MULTI-AGENT LEARNING
IN THE
QUANTUM PRISONER’S DILEMMA

A Msc thesis
SUBMITTED TO THE DEPARTMENT OF
ARTIFICIAL INTELLIGENCE
AND THE COMMITTEE ON GRADUATE STUDIES
OF THE UNIVERSITY OF GRONINGEN
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Maarten Jacobs
August 2009
I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Master of Science.

(prof. dr. Rineke Verbrugge)
Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Master of Science.

(prof. dr. Harry Buhrman)

Approved for the University Committee on Graduate Studies.
Abstract

There has been a recent interest in Quantum Computing as it may offer solutions to ‘classically’ hard problems. When considering Game Theory from the perspective of algorithms, Quantum Game Theory might offer a new way to further understand the intricacies of Quantum Information processing.

The game known as the Prisoner’s Dilemma is often used to illustrate the concepts of traditional Game Theory. It also shows its shortcomings though; Game Theory predicts rational players will play ‘defect’, while in practice human players often ‘cooperate’ in the dilemma. The use of Quantum mathematics tries to solve this issue as its mechanisms intrinsically link the decisions of both players.

Multi-agent learning has been used to mimic, predict and improve upon human strategies in the classical Prisoner’s Dilemma. It has, however, not yet been introduced to the realm of Quantum mathematics. Particularly, the Conditional Joint Action Learner is suitable as it is based on joint probabilities of the actions of the players and Quantum physics is intrinsically based on probability.

The research presented in this thesis focuses on a thorough analysis of the Quantum Prisoner’s Dilemma and the construction of a Quantum Conditional Joint Action Learner playing this game. To this end, an introduction to the field of Quantum Computation and Quantum Information is presented to the reader followed by an overview of the main ideas behind Quantum Game Theory. Finally, the Quantum Conditional Joint Action Learner is proposed for playing an iteration of the Quantum Prisoner’s Dilemma. Experiments with this learner show Pareto Optimal behavior.
Contents

Abstract v

1 Introduction 1

2 Theory 5
  2.1 Quantum Computing ........................................... 5
     2.1.1 Preliminary Linear Algebra ............................... 5
     2.1.2 Quantum Computer Science ............................... 9
     2.1.3 Quantum non-locality and Bell inequalities .............. 17
     2.1.4 Bloch Sphere Representation .............................. 22
     2.1.5 Quantum Algorithms .................................... 27
  2.2 Game Theory .................................................... 31
     2.2.1 Game Theory .............................................. 31
     2.2.2 Prisoner’s Dilemma ..................................... 33
  2.3 Multi-Agent Learning .......................................... 35
     2.3.1 Reinforcement Learning ................................ 35
     2.3.2 Conditional Joint Action Learning ...................... 37
  2.4 Quantum Game Theory .......................................... 41
     2.4.1 Quantum Penny-flip Game ................................ 41
     2.4.2 Non-local games .......................................... 44
  2.5 Quantum Prisoner’s Dilemma .................................. 48
     2.5.1 Unitary Actions .......................................... 51

3 Experimental Setup and Results 58
  3.1 Experimental Methods ......................................... 58
  3.2 A Simple Learner .............................................. 59
  3.3 Quantum Conditional Joint Action Learner .................... 63
     3.3.1 Pure Exploitation ....................................... 65
     3.3.2 Mixed Exploitation-Exploration ......................... 66
     3.3.3 Noisy Actions ........................................... 70
  3.4 Chapter Summary .............................................. 75
List of Tables

2.1 Example Prisoner’s Dilemma ........................................... 33
2.2 Generalized Prisoner’s Dilemma ........................................ 34
2.3 Payoff matrix for Alice in a penny-flip game .......................... 42
2.4 Winning conditions for the CHSH game ............................... 45
2.5 The interaction between pure strategies ............................... 56

3.1 Average number of iterations and its variance before a solution is found with the percentage of runs that terminates with a solution. Statistical data accumulated by executing both experimental settings $N = 5,000$ times. ................................................................. 61
List of Figures

2.1 Spherical Polar Coordinates ................................................. 24
2.2 Bloch sphere representation of a qubit ................................. 25
2.3 The Reinforcement Learning Framework, from [SB98] ............... 36
2.4 The expected payoff for Alice, using the assignment in (2.143), in a separate game. Here, Defection corresponds with $k = 1$, Cooperation with $k = 0$, $Q_1$ with $k = -1$ and $Q_2$ with $k = -2$. .................. 57

3.1 A simple quantum learner playing against an invariable opponent. The probability $\Pr(\text{CC})$, which is the probability that the 2-qubit end state evaluates to $|00\rangle \equiv \text{CC}$, is shown as a solid line versus the left-hand $y$-axis and the actual outcomes are shown as dots labeled by the right-hand $y$-axis. .................. 62
3.2 A simple quantum learner in self-play. The probability $\Pr(\text{CC})$, which is the probability that the 2-qubit end state evaluates to $|00\rangle \equiv \text{CC}$, is shown as a solid line versus the left-hand $y$-axis and the actual outcomes are shown as dots labeled by the right-hand $y$-axis. .................. 63

3.3 Variations in Expected Utilities of the CJAL-algorithm in the classical Prisoner’s Dilemma reproduced from Banerjee et al. [BS07]. .................. 64
3.4 Variations in the Expected Utility of each pure strategy $\{C,D,Q_1,Q_2\}$ for both agents implementing the QCJAL-algorithm with $\epsilon = 0$. .................. 65
3.5 Variations in the Expected Utility of each pure strategy $\{C,D,Q_1,Q_2\}$ for both agents implementing the QCJAL-algorithm with $\epsilon = 0.1$. .................. 67
3.6 Variations in the Expected Utility of each pure strategy $\{C,D,Q_1,Q_2\}$ for both agents implementing the QCJAL-algorithm with $\epsilon = 0.2$. .................. 69
3.7 Variations in the Expected Utility for both agents implementing the QCJAL-algorithm with noise: $U_{\text{noisy}} = U(\theta + \Delta\theta, \phi + \Delta\phi, \alpha + \Delta\alpha)$ where $(0,0,0) \leq (\Delta\theta, \Delta\phi, \Delta\alpha) \leq \frac{1}{4}(\pi, \frac{\pi}{2}, \frac{\pi}{2})$. .................. 72
3.8 Variations in the Expected Utility for both agents implementing the QCJAL-algorithm with agent A subject to noise: $U_{\text{noisy}} = U(\theta + \Delta\theta, \phi + \Delta\phi, \alpha + \Delta\alpha)$ where $(0,0,0) \leq (\Delta\theta, \Delta\phi, \Delta\alpha) \leq \frac{1}{2}(\pi, \frac{\pi}{2}, \frac{\pi}{2})$. .................. 73
3.9 Variations in the Expected Utility for both agents implementing the QCJAL-algorithm with agent A subject to noise: 

\[ U_{\text{noisy}} = U(\theta + \Delta \theta, \phi + \Delta \phi, \alpha + \Delta \alpha) \]

where \( (0, 0, 0) \leq (\Delta \theta, \Delta \phi, \Delta \alpha) \leq \frac{3}{4}(\pi, \pi, \pi). \)
Chapter 1

Introduction

We all play games, whether we like it or not. The term game, in this context, should not be confined to board games, card games, and the like, which are played for simple reasons like enjoyment, but should be considered in a much broader sense. When considering a game as a formalized payoff structure, where payoff can be characterized by a sum of money or something less tangible like happiness, a lot of everyday situations can be defined as games. In fact, every situation where we need to make a decision can be viewed as a game, whether this pertains to choosing which brand of coffee to buy or competing for a job position, making a certain decision results in a more (win) or less (lose) positive outcome.

Game theory and Quantum game theory

Game theory [Mye91, OR94], the theory of decision making in abstract game scenarios, provides mathematical tools for investigating situations in which several parties interact by making decisions according to personal interest. Game theory seeks to analyze how parties would decide in situations which involve contests, rivalry or struggle. The game-theoretical approach to these situations has proven to be relevant to economics, social sciences and biology. Besides finding analogies between situations of conflict within these fields, one can also define the search for a solution to a problem as a game where parties can cooperate and share information in order to reach a particular goal. For example, communication protocols can be put into this context, where the goal of the game amounts to an optimal exchange of information and possibly encryption. Information science can thereby be analyzed in a game theoretic context.

A device for computation that utilizes inherent quantum mechanical features, most importantly superpositions and entanglement, is known as a quantum computer. A recent rise of interest in quantum computing was brought about by the development of quantum algorithms to problems which were believed to be computationally infeasible
on a classical computer. As a consequence, there is an increase in attention to the field of quantum information, the study of information processing tasks using quantum systems.

Less than a decade ago, researchers had the idea of applying the rules of quantum mechanics to game theory. This newly founded field of quantum game theory was conceived to be of interest for further analysis of quantum communication or quantum computing protocols. For example, quantum communication [BBB+92], where an eavesdropper wants to intercept a message, can be viewed as a game. To this same extent, quantum cloning [Wer98] and quantum gambling [GVW99] can be considered as games.

Prisoner’s Dilemma

A well-known game is the so-called Prisoner’s Dilemma, which has been widely studied, experimented and varied upon. In its generic form, the prisoner’s dilemma is a two player game where both players can decide whether they want to ‘cooperate’ or ‘defect’. Consequently, every pair of decisions has a specific payoff of which both players are well aware. The game constitutes a risk versus reward tradeoff in that mutual cooperation yields the best combined payoff for both players, but is risky because playing ‘defect’ against ‘cooperate’ results in the highest possible individual payoff for the defecting player and the lowest payoff for the cooperating player. In the absence of communication or negotiation we have a classic dilemma between personal and mutual good, which can arguably be recognized in a lot of conflict scenarios throughout the world.

The prisoner’s dilemma sketches an abstract situation that has many analogies in the social sciences and biology. However, these real-life situations have living organisms as their players and somehow these players have a tendency to find the more beneficial outcome of mutual cooperation. Even when we put the abstract game into practice by letting human beings play the game against each other and giving them a monetary reward relative to the payoff structure, they often choose irrationally by playing ‘cooperate’. This leads to a fundamental question of whether nature plays the game suboptimally or whether game theory lacks in its descriptiveness.

A different approach to analyzing the prisoner’s dilemma is to recast it in a quantum mathematical framework. Eisert et al. [EWL99, EW00] have proposed a version of the prisoner’s dilemma, where the players have access to a set of quantum strategies. It has been shown that in this setting the fundamental dynamics of the game change and the classical dilemma ceases to exist.

Chen and Hogg [CHH06] have put this quantum prisoner’s dilemma in an experimental setting where people play this game. They show that, even without knowing
the underlying quantum mechanisms, people nearly achieve payoffs predicted by theory (which they claim to be 50% cooperation). Interestingly, this correspondence of actual human play with game theory for the quantum game is closer than that of the classical game. Moreover, people playing successive instances of the game show an increase in average payoff, i.e. they learn to play the game more optimally. This last notion, however, is surprising, as mathematically the quantum prisoner’s dilemma boils down to a four way rock-paper-scissors game\(^1\), i.e. the quantum prisoner’s dilemma constitutes four possible outcomes and four (independent) actions to choose from.

**Multi-agent learning**

In the past, there has been a genuine interest in an optimal strategy for players playing the prisoner’s dilemma repeatedly [Axe84]. To this extent, multi-agent learning algorithms have been devised to computationally find successful strategies. These models are characterized by an (software) agent playing the game against other agents and learning from the resulting outcomes. Particularly, multi-agent learning based on a reinforcement learning framework have been a successful heuristic method.

Specifically, the conditional joint action learner proposed by Banerjee and Sen [BS07] is of interest as it is a probability-based multi-agent learner and quantum mathematics is fundamentally based on probability theory. This property makes for a convenient adoption of this learner into the quantum domain.

This thesis engages itself with a thorough analysis of the quantum prisoner’s dilemma and subsequently explores the possibilities of multi-agent learning within this game. A quantum version of the conditional joint action learner is proposed and is placed in an experimental setting of playing the quantum prisoner’s dilemma repeatedly. Its performance is put into two contexts. First of all it is compared to the empirical results obtained by Banerjee and Sen [BS07], a novel approach to comparing the quantum prisoner’s dilemma with the classical variant. Secondly, the act of constructing and running experiments within the quantum prisoner’s dilemma leads to a deeper understanding of the intrinsic dynamics of this game. This understanding should leave us better equipped in assessing the quantum game and quantum game theory as a whole.

---

\(^1\)This game is also known as Ro-Sham-Bo, and is played by two players. Both players select either rock, paper or scissors and reveal their choices simultaneously. The winner is determined by the rules ‘rock beats scissors’, ‘scissors beats paper’ and ‘paper beats rock’. If both players select the same item, then the game is a tie.


Reading guide

Chapter 2, presents a theoretical background of Quantum Information, Quantum Game Theory, Prisoner’s Dilemma and Multi-Agent Learning. Additionally, this chapter includes an extensive analysis of the Quantum Prisoner’s Dilemma where both players have the full set of quantum strategies at their disposal.

Chapter 3 presents empirical results of two learners playing the quantum prisoner’s dilemma, described in chapter 2. One of these two learners is used to give a first impression of learning in the quantum game and consists of a very basic framework. The second learner is a quantum version of the conditional joint action learner and its performance is compared to the ‘classical’ conditional joint action learner.

Chapter 4 summarizes the contributions of this thesis and discusses the merits of quantum games in general and the quantum prisoner’s dilemma specifically. This chapter is concluded by possible further directions for research.
Chapter 2

Theory

This chapter covers theoretical backgrounds of subjects related to the research presented in this thesis. It offers a general introduction to quantum computation and quantum information followed by some introductory sections about game theory and multi-agent learning. A significant part of the chapter is devoted to a thorough description of the quantum prisoner’s dilemma.

