phi_i\rangle \langle\phi_i|\}$ is a set of $N$ linearly independent pure states in a finite dimensional state space $\mathcal{H}_s$. Let $A$ be the matrix whose $i^{th}$ column is $\phi_i$. Then the square root of the Gram matrix of $A$, $\sqrt{A^\dagger A}$, describes the optimal measurement iff $\sqrt{A^\dagger A}$ has constant diagonal. Then the probability of successful measurement is $|Tr(\sqrt{A^\dagger A})|^2/N$.*

# Chapter 3

# Quantum Concept Learning

## 3.1 Quantum Algorithms

We are now equipped to discuss actual quantum algorithms. A quantum algorithm involves initializing some quantum state, performing certain unitary operations on it, and measuring the result. The result is *a priori* one of a fixed number of states; therefore it is a problem in quantum state discrimination. For most algorithms, measurement in the computational basis suffices.

## 3.2 Concept Learning

A natural setting for the discussion of quantum algorithms is concept learning. This was introduced as a way of interpreting quantum algorithms by [HMP$^+$03], but the field has its roots in computational learning theory. For a discussion of classical concept learning, see [Ang88].

**Definition.** A **concept** is a map $c : X \to \mathbb{Z}_2$. The set $c^{-1}(1)$ is the **extension** of concept $c$.

For example, let the set $X$ be the set of length $n$ binary strings, and a concept $\phi_a$ defined by $\phi_a(x) = 1$ if and only if $x = a$ for some $n$-bit string $a$. This particular example describes the problem of unstructured search and will be addressed later.

A reason concept learning is an applicable framework for quantum algorithms is that it relies fundamentally on the notion of an oracle, or "black box". The task of an algorithm, or *learner* is to determine a concept. It gains information about the concept by querying an oracle, which encodes (or "hides") the concept in some way. Usually, the learner knows that the concept is a member of a set of many concepts.

**Definition.** A **concept class** is a set of concepts $\mathcal{C} = \{c_i | c_i : X \to \mathbb{Z}_2\}$. The task of an algorithm is to identify a **target concept** $c \in \mathcal{C}$.

Sometimes a concept class is a partition of disjoint sets of concepts $\mathcal{C}_i$ and teh goal of the learning algorithm is to determine which partition a target concept is a member of. Typically the members of a concept class are all similar, so that the learner knows

beforehand what *type* of problem it is facing. For example, in the unstructured search problem described above, the concept class would consist of $\mathcal{C} = \{\phi_a | a \in \mathbb{Z}_2^n\}$. Hence a learner would know what form the target concept has. With each concept in a concept class there is an associated oracle, and the quantum algorithm has access to the oracle corresponding to the target concept. This is the foundation for the earlier claim that a quantum algorithm must distinguish between a fixed set of known quantum states. Each state corresponds to final state achieved from performing the algorithm using each of the unique oracles (as described in the formulation of a general quantum algorithm). Deducing the target concept is the same as deciding which of these states is produced, and hence is a problem in quantum state discrimination.

There are many different types of oracles associated with any concept; the most basic are described in [Ang88]. Useful for our purposes is the *membership oracle*, which is equivalent to the concept map.

**Definition 3.2.1.** The **membership oracle** associated with a concept $c$ hides a function $f_c : X \to \mathbb{Z}_2$ is defined by $f_c = 1 \iff c(x) = 1$. As before, the quantum oracle $\mathcal{O}_{f_c}$ that hides $f_c$ acts on the computational basis by $\mathcal{O}_{f_c} |x, b\rangle = |x, b \oplus f_c(x)\rangle$.

Later we will implement oracles that hide more complicated functions $f_c$.

### 3.2.1 The Oracle

The algorithms we consider, and in fact most quantum algorithms, involve the idea of an *oracle* or *black box* which holds a "hidden" function, $f$. The goal of an algorithm is to deduce either the function or some property of that function. The algorithm uses the oracle via queries: it provides an input $x$ and the oracle responds with $f(x)$.

**Definition.** The **query complexity** of an algorithm implementing an oracle is the number of queries submitted to the oracle

To provide comparison with classical algorithms, we want this function just to act on bits, so $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$. Later we will generalize this slightly so that the range of $f$ may be a set larger than $\mathbb{Z}_2$, but for now this suffices. According to the postulates, the oracle must act unitarily on quantum states. To account for this, an oracle is implemented by inputting a quantum state tensored to a *response register* bit. The oracle leaves the original state unchanged but writes output to the register bit( via a unitary transformation). This is described formally below:

**Definition.** Suppose we have a quantum system with a $2^n$ dimensional state space $\mathcal{H}_s \cong \mathbb{C}^{2^n}$. Then the **oracle** that contains a function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ is the operator $\mathcal{O}_f : \mathbb{C}^{2^n} \otimes \mathbb{C}^2$ is defined by its action on the computational basis for $\mathcal{C}^{2^N} \otimes \mathcal{C}^2$:

$$\mathcal{O}_f(|i\rangle \otimes |b\rangle) = |i\rangle \otimes |b \oplus f(i)\rangle$$

where $i \in \mathbb{Z}_2^n$, $b \in \mathbb{Z}_2$, and $\oplus$ denotes addition mod 2.

The oracle is unitary because the oracle is defined by permutations of the computational basis.

Even without knowing anything about the function $f$, there are interesting ways to extract information from the oracle.

**Proposition 3.1.** Suppose we have a quantum system with a $2^n$ dimensional state space $\mathcal{H}_s \cong \mathbb{C}^{2^n}$ and an oracle $\mathcal{O}_f : \mathbb{C}^{2^n} \otimes \mathbb{C}^2 \to \mathbb{C}^{2^n} \otimes \mathbb{C}^2$ hiding a function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$. Let $|i\rangle$ be an element of the computational basis.

$$\mathcal{O}_f(|i\rangle \otimes |-\rangle) = (-1)^{f(i)} \otimes |-\rangle$$

*Proof.*

$$\mathcal{O}_f(|i\rangle \otimes |-\rangle) = |i\rangle \otimes (|0 \oplus f(i)\rangle - |1 \oplus f(i)\rangle)$$

Notice that, if $f(i) = 0$ then

$$|i\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) = |i\rangle \otimes |-\rangle$$

and if $f(i) = 1$ then

$$|i\rangle \otimes \frac{1}{\sqrt{2}}(|1\rangle - |0\rangle) = (-1)\,|i\rangle \otimes |-\rangle$$

Therefore,

$$\mathcal{O}_f(|\eta_n\rangle \otimes |-\rangle) = (-1)^{f(i)}\,|i\rangle \otimes |-\rangle$$

as desired. □

This method of extracting information from using an oracle once is called *phase kickback*.

### 3.2.2 The General Quantum Algorithm

We have described all of the necessary parts to describe a general quantum algorithm. We assume that each algorithm begins with an initialized state $\Psi_0 = |0 \dots 0\rangle$. The operations available to us are oracle queries and unitary operations. Finally, the resulting state is measured. So if the algorithm has access to an oracle $\mathcal{O}_f$, the algorithm consists of interpreting the measurement of the state:

$$|\Phi\rangle = U_k \mathcal{O}_f U_{k-1} \dots U_1 \mathcal{O}_f U_0\,|\Psi_0\rangle$$

where the unitary operators $\{U_k\}$ are fixed by the algorithm.

### 3.2.3 The Deutsch-Jozsa Algorithm

We're ready to put everything together to investigate a quantum algorithm completely unique from any classical counterpart. The first is the Deutsch-Jozsa algorithm, discovered by David Deutsch and Richard Jozsa [NC00]. It uses an oracle to deduce whether $f$ is constant ($f(i) = f(j)\ \forall i, j \in \mathbb{Z}_2^n$) or balanced ($|f^{-1}(0)| = |f^{-1}(1)| = 2^{n-1}$.

**Proposition 3.2.** Suppose we are given an oracle $\mathcal{O}_f : \mathbb{C}^{2^n} \otimes \mathbb{C}^2$ that hides a function $f : \mathbb{Z}^{2^n} \to \mathbb{Z}^2$ that is known to be either constant or balanced. Then there exists a quantum algorithm to determine whether $f$ is constant or balanced using a single query to the oracle.

First, note that classically, any algorithm with access to an oracle that computes values of $f$ (analogous to the action of the quantum oracle), requires at most $2^{n-1}+1$ queries. The proposition remarkably states that only 1 query is required using a quantum algorithm!

*Proof.* We provide a quantum algorithm that determines whether $f$ is constant or balanced with one query to $\mathcal{O}_f$.

**Algorithm**: Deutsch-Jozsa

1. Prepare the state

$$|\Psi_1\rangle = (H^{\otimes n} \otimes H)(|0 \dots 0\rangle \otimes |1\rangle) = |\eta_n\rangle \otimes |-\rangle$$

2. Apply the oracle $\mathcal{O}_f$ to $|\Psi_1\rangle$:

$$|\Psi_2\rangle = \mathcal{O}_f |\Psi_1\rangle = \mathcal{O}_f(|\eta_n\rangle \otimes |-\rangle) = \left( \frac{1}{\sqrt{2^n}} \sum_{z \in \mathbb{Z}_2^n} (-1)^{f(z)} |z\rangle \right) \otimes |-\rangle$$

3. Apply $H^{\otimes n} \otimes I$ to $|\Psi_2\rangle$:

$$|\Psi_3\rangle = (H^{\otimes n} \otimes I) |\Psi_2\rangle = \left( \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} \sum_{z \in \mathbb{Z}_2^n} (-1)^{x \cdot z + f(z)} |x\rangle \right) \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

4. Measure the first $n$ qubits of $|\Psi_3\rangle$ in the computational basis.

The coefficient of $|0 \dots 0\rangle$ is

$$\frac{1}{2^n} \sum_{z \in \mathbb{Z}_2^n} (-1)^{f(z)}$$

This quantity is 0 if $f$ is balanced, and equal to $\pm 1$ if $f$ is constant. Hence the outcome is $|0 \dots 0\rangle$ iff $f$ is constant. Therefore, with one query to $\mathcal{O}_f$ we may determine whether $f$ is balanced or constant. □

**Remark**. In the formulation of a general quantum algorithm, the Deutsch-Jozsa algorithm is the result of measuring (in the computational basis) the the first $n$ qubits of the state

$$|\Phi\rangle = (H^{\otimes n} \otimes I)\mathcal{O}_f(H^{\otimes n} \otimes H) |0\rangle^{\otimes 2n+1}$$

### 3.2.4 Single-Query Learning

Suppose we wish to tell how well a quantum algorithm can determine a target concept using only a single query. Comparison to classical algorithms with this sort of problem obviously is not measured by the query complexity of an algorithm, but rather the probability of success given one query. Since the algorithm is frequently unable to determine the the target concept with probability one with only a single query, this problem is sometimes referred to as *impatient learning*.