2.1 Quantum Computing

The purpose of this section is to give a brief overview of background material in quantum mathematics and computer science in order to understand quantum games based on these techniques. It provides a general introduction to the main ideas and techniques in the field of quantum computation. Most of the information in the following subsections is based on the book *Quantum Computation and Quantum Information* by Nielsen et al. [NC00].

2.1.1 Preliminary Linear Algebra

In order to comprehend the contents of this chapter, a basic understanding of *linear algebra* is essential. The following paragraphs are meant to be a short reminder of some elementary concepts.

**Complex numbers** The letter $\mathbb{C}$ denotes the set of all complex numbers, which are of the form $a + ib$, where both $a$ and $b$ are real numbers and $i$ is the imaginary unit. Addition, multiplication and all other basic operators on complex numbers follow the usual rules of real numbers ($\mathbb{R}$) with added equality $i^2 = -1$.

The complex conjugate of a complex number $c = a + ib$ is $\overline{c} = a - ib$ and the norm (or absolute) is defined as $|c| = \sqrt{\overline{c}c} = \sqrt{a^2 + b^2}$; a real number. Every
complex number $c$ can be written as $|c|\cdot e^{i\theta}$ for some real number $\theta$ (the angle in the complex plane), due to Euler’s formula $e^{i\theta} = \cos \theta + i\sin \theta$.

**Vector spaces** A $d$-dimensional vector space is denoted by $V = \mathbb{C}^d$, which is the set of all column vectors of $d$ complex numbers. A basis for $V$ is a set of $d$ vectors $v_1 \ldots v_d$ such that every vector $v \in V$ can be written as a linear combination of these basis vectors $v = \sum_{i=1}^{d} a_i v_i$ where $a_i \in \mathbb{C}$ is a set of scalars.

**Inner product** The inner product of two vectors $v$ and $w$ is defined as follows:

\[
\langle v, w \rangle = v^T \cdot w = (v_1 \ w_1 \ldots v_d \ w_d) = v_1 w_1 + \ldots + v_d w_d
\]

Note that $v^T$ denotes the complex conjugate transpose of $v$; that is a column vector of elements is mapped to a row vector of complex conjugates of its elements.

**Euclidian Norm** The euclidian norm, or length, of a vector $v$ is defined as the square root of the inner product with itself. That is:

\[
\|v\| = \sqrt{\langle v, v \rangle} = \sqrt{v_1 \overline{v}_1 + \ldots + v_d \overline{v}_d}
\]

**Hilbert space** The Hilbert space is an infinite-dimensional inner product space - a vector space where distances and angles can be measured. All quantum states are defined as a vector in a Hilbert space of a particular dimension. Within the scope of this thesis we can consider Hilbert spaces as being finite-dimensional.

**Tensor product** The tensor product, sometimes referred to as the Kronecker product or outer product, is used to combine two vectors in order to construct a vector of a dimension equal to the summed dimension of these vectors. It is defined as follows:

\[
v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{pmatrix} \quad ; \quad w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix} \quad ; \quad v \otimes w = \begin{pmatrix} v_1 w \\ v_2 w \\ \vdots \\ v_d w \end{pmatrix}
\]

**Operators** Every linear function $A : V \rightarrow V$ corresponds to a $d \times d$ matrix $A$ and
can be applied to a vector $v$ of dimension $d$. In linear algebra, these matrices are called operators.

The identity function is the identity matrix $I_d$, i.e. a $d \times d$ matrix with 1s along its diagonal and 0s elsewhere.

**Transpose and Adjoint** The transpose $A^T$ of a matrix $A$ is the matrix with $A$’s elements swapped around the diagonal, that is, element $a_{ij}$ of $A$ becomes element $a_{ji}$ of $A^T$. If we take the complex conjugates of all elements of $A^T$ we get the conjugate transpose or adjoint of $A$, denoted by $A^\dagger$.

Also, if $v$ is a column vector, then the adjoint of this vector $v^\dagger$ becomes a row vector consisting of complex conjugates of all elements in $v$.

**Determinant** The determinant of a matrix $A$, det($A$), is a scalar factor associated with this matrix. The determinant can be calculated in the following way:

Let $A$ be a $d \times d$ (square) matrix with elements $a_{i,j}$, where $i = 1..d$ and $j = 1..d$. Then the determinant of this matrix is defined inductively by:

\[
\begin{align*}
\text{det}(A) &= a_{11} & \text{if } d = 1 \\
\text{det}(A) &= a_{11}a_{22} - a_{12}a_{21} & \text{if } d = 2 \\
\text{det}(A) &= a_{11}M_{11} - a_{12}M_{12} + ... + (-1)^{d+1}a_{1d}M_{1d} & \text{if } d > 2
\end{align*}
\]

Here, $M_{ij}$ equals the determinant of the $(d-1) \times (d-1)$ submatrix $A_{ij}$, obtained by removing row $i$ and column $j$ from matrix $A$.

**Eigenvalues and Eigenvectors** Given a linear transformation, i.e. applying an operator $A$ (a matrix) on a vector $v$, an eigenvector of that linear transformation is a non-zero vector which may change in length, but not direction, when that transformation is applied to it.

For each eigenvector of a linear transformation, there is a corresponding scalar value called an eigenvalue for that vector, which determines the amount the eigenvector is scaled under the linear transformation.

Let $A$ be an operator defined as a square matrix of (possibly) complex elements and dimension $d$, i.e. $A \in \mathbb{C}^{d \times d}$. Then, a non-zero vector $u \in \mathbb{C}^d$ is an eigenvector of $A$ with corresponding eigenvalue $\lambda \in \mathbb{C}$ if equation (2.4) holds.

\[Au = \lambda u\]  

(2.4)

The eigenvalues $\lambda$ of matrix $A$ can be characterized without the (explicit) notion
of the eigenvectors by solving linear equation (2.5).

$$\det(\lambda I_d - A) = 0$$  \hfill (2.5)

**Normal operators** An operator $A$ is called normal if $A^\dagger A = AA^\dagger$, that is the operator commutes with its adjoint.

**Hermitian operators** An operator $H$ is called *Hermitian* if and only if $H = H^\dagger$, that is the operator is equal to its adjoint. The elements on the main diagonal of any Hermitian matrix is necessarily real.

A special set of Hermitian operators, which are also unitary, are known as the *Pauli* matrices. They consist of a set of four $2 \times 2$ matrices, often referred to as $\{I_2, \sigma_x, \sigma_y, \sigma_z\}$, and are very useful as they formulate an orthonormal basis of the complex Hilbert space of all $2 \times 2$ matrices ($\mathbb{C}^{2\times2}$).

The Pauli matrices are defined as:

$$I_2 \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \sigma_x \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$  \hfill (2.6)

**Unitary operators** An operator $U$ is called *unitary* if and only if $UU^\dagger = U^\dagger U = I$.

Unitary operators are *norm preserving*:

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{pmatrix} \quad U = \begin{pmatrix} u_{11} & \cdots & u_{1d} \\ \vdots & \ddots & \vdots \\ u_{d1} & \cdots & u_{dd} \end{pmatrix} \quad \|Uv\| = \sqrt{\overline{v}^T U^\dagger U}v = \sqrt{\overline{v}^T}v$$

Unitary operators can be thought of as operators that do not change the scale of their argument and are reversible by applying their adjoint.

**Spectral Decomposition** *Spectral decomposition* is part of the spectral theory of operators on a Hilbert space. In a sense, spectral theory is an extension to the theory of eigenvectors and eigenvalues for square matrix operators. Spectral decomposition is a representation theorem for normal operators which states that an operator is normal if and only if it is diagonalizable.

**Theorem 2.1.1** (Spectral Theorem\textsuperscript{1}). Let $A \in \mathbb{C}^{n\times n}$ be a normal matrix, i.e. $AA^\dagger = A^\dagger A$, and let the spectrum be written by $\text{spec}(A) = \{\lambda_1, ..., \lambda_n\}$. Then there exists an orthonormal basis $\{u_1, ..., u_n\}$ of $\mathbb{C}^n$ such that each $u_i$ is an

\textsuperscript{1}Adopted from John Watrous’ lecture notes, see: \url{www.cs.uwaterloo.ca/watrous/lecture-notes.html}
2.1. QUANTUM COMPUTING

eigenvector of $A$ with associated eigenvalue $\lambda_i$, for $i = 1, \ldots, n$. Equivalently, this orthonormal basis is such that equation (2.7) holds.

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^\dagger$$  

(2.7)

The expression of a given normal matrix $A$ in the form of equation (2.7) for some orthonormal basis $\{u_1, \ldots, u_n\}$ is said to be a spectral decomposition of $A$. The spectral decomposition is unique if and only if all of the eigenvalues of $A$ are distinct.

It follows from the Spectral Theorem that the eigenvalues of a Hermitian matrix are always real.

2.1.2 Quantum Computer Science

The bit (binary digit) is the basic unit of information in the classical computer science of computation and information. Quantum computing is based on an analogous concept, namely the quantum bit or qubit for short. Just as a classical bit has a state, namely 0 or 1, a qubit can also be described in states. Two possible states a qubit might be in are $|0\rangle$ and $|1\rangle$, which naturally correspond to the classical states of a bit. The difference, however, is that a qubit can also be in a combination of these states; a so-called superposition of states - one of the main peculiarities found in quantum physics.

Properties of a qubit

The standard notation for states that is used in quantum mechanics is called the Dirac notation. This notation uses $\rangle \langle$ (called ket) to denote a state, and should be thought of as a vector. The classical bits 0 and 1 are defined as:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

By definition, the state of a qubit $\varphi$ is given as:

$$|\varphi\rangle = \alpha |0\rangle + \beta |1\rangle,$$

which makes its vector:

$$|\varphi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$  

(2.8)

The variables $\alpha$ and $\beta$ in (2.8) stand for complex numbers, though in most cases only real numbers are used. Basically, a bit can be categorized as 0 or 1 and a qubit can be categorized as 0 ($\beta = 0$), 1 ($\alpha = 0$) or a superposition of both ($\alpha \neq 0, \beta \neq 0$).
The state of a qubit, for computational purposes, should be considered as being both 0 and 1 at the same time\(^2\).

A qubit can be measured, thereby destroying its quantum state and turning it into a classical bit. The variables \(\alpha\) and \(\beta\) represent the amplitudes of the pure states \(|0\rangle\) and \(|1\rangle\) respectively and are normalized to 1, i.e. equation (2.9) should always hold.

\[
|\alpha|^2 + |\beta|^2 = 1 \tag{2.9}
\]

This definition gives rise to a probability distribution of observing a specific pure state as denoted in (2.10).

\[
P(\varphi = 0) = |\alpha|^2 \quad P(\varphi = 1) = |\beta|^2 \tag{2.10}
\]

A quantum state of one or more qubits should be considered as a vector (in the Hilbert space) with their corresponding classical states as their computational basis states. So, the computational basis states for one qubit are \(|0\rangle\) and \(|1\rangle\), for two qubits \(|00\rangle, |01\rangle, |10\rangle\) and \(|11\rangle\), and so forth. These computational basis states are naturally orthonormal.

**Operators on a qubit**

Given two qubits defined as follows:

\[
|\varphi\rangle = \alpha_1 |0\rangle + \beta_1 |1\rangle \quad |\psi\rangle = \alpha_2 |0\rangle + \beta_2 |1\rangle \tag{2.11}
\]

We can now define:

**Bra-Ket** A ‘ket’ (\(|\rangle\)) denotes a quantum state with according column vector, a ‘bra’ (\(<\langle\rangle\>) denotes the vector dual of this state in a row vector of complex conjugates. This notation is shown in (2.12).

\[
|\varphi\rangle = \begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} \quad <\varphi| = \begin{pmatrix} \alpha_1^* & \beta_1^* \end{pmatrix} \tag{2.12}
\]

\(^2\)Think of it as a ray of light with polarization of 45°, which is split by a prism in two equally large parallel beams with polarization of 90° and 0° respectively. Let’s place another prism interfering both beams again into a single beam of 45°. Now what if this beam only consists of a single photon? Experimental physics has shown that the single photon has a polarization of 45° before the prism, *either* 90° or 0° when measuring between prisms and 45° again after the last prism. How does this work, as it has no other photon to interfere with? The photon interferes with itself, as it is in both states at the same time!
Inner product  The inner product is a function on two vectors and produces a (complex) number as output. It is defined as follows:

$$\langle \varphi \rangle \cdot |\psi\rangle \equiv \langle \varphi |\psi \rangle = \alpha_1 \alpha_2 + \beta_1 \beta_2$$

$$\langle \varphi |\psi \rangle = (\alpha_1 \beta_1) \cdot \begin{pmatrix} \alpha_2 \\ \beta_2 \end{pmatrix}$$  \hspace{1cm} (2.13)$$

Note that the inner product is a linear function and that the inner product of two states equals 0 if and only if these states are orthogonal, and the inner product equals 1 if and only if these states are the same.

Tensor product  The tensor product is used when describing the joint state of qubits. The $\otimes$ operator is used to denote this product. The joint state of two single qubits is defined as follows:

$$|\varphi\rangle \otimes |\psi\rangle = \alpha_1 \alpha_2 |00\rangle + \alpha_1 \beta_2 |01\rangle + \beta_1 \alpha_1 |10\rangle + \beta_1 \beta_2 |11\rangle$$  \hspace{1cm} (2.14)$$

Equivalently, these states can be described as vectors in the Hilbert space using the Kronecker-product$^3$.

$$|\varphi\rangle = \begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} \hspace{1cm} |\psi\rangle = \begin{pmatrix} \alpha_2 \\ \beta_2 \end{pmatrix}$$  \hspace{1cm} (2.15)$$

$$|\varphi\rangle \otimes |\psi\rangle = \begin{pmatrix} \alpha_1 |\psi\rangle \\ \beta_1 |\psi\rangle \end{pmatrix} = \begin{pmatrix} \alpha_1 \alpha_2 \\ \alpha_1 \beta_2 \\ \beta_1 \alpha_2 \\ \beta_1 \beta_2 \end{pmatrix}$$  \hspace{1cm} (2.16)$$

Furthermore, the tensor product is used in constructing multiple qubit operators from their (standard) single qubit form. Consider $A$ and $B$ as single qubit

---

$^3$The Kronecker-product is a special case of the tensor product and corresponds to the abstract tensor product of linear maps. Within the context of quantum computation in this thesis, it is sufficient to consider the tensor product as the composition of quantum states where the Kronecker product is the actual operation on these quantum states when represented as vectors and on quantum operator when represented as matrices.
operators, then their tensor product becomes:

\[ A \otimes B = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \otimes \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \]

\[ = \begin{bmatrix} a_{1,1}b_{1,1} & a_{1,1}b_{1,2} & a_{1,2}b_{1,1} & a_{1,2}b_{1,2} \\ a_{1,1}b_{2,1} & a_{1,1}b_{2,2} & a_{1,2}b_{2,1} & a_{1,2}b_{2,2} \\ a_{2,1}b_{1,1} & a_{2,1}b_{1,2} & a_{2,2}b_{1,1} & a_{2,2}b_{1,2} \\ a_{2,1}b_{2,1} & a_{2,1}b_{2,2} & a_{2,2}b_{2,1} & a_{2,2}b_{2,2} \end{bmatrix} \]

By definition, the tensor product satisfies the following basic properties:

1. For arbitrary scalar \( z \) and arbitrary states \( \varphi \) and \( \psi \):

\[ z(|\varphi\rangle \otimes |\psi\rangle) = (z|\varphi\rangle) \otimes |\psi\rangle = |\varphi\rangle \otimes (z|\psi\rangle) \]

2. For arbitrary states \( \varphi \), \( \phi \) and \( \psi \):

\[ (|\varphi\rangle + |\phi\rangle) \otimes |\psi\rangle = |\varphi\rangle \otimes |\psi\rangle + |\phi\rangle \otimes |\psi\rangle \]

**Evolution** All systems in quantum computing evolve unitarily, which means that any operator \( U \) on a quantum state must be unitary. An operator \( U \) is unitary if and only if (2.21) holds.

\[ UU^\dagger = I \]

here \( U^\dagger \) denotes the complex conjugate transpose of \( U \) and \( I \) the identity(-matrix) (2.21)

Because every operator in a quantum system is unitary, the following properties are true for quantum systems:

1. The probability interpretation as defined in (2.10) is preserved throughout application of unitary operators, i.e. inner product is preserved.
2. The system is reversible and therefore can move forward or backward in time.
3. Information (states) cannot be erased or copied throughout computation.
Quantum computation has a specific set of basic operators of which the Hadamard transform is specifically interesting. The Hadamard transform, when applied to a state \( |\phi\rangle \), produces an equal superposition of all computational basis states. Moreover, it does so extremely efficiently, producing a superposition of \( 2^n \) states in just \( n \) operations. The Hadamard transform on a single qubit is given as a matrix in (2.22).

\[
H = \begin{pmatrix}
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}
\end{pmatrix}
\] (2.22)

For example, application of the Hadamard transform on \( |0\rangle \):

\[
H |0\rangle = \begin{pmatrix}
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}
\end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle)
\]

and backwards:

\[
H(\frac{1}{\sqrt{2}} (|0\rangle + |1\rangle)) = \begin{pmatrix}
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}
\end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |0\rangle
\]

**Measurement** The act of measuring a qubit entails destroying its quantum state and observing a classical bit, as mentioned earlier. This type of measurement is often referred to as *measurement of a qubit in the computational basis*. Formally, measurement of a single qubit in the computational basis is defined by the two measurement operators:

\[
M_0 = |0\rangle \langle 0| \quad \text{and} \quad M_1 = |1\rangle \langle 1|
\]

Now, the probability for observing result \( |m\rangle \) in quantum state \( |\varphi\rangle = \alpha |0\rangle + \beta |1\rangle \) is given by equation (2.23) and the state of this system after the measurement \( |\varphi'\rangle \) will become as defined in equation (2.24).

\[
Pr(m) = \langle \varphi | M_m^\dagger M_m | \varphi \rangle = \begin{cases} 
|\alpha|^2 & \text{for } m = 0 \\
|\beta|^2 & \text{for } m = 1
\end{cases} 
\] (2.23)

\[
|\varphi'\rangle = \frac{M_m |\varphi\rangle}{\sqrt{\langle \varphi | M_m^\dagger M_m | \varphi \rangle}} = \begin{cases} 
|0\rangle & \text{for } m = 0 \\
|1\rangle & \text{for } m = 1
\end{cases} 
\] (2.24)

From here, it should not be hard to understand that it is also possible to measure

\footnote{Note that a Hadamard transform on a two-qubit state can be performed using \( H \otimes H \).}
qubits in different basis states. An often used alternative of measurement is measuring in the orthonormal basis of \(|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\) and \(|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\).

An important special case of measurement is that of projective measurement. A projective measurement is described by an observable, \(H\), which is a Hermitian operator on the state space of the system being observed. This observable can be defined by its eigenvalues \(\{\lambda_1, ..., \lambda_n\}\) and associated eigenvectors \(\{u_1, ..., u_n\}\) according to spectral decomposition, as defined in equation (2.25).

\[
H = \sum_{i=1}^{n} \lambda_i |u_i \rangle \langle u_i |
\]  

(2.25)

It follows from spectral theory that \(|u_i \rangle \langle u_i |\) is the projector onto the eigenspace of \(H\) with eigenvalue \(\lambda_i\). The possible outcomes of a measurement, \(m\), correspond to the eigenvalues, \(\{\lambda_1, ..., \lambda_n\}\), of the observable and depend on the associated projector \(P_m\). Upon measuring a quantum state \(|\varphi\rangle\), the probability of getting result \(m\) is given by equation (2.26) with \(m = \{\lambda_1, ..., \lambda_n\}\) and associated projectors \(P_m = \{|u_1 \rangle \langle u_1 |, ..., |u_n \rangle \langle u_n |\}\).

\[
Pr(m) = \langle \varphi | P_m | \varphi \rangle
\]

(2.26)

Given that outcome \(m\) occurred, the state of the quantum system after the measurement \(|\varphi'\rangle\) becomes as defined in equation (2.27).

\[
|\varphi'\rangle = \frac{P_m |\varphi\rangle}{\sqrt{Pr(m)}}
\]

(2.27)

Consider the following example of a projective measurement on a single qubit state:

**Example** We take our observable to be the Pauli matrix \(\sigma_z\):

\[
\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.
\]

The eigenvalues and associated eigenvectors of this observable are:

\[
\lambda_1 = 1 \quad \text{and} \quad \lambda_2 = -1.
\]

\[
u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} (\equiv |0\rangle) \quad \text{and} \quad \nu_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} (\equiv |1\rangle).
\]
We have an arbitrary single qubit state in:

$$|\varphi\rangle = \alpha |0\rangle + \beta |1\rangle.$$ 

Thus, the probability for measuring either +1 or −1 follows:

$$Pr(m) = \langle \varphi | P_m | \varphi \rangle$$

$$= \begin{cases} 
    \langle \varphi | 0 \rangle \langle 0 | \varphi \rangle = |\alpha|^2 & \text{for } m = 1 \\
    \langle \varphi | 1 \rangle \langle 1 | \varphi \rangle = |\beta|^2 & \text{for } m = -1 
\end{cases}$$

and after measurement, the state $|\varphi\rangle$ becomes:

$$|\varphi'\rangle = \frac{P_m}{\sqrt{Pr(m)}}$$

$$= \begin{cases} 
    \frac{\alpha}{\sqrt{|\alpha|^2}} |0\rangle = |0\rangle & \text{for } m = 1 \\
    \frac{\beta}{\sqrt{|\beta|^2}} |1\rangle = |1\rangle & \text{for } m = -1 
\end{cases}$$

This same type of projective measurement can be performed for each of the Pauli matrices, which all have the same eigenvalues but different eigenvectors. More generally, the eigenvectors of the Pauli matrices together form an orthonormal basis in $\mathbb{C}^2$ and are often referred to as eigenspinors\(^5\). It is possible to perform a ‘measurement of spin along the $v$ axis’, by taking any real three-dimensional unit vector $v$ and defining an observable:

$$v \cdot \sigma \equiv v_1 \sigma_x + v_2 \sigma_y + v_3 \sigma_z.$$ 

Now, projective measurement has many useful properties. In particular, it offers an easy way to calculate average values for measurements by calculating the expected value as defined in equation (2.28), which reduces to an expression for the average value of a specific observable $H$ in equation (2.29).

$$E(H) = \sum_m m Pr(m) \quad (2.28)$$

$$= \sum_m m \langle \varphi | P_m | \varphi \rangle$$

$$= \langle \varphi | \left( \sum_m m P_m \right) | \varphi \rangle$$

$$= \langle \varphi | H | \varphi \rangle \equiv \langle H \rangle \quad (2.29)$$

\(^5\)The term eigenspinors stems from basis spin directions a particle can take.
Entanglement is a property associated with composite quantum systems in which two or more states are linked in such a way that one state can no longer be adequately described without the full mention of the other state(s). By definition, an entangled state of \( n \) qubits cannot be factorized into single qubit states. That is, for a two qubit state \( |\phi\rangle \) it cannot be rewritten in terms of \( |\varphi\rangle \otimes |\psi\rangle \), where both \( \varphi \) and \( \psi \) are single qubit states.

The classical example of an entangled quantum state is the so-called EPR-pair (named after Albert Einstein, Boris Podolsky and Nathan Rosen [EPR35]). The quantum state of an EPR-pair, as in (2.30), consists of two qubits maximally correlated with each other.

\[
|\varphi\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle)
\]  

In theory, when measuring these two qubits, one can either observe \( |00\rangle \) with probability \( \frac{1}{2} \) or \( |11\rangle \) with probability \( \frac{1}{2} \). As a result, a measurement of the first qubit always gives the same result as the measurement of the second qubit. That is, the measurement outcomes are correlated.

These correlations have been of great interest in the investigation of the foundations of quantum physics as it poses an inherent paradox: ‘What if we would only measure one qubit? Would this have a non-local effect on the other qubit?’.

Let’s say we are able to separate these two qubits and place them thousands of miles apart, then in theory by measuring one qubit we would be able to tell the state of the other qubit because of its entanglement. This would either suggest that quantum mechanics somehow is able to ‘predict’ the outcome of a measurement yet to take place or that there exists an exchange of information between the two qubits with a speed faster than light. Both these explanations violate fundamental laws in physics.

This paradox, devised by Einstein, Podolsky and Rosen [EPR35] in order to show that quantum physics is an incomplete theory, was further investigated by John Bell [Bel65]. Bell proved that the measurements in the EPR-pair are stronger than could ever exist between classical systems. These results were the first to imply that quantum physics allows information processing beyond what is possible from the classical point of view.

Bell’s experiment and the demonstration of quantum non-locality is described in more detail in the next section, section 2.1.3.
2.1.3 Quantum non-locality and Bell inequalities

From the classical point of view, it is assumed that physical properties of objects have an existence independent of observation. Measurements on objects simply reveal these physical properties. For example, the speed of a car can be measured by the reflections of multiple infra-red beams of light and calculating the distance this light has travelled.

According to quantum physics, however, an unobserved particle does not possess physical properties that exist independent of observation. Rather, such physical properties arise as a consequence of measurements performed upon the system.

At first, this quantum mechanical view of nature was rejected by many physicists, most prominently by Albert Einstein. As a consequence, Einstein, Podolsky and Rosen proposed a thought experiment [EPR35] which was meant as a demonstration that quantum mechanics is not a complete theory of nature. The essence of their argument is based on what they termed ‘elements of reality’ and their belief was that any such element must be represented in any complete physical theory. By identifying elements of reality that were not included in quantum mechanics, they intended to show that quantum mechanics is not a complete physical theory. The way they attempted to do this was by introducing a ‘sufficient condition’ for a physical property to be an element of reality, namely, that it be possible to predict with certainty the value that property will have, immediately before measurement.

Let us consider the fully entangled EPR-pair, as in equation (2.30), and suppose two parties, Alice and Bob, both possess one of these qubits each. For the sake of the argument, we should consider Alice and Bob to be a long distance away from one another. Suppose that Alice performs a measurement on her qubit and that she measured a 1. Then according to the entangled quantum state, she is able to predict with certainty that Bob will also measure a 1 on his qubit. Because it is always possible for Alice to predict the outcome of Bob’s measurement, that physical property must correspond to an element of reality, by the EPR criterion, and should be represented in any complete physical theory. However, quantum mechanics, as it is presented, merely dictates how to calculate the probabilities of the respective measurement outcomes and certainly does not include any fundamental element intended to represent the value of the measurement. As a result Einstein, Podolsky and Rosen argued that quantum mechanics is incomplete as it lacked some essential ‘element of reality’, by their criterion.

However, this argument was not deemed to be very convincing and as a result an experimental test was proposed by John Bell [Bel65] that could be used to check whether there was any truth in this argument. Bell’s experiment and its resulting theorem is meant to prove that local hidden variable theories, theories which explain the lacking ‘element of reality’ with a hidden variable, cannot remove the statistical nature of quantum mechanics. Furthermore, Bell’s theorem implies that nature,
according to quantum mechanics, is not locally deterministic.

The specifics of the thought experiment proposed by Bell will be described in two parts. First of all, the experiment is represented in its classical form. That is, we ignore the possible existence of quantum phenomena like entanglement and analyze the possible outcomes of the experiment using general probability theory. This analysis yields a set of inequalities, known as the Bell inequalities, which offer a boundary to the outcomes of the experiment. The second part will describe this same experiment, but now with the introduction of quantum entanglement, which is shown to violate the Bell inequalities defined earlier.