An advantage of studying single-query learning is that the general algorithm is much simpler. Any single-query algorithm now consists in measuring the state

$$|\Phi\rangle = U_1 \mathcal{O}_f U_0 \, |0 \dots 0\rangle$$

We simplify this further by restricting our attention to algorithms which subject the state $\frac{1}{\sqrt{|X|}} \sum_{i \in X} |i\rangle \otimes |-\rangle$ to the oracle. This seems like a good idea if we are searching for a target concept in a sufficiently symmetric concept class (although we will later generalize the algorithm to include different response registers). We therefore simplify the general single-query algorithm to measuring the state

$$|\Phi\rangle = \frac{1}{\sqrt{|X|}} \sum_{i \in X} (-1)^{c(i)} \, |i\rangle$$

**Definition 3.2.1.** Let $\mathcal{C}$ be a concept class over a set $X$. Then the **membership learning matrix** $M_{\mathcal{C}}$ is the $|X| \times |\mathcal{C}|$ matrix with entries

$$(M_{\mathcal{C}})_{i,j} = (-1)^{c_j(i)}$$

where $c_j \in \mathcal{C}$.

The main results from Chapter 2 tell us that if we know that if these columns are orthogonal, then there automatically exists a complete von Neumann measurement that discriminates between these states with probability 1 (this measurement is the set of projectors on those states). If the states are linearly independent and the Gram matrix $G = M_{\mathcal{C}}^{\dagger} M_{\mathcal{C}}$ has constant diagonal, then the square root Gram matrix $\sqrt{G}$ describes the optimal measurement. If the columns span a $k < |X|$ dimensional subspace, then any measurement is successful with probability at most $\frac{k}{n}$. Generally, the measurement query matrix provides a good way to study the single-query strategy of passing the state $|\eta_n\rangle \otimes |-\rangle$ to the oracle.

The choice of response register is important. If $|b\rangle = |0\rangle$, then the oracle writes the value $f_c(x)$ to the register. If $|b\rangle = |+\rangle$, then the oracle acts by the identity. If $|b\rangle = |-\rangle$, the oracle acts via a phase kickback on the input state:

$$\mathcal{O}_c(|\eta\rangle \otimes |-\rangle) = \frac{1}{\sqrt{|X|}} \sum_{x \in X} (-1)^{f_c(x)} \, |x\rangle \otimes |-\rangle$$

Many algorithms use $|-\rangle$ as the response register. However it is not always the optimal register.

## 3.3 Grover's Algorithm

Let's consider now the problem of unstructured search mentioned before. The algorithm provided below was first provided by Lov Grover [NC00]. For simplicity, we suppose that we are searching for a single element of $\mathbb{Z}_2^N$. We are provided with the concept class

$$\mathcal{G}^N = \{c_a : \mathbb{Z}_2^N \to \mathbb{Z}_2 \mid c_a(x) = \delta_{x,a}\}$$

Denote by $\mathcal{O}_a$ the oracle associated with $c_a \in \mathcal{G}^N$ which acts by

$$\mathcal{O}_a |x, b\rangle = |x, b \oplus \delta_{x,a}\rangle$$

Clearly a classical algorithm requires $2^N$ queries to deduce $c_a$ in the worst case. However, a quantum algorithm requires only $O(\sqrt{2^N})$ queries to deduce $c_a$ with high probability![1]

**Theorem 3.3.1.** *A quantum algorithm deduces a target concept $c_a \in \mathcal{G}^N$ with success probability $O(1)$ with $\sqrt{2^N}$ oracle queries.*

*Proof.* Suppose $c_a(x) = \delta_{a,x}$ is the target concept. Let $|\Psi\rangle = \sum_{i \in \mathbb{Z}_2^N} \alpha_i |i\rangle$ be an arbitrary unit vector in $\mathbb{C}^{2^N}$. Notice the effect of submitting $|\Psi\rangle |-\rangle$ to the membership oracle $\mathcal{O}_{c_a}$:

$$\mathcal{O}_{c_a} |\Psi, -\rangle = \left( \sum_{i \in \mathbb{Z}_2^N} (-1)^{\delta_{a,i}} \alpha_i |i\rangle \right) \otimes |-\rangle$$

The effect on the first $N$ qubits has a geometric interpretation of reflecting $|\Psi\rangle$ over the hyperplane orthogonal to $a$ in the space $\mathbb{C}^{2^N}$, as it negates the coefficient of $|a\rangle$ in $|\Psi\rangle$ and leaves the rest unchanged. Intuitively, our goal is to manipulate the vector $|\eta_n\rangle$ using this operation and additional unitary operations so that it is close to $|a\rangle$. Remember that measurement in the computational basis of a state $|\Phi\rangle$ will result in outcome $a$ with probability $\| \langle \Phi | a \rangle \|^2$, so the probability of correct measurement increases as $|\Phi\rangle$ approaches $|a\rangle$. To that end, we introduce the operator

$$R \equiv I - 2 |\eta_n\rangle \langle \eta_n|$$

R is reflection over the hyperplane orthogonal to $|\eta_n\rangle$. To see this, let an arbitrary state

$$|\Psi\rangle = |\eta_n\rangle + |\eta_n\rangle_\perp$$

where $|\eta_n\rangle$ is the component of $|\Psi\rangle$ orthogonal to $|\eta_n\rangle$. Then

$$R |\Psi\rangle = |\eta_n\rangle + |\eta_n\rangle_\perp - 2 |\eta_n\rangle = - |\eta_n\rangle + |\eta_n\rangle_\perp$$

So R negates the component of $|\Psi\rangle$ parallel to $|\eta_n\rangle$ and leaves the rest unchanged. Hence R is reflection over the hyperplane orthogonal to $|\eta_n\rangle$ (and is clearly unitary). Let

$$\theta' = \cos^{-1}(\langle \eta_n | a \rangle) = \cos^{-1}(\frac{1}{\sqrt{2^N}})$$

[1] I assume familiarity with the notation $O(f)$, or *big-O* notation. Roughly, if $f, g$ are functions from $\mathbb{Z}^+ \to \mathbb{R}$, then $f$ is $O(g)$ iff there exists $k \in \mathbb{Z}^+$ and $c \geq 0$ such that $f(n) \leq cg(n)$ for all $n \geq k$.

denote the angle between $|\eta_n\rangle$ and $|a\rangle$. Consider the real 2-dimensional subspace spanned by $|a\rangle$ and $|\eta_n\rangle$. Clearly the effect of $R\mathcal{O}_a$ is rotation by $2\theta'$ in this subspace, as the composition of two reflections is a rotation of twice the angle between the vectors specifying the hyperplanes over which the reflections occur. Now let $\theta \equiv \frac{\pi}{2} - \theta'$. Then $R\mathcal{O}_a$ affects a rotation of $2\pi - 2\theta$, or $-2\theta$. Therefore $-R\mathcal{O}_a$ is a rotation of $2\theta$. We call the operator $-R\mathcal{O}_a$ the *Grover iteration*.

If we start with $|\eta_n\rangle$, we see that one application of the Grover iteration rotates $|\eta_n\rangle$ an angle of $2\theta$ towards $|a\rangle$. This is because the $|a\rangle$ component of $|\eta_n\rangle$ is positive, so the first rotation $\mathcal{O}_a$ rotates $|\eta_n\rangle$ away from $|a\rangle$, and ultimately negating $R\mathcal{O}_a$ causes a rotation towards $|a\rangle$. Figure 3.1 shows the real subspace spanned by $|\eta_n\rangle$ and $|a\rangle$ (the bold lines represent the hyperplanes orthogonal to $|a\rangle$ and $|\eta_n\rangle$) and shows the effect of one Grover iteration.

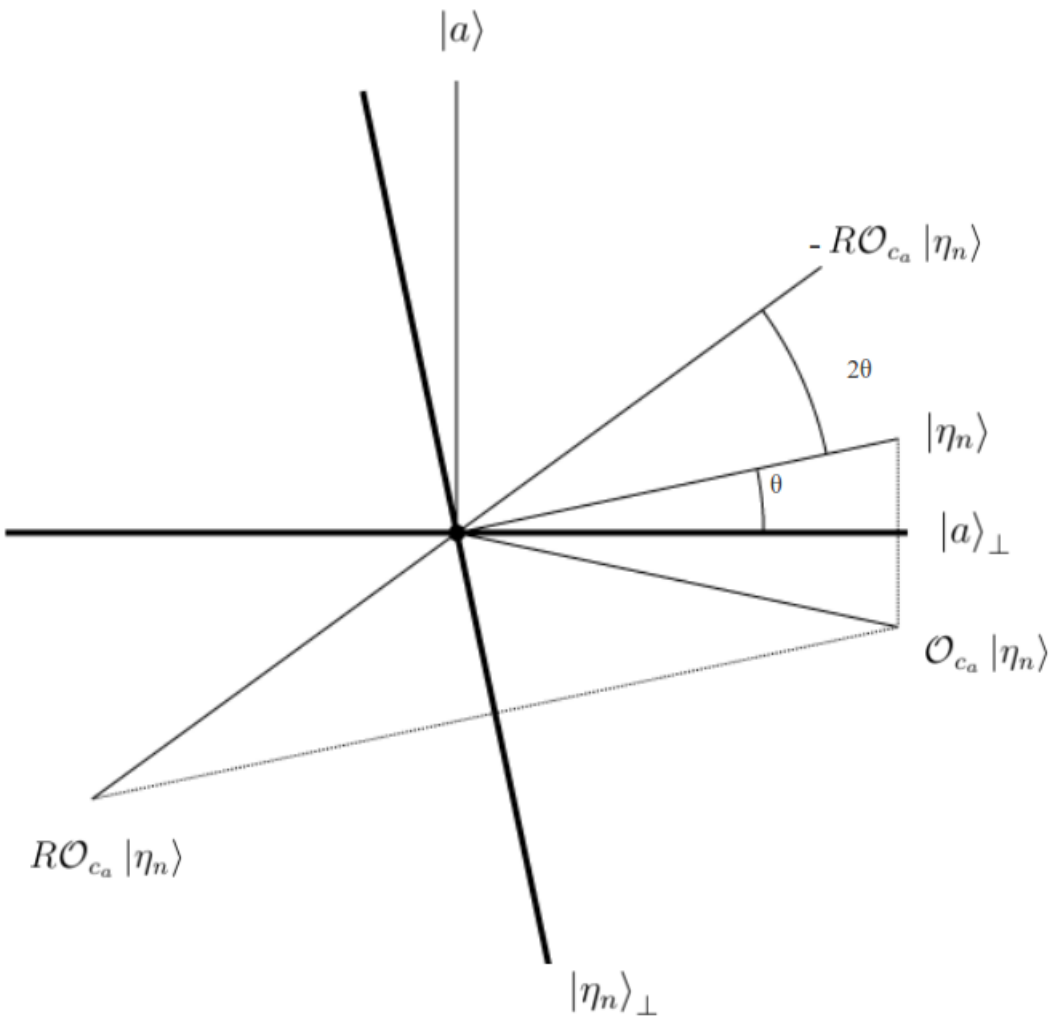


Figure 3.1: Effect of Grover iteration on $|\eta_n\rangle$

We now have the machinery to rotate $|\eta_n\rangle$ towards $|a\rangle$ by $2\theta$ any number of times.