Bell’s experiment (classically): Suppose we prepare two objects in a repeatable procedure, that is it doesn’t matter how these objects are prepared, just that we are able to repeat this experimental procedure. These two objects are being separated and sent to two parties; Alice and Bob. Both Alice and Bob are then allowed to simultaneously perform some measurement on these objects. Alice possesses two apparatuses, which we shall label $P_Q$ and $P_R$, and Bob also has two, which we shall label $P_S$ and $P_T$. These devices have just the simple purpose of revealing a certain objective property of the object, that is Alice’s apparatus $P_Q$ reveals whether the object has property $Q$, $P_R$ reveals whether the object has property $R$ and the same for Bob’s $P_S$ and $P_T$ and his object. Let us assign an output of $+1$ for each apparatus when the object has its respective property and an output of $-1$ when it doesn’t.

Now, for the actual experiment, we consider both Alice and Bob to perform a measurement with only a single apparatus and that they decide randomly on which apparatus to use, e.g. both Alice and Bob flip a coin for deciding on using $P_Q$ or $P_R$ and $P_S$ or $P_T$ respectively. So, Alice chooses to measure either $Q$ or $R$ which results in a value of $±1$ for the respective property and Bob chooses to measure either $S$ or $T$ which also results in a value of $±1$ for the respective property. The timing of this experiment is arranged so that Alice and Bob perform their measurement at the same time and are a long distance away from each other. Therefore, according to the laws of relativity, neither Alice nor Bob can disturb the results of the other’s measurements, i.e. they are causally disconnected.

The variables of note in this experiment are \(\{Q, R, S, T\}\) and each of them has a value of $±1$, depending on how the objects are actually prepared. Considering these values, equation (2.31) will always hold because from $R, Q \pm 1$ it follows that either $Q + R = 0$ or $R - Q = 0$.

\[
Q \cdot S + R \cdot S + R \cdot T - Q \cdot T = (Q + R)S + (R - Q)T = \pm 2 \tag{2.31}
\]
Suppose that $Pr(q, r, s, t)$ is the probability that the system is in a state where $Q = q$, $R = r$, $S = s$ and $T = t$. These probabilities may depend on how the actual objects are prepared and (eventually) on experimental noise. We can now describe the expected value, or statistical mean, by considering:

$$E(QS + RS + RT - QT) = \sum_{qrst} Pr(q, r, s, t)(qs + rs + rt - qt)$$

$$\leq \sum_{qrst} Pr(q, r, s, t) \cdot 2 = 2$$

There also follows that:

$$E(QS + RS + RT - QT) = \sum_{qrst} Pr(q, r, s, t)qs + \sum_{qrst} Pr(q, r, s, t)rs + \sum_{qrst} Pr(q, r, s, t)rt - \sum_{qrst} Pr(q, r, s, t)qt$$


When we compare equation (2.32) with (2.33), we obtain the *Bell inequality* as denoted in inequality (2.34).

$$E(QS) + E(RS) + E(RT) - E(QT) \leq 2$$

This experiment can be repeated many times, thereby allowing Alice and Bob to determine each quantity on the left side of the Bell inequality. For example, after a (preferably large) set of experiments, Alice and Bob analyze their data together. They look at all the experiments where Alice measured $Q$ with her apparatus $P_Q$ and Bob measured $S$ with his apparatus $P_S$. By multiplication, they obtain a sample of values for $QS$ and by averaging, an estimate of $E(QS)$ to an accuracy only limited by the number of experiments. In a similar fashion, they are able to estimate the other quantities on the left hand side of inequality (2.34) and thereby check whether the inequality holds.

John Bell’s theory behind this thought experiment is that the Bell inequality (2.34) must be obeyed under any *local hidden variable theory*. The Bell experiment serves to investigate the validity of the entanglement effect as described by quantum

---

6This specific inequality is also often known as the *CHSH inequality*, after Clauser, Horne, Shimony and Holt [CHSH69]. It is, however, part of a larger set of inequalities generally known as Bell inequalities since the first was found by Bell.

7Local hidden variable theory, in general, states that distant events have no instantaneous (or at
mechanics instead of finding its explanation in hidden variable theory. To this end, consider the following example, in which quantum entanglement is introduced in the experiment.

**Quantum mechanics in the Bell experiment** Instead of the unknown preparation of the particles we now introduce a quantum system of two qubits in the following state:

\[ |\varphi\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \]

That is, the objects mentioned in the experiment are now the two qubits from this quantum system, where Alice receives the first qubit and Bob the second. This two qubit state is anti-correlated in that measuring the first qubit would result in a definite opposite value for the second qubit.

Now, Alice and Bob have their measurement apparatuses defined as projective measurements (see section 2.1.2) of the following observables:

\[
Q = \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad S = \frac{-\sigma_z - \sigma_x}{\sqrt{2}} = \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \\
R = \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad T = \frac{\sigma_z - \sigma_x}{\sqrt{2}} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.
\]

Performing these projective measurements on quantum state \( |\varphi\rangle \) will result in an outcome of \( \pm 1 \), similarly to the ‘classical’ experiment.

Alice and Bob still randomly decide which apparatus to use and perform a projective measurement on their respective qubit. In order to consider the effect of these measurements on the two-qubit state \( |\varphi\rangle \) we need to consider measurement of the following four observables:

\[
Q \otimes S = \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \quad Q \otimes T = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.
\]

least faster-than-light) effect on local ones. That is, this local hidden variable theory attempts to explain quantum mechanical effects by assuming a hidden variable present, a variable or property which is not detected or measured, that causes correlation.
2.1. QUANTUM COMPUTING

\[
R \otimes S = \begin{pmatrix}
0 & 0 & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\
0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\
-\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\
-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0
\end{pmatrix}
\quad R \otimes T = \begin{pmatrix}
0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\
0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\
-\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0
\end{pmatrix}.
\]

From these observables, we are able to calculate expected values:

\[
E(Q \otimes S) = \langle \phi | Q \otimes S | \phi \rangle = \frac{1}{\sqrt{2}}
\]

\[
E(Q \otimes T) = \langle \phi | Q \otimes T | \phi \rangle = -\frac{1}{\sqrt{2}}
\]

\[
E(R \otimes S) = \langle \phi | R \otimes S | \phi \rangle = \frac{1}{\sqrt{2}}
\]

\[
E(R \otimes T) = \langle \phi | R \otimes T | \phi \rangle = \frac{1}{\sqrt{2}}.
\]

And thus for this experiment, equation (2.35) holds.

\[
E(QS) + E(RS) + E(RT) - E(QT) = 2\sqrt{2}
\]

However, this violates inequality (2.34) from the Bell experiment, which dictated that the sum of average values should never exceed 2.

The difference in average measurement outcomes between the ‘classical’ approach (the Bell inequality (2.34)) and the quantum mechanical approach (equation (2.35)) has led to much debate. There have been numerous attempts to physically implement the experiment, most notably the experiments conducted by Aspect, Grangier and Roger [AGR82] confirmed the quantum mechanical predictions more dramatically than any of its predecessors. However, even though most experiments conducted so far have favored quantum mechanics, they are not entirely decisive because of certain loopholes\(^8\).

The violation of the Bell inequality refutes the argument by Einstein, Podolsky and Rosen. However, the derivation of the Bell inequality is based on two assumptions, namely:

1. The assumption that the physical properties \(\{Q, R, S, T\}\) have definite values which exist independent of observation. This assumption is called the assumption of realism.

\(^8\)These loopholes are known as the ‘detection loophole’ and the ‘communication loophole’, both of which are still under extensive research. For further reading, the Stanford Encyclopedia of Philosophy offers a concise overview of the matter: [http://plato.stanford.edu/entries/bell-theorem/]
2. The assumption that either of the parties performing their measurement does not influence the result of the measurement of the other party. This assumption is called the assumption of *locality*.

These two assumptions together are known as the assumptions of *local realism* and can be characterized as being intuitively plausible, especially if we consider the world to behave *deterministically*. However, the Bell inequalities show that at least one of these assumptions is incorrect, thereby stating that the world is not locally realistic.

The Bell inequalities function as a cornerstone in quantum mechanics and subsequently in quantum computing and quantum information. On one hand, they support quantum theory fundamentally and on the other hand they show exactly where the power of quantum theory lies: within entanglement. In a way, entanglement is a new resource with possibilities beyond classical resources and quantum computing and quantum information are positioned to exploit this new resource.

On a further note, the (hypothetical) Bell experiment can also be formulated as a game. The game formulation is natural and appealing for theoretical computer science, as it is more aimed at analyzing how different parties can optimize their performance in various abstract settings. A quantum game variant based on the experiment in this section is described extensively in section 2.4.2.

### 2.1.4 Bloch Sphere Representation

The Bloch sphere is a geometrical representation of single qubit states as points on the surface of a unit sphere. Most operators on single qubits can be neatly described within the Bloch sphere picture. Points on a unit sphere can be accurately described by the angles they make with the cartesian axes and thereby operators can be defined by the rotations they pose on these angles.

Given an arbitrary single qubit state $|\varphi\rangle$, which can be written as:

$$|\varphi\rangle = \alpha |0\rangle + \beta |1\rangle.$$  \hspace{1cm} (2.36)

Here, $\alpha, \beta \in \mathbb{C}$ denote the amplitudes of the pure states $|0\rangle$ and $|1\rangle$, that is this single qubit state is in a superposition of these pure states if and only if both amplitudes are non-zero. Any qubit state $|\varphi\rangle$ is subjected to the normalization constraint that the inner-product with itself is equal to 1 ($\langle \varphi | \varphi \rangle = 1$), which requires that:

$$|\alpha|^2 + |\beta|^2 = 1.$$ \hspace{1cm} (2.37)

---

9Determinism is based on the philosophical proposition that every event is necessitated by antecedent events and conditions together with the laws of nature. Whether this proposition is true has led to large debates, which also include the question of free will.
Because of this normalization, $|\alpha|^2$ represents the probability of this single qubit state collapsing into $|0\rangle$ and $|\beta|^2$ represents the probability of collapsing into $|1\rangle$ upon measurement of this single qubit. This type of collapsing is one of the fundamental postulates of quantum physics (see equation (2.10)).

As we know from complex number analysis, a complex number can be represented two-dimensionally by its real and imaginary part. This makes an arbitrary single qubit state four-dimensional, or expressed in polar coordinates:

$$|\phi\rangle = r_\alpha e^{i\phi_\alpha} |0\rangle + r_\beta e^{i\phi_\beta} |1\rangle.$$  \hspace{1cm} (2.38)

Here, $r_\alpha, r_\beta, \phi_\alpha, \phi_\beta \in \mathbb{R}$. However, the only measurable quantities of a single qubit state are the probabilities $|\alpha|^2$ and $|\beta|^2$, so multiplying the state by an arbitrary factor $e^{i\gamma}$ (a global phase) has no observable consequences, because:

$$|e^{i\gamma}\alpha|^2 = (e^{-i\gamma}\alpha)(e^{i\gamma}\alpha) = e^{-i\gamma i\gamma}\alpha\alpha = |\alpha|^2.$$  \hspace{1cm} (2.39)

Similarly for $|\beta|^2$. Therefore, we are free to multiply state $|\phi\rangle$ with $e^{-i\phi_\alpha}$ giving:

$$|\phi'\rangle = e^{-i\phi_\alpha} |\phi\rangle = r_\alpha |0\rangle + r_\beta e^{i(\phi_\beta - \phi_\alpha)} |1\rangle = r_\alpha |0\rangle + r_\beta e^{i\phi} |1\rangle.$$  \hspace{1cm} (2.40)

This now only has three variables $r_\alpha, r_\beta, \phi \in \mathbb{R}$ (with $\phi = \phi_\beta - \phi_\alpha$). Additionally, the normalization constraint $\langle \phi' | \phi' \rangle = 1$ should still hold. Switching back to cartesian representation for the amplitude of $|1\rangle$:

$$|\phi'\rangle = r_\alpha |0\rangle + r_\beta e^{i\phi} |1\rangle = r_\alpha |0\rangle + (x + iy) |1\rangle$$  \hspace{1cm} (2.41)

and the requirement of the normalization constraint:

$$|r_\alpha|^2 + |x + iy|^2 = 1$$  \hspace{1cm} (2.42)

$$r_\alpha^2 + (x + iy)(x + iy) = 1$$  \hspace{1cm} (2.43)

$$r_\alpha^2 + (x - iy)(x + iy) = 1$$  \hspace{1cm} (2.44)

$$r_\alpha^2 + x^2 + y^2 = 1.$$  \hspace{1cm} (2.45)

Here, the last equality is the equation of a unit sphere in the real 3-dimensional space with cartesian coordinates $(x, y, r_\alpha)$. Any point in the cartesian coordinate system is expressed by its $x$, $y$ and $z$ coordinates on their respective axis, but can also be expressed using polar coordinates as shown in figure 2.1. The cartesian coordinates

\hspace{1cm}

\footnote{Let $z \in \mathbb{C}$ then $z = r(\cos \theta + i \sin \theta) = re^{i\theta}$}
are related to polar coordinates by:

\[ x = r \sin \theta \cos \phi \]  \hspace{1cm} (2.46)

\[ y = r \sin \theta \sin \phi \]  \hspace{1cm} (2.47)

\[ z = r \cos \theta. \]  \hspace{1cm} (2.48)

Figure 2.1: Spherical Polar Coordinates

We can now rename \( r, x, \) and \( y \) in (2.41) with their polar coordinates, while noting that \( r = 1 \) by definition.

\[ |\varphi\rangle = z |0\rangle + (x + iy) |1\rangle \]  \hspace{1cm} (2.49)

\[ = \cos \theta |0\rangle + (\cos \phi + i \sin \phi) \sin \theta |1\rangle \]  \hspace{1cm} (2.50)

\[ = \cos \theta |0\rangle + e^{i\phi} \sin \theta |1\rangle \]  \hspace{1cm} (2.51)

This leaves us with a formula of just two parameters defining points on a unit sphere (figure 2.2). Notice that \( \theta = 0 \) results in \( |\varphi\rangle = |0\rangle \) and \( \theta = \frac{\pi}{2} \) results in \( |\varphi\rangle = |1\rangle \), which suggests that \( 0 \leq \theta \leq \frac{\pi}{2} \) may generate all points on the Bloch sphere. In order to verify this statement we should consider two points on opposite sides of the sphere. The opposite point of a given point \( |\varphi\rangle \) with polar coordinates \((1, \theta, \phi)\) has polar coordinates \( (1, \pi - \theta, \phi + \pi) \) (see figure 2.2). We can rewrite this opposite state
\[ |\varphi'_o\rangle: \]

\[ |\varphi'_o\rangle = \cos(\pi - \theta) |0\rangle + e^{i(\phi + \pi)} \sin(\pi - \theta) |1\rangle \tag{2.52} \]

\[ = - \cos(\theta) |0\rangle + e^{i\phi} e^{i\pi} \sin(\theta) |1\rangle \tag{2.53} \]

\[ = - \cos(\theta) |0\rangle - e^{i\phi} \sin(\theta) |1\rangle \tag{2.54} \]

\[ = - |\varphi\rangle. \tag{2.55} \]

So, it is only necessary to consider the upper hemisphere \(0 \leq \theta \leq \frac{\pi}{2}\), as opposite points in the lower hemisphere differ only by a phase factor \(-1\) and are therefore equal for all measurable qualities of a qubit. In order to map the points on the upper hemisphere onto points on a sphere we simply substitute \(\theta\) with \(\frac{\theta}{2}\) and we now have:

\[ |\varphi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle \tag{2.56} \]

where \(0 \leq \theta \leq \pi\), \(0 \leq \phi \leq 2\pi\) are the coordinates of points on the Bloch sphere (see figure 2.2).

Proposition 2.1.2. Opposite points on the Bloch sphere correspond to orthogonal qubit states.
**Proof.** Consider a general qubit state $|\varphi\rangle$:

$$|\varphi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle.$$  \hspace{0.5cm} (2.57)

Also, consider $|\chi\rangle$ corresponding to the opposite point on the Bloch sphere ($\theta_\chi = \pi - \theta_\varphi$ and $\phi_\chi = \phi_\varphi + \pi$)

$$|\chi\rangle = \cos \left(0 \frac{\pi - \theta}{2}\right) |0\rangle + e^{i(\phi + \pi)} \sin \left(0 \frac{\pi - \theta}{2}\right) |1\rangle$$ \hspace{0.5cm} (2.58)

$$= \cos \left(0 \frac{\pi - \theta}{2}\right) |0\rangle - e^{i\phi} \sin \left(0 \frac{\pi - \theta}{2}\right) |1\rangle.$$ \hspace{0.5cm} (2.59)

So,

$$\langle \chi | \varphi \rangle = \cos \left(\frac{\theta}{2}\right) \cos \left(0 \frac{\pi - \theta}{2}\right) - e^{-i\phi} e^{i\phi} \sin \left(\frac{\theta}{2}\right) \sin \left(0 \frac{\pi - \theta}{2}\right)$$ \hspace{0.5cm} (2.60)

$$= \cos \left(\frac{\theta}{2}\right) \cos \left(0 \frac{\pi - \theta}{2}\right) - \sin \left(\frac{\theta}{2}\right) \sin \left(0 \frac{\pi - \theta}{2}\right)$$ \hspace{0.5cm} (2.61)

$$= \cos \left(\frac{\theta}{2}\right) \sin \left(\frac{\theta}{2}\right) - \sin \left(\frac{\theta}{2}\right) \cos \left(\frac{\theta}{2}\right)$$ \hspace{0.5cm} (2.62)

$$= 0.$$ \hspace{0.5cm} (2.63)

Opposite points correspond to orthogonal qubit states.\(^{11}\)

Every possible single qubit state is represented by the Bloch vector $(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$ and every unitary operation on this qubit can be decomposed into a linear combination of rotations around the $x$, $y$ and $z$ axes. It is therefore useful to define these rotation operators specifically. The Pauli matrices $\sigma_x$, $\sigma_y$ and $\sigma_z$ give rise to these rotation operators.

**Definition** Let $\theta$ be the angle we want to rotate around our axis then the following

---

\(^{11}\) We used the derivation of the Bloch sphere using half angles $\theta/2$, meaning that two points are also orthogonal $-90^\circ$ apart.
functions define the respective unitary operation:

\[ R_x(\theta) \equiv e^{-i\theta \frac{\sigma_x}{2}} = \cos \frac{\theta}{2} I - i \sin \frac{\theta}{2} \sigma_x = \begin{pmatrix} \cos \frac{\theta}{2} & -i \sin \frac{\theta}{2} \\ -i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix} \] (2.64)

\[ R_y(\theta) \equiv e^{-i\theta \frac{\sigma_y}{2}} = \cos \frac{\theta}{2} I - i \sin \frac{\theta}{2} \sigma_y = \begin{pmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix} \] (2.65)

\[ R_z(\theta) \equiv e^{-i\theta \frac{\sigma_z}{2}} = \cos \frac{\theta}{2} I - i \sin \frac{\theta}{2} \sigma_z = \begin{pmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{pmatrix}. \] (2.66)

Corollary 2.1.3. Any single qubit unitary operation \( U \) is closed under decomposition of \( R_x, R_y \) and \( R_z \) and therefore the set \( S = \{ R_x, R_y, R_z \} \) is functionally complete considering all single qubit unitary operations.

Lemma 2.1.4 (Z-Y decomposition for a single qubit; Theorem 4.1 in [NC00]). Suppose \( U \) is a unitary operation on a single qubit. Then there exist real numbers \( \alpha, \beta, \gamma \) and \( \delta \) such that

\[ U = e^{i\alpha} R_z(\beta) R_y(\gamma) R_z(\delta) \] (2.67)

Proof. Since \( U \) is unitary, the rows and columns of \( U \) are orthonormal, from which it follows that there exist real numbers \( \alpha, \beta, \gamma \) and \( \delta \) such that

\[ U = \begin{pmatrix} e^{i(\alpha - \frac{\beta}{2} + \frac{\gamma}{2})} \cos \frac{\gamma}{2} & -e^{i(\alpha - \frac{\beta}{2} + \frac{\gamma}{2})} \sin \frac{\gamma}{2} \\ e^{i(\alpha + \frac{\beta}{2} - \frac{\gamma}{2})} \sin \frac{\gamma}{2} & e^{i(\alpha + \frac{\beta}{2} + \frac{\gamma}{2})} \cos \frac{\gamma}{2} \end{pmatrix} \] (2.68)

Equation (2.67) now follows immediately from the definition of the rotation matrices and matrix multiplication.

2.1.5 Quantum Algorithms

Even though quantum computers do not exist as of today, numerous algorithms have already been defined by quantum computer scientists. They solve certain problems much faster than any classical computer could. Some of these algorithms merely show the potency of quantum computers by having a much lower computational complexity in solving abstract (mathematical) problems. Others, however, have a significant impact on computer science and information technology if they would ever be successfully implemented on a quantum computer. One algorithm of the latter class, called Shor’s algorithm [Sho97], goes even so far as making the widely used cryptographic scheme RSA obsolete.

The following section describes the Deutsch-Jozsa algorithm [DJ92], which is one of the first examples of a quantum algorithm that is more efficient than any possible
CHAPTER 2. THEORY

classical algorithm. Although this algorithm is of little practical use, it shows the foundations of a quantum algorithm quite elegantly and gives a general idea of how these algorithms are conceived.

The Deutsch-Jozsa algorithm

Consider having a device that computes some function \( f : \{0,1\}^n \to \{0,1\} \), i.e. we input \( n \) bits and receive 1 bit as an answer. For the purpose of the present investigation this device should be thought of as a black box, meaning that we cannot look inside the device to see how it works. The only way to gain information about the function \( f \) is to give it some input \( a \in \{0,1\}^n \) and allow it to output \( f(a) \).

Now, for the Deutsch-Jozsa problem, we are promised that the function computed by the device is either balanced or constant, i.e. either the function returns the same output for all inputs or produces a 0 for half of the input domain and a 1 for the other half as shown in equation (2.69).

\[
\begin{align*}
\text{constant:} & \quad \forall x : f(x) = 0 \quad (\text{or } 1) \\
\text{balanced:} & \quad ||\{x|f(x) = 1\}|| = ||\{x|f(x) = 0\}|| = \frac{n}{2} \tag{2.69}
\end{align*}
\]

In the classical case, for an algorithm to prove that the promised function is either balanced or constant, we need up to \( 2^n - 1 + 1 \) evaluations of different inputs. That is, in the worst case of having a constant function \( f \), we would need to evaluate just over half of the input domain before we can be correct in deciding that this function actually is constant. On the other hand, in the best case of discovering two different outputs with only two different inputs, we can already correctly decide that the function \( f \) is balanced. In case of Deutsch-Jozsa’s quantum algorithm [DJ92], however, an answer is provided that is always correct with only a single evaluation of \( f \). The algorithm is defined as follows:

**Definition** Start with the quantum state \( |0\rangle^\otimes n |1\rangle \), i.e. the first \( n \)-qubits are (classical) zeros and the final qubit is a (classical) one. Next, a Hadamard transform is applied to this quantum state, which results in:

\[
H^{\otimes n+1} |0\rangle^{\otimes n} |1\rangle = \sum_{x \in \{0,1\}^n} \frac{|x\rangle}{\sqrt{2^n}} \cdot \frac{1}{\sqrt{2}} (|0\rangle - |1\rangle) \tag{2.70}
\]

The function \( f \) is implemented as a unitary transform in the quantum oracle \( U_f : |x,y\rangle \to |x, y \oplus f(x)\rangle \), where \( \oplus \) denotes the exclusive-OR operation (addition modulo 2; XOR). Evaluation of the Hadamard-transformed input on the quantum oracle gives
us:

\[ U_f H^{\otimes n+1} |0\rangle^{\otimes n} |1\rangle = \sum_{x \in \{0,1\}^n} \frac{|x\rangle}{\sqrt{2^n}} \cdot \frac{1}{\sqrt{2}} (f(x) - |1 \oplus f(x)\rangle) \]  \hspace{1cm} (2.71)

\[ = \sum_{x \in \{0,1\}^n} (-1)^{f(x)} \frac{|x\rangle}{\sqrt{2^n}} \cdot \frac{1}{\sqrt{2}} (|0\rangle - |1\rangle) \]  \hspace{1cm} (2.72)

The set of qubits in equation (2.72) has the result of the function evaluation stored in the amplitude of the qubit superposition state. We can now infer the terms in superposition by using a Hadamard transform on the first \( n \)-qubits. In order to clearly show the effect of this Hadamard transform, it is useful to first determine the effect of the Hadamard on some state \( |x\rangle \). By checking the cases \( x = 0 \) and \( x = 1 \) separately we see that for a single qubit \( H |x\rangle = \sum_{z \in \{0,1\}} (-1)^{xz} \frac{1}{\sqrt{2}} |z\rangle \) and thus for a \( n \)-qubit state:

\[ H^{\otimes n} |x_1, ..., x_n\rangle = \frac{1}{\sqrt{2^n}} \sum_{z_1, ..., z_n} (-1)^{x_1 z_1 + ... + x_n z_n} |z_1, ..., z_n\rangle \]  \hspace{1cm} (2.73)

\[ H^{\otimes n} |x\rangle = \frac{1}{\sqrt{2^n}} \sum_{z \in \{0,1\}^n} (-1)^{x \cdot z} |z\rangle \]  \hspace{1cm} (2.74)

In equation (2.74) the term \( x \cdot z \) denotes the bitwise inner product of \( x \) and \( z \), modulo 2. Using this equation we can now express our end state as defined in equation (2.75).

\[ H^{\otimes n} U_f H^{\otimes n+1} |0\rangle^{\otimes n} |1\rangle = \frac{1}{2^n} \sum_{z \in \{0,1\}^n} \sum_{x \in \{0,1\}^n} (-1)^{x \cdot z} (-1)^{f(x)} |z\rangle \frac{1}{\sqrt{2}} (|0\rangle - |1\rangle) \]  \hspace{1cm} (2.75)

Now, we are only interested in the amplitudes of the first \( n \)-qubit state and we ignore the last qubit. Note that the amplitude for the state \( |0\rangle^{\otimes n} \) can be deduced from equation (2.75) by choosing \( z = \{0\}^n \), which amounts to:

\[ \frac{1}{2^n} \sum_{x \in \{0,1\}^n} (-1)^{f(x)} \]  \hspace{1cm} (2.76)

Considering our two cases of either a constant function or a balanced function, we
observe:

\[
f(x) \overset{\text{def}}{=} \text{constant} \Rightarrow \frac{1}{2^n} \sum_{x \in \{0,1\}^n} (-1)^{f(x)} = \begin{cases} 
1 & \text{if } f(x) = 0 \\
-1 & \text{if } f(x) = 1
\end{cases}
\]

\[
f(x) \overset{\text{def}}{=} \text{balanced} \Rightarrow \frac{1}{2^n} \sum_{x \in \{0,1\}^n} (-1)^{f(x)} = 0
\]

A quantum state is always of unit length, from which it follows that observing an amplitude of 1 (or -1) for the state \(|0\rangle \otimes n\) results in an amplitude of 0 for all other states and that observing an amplitude of 0 results in a non-zero amplitude for either of the other possible states \(|x\rangle\) with \(x = \{0,1\}^n/\{0\}^n\). Thus, when measuring the first \(n\) qubits, we find \(n\) zeros when \(f\) is constant and anything else when \(f\) is balanced.