We wish to apply the rotation $k$ times, where $k$ is the smallest integer that minimizes

$$|2k\theta + \theta - \frac{\pi}{2}|$$

so $k = \lfloor \frac{\pi}{4\theta} - \frac{1}{2} \rceil$ (where $\lfloor x \rceil$ denotes closest integer to $x$). Then we may produce a state $|\Phi\rangle$ within $\frac{\theta}{2}$ of $|a\rangle$. The probability of measuring $a$ is:

$$P(a) = \| \langle \Phi | a \rangle \|^2 \geq \cos^2(\frac{\theta}{2}) = 1 - \sin^2(\frac{\theta}{2})$$

Remember that $\theta = \frac{\pi}{2} - \theta'$ and hence

$$\sin\theta = \sin(\frac{\pi}{2} - \cos^{-1}(\frac{1}{\sqrt{2^N}})) = \frac{1}{\sqrt{2^N}}$$

Therefore, performing the Grover iteration on $|\eta_n\rangle$ $k$ times and measuring the resulting state, we will measure $a$ with probability

$$P(a) = 1 - \sin^2(\frac{\theta}{2}) \geq 1 - \sin^2(\frac{\theta}{2}) = 1 - \frac{1}{2^N}$$

which tends to 1 as $N$ increases. However, the algorithm must know beforehand how many times to apply the Grover iteration. It cannot "try out" a measurement and then continue with the algorithm because measurement alters the state of the system. Therefore, we note that for$N \gg 1$, $\theta \approx \sin\theta = \frac{1}{\sqrt{2^N}}$. We may then calculate

$$k = \lfloor \frac{\pi}{4\theta} - \frac{1}{2} \rceil \approx \lfloor \frac{\pi}{4}\sqrt{2^N} \rceil$$

Hence the algorithm requires $O(\sqrt{2^N})$ oracle queries to find $a$ with high probability.

Let us summarize the algorithm:

**Algorithm:** Grover

1. Prepare the state
$$|\Psi_1\rangle = |\eta_N\rangle \otimes |-\rangle$$
2. Apply the operator $-R\mathcal{O}_{c_a}$ (where $R = I - 2|\eta_N\rangle\langle\eta_N|$) to $|\Psi_1\rangle$ a total of $k = \lfloor \frac{\pi}{4}\sqrt{2^N} \rceil$ times to produce:
$$|\Psi_2\rangle = (-R\mathcal{O}_a)^k |\Psi_1\rangle$$
3. Measure $|\Psi_2\rangle$ in the computational basis to achieve outcome $a$ with probability $P(a) \geq 1 - \frac{1}{2^N}$

□

### 3.3.1 Discussion of Grover's Algorithm

It is worth briefly discussing robustness of Grover's Algorithm. For small values of $N$, it is not actually very accurate. For example, with $N = 1$, we have $\theta = \frac{\pi}{4}$, which means that the Grover iteration repeatedly rotates $|\eta_n\rangle$ by $\frac{\pi}{2}$; clearly, there can be no improvement on a success probability $P = \frac{1}{2}$, so Grover's algorithm does no better than straight guessing!

It is an obvious question to determine how well a quantum algorithm will perform with just one query to the membership oracle. To that end, we examine the quantum learning matrix $\mathcal{M}_{\mathcal{O}}$ with response register $|-\rangle$, whose $i^{th}$ column is the first $n$ bits of

$$\mathcal{O}_a |\eta_n\rangle \otimes |-\rangle = \frac{1}{\sqrt{2^n}} \sum_{i \in \mathbb{Z}_2^n} (-1)^{\delta_{a,i}} |i\rangle$$

Hence

$$(\mathcal{M}_{\mathcal{O}})_{x,y} = \begin{cases} -1 & \text{for } x = y \\ 1 & \text{for } x \neq y \end{cases}$$

When $N = 2$, the columns of $\mathcal{M}_{\mathcal{O}}$ are orthogonal, so there is a complete von Neumann measurement that distinguishes these states with probability 1. Therefore, a quantum algorithm may identify a target concept $c_a$ with a single oracle query.

Now considering Grover's algorithm for $N = 2$, we see that $\theta = \frac{\pi}{6}$ and after 1 iterations, $|\eta_n\rangle$ is within $\frac{\pi}{2} - (\frac{\pi}{6} + \frac{\pi}{3}) = 0$ of $|a\rangle$. Therefore a measurement will return outcome $a$ with probability 1. However, the instructions for the algorithm are to iterate

$$k = \lfloor \frac{\pi}{2} \rceil = 2$$

which rotates $|\eta_n\rangle$ within $\frac{\pi}{2} - (\frac{\pi}{6} + 2\frac{\pi}{3}) = -\frac{\pi}{3}$ and will measure $|a\rangle$ with a probability $P(a) = cos^2(\frac{\pi}{3}) = \frac{1}{4}$. This illustrates that Grover's algorithm must be used carefully, with large enough $N$, to ensure that the calculated value of $k$ supplies the optimal number of iterations.

## 3.4 The Bernstein-Vazirani Algorithm

We consider a famous example of a structured search problem. This problem and its solution were first provided by Umesh Vazirani and Ethan Bernstein [BV93]. We consider the concept class

$$\mathcal{BV}^n = \{c_a : \mathbb{Z}_2^n \to \mathbb{Z}_2 \mid c_a(x) = x \cdot a\}$$

Remember that $x \cdot a$ is the binary dot product mod 2. Classically, the worst case requires $n$ oracle queries because each subsequent query eliminates at worst half of the remaining concepts. The quantum version, however, does much better, learning a target concept with just one oracle query in every case.

**Theorem 3.1.** *A quantum algorithm can learn a target concept $c_a \in \mathcal{BV}^n$ with a single oracle query.*

*Proof.* Let $\mathcal{O}_a$ denote the oracle associated with a concept $c_a \in \mathcal{BV}^n$, acting on the computational basis of $\mathbb{C}^{2^n} \otimes \mathbb{C}^2$ by $\mathcal{O}_a |x, b\rangle = |x, b \oplus x \cdot a\rangle$. Consider the membership oracle matrix $\mathcal{M}_{\mathcal{BV}}$. Its columns are

$$\frac{1}{\sqrt{2^n}} \sum_{x \in \mathcal{Z}_2^n} (-1)^{x \cdot a} |x\rangle$$

We may therefore write

$$\mathcal{BV}^n = \frac{1}{\sqrt{2^n}} \sum_{x,a \in \mathcal{Z}_2^n} (-1)^{x \cdot a} |x\rangle \langle a|$$

which we recognize as $H^{\otimes n}$. Therefore we have the following algorithm:

**Algorithm:** Bernstein-Vazirani

1. Prepare the state

$$|\Psi_1\rangle = H^{\otimes n+1} |0 \ldots 01\rangle = |\eta_n\rangle \otimes |-\rangle$$

2. Query the membership oracle $\mathcal{O}_a$:

$$|\Psi_2\rangle = \mathcal{O}_a |\Psi_1\rangle = \left( \frac{1}{\sqrt{2^n}} \sum_{i \in \mathbb{Z}_2^n} (-1)^{a \cdot i} |i\rangle \right) \otimes |-\rangle$$

3. Apply $H^{\otimes n+1}$:

$$|\Psi_3\rangle = H^{\otimes n+1} |\Psi_2\rangle = |a\rangle \otimes 0$$

4. Measure $\Psi_3$ and receive outcome $a0$ with certainty and hence determine $a$.

The algorithm is summarized as:

$$|a0\rangle = H^{\otimes n+1} \mathcal{O}_a H^{\otimes n+1} |0 \ldots 0\rangle$$

□

# Chapter 4

# Hamming Distance Oracles

We now consider oracles that operate as a function of the Hamming distance between bit strings. The Hamming distance between two bit strings is the number of bits at which they differ. This problem was studied first in [HM02] and examined further in [MP09].

## 4.1 Hamming Distance

As in earlier sections, we think if elements of $\mathbb{Z}_2^n$ as bit strings of length $n$.

**Definition.** Suppose $a, x \in \mathbb{Z}_2^n$. Then the **Hamming distance of** $a, x$ is $dist(a, x) = |\{i|\ a_i \neq x_i\}|$. The **Hamming weight** of $x$ is $wt(x) = dist(x, 0)$.

Clearly $dist(a, x) \in \{0, 1, \ldots n\}$. For $a, x \in \mathbb{Z}_2^n$ we use $a + x$ to denote the usual group operation on $\mathbb{Z}_2^n$, i.e. coordinate-wise addition mod 2. For example, $01101 + 11100 = 10001$.

**Definition.** The **complement** of $a \in \mathbb{Z}_2^n$ is written $\bar{a}$ and defined by $\bar{a} = \{11 \ldots 1\} + a$.