As we see, the Deutsch-Jozsa algorithm provides a solution to the stated problem in just a single query. In contrast, any classical algorithm would require over half of the input domain in queries to always be correct. This appears to be impressive, but there are a few strings attached. First of all, the Deutsch-Jozsa problem has no known application; it is just a simplified mathematical problem designed specifically for solving it via quantum computation. Therefore this specific design of the problem leads to an unfair comparison between classical and quantum computation. And lastly, the comparison is based on a deterministic approach. If, for example, one would allow a small margin for error, then the problem could be solved classically with only a handful of queries\(^{12}\).

However, the Deutsch-Jozsa algorithm inspired more impressive quantum algorithms like Shor’s algorithm, and it is enlightening in understanding the principles behind quantum algorithms. This is especially true in that it showcases how a quantum computer is able to process a query task like this in a sort of parallel fashion, i.e. evaluation of the unknown function \(f\) through the use of qubits is like evaluating the whole input domain all at once.

\(^{12}\) Let’s say we randomly query the black box 5 times and receive 5 equal outcomes. In this case, the estimated error in deciding that function \(f\) is constant while it actually is balanced is merely the probability of observing this pattern given that the function actually is balanced, i.e. \(\epsilon = 2 \cdot \left(\frac{1}{2}\right)^5 \approx 0.06\).
2.2 Game Theory

The following section provides an introduction to the field of game theory. The first paragraphs provide some scientific background and the philosophy behind game theory, followed by a description of the prisoner’s dilemma.

2.2.1 Game Theory

John von Neumann, founder of game theory as we know it today, realized that winning a game of poker was not guided by probability theory alone, but more importantly by a player’s rational approach to the behavior of the other players in the game. A simple game of poker can be decided by correctly interpreting the action made by a player. Von Neumann set out to formalize the idea of ‘bluffing’; a strategy of deceiving the other players and hiding information from them. Von Neumann’s first approach to analyzing games was his proof of the minimax (minimizing the maximum possible loss) theorem [vN28]. He believed that his findings in the study of games could be invaluable to the field of economics, which eventually led to the book Theory of Games and Economic Behaviour [vNM44] in cooperation with the economist Oskar Morgenstein. This book laid the foundations for the field of game theory.

By defining a game as a formalized payoff structure, a wide variety of situations involving interactions can be characterized as a game. Game theory, subsequently, hands us a bag of analytical tools in order to understand the phenomena observed when decision-makers interact. It can be defined as a mathematical study of models of conflict and cooperation between intelligent and rational interactors, also called agents (see section 2.3 for a further description of multi-agent structures).

Strategic Games

Games need to be mathematically well-defined in order to study them in a game theoretic context. For the scope of this thesis, we will only be dealing with strategic games. A strategic game consists of a set of players, a set of actions (or strategies) available to these players, and a specification of payoffs for each combination of strategies. These games are usually represented by a matrix showing the players, strategies and payoffs (see table 2.1 for a two player example). More generally, it can be represented by any function that associates a payoff for each player with every possible combination of actions.

Pure and Mixed Strategies

A pure strategy provides a complete definition of how a player will play a game. In particular, it determines the move a player will make for any situation they could
face. A player’s strategy set is the set of pure strategies available to that player.

A mixed strategy is an assignment of a probability to each pure strategy. This allows for a player to randomly select a pure strategy. Since probabilities are continuous, there are infinitely many mixed strategies available to a player, even if their strategy set is finite.

Of course, one can regard a pure strategy as a degenerate case of a mixed strategy, in which that particular pure strategy is selected with probability 1 and every other strategy with probability 0.

**Zero-sum Games**

Initially, game theory has been developed to analyze competitions between individuals in which one individual does better at the expense of another. Almost all classical table games can be characterized as such, e.g. all money won (or lost) in a game of poker comes at the expense of another player losing (or winning). Also board games like chess have one player winning at the expense of the other player. This class of games is called zero-sum games because the total benefit (payoff) for these games always adds up to zero.

John von Neumann proved with his minimax algorithm [vN28] that every zero-sum game has a rational solution, i.e. if one would know all possible action sequences in the game then one can always play optimally by minimizing the maximum possible loss\(^\text{13}\). Many games studied in game theory, however, are non-zero-sum games. This includes the famous prisoner’s dilemma; the game discussed in this thesis.

**Nash-equilibrium and Pareto optimality**

John Nash revolutionized the field of Game Theory with his study of equilibria (or solutions) in non-zero-sum games. His research led to the proposition of the Nash equilibrium [Nas50], which is a solution concept of a game involving two or more players where each player is assumed to know the possible strategies of the other players, and no player has anything to gain by changing only his own strategy unilaterally. To be more specific, if each player has chosen a strategy and no player can benefit by changing his strategy while the other players keep theirs unchanged, then the current set of strategy choices and corresponding payoffs constitutes a Nash equilibrium.

Within a non-zero-sum game, the interests of the players are not completely opposite by definition. In some settings, cooperation between players might be beneficial to all players involved. Generally speaking, it is hard to distinguish a rational solution to these kinds of games. A Nash-equilibrium gives us a possible solution to this kind

\(^{13}\)Nevertheless, this is not very plausible in most games as they are simply too complex. In chess, for example, even super-computers can’t calculate all possible action sequences. Simple games, however, have been solved in this way, e.g. tic-tac-toe.
of situations as it proposes a choice of strategies in such a way that no player regrets his chosen action. It is proven that each non-zero-sum game possesses at least one Nash equilibrium, either in a pure or in a mixed strategy.

However, some Nash equilibria can be intuitively suboptimal as they are not Pareto optimal, i.e. the set of strategies which yields the highest possible \textit{summed} payoff. That is, some games incorporate a Pareto optimal set of strategies from which a player can unilaterally change his strategy in order to receive a higher \textit{individual} payoff at the expense of another player. This notion is especially important to the prisoner's dilemma, which will be discussed next.

\subsection{Prisoner’s Dilemma}

The prisoner’s dilemma is one of the most studied games in the literature of Game Theory. The story behind this game can be described as follows [Mye91]. The two players, involved in this game, are accused of conspiring in two crimes, one minor crime for which their guilt can be proved without any confession, and one major crime for which they can be convicted only if at least one confesses. The prosecutor promises that, if exactly one confesses, the confessor will go free now but the other will go to jail for 6 years. If both confess, then both go to jail for 5 years. If neither confesses then they will go both to jail for only 1 year. So, each player has two options; either to confess or not to confess. Generally these choices are called Defect (D) or Cooperate (C) respectively, the naming stems from its relativity to the other player (not confessing is cooperating with your partner-in-crime). Schematically this scenario yields Table 2.1.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
\textbf{player A} & \textbf{C} & \textbf{D} \\
\hline
\textbf{C} & -1,-1 & -6,0 \\
\textbf{D} & 0,-6 & -5,-5 \\
\hline
\end{tabular}
\caption{Example Prisoner’s Dilemma}
\end{table}

This specific dilemma can be summarized in a generalized form in order to expose the skeleton of the game. First of all, instead of talking about specific payoffs (number of years of prison sentence) these can be categorized in \textit{Reward for mutual cooperation} (\(R\)), \textit{Temptation to defect} (\(T\)), \textit{Punishment for mutual defection} (\(P\)), and the \textit{Sucker’s payoff} (\(S\)). This result schematically in Table 2.2.

In order for this game to be a \textit{Prisoner’s Dilemma}, the following inequalities most
Table 2.2: Generalized Prisoner’s Dilemma

<table>
<thead>
<tr>
<th></th>
<th>player B</th>
</tr>
</thead>
<tbody>
<tr>
<td>player A</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>R,R</td>
</tr>
<tr>
<td>D</td>
<td>T,S</td>
</tr>
</tbody>
</table>

hold true:

\[
T > R > P > S \quad \text{and} \quad R > \frac{T + S}{2} \quad (2.77) \\
\]

In general any game satisfying (2.77) and (2.78) is called a Prisoner’s Dilemma. The first inequality, equation 2.77, illustrates the payoff preference for each player and the second inequality, equation 2.78, makes sure that in repeated play it is not rewarding to alternately betray and be betrayed.

Analyzing the Prisoner’s Dilemma

The prisoner’s dilemma is designed to highlight irrational equilibria and to analyze their consequences. Upon closer observation of the payoff matrix in table 2.1, we distinguish a Pareto optimal outcome of mutual cooperation. That is, given the payoff structure proposed in the prisoner’s dilemma, the most beneficial outcome for the two criminals together is that of a situation in which neither of the criminals confesses. However, when one of the criminals already knows that the other criminal is not going to confess then he can improve his payoff by confessing, i.e. defecting. The other criminal, wanting to avoid this risk, can thus conclude that it might be better to also defect. This situation of both players defecting is the Nash equilibrium in this game as both players cannot unilaterally change their strategy from defection to cooperation without negatively effecting their payoff.

This specific difference in the Pareto optimal strategy choice versus the Nash equilibrium constitutes the dilemma in the game.
2.3 Multi-Agent Learning

Multi-agent learning, the intersection of Multi-agent Systems and Machine Learning, is a framework used to describe the problem of learning from interaction in order to achieve a certain goal. The learner and decision maker in this setting is called the *agent*; in case of Multi-agent learning we have multiple agents. The following section will give a general description of reinforcement learning [SB98], commonly used in Multi-agent learning, and a detailed description of a specific reinforcement learning algorithm, namely the Conditional Joint Action Learner [BS07].

2.3.1 Reinforcement Learning

Reinforcement learning is a term used to describe a (Machine) learning process, where there is no predefined mapping of inputs (questions) to outputs (answers). It is based on an *agent* interacting with an environment (with or without other agents), where the *agent* can take an action resulting in some outcome. Every outcome brings a (numerical) reward, describing how much this outcome is wanted.

Specifically, the agent interacts with its environment at each of a sequence of discrete time steps, $t = \{0, 1, 2, 3, \ldots\}$. At each time step $t$, the agent receives some representation of the environment called the *state*, $s_t \in S$, where $S$ is the set of possible states. On the basis of this state $s_t$, the agent selects an *action* $a_t \in A(s_t)$, where $A(s_t)$ is the set of possible actions available in state $s_t$. As a consequence of its action, the agent receives a reward $r_{t+1} \in R$, where $R$ is the set of rewards associated with the combination of a specific action in a certain state, and finds itself in a new state $s_{t+1}$. The idea behind the reward is that the actions which yield a favorable result in a specific state should be strengthened, i.e. associated with a high reward thereby *reinforcing* this action, and the actions with unfavorable results should be weakened. This agent-environment interaction is shown schematically in figure (2.3).

At each time step, the agent must decide on a mapping from states to actions, generally described by probabilities. This mapping is called the agent’s *policy* and is denoted by $\pi_t : S \rightarrow A$, where $\pi_t(s, a)$ is the probability of selecting action $a$ in state $s$ at time step $t$. The policy $\pi_t$ of the agent is based on the choice of actions and corresponding rewards it received in the previous time steps $< t$ and is updated every time step. The agent’s goal is to decide on a mapping $\pi$ which maximizes the amount of reward it receives, i.e. maximizes expected payoff.

One of the main characteristics of reinforcement learning is the lack of knowledge of the agent at the start of the learning process. The agent has knowledge of all its possible actions, but is uncertain of the outcome of these actions. General reinforcement learning is therefore characterized by an *explorative* phase, a phase in
which the agent (randomly) tries different actions in different states in order to gain knowledge about which action are profitable. That is, the agent receives feedback on its action by receiving the associated reward and as a result reinforces this specific action proportional to this reward. This information can subsequently be used in an exploitative phase, where the agent selects its actions based on the expected reward of selecting a particular action in a specific state.

Generally speaking, an agent within the reinforcement learning framework is faced by the dilemma of exploration versus exploitation. After some exploration the agent might have found a set of apparently rewarding actions. However, how can the agent be sure that these actions are actually best? Hence, when should an agent continue to explore and when should it just exploit its current knowledge? Therefore, a reinforcement learning algorithm should address this dilemma by allowing the agent to sufficiently explore the action-state space. This can be done by letting the explorative phase last a specific number of time steps or by more sophisticated strategies like reducing random behavior over time.

The environment within a reinforcement learning framework is typically formulated as a Markov Decision Process (henceforth MDP); a discrete time stochastic control process characterized by a set of states. The transition from a state $s$ to the next state $s'$ is expressed by a probabilistic function $P_{ss'}^a$, which fully describes the probability of ending up in state $s'$ after selecting strategy $a$ while being in state $s$. All Markov Decision Processes possess the Markov property, meaning that transition from one state to another only depends on the current state and the action selected in this state; it is independent of any previously encountered states or actions.

Although this Markov property makes it easier for a reinforcement learning algorithm to operate, it is not a necessity. For example, reinforcement learning algorithms have been put to good use in the Iterated Prisoner’s Dilemma [SC96] which is arguably not a MDP because playing defection during one of the earlier stages of the
2.3. MULTI-AGENT LEARNING

iteration may have its effect in a (much) later state and should not be ignored by a sophisticated learning algorithm.

The next section will describe a specific reinforcement learning algorithm, designed for playing the iterated prisoner’s dilemma, followed by a concrete example of this algorithm in practice.

2.3.2 Conditional Joint Action Learning

This section presents the details of the Conditional Joint Action Learner, henceforth CJAL [BS07], specifically for the classical Prisoner’s Dilemma.

Consider a set of agents $S = \{Alice, Bob\}$ where each agent $i \in S$ has a fixed set of actions $A_i = \{C, D\}$ and repeatedly plays the Prisoner’s Dilemma. In every iteration each agent chooses an action $a_i \in A_i$. The expected utility of an agent $i$ at iteration $t$ for an action $a_i$ is denoted as $E^i_t(a_i)$.