Immediately from the definitions come the following relations:

**Proposition.** *Let $a, x \in \mathbb{Z}_2^n$. Then:*

1. $dist(a, x) = dist(x, a) = wt(a + x)$
2. $wt(a) = wt(a + 0 \ldots 0)$
3. $dist(a, x) = dist(a + x, 0 \ldots 0)$
4. $dist(a, x) + dist(\bar{a}, x) = n$

*Proof.* The proofs are similar; we prove (4) as an example: $x$ agrees with $a$ or $\bar{a}$ at each bit, but not both; summing over each bit gives the desired identity. □

## 4.2 The Hamming Distance Oracle

We wish to consider concept classes which hide functions of the Hamming distance. According to our description of concept learning, a valid concept has range $\mathbb{Z}_2^n$. Therefore, we are looking for concepts $c_a : \mathbb{Z}_2^n \to \mathbb{Z}_2$ such that $c_a(x) = f(dist(a, x))$ where $f : 0, 1, \ldots n \to \mathbb{Z}_2$ and $dist(a, x)$ is the Hamming distance between the input string and the hidden string $a$. A natural first choice for the concept class is the mod 2 function:

$$\mathcal{BH}_{n,2} = \{c_a : \mathbb{Z}_2^n \to \mathbb{Z} \mid c_a = dist(a, x) \mod 2\}$$

However, due to the next lemma, the oracles associated with this concept class are nearly useless.

**Lemma (from [HM02]]) 1.** *Suppose $a, a' \in \mathbb{Z}_2^n$. If $dist(a, a') \equiv 0 \mod 2$ then $dist(a, x) \equiv dist(a', x) \mod 2$ for all $x \in \mathbb{Z}_2^n$.*

*Proof.* The bits where $a$ and $a'$ agree contribute equally to $dist(a, x)$ and $dist(a', x)$. There are an even number of bits where $a$ and $a'$ differ. Each of these contributes 1 to $dist(a, x)$ or $dist(a', x)$ but not both. Therefore, over all such bits the parity of $dist(a, x)$ changes by the same amount as the parity of $dist(a', x)$. □

The lemma shows that the most information about a target concept $c_a \in \mathcal{BH}_{n,2}$ than any algorithm can learn, quantum or classical, is $wt(a) \mod 2$. With the foresight that it will be successful, we are compelled to find other functions of the Hamming distance that may provide information that a quantum algorithm can use. Indeed, one such function is presented below that allows single-query learning with certainty when $n$ is even.

## 4.3 The Mod 4 Hamming Oracle

Lemma (4.2) shows that for $\mathcal{H}_{n,2}$, $f = dist(a, x) \mod 2$ provides a (mostly) useless oracle. However, [MePo 2009] show that a concept class hiding a function that returns the second least significant bit of $dist(a, x)$ provides a learning problem which requires only one oracle query. The least significant bit $b(d)$ is defined as

$$b(d) = \begin{cases} 0 & \text{for } d \equiv 0, 1 \pmod 4 \\ 1 & \text{for } d \equiv 2, 3 \pmod 4 \end{cases}$$

Hence we are considering the concept class

$$\mathcal{H}_{n,2}^b = \{c_a : \mathbb{Z}_2^n \to \mathbb{Z}_2 \mid c_a(x) = b(dist(a, x))\}$$

This concept class comes with a caveat: if $n \equiv 1 \mod 4$, then there are only $2^{n-1}$ distinct concepts in $\mathcal{H}_{n,2}^b$.

**Proposition.** *If $n \equiv 1 \mod 4$ then $b(dist(a, x)) = b(dist(\bar{a}, x))$*

*Proof.* See [MP09]. □

Therefore $c_a$ is equivalent to $c_{\bar{a}}$. However, if $n$ is even, there exists a quantum algorithm to identify a target concept with a single query, provided by [MP09].

**Theorem 4.1.** *Let the concept class $\mathcal{H}^b_{n,2} = \{c_a : \mathbb{Z}_2^n \to \mathbb{Z}_2 \mid c_a(x) = b(dist(a,x))\}$. Let $n$ be even. Then a quantum algorithm can determine a target concept $c_a \in \mathcal{H}^b_{n,2}$ with a single oracle query.*

The proof relies on the following lemma, the proof of which are found in [MePo 2009]. For $x \in \mathbb{Z}_2^n$, define $\hat{x}$ by

$$\hat{x} = \begin{cases} x & \text{if } wt(x) \text{ is even} \\ \bar{x} & \text{if } wt(x) \text{ is odd} \end{cases}$$

When $n$ is even, the map $x \to \hat{x}$ is a bijection, and we may define a matrix $P : \mathbb{C}^{2^n}$ by its action on the computational basis:

$$P(x) = \hat{x}$$

$P$ is clearly unitary, and hence a valid operation to use in a quantum algorithm.

**Lemma.** *Let $a, x \in \mathbb{Z}_2^n$. Then*

$$(-1)^{b(dist(a,x))} = (-1)^{wt(a)}(-1)^{wt(x)}(-1)^{a\cdot\hat{x}}$$

*Proof.* (Theorem 4.1)
An oracle associated with a concept $c_a$, denoted $\mathcal{O}_a$ acts on the computational basis by

$$\mathcal{O}_a |x, r\rangle \to |x, r \oplus b(dist(ax))\rangle$$

The theorem is proved by providing the following algorithm, from [MePo 2009]:

**Algorithm:** Hamming Concept Class for even $n$

1. Prepare the state

$$|\Phi_1\rangle = (H^{\otimes n} \otimes H)(|0 \ldots 0\rangle \otimes |1\rangle) = |\eta_n\rangle \otimes |-\rangle$$

2. Apply the the operator $D \otimes I$ which acts on the computational basis of $\mathbb{C}^{2^n}$ as $D|x\rangle = (-1)^{b(wt(x))}|x\rangle$ to $|\Phi_1\rangle$:

$$|\Phi_2\rangle = D|\Phi_1\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \mathbb{Z}_{\not\bowtie}^{\bowtie}} (-1)^{b(wt(x))} |x\rangle \otimes |-\rangle$$

3. Apply the oracle $\mathcal{O}_a$ to $|\Phi_2\rangle$:

$$|\Phi_3\rangle = \mathcal{O}_a |\Phi_2\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \mathbb{Z}_{\not\bowtie}^{\bowtie}} (-1)^{b(wt(x))}(-1)^{b(dist(a,x))} |x\rangle \otimes |-\rangle$$

By the lemma, we may rewrite $|\Phi_3\rangle$:

$$|\Phi_3\rangle = \frac{(-1)^{b(wt(a))}}{\sqrt{2^n}} \sum_{x\in\mathbb{Z}_2^n} (-1)^{b(wt(\hat{x}))}(-1)^{a\cdot\hat{x}} |x\rangle \otimes |-\rangle$$

4. Apply $P \otimes I$ to $|\Phi_3\rangle$:

$$|\Phi_4\rangle = P|\Phi_3\rangle = \frac{(-1)^{b(wt(a))}}{\sqrt{2^n}} \sum_{x\in\mathbb{Z}_2^n} (-1)^{b(wt(\hat{x}))}(-1)^{a\cdot\hat{x}} |\hat{x}\rangle \otimes |-\rangle$$

since $P$ is a bijection,

$$|\Phi_4\rangle = \frac{(-1)^{b(wt(a))}}{\sqrt{2^n}} \sum_{x\in\mathbb{Z}_2^n} (-1)^{b(wt(x))}(-1)^{a\cdot x} |x\rangle \otimes |-\rangle$$

5. Apply $H^{\otimes n+1}$ to $|\Phi_4\rangle$:

$$|\Phi_5\rangle = (H^{\otimes n+1})|\Phi_4\rangle = (-1)^{b(wt(a))}|a\rangle \otimes |0\rangle$$

6. Measure $|\Phi_5\rangle$ in the computational basis to observe the state $|a0\rangle$ with probability 1.

The algorithm is summarized as:

$$|a0\rangle = H^{\otimes n+1}(P \otimes I)\mathcal{O}_a(D \otimes I)H^{\otimes n+1}|0\cdots0\rangle|1\rangle$$

□

## 4.4 Y-Valued Concept Learning

Although the simplest Hamming oracle fails to provide a suitable learning environment, [HM02] and [MP09] show that it is possible to use the Hamming distance to identify a string with a single query. In the first paper, the Hamming distance is calculated mod 4. However, this operation returns an element of $\mathbb{Z}_4$ rather than $\mathbb{Z}_2$, and thus involves oracles that act on a 4-bit response register. We are compelled then to generalize the range of concepts to sets larger than $\mathbb{Z}_2$.

**Definition.** Let $Y$ be a finite set. A **Y-valued concept** is a map $c : X \to Y$. A **Y-valued concept class** is a set of concepts $\mathcal{C}^Y = \{c_x : X \to Y\}$. The goal of a concept learning algorithm is to identify a **target concept** $c \in \mathcal{C}^Y$.

When $|Y| = 2$ the definition specifies to the previous notion of concept learning. The oracles associated with $Y$-valued concept classes must act on Y dimensional response registers. Recall that for $|Y| = 2$, the action of the oracle on a computational basis element $|x, b\rangle \in \mathbb{C}^{2^n} \otimes \mathbb{C}^2$ was $|x, b \oplus f(x)\rangle$. However, the operation $\oplus$ no makes sense on a $|Y|$ dimensional response register because $Y$ may not be endowed with a group operation, and we must consider a much larger set of possible operations on the register. We therefore have an admittedly vague definition for a Y-valued oracle:

**Definition 1.** A **Y-valued oracle** associated with a concept $c \in \mathcal{C}^Y = \{c_x : \mathbb{Z}_2^n \to \mathbb{Z}_Y\}$ is an oracle $\mathcal{O}^Y$ which acts on a computational basis element $|x, b\rangle \in \mathbb{C}^{2^n} \otimes \mathbb{C}^Y$ by

$$\mathcal{O}^Y(|x, b\rangle) = |x, g_{c(x)}(b)\rangle$$

Where $g_{c(x)} : \mathbb{Z}_Y \to \mathbb{Z}_Y$ is some function acting on the response register, determined by the concept $c$.

We are able to describe the family of concept classes which constitutes the main investigation of this paper.

**Definition.** Let $n, r \in \mathbb{Z}^+$. An **(n, r) Hamming concept class** is a set of $r$-valued concepts $\mathcal{H}_{n,r} = \{c_a : \mathbb{Z}_2^n \to \mathbb{Z}_r\}$ such that $c_a(x) = f(dist(a, x))$ for some $f : \mathbb{Z}_{n+1} \to \mathbb{Z}_r$.