The introduction of some notations and definitions is needed to build the framework for CJAL learning. The probability that agent $i$ plays action $a_i$ at iteration $t$ is denoted as $Pr^i_t(a_i)$ and the conditional probability that the other agent, $j$, will play $a_j$ given that the $i$th agent plays $a_i$ at iteration $t$ as $Pr^i_t(a_j|a_i)$. The joint probability of an action pair $(a_i, a_j)$ (the probability of two actions selected in conjunction) at iteration $t$ is given by $Pr_t(a_i, a_j)$. Each agent $i$ maintains a history $H^i_t$ of interactions at every iteration $t$ as defined in equation (2.79). Typically, the history is represented as a matrix.

$$ H^i_t = \bigcup_{a_i \in A_i, a_j \in A_j} n^i_t(a_i, a_j) $$

(2.79)

Here, $n^i_t(a_i, a_j)$ denotes the number of times agent $i$ has played action $a_i$ while agent $j$ has played action $a_j$ from the beginning up to iteration $t$. The number of times agent $i$ has played action $a_i$ up to iteration $t$ is equal to the sum of all interactions with agent $j$ so far, as defined in equation (2.80).

$$ n^i_t(a_i) = \sum_{a_j \in A_j} n^i_t(a_i, a_j) $$

(2.80)

**Definition** A CJAL learner is an agent $i$ who at any iteration $t$ chooses an action $a_i \in A_i$ with a probability proportional to $E^i_t(a_i)$ according to equation (2.81) where $U^i_t(a_i, a_j)$ denotes the utility (or pay-off) of agent $i$ when action pair $(a_i, a_j)$ is played:

$$ E^i_t(a_i) = \sum_{a_j \in A_j} U^i_t(a_i, a_j) \cdot Pr^i_t(a_j|a_i) $$

(2.81)
Using the definition of conditional probability, equation (2.81) can be rewritten as equation (2.82):

\[
E_i(a_i) = \sum_{a_j \in A_j} U_i(a_i, a_j) \cdot \frac{Pr^i_t(a_i, a_j)}{Pr^i_t(a_i)}
\]  

(2.82)

The probability of an event can be approximated by the fraction of times the event has occurred in the past and thereby \(E_i(a_i)\) is approximated by \(\tilde{E}_i(a_i)\) as denoted in equation (2.83):

\[
\tilde{E}_i(a_i) = \sum_{a_j \in A_j} U_i(a_i, a_j) \cdot \frac{n^i_{t-1}(a_i, a_j)}{n^i_{t-1}(a_i)}
\]  

(2.83)

Unlike a JAL (Joint Action Learner), a CJAL does not assume the probability of the others player’s action to be independent of its own action. A CJAL agent learns the correlation between its actions and the other agent’s actions and uses the (estimated) conditional probabilities instead of marginal probabilities to calculate the expected utility of an action. Therefore, a CJAL splits the marginal probabilities of an action \(a_j\) taken by the other player into conditional probabilities and considers them as the probability distribution associated with the joint action event \((a_i, a_j)\).

An intuitive reason behind this choice of probability distribution can be obtained by considering each agent’s point of view. If each agent views the simultaneous move game as a sequential move game where it is the first one to move, then to calculate the expected utility of its action, it must try to find the probability of the other player’s action given its own action. This is basically the conditional probability described above.

**Conditional Joint Action Learning in practice**

Banerjee and Sen [BS07] use two distinct exploration phases in CJAL. The first phase assumes that the agents explore actions randomly for \(N\) initial interactions and it is followed by the second phase where the agent uses an \(\epsilon\)-greedy exploration. In this \(\epsilon\)-greedy exploration phase, an agent chooses the action associated with the highest expected utility (see equation (2.83)) with a probability \(1 - \epsilon\) and explores other actions randomly with probability \(\epsilon\). To summarize, the probability at iteration \(i\) that agent \(i\) will choose action \(a_i\), namely \(Pr^i_t(a_i)\) is given in equation (2.84).

\[
Pr^i_t(a_i) = \begin{cases} 
\frac{1}{|A_i|} & \text{if } t < N \\
1 - \epsilon & \text{if } t > N \text{ and } a_i = a_{\text{max}} \\
\epsilon & \text{if } t > N \text{ and } a_i \neq a_{\text{max}} \end{cases}
\]  

(2.84)
2.3. MULTI-AGENT LEARNING

Here, $a_{\text{max}}$ is the action with the highest associated utility as defined in equation (2.85).

$$a_{\text{max}} = \arg \max_{a_i \in A_i} E_{i-1}^i(a_i)$$ (2.85)

The dynamics of the CJAL mechanism can intuitively be analyzed in self-play. Consider two CJAL agents playing the Prisoner’s Dilemma. The following paragraph is an analytical prediction of the sequence of actions the agents would take with time.

**Example** In the Prisoner’s Dilemma game, the possible actions are defined as $A_i = \{C, D\}$ and the agents as $i = \{Alice, Bob\}$. The utilities are denoted as $U_i(C, C) = R$, $U_i(C, D) = S$, $U_i(D, C) = U_i(D, D) = T$ and $U_i(D, D) = P$. In the exploration phase, both agents choose their actions randomly from a uniform distribution. Therefore, the expected number of occurrences of each outcome will be $\frac{N}{2}$ after $N$ iterations. Also, the expected number of times an agent would play each of its two actions is $\frac{N}{2}$. So, with a sufficiently large $N$, the expected values of the conditional probabilities for both players will be equal, as in:

$$\forall i : Pr_{i}^N(C|C) = Pr_{i}^N(C|D) = Pr_{i}^N(D|C) = Pr_{i}^N(D|D) = \frac{1}{2}$$

From this notion, it can be expected that the expected utility of the two actions after $N$ iterations will converge accordingly (equations (2.86) and (2.87)).

$$E_{i}^N(C) = U_i(C, C) \cdot Pr_{i}^N(C|C) + U_i(C, D) \cdot Pr_{i}^N(D|C) = \frac{R + S}{2}$$ (2.86)

$$E_{i}^N(D) = U_i(D, C) \cdot Pr_{i}^N(C|D) + U_i(D, D) \cdot Pr_{i}^N(D|D) = \frac{T + P}{2}$$ (2.87)

Based on the constraint on the payoffs in an actual Prisoner’s Dilemma game, it should be noted that $E_{i}^N(D) > E_{i}^N(C)$, i.e., defection is preferable to cooperation. Therefore, if both agents choose their action greedily ($\epsilon = 0$), they will choose to play Defect. They will continue to play Defect at iteration $t$ as long as $E_{i}^t(D) > E_{i}^t(C)$, however continuously playing Defect will decrease the probability $Pr_{i}^t(C|D)$ (the other agent choosing cooperating while one keeps defecting) and increase the probability $Pr_{i}^t(D|D)$ (the other agent also choosing defect). This change in probabilities will affect the expected utility of defecting (equation (2.87)) negatively; it will decrease towards $P$. So, during one point in the iteration the expected utility of defecting will be less than the expected utility of cooperating and both players will switch to the strategy Cooperate. While doing so the agents receive the reward payoff $R$.

---

14 These payoffs are generally referred to as the reward ($R$), the sucker’s payoff ($S$), the temptation ($T$) and the punishment ($P$) respectively.

15 The actual dilemma only exists if the payoffs are constrained by $R + S > 2P$. 


reaffirming the expected payoff for cooperating and remaining in the situation where \( E_i(C) > E_i(D) \). This analytical outcome is, however, based on the assumption that after \( N \) iterations each outcome will occur approximately \( \frac{N}{4} \) times.

This scenario is not so simple for the case \( \epsilon > 0 \). Banerjee and Sen [BS07] have shown in their experiments that convergence to mutual cooperation is achieved with \( \epsilon = 0 \), but when \( \epsilon = \frac{1}{10} \) convergence to the Pareto Optimal outcome \((C, C)\) cannot be achieved.
2.4 Quantum Game Theory

There is a common link between quantum computing and game theory: they are both concerned with information. On the one hand, quantum computation processes information that is represented by qubits, while on the other hand, game theory analyzes the exchange of information between players in a game. Quantum game theory recasts classical game theory using quantum probability amplitudes (see equation (2.8)). It is a study of the effects of quantum superposition, interference and entanglement on the optimal strategies of players in a game.

Identifying strategic moves with quantum operations appears to be fruitful in at least two ways. First of all, several proposed applications of quantum information theory can already be conceived as competitive situations where several parties with possibly opposed motives interact. For example, quantum teleportation [NC00, PV98] can be defined as a game between two players, where they share the common goal to communicate the unknown state of a qubit without destroying it, i.e. measuring it. Similarly, quantum cloning has been formulated as a game between two players [SIGA05]. In this same context, quantum cryptography [BBB+92] can be regarded as a game between an eavesdropper and the sender. Secondly, a generalization of the theory of decisions into the domain of quantum probabilities seems interesting, as the roots of game theory are partly in probability theory. In this context it is of interest to investigate what solutions are attainable if superpositions of strategies are allowed for.

On a further note, classical game theory does not explicitly concern itself with how the information is transmitted once a decision is taken. Yet, it should be clear that the practical implementation of any (classical) game inevitably makes use of the exchange of information via some medium. In the prisoner’s dilemma, for example, the two parties have to communicate their decision to an impartial third party in order to retain unconditional decision making. Keeping in mind that a game is also about the transfer of information, it becomes legitimate to ask what happens if these carriers are taken to be quantum systems, quantum information being a fundamental notion of information.

2.4.1 Quantum Penny-flip Game

Mathematician David Meyer was first to introduce the concept of a quantum game and to define a specific quantum strategy in this game [Mey99]. He considered a simple penny-flip game and extended it to a quantum game. Let’s first consider this classical penny-flip game.

**Penny-flip game** Two players, who we call Alice and Bob for convenience, are playing a game involving a single penny. First, Alice prepares the penny, putting heads
up, and hides it in a closed box. Then the game begins with Bob, unaware of
the outcome of Alice’s action, who decides whether he wants to flip the coin
or not. Next move is for Alice, not knowing Bob’s action, who can also decide
whether she wants to flip the coin or not. Bob has the last move and is able to
decide if he wants to flip the coin a last time or not. The box is opened and
Alice wins when tails shows up and Bob wins when heads shows up. This is
a so called zero-sum game [Mye91], which means the gain of one player equals
the loss of the other player(s). This is summarized in table 2.3. Considering

<table>
<thead>
<tr>
<th>Alice’s Actions</th>
<th>Bob’s Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>No-flip, No-flip</td>
<td>-1</td>
</tr>
<tr>
<td>No-flip, Flip</td>
<td>1</td>
</tr>
<tr>
<td>Flip, No-flip</td>
<td>1</td>
</tr>
<tr>
<td>Flip, Flip</td>
<td>-1</td>
</tr>
</tbody>
</table>

Table 2.3: Payoff matrix for Alice in a penny-flip game

this table Alice tries to determine a strategy. Suppose she doesn’t flip the coin,
then if Bob flips the coin exactly once, she will win. On the other hand, if Bob
doesn’t flip the coin or does so twice, she will lose. Similarly, if she does flip
the coin, she only wins if Bob doesn’t flip the coin or does so twice. Thus this
penny-flip game has no deterministic solution and no Nash Equilibrium; there is
no pair of pure strategies, one for each player, such that neither player improves
his result by changing his strategy while the other does not. There is, however,
a probabilistic solution; the pair of mixed strategies consisting of Alice flipping
the penny with probability \( \frac{1}{2} \) and Bob playing each of his four possibilities with
probability \( \frac{1}{4} \) each. The expected payoff in this case is 0 for both players, which
is a probabilistic Nash Equilibrium [Mye91].

As one might have guessed, the example above is already prepared for extension
to a quantum game. The trick is quite simple: just substitute the penny in the
example above by a qubit and define heads as the zero-state (\(|0\rangle\)) and tails as the
one-state (\(|1\rangle\)). The penny was hidden in a closed box and it could be either heads
up or tails up: a superposition of states. This superposition only collapses when we
‘measure’ the qubit by observing the actual state of the penny. We already know that
operations on qubits should be unitary, we therefore still need to define the ‘flip’- and
‘no-flip’ actions. No-flip is trivial; just apply the identity matrix. Flipping the penny
is done by applying one of the Pauli spin matrices (namely \( \sigma_x \)), as noted in equation
2.88.

\[
\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}
\] (2.88)
Defining this penny-flip game as a quantum game has some very specific consequences. That is, we can now ‘cheat’ by allowing non-deterministic quantum strategies. This is illustrated in the example below where Bob ‘cheats’ by adopting a quantum strategy.

A quantum strategy Let’s assume that Alice is unaware of Bob’s new quantum strategy and still prepares the penny (qubit from now on) by putting heads up: we start with a quantum zero-state $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Bob is now allowed to act first and instead of flipping he applies the Hadamard transform, which results in (2.89).

$$H |0\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$ (2.89)

Alice, still playing deterministically, decides to flip the coin, which results in (2.90).

$$\sigma_x \left( \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$ (2.90)

That is, whatever Alice decides to do nothing happens! Now Bob can again perform the Hadamard as his last action, which always results in the winning state for Bob as is shown in (2.91).

$$H \left( \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right) = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = |0\rangle$$ (2.91)

On the surface, this example looks like cheating as Bob is allowed some deviating move, which gives him an expected payoff of 1. We can think of his move as putting the penny standing on edge (neither heads nor tails) and flipping a penny won’t change it - it will still be on his edge. After Alice’s move, Bob can then reverse his action and put the penny heads up.