The previous algorithm shows that concepts in $\mathcal{H}_{2k,2}$ are learned in a single query with a quantum algorithm (for any $k \in \mathbb{Z}^+$) with a suitable choice of $f_a(x) = b(dist(a, x))$.

### 4.4.1 The 2-bit Register Hamming Concept Class

The next algorithm, found in [HM02], shows that concepts in $\mathbb{H}_{n,4}$ are identified in a single query using $f_a(x) = dist(a, x) \mod 4$. The algorithm uses the 4-dimensional discrete Fourier transform $\mathcal{F}_4$, and the "bit shift" operator $T_d : \mathbb{C}^d \otimes \mathbb{C}^d$ which acts on the computational basis by $T_d |x\rangle = T_d |x \oplus 1\rangle$. Here $\oplus$ denotes addition mod $d$. We may calculate, from the definition of the discrete Fourier transform, the matrix representation of $\mathcal{F}_4$ in the computational basis:

$$\mathcal{F}_4 = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}$$

The algorithm also implements $S = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$, due to its action on the computational basis:

**Lemma 1.** Suppose $x \in \mathbb{Z}_2$. Then

$$S^{\otimes n} |x\rangle = \frac{1}{\sqrt{2^n}} \sum_{y \in \mathbb{Z}_2^n} (i)^{dist(x,y)} |y\rangle$$

*Proof.* We express $S^{\otimes n}$ in outer product notation:

$$S^{\otimes n} |x\rangle = \left(\frac{1}{\sqrt{2}} \left((|0\rangle + i |1\rangle) \langle 0| + (i |0\rangle + |1\rangle) \langle 1|\right)\right)^{\otimes n} |x\rangle \tag{4.1}$$

If we group the tensor products together, we see that each computational basis element $|y\rangle : y \in \mathbb{Z}_2^n$ is multiplied by a factor of $i$ for each bit in which $x$ and $y$ differ, which is what the lemma states. □

To each concept $c_a \in \mathcal{H}_{n,4}$ is associated an oracle which hides the function $f = dist(a,x) \mod 4$). Therefore the oracle requires a 2-bit register to hold the output. Hence each oracle $\mathcal{O}_a$ acts on $\mathcal{C}^{2^n} \otimes \mathcal{C}^4$. We index the computational basis of $\mathcal{C}^4$ by $\{|0\rangle, |1\rangle, |2\rangle, |4\rangle\}$. Therefore we may express the action of the oracle as

$$\mathcal{O}_a |x, b\rangle = |x, b \oplus (dist(a,x) \mod 4)\rangle = (I^{\otimes n} \otimes T_4^{dist(a,x)}) |x, b\rangle$$

We present the algorithm:

**Algorithm**: 2-bit Register Hamming Learning

1. Apply $\mathcal{H}^{\otimes n} \otimes \mathcal{F}_4$ to the state $|0 \dots 0\rangle |1\rangle$:

$$|\Phi_1\rangle = (\mathcal{H}^{\otimes n} \otimes \mathcal{F}_4) |0 \dots 0\rangle |1\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \mathbb{Z}_2^n} |x\rangle \otimes \frac{1}{2}(|0\rangle - |1\rangle + i |2\rangle - i |3\rangle)$$

2. Apply the oracle $\mathcal{O}_a$ to $|\Phi_1\rangle$. First, note that

$$T(|0\rangle - |1\rangle + i |3\rangle - i |4\rangle) = -i(|0\rangle - |1\rangle + i |3\rangle - i |4\rangle)$$

   Therefore

$$\mathcal{O}_a |\Phi_1\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \mathbb{Z}_2^n} (-i)^{dist(a,x)} |x\rangle \otimes \frac{1}{2}(|0\rangle - |1\rangle + i |2\rangle - i |3\rangle)$$

3. Apply $S^{\otimes n} \otimes I$ to $|\Phi_2\rangle$. By the lemma,

$$|\Phi_3\rangle = (S^{\otimes n} \otimes I) |\Phi_2\rangle = \frac{1}{2^n} \sum_{x,y \in \mathbb{Z}_2^n} (i)^{dist(x,y)} (-i)^{dist(a,x)} |y\rangle \otimes \frac{1}{2}(|0\rangle - |1\rangle + i |2\rangle - i |3\rangle)$$

   Note that when summing over $y$, if $y = a$, then the summand becomes

$$\sum_{x \in \mathbb{Z}_2^n} (i)^{dist(x,a)} (i)^{dist(a,x)} |a\rangle = 2^n |a\rangle$$

   Therefore the coefficient of $a$ is 1; hence the first tensor factor is simply $|a\rangle$ since it is a unit vector. Hence

$$|\Phi_3\rangle = |a\rangle \otimes \frac{1}{2}(|0\rangle + i |1\rangle - |2\rangle - i |3\rangle)$$

4. Apply the operation $I \otimes \mathcal{F}_4^{-1}$ to $|\Phi_3\rangle$ to obtain

$$|\Phi_4\rangle = (I \otimes \mathcal{F}_4^{-1}) |\Phi_3\rangle = |a, 1\rangle$$

   And measure the resulting state in the computational basis to receive outcome $a1$ with probability 1.

The algorithm is summarized (letting $I$ represent the identity transformation in both $\mathbb{C}^{2^n}$ and $\mathbb{C}^4$):

$$|a1\rangle = (I \otimes \mathcal{F}_4^{-1})(S^{\otimes n} \otimes I)\mathcal{O}_a(H^{\otimes n} \otimes \mathcal{F}_4) |0 \dots 0\rangle |1\rangle$$

## 4.5 The Permutation Model

We have seen that the $(2m, 2)$ and $(m, 4)$ Hamming distance concept classes (for any $m \in \mathbb{Z}^+$) allow perfect single query learning. A next question is to determine how well a single-query algorithm can learn concepts in a $(2m + 1, 2)$ Hamming distance concept class. For all algorithms, we assume that we pass the equal superposition state to an oracle associated with a concept $c_a$, denoted $\mathcal{O}_a$. Therefore, the problem of determining how well a quantum algorithm will learn from a concept class consists of two parts: first, the concept class must encode the "best" function of the Hamming distance. Recall that a concept in a Hamming distance concept class $c_a(x) = f(dist(a, x))$. Second, we must determine the "best" response register $|b\rangle$ to send through the oracle (tensored to the equal superposition state). For all of algorithms with 1 dimensional response registers considered so far, $|b\rangle = |-\rangle$ provides the best choice of response register. In both cases, the word "best" refers to a configuration that will identify a target concept in a single query with the highest possibility possible.

For a 1-dimensional response register, an oracle acted on a computational basis element $|x, b\rangle \in \mathbb{C}^{2^n} \otimes \mathbb{C}^2$ by

$$\mathcal{O}_a |x, b\rangle = |x, b \oplus f_{c_a}(x)\rangle$$

We wish to generalize the action of the oracle on the response register beyond the $\oplus$ operation. Y-valued Hamming distance oracles hide functions $f_{c_a} : \{0, 1, \ldots n\} \to \mathbb{Z}_r$, where $f_{c_a}(x)$ is a function of $dist(a, x)$. All possible actions on the range $\mathbb{Z}_r$ are described by the symmetry group $S_r$. To that end, we associate each possible $f_{c_a}$ that maps the computational basis of $\mathbb{C}^r$ into itself with a map $\sigma : \mathbb{Z}_n \to S_r$. For example, just adding the Hamming distance to the response register corresponds to the map $d \to (12)^d$. The 2-bit Hamming distance oracle corresponds to the map $d \to (12)^{b(d)}$.

**Remark** This generalization still does not describe every possible oracle action on a $Y$ dimensional response register – only those which map the computational basis of the response register into itself. The most general quantum oracle would allow any unitary transformation of the response register. However, the permutation model retains analogy with a classical oracle, which, lacking superposition, may only map $\mathbb{Z}_r \to \mathbb{Z}_r$.

We redescribe the action of the oracle formally: suppose $\mathcal{C}$ is an $(n, r)$ Hamming distance concept class. For each possible Hamming distance $d \in \mathcal{Z}_r$ we associate a permutation $\sigma_d \in S_r$. Then the oracle associated with $c_a$ acts on the computational basis of $\mathcal{C}^{2^n} \otimes \mathbb{C}^r$ by

$$\mathcal{O}_a |x, b\rangle \to |x\rangle \otimes |\sigma_{dist(a,x)}(b)\rangle$$

We generalize the notion of the membership query matrix to a *quantum learning matrix*, whose $a^{th}$ column is the state produced by querying $O_a$ with the input state $|\eta_n\rangle \otimes |phi\rangle$.

**Definition.** Suppose $\mathcal{C}$ is an $(n, r)$ Hamming distance concept class. Suppose $|\phi\rangle \in \mathbb{C}^r$ is a unit vector chosen to be the initial state of the response register. Then the **quantum learning matrix** $A$ is an $2^n r \times 2^n$ matrix whose $a^{th}$ column is

$$\mathcal{O}_a(|\eta_n\rangle \otimes |\phi\rangle).$$

We would like to know the probability of successfully learning a target concept with a single guess from a Hamming distance oracle. Measuring this probability requires using some measurement; we know that in many cases the square-root of the Gram matrix $G = A^\dagger A$ of provides the optimal measurement. Calculating the square-root matrix may be difficult if $n$ is large. However, we are equipped with a useful theorem that shows that the Gram matrix of a quantum learning matrix of any Hamming concept class is diagonalized by a Hadamard matrix.

More generally, if a matrix exhibits the symmetry that the $(a, b)$ matrix element is equal to the $(a + b, 0)$ matrix element, then it is diagonalized by the Hadamard matrix. Note that the matrix is indexed by elements of $\mathbb{Z}_2^n$.

**Theorem 4.2.** *Suppose $G$ is a $2^n \times 2^n$ complex matrix such that $G_{a,b} = G_{a+b,0)}$. Then $H^{\otimes n}$ diagonalizes $G$.*

The proof makes use of the following lemma:

**Lemma 1.** *Suppose $j \in \mathbb{Z}_2^n$. Then*

$$\frac{1}{2^n} \sum_{x \in \mathbb{Z}^2} (-1)^{x \cdot j} = \begin{cases} 1 \text{ if } j = 0 \dots 0 \\ 0 \text{ otherwise} \end{cases}$$

*Proof.* First suppose $j = 0 \dots 0$. Then $x \cdot j = 0$ for all $x \in \mathbb{Z}_2^n$, so each summand is 1, proving the first case. Therefore, suppose $wt(j) = k > 0$. Then $x \cdot j \equiv 1 \mod 2$ iff $x$ and $j$ share an odd number of ones. Hence, by counting the number of strings which share an even and odd number of 1's with $j$, we write the sum:

$$\frac{1}{2^n} 2^{n-k} \sum_{i=0}^{k} (-1)^i \binom{k}{i} = 0$$

This is a well-known combinatorial identity. □

*Proof.* We calculate the matrix $H^{\otimes n} G H^{\otimes n}$.

$$(H^{\otimes n} G H^{\otimes n})_{i,j} = \sum_{x,y \in \mathbb{Z}_2^n} H_{i,x}^{\otimes n} G_{x,y} H_{y,j}^{\otimes n} = \frac{1}{2^n} \sum_{x,y \in \mathbb{Z}_2^n} (-1)^{i \cdot x} (-1)^{y \cdot j} G_{x,y}$$

For a fixed $x \in \mathbb{Z}_2^n$, the map $y \to y + x$ is a bijection. Furthermore, $(-1)^{(y+x) \cdot j} = (-1)^{x \cdot j} (-1)^{y \cdot j}$. Finally, due to the symmetry assumption on $G$, $G_{x,y+x} = G_{x+x,y} = G_{0,y}$ Therefore we rewrite the sum:

$$\frac{1}{2^n} \left( \sum_{y \in \mathbb{Z}_2^n} (-1)^{y \cdot j} G_{0,y} \right) \sum_{x \in \mathbb{Z}_2^n} (-1)^{x \cdot (i+j)}$$

By the lemma, this is 0 unless $i + j = 0 \dots 0$, that is, $i = j$. Therefore the theorem is proved. □

We shall see that the quantum learning matrix for any Hamming distance concept class admits the symmetry described above, and is therefore diagonalized by a Hadamard matrix.

**Theorem 4.3.** *The Gram matrix of the quantum learning matrix $G = A^\dagger A$ for any Hamming concept class satisfies $G_{i,j} = G_{i+j,0}$.*

*Proof.* An oracle $\mathcal{O}_{c_a}$ acts on an arbitrary $|x\rangle \otimes |\phi\rangle \in \mathbb{C}^{2^n} \otimes \mathbb{C}^r$ where $|x\rangle$ is a computational basis element by

$$\mathcal{O}_{c_a}(|\Psi\rangle \otimes |\phi\rangle) = |\Psi\rangle \otimes U_{dist(a,x)} |\phi\rangle$$

For some unitary operator $U_{dist(a,x)}$. Therefore we may calculate the Gram matrix:

$$\begin{aligned} G_{i,j} &= (\mathcal{O}_{c_i}(|\eta_n\rangle \otimes |b\rangle))^\dagger \left(\mathcal{O}_{c_j}(|\eta_n\rangle \otimes |b\rangle)\right) \\ &= \left(\sum_{x\in\mathbb{Z}_2^n} |x\rangle \otimes U_{dist(i,x)} |b\rangle\right)^\dagger \left(\sum_{y\in\mathbb{Z}_2^n} |y\rangle \otimes U_{dist(y,j)} |b\rangle\right) \\ &= \sum_{x\in\mathbb{Z}_2^n} \left(U_{dist(i,x)} |b\rangle\right)^\dagger \left(U_{dist(j,x)} |b\rangle\right) \end{aligned}$$

For a fixed $c \in \mathbb{Z}_2^n$, the map $x \to x + sc$ is a bijection. Hence we rewrite the sum:

$$\begin{aligned} \sum_{x\in\mathbb{Z}_2^n} \left(U_{dist(i,x+c)} |b\rangle\right)^\dagger \left(U_{dist(j,x+c} |b\rangle\right) &= \sum_{x\in\mathbb{Z}_2^n} \left(U_{dist(i+c,x)} |b\rangle\right)^\dagger \left(U_{dist(j+c,x)} |b\rangle\right) \\ &= G_{i+c,j+c} \end{aligned}$$

By the same calculation. Finally, using $c = j$,

$$G_{i,j} = G_{i+sj,0}$$

as desired. □

### 4.5.1 Numerical Results

The following proposition is a conglomeration of previous results discussed in quantum state discrimination and quantum learning, and allows for a simple implementation of numerical methods for determining the best permutations and response register for $(n, r)$ Hamming distance concept classes.

**Proposition 1.** *Let $\mathcal{C}$ be an $(n, r)$ Hamming distance concept class. To each $c_a \in \mathcal{C}$ fix a permutation $\sigma_a$ of $\mathbb{Z}_r$. Also fix the response register $|b\rangle \in \mathbb{C}^r$. Let $B$ be the $nr \times n$ matrix whose $a^{th}$ column is the state $\mathcal{O}_{c_a} |\eta_n\rangle \otimes |b\rangle$. If the columns of $B$ are linearly independent, then the square root of the Gram matrix $G = B^\dagger B$ describes the optimal measurement, with success probability $|Tr(\sqrt{G})|^2/2^n$.*

*Proof.* This is a reformulation of the statement that the square-root Gram matrix provides the optimal measurement when the states to be distinguished are linearly independent and $\sqrt{G}$ has constant diagonal. Hence is suffices to show that $\sqrt{G} = \sqrt{B^\dagger B}$ has constant diagonal. We know from Theorem 4.2 that $H^{\otimes n}$ diagonalizes $G$, so $D = H^{\otimes n} G H^{\otimes n}$. Therefore $\sqrt{G} = H^{\otimes n}\sqrt{D}H^{\otimes n}$. We calculate the $i^{th}$ diagonal entry of $\sqrt{G}$:

$$\begin{aligned}\sqrt{G}_{i,i} &= \frac{1}{2^n}\sum_{x\in\mathbb{Z}_2^n}\sum_{y\in\mathbb{Z}_2^n}(-1)^{i\cdot x}\sqrt{D}_{x,y}(-1)^{y\cdot i} \\ &= \frac{1}{2^n}\sum_{x\in\mathbb{Z}_2^n}\sqrt{D}_{x,x} \qquad (4.2)\end{aligned}$$

Therefore each $\sqrt{G}_{i,i}$ is the same. □

This justifies the numerical method used by [MePo 2009] to study the $(3,2),(5,2)$ and $(5,3)$ Hamming distance concept classes with the permutation model to find both the optimal permutations and optimal response register:

**Numerical Method**:

1. Fix $n$ and $r$.
2. Repeat steps 3 and 4 for all possible permutations $\sigma_a : \mathbb{Z}_n \to S_r$:
3. Choose a initial response register unit vector $|b\rangle \in \mathbb{C}^r$. By the above proposition, we may calculate the success probability of inputting the state $|\eta_n\rangle \otimes |b\rangle$ to an oracle $\mathbb{O}_{c_a}$.
4. Maximize the success probability over all $|b\rangle \in \mathbb{C}^r$.

**Remark.** The numerical method, and technique used in the following sections, measure the states of the quantum learning matrix with the square-root Gram matrix. However, this is the optimal measurement only if those states are linearly independent. Recall that if the states span a $k < 2^n$ dimensional subspace of $\mathcal{H}_s$, the maximum success probability is $\frac{k}{2^n}$. For general $n$ and $r$, it is conjectured that optimal permutation assignments will result in linearly independent columns of $A$, but has not been proven.

## 4.6 Analysis of the $(n,2)$ Permutation Model for odd $n$

The ultimate goal is to determine analytically the optimal permutations and best response register, without the numerical approximation. For now, we consider only $(n,2)$ Hamming distance concept classes (for odd $n$). This requires an explicit calculation of the success probability $P = |Tr(\sqrt{G})|^2/2^n$.

**Theorem 4.4.** *Suppose $\mathcal{C}$ is an $(n, 2)$ Hamming distance concept class where $n$ is odd. Let $\{\sigma_d : \sigma_d = (1) \text{ or } (12)\}$ be a set of permutations indexed by the Hamming distance $d \in \mathbb{Z}_r$ so that the oracle $\mathcal{O}_{c_a}$ associated with a concept $c_a \in \mathcal{C}$ acts on a computational basis element $|x, b\rangle \in \mathbb{C}^{2^n} \otimes \mathbb{C}^2$ by*

$$\mathcal{O}_{c_a} |x, b\rangle = |x, \sigma_{dist(a,x)}(b)\rangle$$

*Fix a response register $|\phi\rangle \in \mathbb{C}^2$. Then the success probability can be expressed as a function of a real variable*

$$P(\gamma) = \frac{1}{2^n\sqrt{2^n}} \left(\sqrt{a + b\gamma} + c\sqrt{(1 - \gamma)}\right)^2$$

*where $\gamma = \langle X\phi|\phi\rangle \in [-1, 1]$ and $a, b \in \mathbb{Z}^+$ such that $a + b = 2^{2n}$ and $c \in \mathbb{R}^+$.*

Here $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ denotes the "bit-flip" unitary transformation.

*Proof.* Let $A$ be the quantum learning matrix for $\mathcal{C}$, i.e. the matrix whose $i^{th}$ column is

$$\mathcal{O}_{c_i}(|\eta_n\rangle \otimes |\phi\rangle) = \sum_{x \in \mathbb{Z}_2^n} |x\rangle \otimes X^{\sigma_{dist(i,x)}} |\phi\rangle$$

where

$$X^{\sigma_{dist(i,x)}} = \begin{cases} X \text{ if } \sigma_{dist(i,x)} = (12) \\ I \text{ if } \sigma_{dist(i,x)} = (1) \end{cases}$$

Let $G = A^\dagger A$ be the Gram matrix of $A$. From Eq. 4.2 we see that

$$P = \frac{1}{2^n} \left( \sum_{x \in \mathbb{Z}_2^n} \sqrt{D}_{x,x} \right)^2 \tag{4.3}$$

So we must calculate the values $\sqrt{D}_{x,x}$, which is the square root of the diagonal matrix of $G$. We know that the Hadamard matrix diagonalizes $G$, so

$$\begin{aligned} D_{i,i} &= \sum_{x \in \mathbb{Z}_2^n} \sum_{y \in \mathbb{Z}_2^n} H_{i,x}^{\otimes n} G_{x,y} H_{y,i}^{\otimes n} \\ &= \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} \sum_{y \in \mathbb{Z}_2^n} (-1)^{i \cdot x} G_{x+y,0} (-1)^{y \cdot i} = \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} \sum_{y \in \mathbb{Z}_2^n} (-1)^{(x+y) \cdot i} G_{x+y,0} \\ &= \sum_{j \in \mathbb{Z}_2^n} (-1)^{i \cdot j} G_{j,0} \end{aligned} \tag{4.4}$$

We see that calculating the success probability boils down to calculating the first

column of the Gram matrix of $A$. Consider one such entry, $G_{j,0}$.

$$\begin{aligned} G_{j,0} &= \left(\mathcal{O}_{c_j}(|\eta_n\rangle \otimes |\phi\rangle)\right)^\dagger \left(\mathcal{O}_{c_0}(|\eta_n\rangle \otimes |\phi\rangle)\right) \\ &= \frac{1}{2^n}\left(\sum_{x\in\mathbb{Z}_2^n} |x\rangle \otimes X^{dist(x,j)}\,|\phi\rangle\right)^\dagger \left(\sum_{y\in\mathbb{Z}_2^n} |y\rangle \otimes X^{dist(y,0)}\,|\phi\rangle\right) \\ &= \frac{1}{2^n}\sum_{x\in\mathbb{Z}_2^n} \langle\phi|(X^{dist(j,x)})^\dagger X^{dist(x,0)}|\phi\rangle \end{aligned}$$

Since $X^\dagger X = XX = I$, the summand is equal to $\langle\phi|\phi\rangle = 1$ if $X^{dist(j,x)} = X^{dist(x,0}$ and is equal to $\gamma = \langle X\phi|\phi\rangle$ if $X^{dist(j,x)} \neq X^{dist(x,0)}$. Note that if $\phi = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$, then $\gamma = \langle X\phi|\phi\rangle = \alpha^*\beta + \beta^*\alpha \in \mathbb{R}$. Also, the Cauchy Schwarz inequality tells us

$$\| \langle\phi|X^{dist(j,x)}\rangle \|^2 \leq \langle\phi|\phi\rangle \langle X\phi|X\phi\rangle = 1$$

Therefore $\gamma \in [-1, 1]$, and $\gamma$ achieves the endpoints of the interval at $|\phi\rangle = |-\rangle$ and $|\phi\rangle = |+\rangle$.

We return to the previous sum. If $X^{dist(j,x)} = X^{dist(x,0)}$ for all $x \in \mathbb{Z}_2^n$ then the sum is equal to $2^n$. We therefore define the integer $f_j$ to be the number of times in the summand that $X^{dist(j,x)} \neq X^{dist(x,0)}$. Therefore the sum may be written as:

$$G_{j,0} = \frac{1}{2^n}(2^n + f_j(\gamma - 1))$$

for some $f_j \in \mathbb{Z}$. It's time to return to the elements of $D$. From Eq. 4.4,

$$D_{i,i} = \sum_{j\in\mathbb{Z}_2^n} (-1)^{i\cdot j} G_{j,0}$$

First, note that

$$\begin{aligned} D_{0,0} = \sum_{j\in\mathbb{Z}_2^n} G_{j,0} &= \sum_{j\in\mathbb{Z}_2^n} \frac{1}{2^n}(2^n + f_j(\gamma-1)) = \frac{1}{2^n}(2^{2n} - (1-\gamma)\sum_{j\in\mathbb{Z}_2^n} f_j) \\ &= \frac{1}{2^n}(a + b\gamma) \end{aligned}$$

For $a = 2^{2n} - \sum_{j\in\mathbb{Z}_2^n} f_j$ and $b = \sum_{j\in\mathbb{Z}_2^n} f_j$. Clearly $a, b \in \mathbb{Z}$ and $a, b \geq 0$ since $f_j \leq 2^n$. Furthermore, $a + b = 2^{2n}$.

Let us now consider diagonal elements $D_{i,i}$ such that $wt(i) \geq 0$. Then Eq. 4.4 shows that

$$\begin{aligned} D_{i,i} &= \frac{1}{2^n}\sum_{j\in\mathbb{Z}_2^n} (-1)^{i\cdot j}(2^n + f_j(\gamma-1)) \\ &= \frac{1}{2^n}\left(\sum_{j\in\mathbb{Z}_2^n} (-1)^{i\cdot j}2^n + (\gamma-1)\sum_{j\in\mathbb{Z}_2^n} (-1)^{i\cdot j} f_j\right) \\ &= \frac{1}{2^n}(\gamma-1)\sum_{j\in\mathbb{Z}_2^n} (-1)^{i\cdot j} f_j \end{aligned}$$

Remember that $D_{i,i}$ is an eigenvalue of the positive operator $G = A^\dagger A$ and is therefore positive; $(\gamma - 1)$ is always negative, so we may write

$$D_{i,i} = \frac{1}{2^n}(c_i(1-\gamma))$$

for some $c_i \in \mathbb{Z}^+$ when $wt(i) \neq 0$. Let us substitute all the results into Eq. 4.3:

$$\begin{aligned} P(\gamma) = \frac{1}{2^n}\left(\sum_{i\in\mathbb{Z}_2^n}\sqrt{D}_{i,i}\right)^2 &= \frac{1}{2^n}\left(\sqrt{D}_{0,0} + \sum_{\substack{i\in\mathbb{Z}_2^n \\ i\neq 0}}\sqrt{D}_{i,i}\right)^2 \\ &= \frac{1}{2^n\sqrt{2^n}}\left(\sqrt{a+b\gamma} + \sum_{\substack{i\in\mathbb{Z}_2^n \\ i\neq 0}}\sqrt{c_i(1-\gamma)}\right)^2 \\ &= \frac{1}{2^n\sqrt{2^n}}\left(\sqrt{a+b\gamma} + c\sqrt{1-\gamma}\right)^2 \end{aligned}$$

where $c = \sum_{i=1}^{2^n}\sqrt{c_i} \in \mathbb{R}^+$. Hence we have proved the theorem. □

We know wish to maximize $P(\gamma)$. Luckily, there are only two radicals in $\sqrt{P(\gamma)}$, so we may easily calculate the derivative.

**Proposition.** *Suppose $\sqrt{P(\gamma)}$ is function of a real-variable of the form*

$$\sqrt{P(\gamma)} = \frac{1}{2^n\sqrt{2^n}}\left(\sqrt{a+b\gamma} + c\sqrt{1-\gamma}\right)$$

*for $a, b, c \geq 0$, $n$ an odd positive integer, and $a + b = 2^n$. Then $\sqrt{P(\gamma)}$ has a single critical point, at*

$$\gamma' = \frac{b^2 - c^2 a}{bc^2 + b^2}$$

*and*

$$\sqrt{P(\gamma')} = \sqrt{\frac{b+c^2}{2^n b}}$$

*Proof.* Differentiating $\sqrt{P(\gamma)}$ and setting it equal to 0 we see

$$b^2(1-\gamma) = c^2(a+b\gamma)$$

So $\sqrt{P(g)}$ has one critical point. We solve for $\gamma$:

$$\gamma(bc^2 + b^2) = b^2 - c^2 a \iff \gamma' \equiv \frac{b^2 - c^2 a}{bc^2 + b^2}$$

Then

$$\sqrt{2^{3n}}\sqrt{P(\gamma')} = \sqrt{\frac{ab+b^2}{c^2+b}} + c\sqrt{\frac{bc^2+c^2a}{bc^2+b^2}}$$
$$= \frac{2^n}{\sqrt{b+c^2}}(\sqrt{b} + \frac{c^2}{\sqrt{b}}) = 2^n\frac{\sqrt{b+c^2}}{\sqrt{b}}$$

□

Since this is the only critical point of $P(\gamma)$ for $\gamma \in [-1,1]$; therefore $P(\gamma)$ is maximized at $\gamma = \gamma'$, 1, or -1.

**Corollary.** *At the critical point $\gamma'$,*

$$P(\gamma') = \frac{b+c^2}{2^n b}$$

*The maximum value of $P(\gamma)$ over the interval $[-1,1]$ is:*

$$\max\{\frac{b+c^2}{2^n b}, P(1), P(-1)\} = \max\{\frac{b+c^2}{2^n b}, \sqrt{a-b} + c\sqrt{2}\}$$

*unless $\gamma' \notin [-1,1]$ when the success probability is just $\sqrt{a-b} + c\sqrt{2}$*

### 4.6.1 Summary of Results

The purpose of the last section is to provide a first step towards an analytic description of the best possible success possibilities for a the $(n,2)$ Hamming distance concept classes. This involves knowing the best permutation assignment for a given $n$, which requires a grasp of the integers $f_j$ used in the proof of Theorem 4.4. Using the combinatorial method of determining $P(\gamma)$, (see Appendix A for code), we have extended the numerical results of [MP09] to find the best success probability for the $(7,2)$ Hamming distance concept class. We also present the redone results for the $(3,2)$ and $(5,2)$ Hamming distance concept classes.

**(3, 2) Hamming distance concept class: Numerical Results**. The best success probability $P \approx .8000$ is achieved with the permutation assignments $\sigma_0 = \sigma_1 = \sigma_3 = (1)$ while $\sigma_2 = (12)$, and $\sigma_0 = \sigma_2 = \sigma_3 = (1)$ while $\sigma_1 = (12)$. In both cases $\gamma \approx -.8000$. Using the method of calculating the success probability described in the previous section, we may prove (not included):

**Proposition.** *For the permutation assignments describe above, the best success probability $P = \frac{4}{5}$ at $\gamma' = -\frac{4}{5}$.*

**(5, 2) Hamming distance concept class: Numerical Results**. The best success probability $P \approx .7206$ is achieved with the permutation assignments

1.

$$\sigma_0 = \sigma_3 = \sigma_4 = \sigma_5 = (1)$$
$$\sigma_1 = \sigma_2$$

2.

$$\sigma_0 = \sigma_1 = \sigma_2 = \sigma_5$$
$$\sigma_3 = \sigma 4$$

In both cases $\gamma \approx -.9208$.

**(7, 2) Hamming distance concept class: Numerical Results**. The best success probability $P \approx .7052$ is achieved with the permutation assignments:

1.

$$\sigma_0 = \sigma_4 = \sigma_5 = \sigma_7 = (1)$$
$$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_6 = (12)$$

2.

$$\sigma_0 = \sigma_2 = \sigma_3 = \sigma_7 = (1)$$
$$\sigma_1 = \sigma_4 = \sigma_5 = \sigma_6 = (12)$$

In both cases $\gamma \approx -.9954$.

Examination of the data leads us to the following conjectures:

**Conjecture 1** As $n \to \infty$, $\gamma \to -1$.

We see that for $n = 7$ the optimal response register is already very close to $\gamma = -1$. Remember $\gamma = -1$ corresponds to the usual response regsiter $|-\rangle$.

**Conjecture 2** Using the notation of Theorem 4.4, we conjecture that the optimal permutation assignment has the highest possible value of $c$.

This claim is justified somewhat because it appears that a permutation can be chosen such that $c \gg b$; therefore the best probability would have a high value of $c$.

**Conjecture 3** For a $(n, 2)$ Hamming distance concept class, there is always some permutation assignment with success probability $P \geq \frac{1}{2}$.

For even $n$, this is clearly the case ($P = 1$). For odd $n$, the data indicates that there are permutation assignments that are similar to to the "second least significant bit" permutation used in the algorithm for $n$ even, suggesting that it is possible to simulate that algorithm on some even subset of bits, and produce an algorithm that succeeds with $P = \frac{1}{2}$.

# Conclusion

There are still many unanswered questions concerning the Hamming distance concept classes. The ultimate goal is to analytically determine the best success probability and therefore best permutation assignment for the $(n, 2)$ Hamming distance concept class. Theorem 4.4 suggests that this may be possible with a sufficient understanding of the combinatorics of the permutation assignments. Immediately, Conjectures 1-3 may be examined. The rank of the matrix $A$ should also be studied; we remarked that the best success probability intuitively requires the columns of $A$ to be linearly independent, but this is unproven.

# Appendix A

# Code

The code used to generate the numerical results is included here. It is written in Python and requires the use of the numpy and sympy packages.

```
# Single-Query Hamming Distance Learning

# This program is designed to provide insight into a problem of
# quantum concept learning: determining a hidden string "a"
# given an oracle which permutes the response register in some way
# based on the Hamming distance between the input and the string "a".
#
#
# Running the command HammingProb(n) will generate a file n=-.dat
# that contains the following data for every permutation assignment:
# - optimal response regsiter
# - best success probability
# - value of g (= \gamma in the thesis) at the critical point
# - eigenvalues of the Gram matrix, sorted by Hamming weight

from sympy.matrices import *
import sympy as sp
import numpy as np

# Creates the n-bit Hadamard gate
def Hadamard(n):
    h = Matrix([[1, 1], [1, -1]])
    H = h
    for i in range(n-1):
        H = np.kron(h, H)
    return H

# Calculates the hamming-distance of 2 vectors
def hamming(u, v):
    x = []
```

```
    y = []
    for c in u:
        x.append(int(c))
    for e in v:
        y.append(int(e))
    if len(x) == len(y):
        d = 0
        for i in range(len(u)):
            if x[i] != y[i]:
                d = d+1
        return d
    else:
        print x
        print u
        print y
        return None

# determines the Hamming weight of an integer
def wt(u):
    wt = 0
    for digit in np.binary_repr(u):
        if digit == '1':
            wt = wt + 1
    return wt

# Creates the n-bit "Hamming-distance matrix" = HD
# where HD(i,j) = dist(binary rep'n of i, binary rep'n of j)
def HDistMat(n):
    return Matrix(n, n, lambda i,j: hamming(np.binary_repr(i, width=n),
     np.binary_repr(j,width=n)))

# The following function takes a permutation of the response register
# (which is a subset of {0, 1, ... n} for a two-bit register; 1 denotes
# a flip of the response register at that bit. Therefore, the function places
# a one whenver it finds a number contained in the permutation (subset)
# and a 0 if it is not contained there.

def PermuteAffect(sigma, M):
    def f(x):
        if x in sigma: return 1
        else: return 0
    return M.applyfunc(f)

# We are now able to calculate the (x,y)th entry of the Gram matrix G
# given a permutation sigma. G(x,y) = G(x+y,0) which is <Psi(x+y), Psi(0)>
```

```
# = 2*g*dist(PermuteAffect(x+y), PermuteAffect(0)) (the x+yth and 0th rows)
g = sp.Symbol('g') #This is a placeholder for the variable gamma
def GTerm(M, a):
    a = a % sp.sqrt(len(list(M.vec())))
    u = list(M[0:1, 0:])
    v = list(M[a:a+1, 0:])
    d = hamming(u,v)
    return (d*g + len(u) - d)

# Finally we can calculate a diagonal entry of the square-root
# Gram matrix, which depends only on GTerm(M, x+y)

def dotsum(i, x, y):
    return wt(np.bitwise_and(i, x)) + wt(np.bitwise_and(i, y))

def bitsum(x, y, n):
    u = np.binary_repr(x, width=n)
    v = np.binary_repr(y, width=n)
    diff = []
    s = 0
    for i, dum in enumerate(u):
        diff.append(np.bitwise_xor(int(u[n-i-1]), int(v[n-i-1])))
    for i, dum in enumerate(diff):
        s = (2**(i))*(diff[i]) + s
    return s

def SuccessProbFunc(sigma, n):
    N = 2**n
    M = HDistMat(N)
    M = PermuteAffect(sigma, M)
    d = []
    D = 0
    s0 = 0
    s2 = 0
    w = 1
    e = []
    GTerms = []
    # First calculate the first eigenvalue D0
    for j in range(n+1):
        GTerms.append(GTerm(M, 2**j - 1))
    for x in range(N):
        s0 = s0 + GTerms[wt(x)]
    d.append(s0)
    D = sp.sqrt(s0)
    # Calculate the rest of the eigenvalues, keeping track of the
```

```
    # the coefficients of g so that the final values a, b, c are recorded
    # properly
    for i in range(N-1):
        s1 = 0
        for x in range(N):
            k = ((-1)**(dotsum(i+1, x, 0)))*GTerms[wt(x)]
            s1 = s1 + k
        s2 = s2 + sp.sqrt(sp.simplify(s1/(1-g)))
        d.append(s1)
    # This is to make the output more readable
    D = D + s2*sp.sqrt(1-g)
    h = sp.simplify((D - sp.sqrt(d[0]))/(sp.sqrt(1-g)))
    D = sp.sqrt(d[0]) + sp.sqrt((h**2)*(1-g))
    p1 = (float(D.subs(g, -1.0)))**2/(N*N*N)

    # Calculating the critical probabilities and probs at g = -1
    a = float(d[0].subs(g, 0.0))
    b = float(N*N - a)
    c = float(h)*float(h)
    crit = (b*b - c*a)/(b*c + b*b)
    pbest = (1.0 +  (float(h)*float(h))/b)/float(N)
    for i in range(n+1):
        e.append(d[2**i - 1])
    return D, e, p1, pbest, crit

def HammingProb(n):
    filename = 'n=' + str(n) + '.dat'
    P = genPerms(n)
    FILE = open(filename, 'w')
    for sigma in P:
        if sigma == []: continue
        else:
            f, d, p1, pbest, crit  = SuccessProbFunc(sigma, n)
            FILE.write('Permutation =' + str(sigma)+ '\n')
            FILE.write('(Sqrt Of) Success Prob Func:\n')
            FILE.write(str(f)+ '\n')
            FILE.write('Prob of success at g = -1:      ' + str(p1) + '\n')
            FILE.write('Prob of success at g = gcrit:     ' + str(pbest) + '\n')
            FILE.write('Critical point g =        ' + str(crit) + '\n')
            FILE.write('Eigenvalues of Gram Matrix (sorted by wt(i)):\n')
            for eig in d:
                FILE.write(str(eig)+'\n')
            FILE.write('\n')
    FILE.close()
```

# References


[Ang88] Dana Angluin. Queries and concept learning. *Machine Learning 2*, pages 319 – 342, 1988.

[BV93] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. In *in Proc. 25th Annual ACM Symposium on Theory of Computing, ACM*, pages 11–20, 1993.

[CDS08] G. Chirbella, G. M. D'Ariano, and D.-M. Schilngemann. Realization of continuous-outcome measurements on finite dimensional quantum systems, http://arxiv.org/abs/quant-ph/0703110v. 2008.

[EJ01] Yonina C. Eldar and G. David Forney Jr. On quantum detection and the square root measurement. *IEEE Transactions on Information Theory*, 47(3):858 – 872, 2001.

[EMV08] Yonina C. Eldar, Alexandre Megretski, and George C. Verghese. Designing optimal quantum detectors via semidefinite programming, http://arxiv.org/abs/quant-ph/0205178v1. 2008.

[FIS03] Stephen H. Friedberg, Arnold J. Insel, and Lawrence E. Spence. *Linear Algebra, 4th ed.* Pearson Education Inc., Upper Saddle River, New Jersey, 2003.

[Gri04] David Griffiths. *Introduction to Quantum Mechanics, 2nd ed.* Prentice Hall, USA, 2004.

[Hel76] Carl W. Helstrom. *Quantum Detection and Estimation Theory.* Academic Press, Inc., New York, NY, 1976.

[HM02] Markus Hunziker and David A. Meyer. Quantum algorithms for highly structured search problems. In *Quantum Inform. Processing*, pages 145–154. Meyer, 2002.

[HMP+03] Markus Hunziker, David A. Meyer, Jihun Park, James Pommersheim, and Mitch Rothstein. The geometry of quantum learning, http://arxiv.org/abs/quant-ph/0309059v1. 2003.

[Ken74] Robert S. Kennedy. Uniqueness of the optimum receiver for the m-ary pure state problem. *MIT, Quart. Prog. Rep*, (113):129 – 130, 1974.

[MP09] David A. Meyer and James Pommersheim. Single-query learning from abelian and non-abelian hamming distance oracles, http://arxiv.org/abs/0912.0583v1. 2009.

[NC00] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, UK, 2000.

[SKIH97] Masahide Sasaki, Kentaro Kato, Masayuki Izutsu, and Osamu Hirota. Quantum channels showing superadditivity in capacity, http://arxiv.org/abs/quant-ph/98010asdf. 1997.

[YKL75] Horace P. Yuen, Robert S. Kennedy, and Melvin Lax. Optimum testing of multiple hypotheses in quantum detection theory. *IEEE Transactions on Information Theory*, 21(2):125–134, March 1975.