Naturally, we should ask ourselves what would happen if both players would adopt quantum strategies. Consider the same quantum penny-flip game but with both players allowed any unitary operation on the quantum penny. As it turns out, we now have the same problem as with the classical game: there is no solution to this game denoted by pure quantum strategies and also no Nash equilibrium. Or to generalize:

**Theorem 2.4.1.** A two-person zero-sum game need not have a (quantum,quantum) Nash equilibrium [Mey99].

**Proof.** Consider an arbitrary pair of quantum strategies $([U_A], [U_{B1}, U_{B2}])$, that is Bob first plays $U_{B1}$ then Alice plays $U_A$ followed by Bob’s last move $U_{B2}$. The game starts
with the penny heads up denoted as $|0\rangle$ and the outcome of the game can be denoted by a final state $U_{B2}U_AU_{B1}|0\rangle$. Now suppose that Bob doesn’t win with probability 1, that is $U_{B2}U_AU_{B1}|0\rangle \neq |0\rangle$. Then Bob can improve his expected payoff (to 1) by changing his strategy, replacing $U_{B2}$ with $U_{B1}^\dagger U_A^\dagger$ which is also a unitary operation since $U_{B1}$ and $U_A$ are, and any unitary operation is a valid quantum strategy. This results in the following outcome:

$$U_{B2}U_AU_{B1}|0\rangle = U_{B1}^\dagger U_A^\dagger U_A U_{B1}|0\rangle = U_{B1}^\dagger U_{B1}|0\rangle = |0\rangle \quad (2.92)$$

Similarly, Alice can win with probability 1 if she would change her strategy to $U_{B2}^\dagger \sigma_x U_{B1}^\dagger \text{ag}$. 

$$U_{B2}U_AU_{B1}|0\rangle = U_{B2}U_{B2}^\dagger \sigma_x U_{B1}^\dagger U_{B1}|0\rangle = \sigma_x |0\rangle = |1\rangle \quad (2.93)$$

Since $U_{B2}U_AU_{B1}|0\rangle$ cannot equal both $|0\rangle$ and $|1\rangle$, at least one of the players can improve his expected payoff by changing his/her strategy while the other does not. Thus $([U_A], [U_{B1}, U_{B2}])$ cannot be an equilibrium for any pure quantum strategy and therefore this game has no Nash equilibrium.

Along these same lines we can construct mixed quantum strategies. Allowing mixed strategies results in a probabilistic solution of this quantum game similarly to the classical one. That is, even though we change the rules of the game quite radically, game-theoretic analysis yields similar results.

### 2.4.2 Non-local games

Non-local games are constructed with two or more players who are cooperating in order to reach a common goal. The game is run by a referee and all communication in the game is between the players and the referee. The referee in question (randomly) selects a ‘question’ for each player and sends them to the appropriate player. The task of the players consist of finding an ‘answer’ to the question and send it back to the referee. The referee determines whether the players win or lose, based on the answers he received.

This kind of games is non-local in the sense that the players are unable to (directly) communicate with each other. However, through the use of entanglement, the players are able to outperform any classical strategy in a similar setting.

The following game is directly derived from the Bell experiment by Clauser et al.[CHSH69] as described in section 2.4.2. The description of the CHSH game is adopted from John Watrous’ lecture notes\textsuperscript{16} and from Cleve et al. [CHTW04]. The

\textsuperscript{16}see www.cs.uwaterloo.ca/ watrous/lecture-notes.html
review article by Buhrman et al. [BCMdW] also offers a comprehensive account of non-local games and communication complexity.

**The CHSH game** Consider two players, Alice and Bob, playing a non-local game. A referee in this game will ask both players a question to which they have to respond with an answer. The question as well as the answer is represented by a single bit.

The referee chooses a two bit string $rs$ uniformly from the set \{00, 01, 10, 11\}, sends $r$ to Alice and $s$ to Bob. The answers must be bits: $a$ from Alice and $b$ from Bob. The game is won if $a \oplus b = r \land s$, i.e. the logical and of questions $r$ and $s$ must be equal to the sum of answers $a$ and $b$ modulo 2. The winning conditions are represented in table 2.4. In order to determine the maximum probability of winning we first consider deterministic strategies, where each answer is a function of the question received and no randomness is used by the players. Let $a_0$ and $a_1$ be the answers Alice gives to questions $r = 0$ and $r = 1$ respectively. Similarly $b_0$ and $b_1$ are the deterministic strategies for Bob. Consequently, the winning conditions can be expressed by four equations:

- $a_0 \oplus b_0 = 0$,
- $a_0 \oplus b_1 = 0$,
- $a_1 \oplus b_0 = 0$,
- $a_1 \oplus b_1 = 1$. (2.94)

It is impossible to satisfy all four equations simultaneously, since summing both the left-hand side and the right-hand side modulo 2 yields $0 = 1$: a contradiction. This means it is not possible for a deterministic strategy to win every time, but we can satisfy any three of these four equations, so the probability of winning can at most be $\frac{3}{4}$. Probabilistic strategies are nothing more than a probability distribution over some set of deterministic strategies and therefore can not exceed a winning chance of $\frac{3}{4}$ either.

---

17 For example, strategy $a_0 = a_1 = b_0 = b_1 = 1$ wins in $\frac{3}{4}$ of the possible cases.
Now, if we consider the same problem except that Alice and Bob are each supplied with a qubit from the quantum state $|\varphi\rangle$ in equation (2.95).

$$|\varphi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$ (2.95)

Alice and Bob each decide on two observables\(^{18}\) for projective measurements on their respective qubit. The observables for Alice are defined in equation (2.96) and for Bob in equation (2.97).

$$A_0 = \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$
$$A_1 = \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$ (2.96)

$$B_0 = \frac{1}{\sqrt{2}}(\sigma_z + \sigma_x) = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$
$$B_1 = \frac{1}{\sqrt{2}}(\sigma_z - \sigma_x) = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$ (2.97)

That is, Alice does a projective measurement on her qubit for observable $A_0$ if her question $r = 0$ and $A_1$ if her question is $r = 1$. The same goes for Bob’s question $s \in \{0, 1\}$. The outcome of their measurements is always $\pm 1$, so we will identify an outcome of $-1$ with an actual answer of 0 and an outcome of 1 with actual answer 1.

For the questions $r, s = \{00, 01, 10\}$ the outcome of their measurements must agree and for question $r, s = \{11\}$ the outcome of their measurement must disagree in order to win the game. When the measurement outcomes agree, the product of the outcome is equal to 1 and if they disagree the outcome is equal to $-1$. We can find a probability of this by looking at the expected value of

---

\(^{18}\)An observable is a Hermitian operator with eigenvalues \{+1, −1\} and corresponds to a projective measurement with respect to the two eigenspaces, see section 2.1.2.
2.4. QUANTUM GAME THEORY

For $A, B$, it turns out that:

\[
E(A_0 B_0) = \langle \varphi | A_0 \otimes B_0 | \varphi \rangle = \frac{1}{\sqrt{2}}
\]

\[
E(A_0 B_1) = \langle \varphi | A_0 \otimes B_1 | \varphi \rangle = \frac{1}{\sqrt{2}}
\]

\[
E(A_1 B_0) = \langle \varphi | A_1 \otimes B_0 | \varphi \rangle = \frac{1}{\sqrt{2}}
\]

\[
E(A_1 B_1) = \langle \varphi | A_1 \otimes B_1 | \varphi \rangle = -\frac{1}{\sqrt{2}}.
\]

From these expected values, we can calculate probabilities:

\[
Pr(A_0 = B_0) = Pr(A_0 = B_1) = Pr(A_1 = B_0) = Pr(A_1 \neq B_1) = \frac{1}{2} + \frac{1}{2\sqrt{2}}.
\]

Therefore, the probability of satisfying any of the four winning conditions from equations (2.94) is $\frac{1}{2} + \frac{1}{2\sqrt{2}} \approx 0.853$ each. This forms significant improvement over the classical win probability of $\frac{3}{4}$.

The CHSH game is a typical example of a non-local game of interest to the field of quantum communication complexity. Strategies in this kind of games have applications in for example interactive proof systems [CHTW04], but are outside the scope of this thesis.
2.5 Quantum Prisoner’s Dilemma

The set up for the Quantum Prisoner’s dilemma is as follows: two players, Alice and Bob, play the game. The choice of action is modelled by a single qubit each. This qubit can be a classical 0, representing cooperate, a classical 1, representing defect, or a superposition of both. The game starts with the creation of a qubit for each player in the classical zero state: $|0\rangle$. The starting state of the game becomes:

$$|0_a\rangle \otimes |0_b\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$ (2.98)

Here, the vector $(1, 0, 0, 0)$ reflects the amplitudes of the possible combinations of values for the two bits, i.e., $|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$. So, initially only the state $|00\rangle$ has an amplitude and therefore measuring these qubits will result in the classical state $00$ with probability $|1|\sqrt{2}|^2 = 1$.

The next step is to bring these two qubits in full entanglement by performing unitary operation $J$ on the two-qubit state, yielding the following result:

$$J = \frac{1}{\sqrt{2}} (I + i\sigma_x \otimes \sigma_x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & i \\ 0 & 1 & i & 0 \\ 0 & i & 1 & 0 \\ i & 0 & 0 & 1 \end{pmatrix}$$ (2.99)

$$J |00\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & i \\ 0 & 1 & i & 0 \\ 0 & i & 1 & 0 \\ i & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ i \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ i \end{pmatrix} = \frac{1}{\sqrt{2}} (|00\rangle + i |11\rangle)$$ (2.100)

After performing operation $J$, these qubits are in full entanglement in the sense that measurement of these qubits will result in $00$ with probability $|\frac{1}{\sqrt{2}}|2 = \frac{1}{2}$ and $11$ with probability $|\frac{1}{\sqrt{2}}i|^2 = \frac{1}{2}$ and measuring one qubit automatically results in knowledge of the exact state of the other qubit.\footnote{This two qubit state is commonly known as an EPR-pair, named after a famous paper by Einstein, Podolsky and Rosen who were the first to point out this strange correlation between qubits.}

Both players receive one of these qubits and are allowed to perform a unitary action on them. Alice and Bob are allowed any (unitary) quantum operation on a
single qubit, which will be defined by:

\[
U(\theta, \phi, \alpha) = \begin{pmatrix}
    e^{-i\theta} \cos \frac{\theta}{2} & e^{i\phi} \sin \frac{\theta}{2} \\
    e^{-i\phi} \sin \frac{\theta}{2} & e^{i\theta} \cos \frac{\theta}{2}
\end{pmatrix}
\] (2.101)

We can model these actions on the starting state in the following way:

\[
(U_a \otimes U_b)J |00\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix}
    \cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} e^{i(\phi_a + \phi_b)} & \cos \frac{\theta_a}{2} \sin \frac{\theta_b}{2} e^{i(\phi_a - \phi_b)} & \sin \frac{\theta_a}{2} \cos \frac{\theta_b}{2} e^{i(\phi_a - \phi_b)} & \sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} e^{i(\phi_a + \phi_b)} \\
    -\cos \frac{\theta_a}{2} \sin \frac{\theta_b}{2} e^{i(\phi_a + \phi_b)} & -\sin \frac{\theta_a}{2} \cos \frac{\theta_b}{2} e^{i(\phi_a - \phi_b)} & -\sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} e^{i(\phi_a - \phi_b)} & -\cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} e^{i(\phi_a + \phi_b)} \\
    -\sin \frac{\theta_a}{2} \cos \frac{\theta_b}{2} e^{-i(\phi_a + \phi_b)} & -\sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} e^{-i(\phi_a - \phi_b)} & -\cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} e^{-i(\phi_a - \phi_b)} & -\sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} e^{-i(\phi_a + \phi_b)} \\
    \sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} e^{-i(\phi_a + \phi_b)} & \cos \frac{\theta_a}{2} \sin \frac{\theta_b}{2} e^{-i(\phi_a - \phi_b)} & \cos \frac{\theta_a}{2} \sin \frac{\theta_b}{2} e^{-i(\phi_a - \phi_b)} & \cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} e^{-i(\phi_a + \phi_b)}
\end{pmatrix}
\] (2.102)

The final two steps of the game involve performing a disentanglement operation \( J^\dagger \), the adjoint of \( J \), followed by measuring the two qubits. A single execution of the game can thus be described as:

\[
J^\dagger(U_a \otimes U_b)J |00\rangle = \begin{pmatrix}
    -\sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \sin (\phi_a + \phi_b) + \cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \cos (\phi_a + \phi_b) \\
    i(\cos \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \sin (\phi_a + \phi_b) + \sin \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \cos (\phi_a + \phi_b)) \\
    i(\cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \cos (\phi_a + \phi_b) + \sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \sin (\phi_a + \phi_b)) \\
    \sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \cos (\phi_a + \phi_b) - \cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \sin (\phi_a + \phi_b)
\end{pmatrix}
\] (2.103)

This yields the following probabilities for measuring the states 00, 01, 10 and 11 respectively:

\[
Pr(|00\rangle) = \left( \cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \cos (\phi_a + \phi_b) - \sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \sin (\phi_a + \phi_b) \right)^2
\] (2.104)
\[ Pr(|01\rangle) = \left( \cos \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \sin (\phi_a + \alpha_b) + \sin \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \cos (\alpha_a + \phi_b) \right)^2 \]  
(2.105)

\[ Pr(|10\rangle) = \left( \cos \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \cos (\phi_a + \alpha_b) + \sin \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \sin (\alpha_a + \phi_b) \right)^2 \]  
(2.106)

\[ Pr(|11\rangle) = \left( \cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \sin (\phi_a + \phi_b) - \sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \cos (\alpha_a + \alpha_b) \right)^2 \]  
(2.107)

One should note that these equations differ from the paper [CHH06], where this same game is described\(^{20}\).

We can use these equations to deduce perfect counter-strategies for Bob given a specific action by Alice. These counter-strategies can be represented in linear equations as follows\(^{21}\):

**Outcome 00** For Bob, in order to reach the state of mutual cooperation with probability 1, he must play either of the following:

\[
\begin{pmatrix}
\theta_b \\
\phi_b \\
\alpha_b
\end{pmatrix} = \begin{pmatrix}
\theta_a \\
\phi_a \\
\alpha_a
\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}
-\theta_a \\
-\phi_a \\
-\alpha_a
\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}
\theta_a \\
\pi - \phi_a \\
\frac{1}{2} \pi - \alpha_a
\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}
-\theta_a \\
\pi - \phi_a \\
-\frac{1}{2} \pi - \alpha_a
\end{pmatrix}
\]  
(2.108)

**Outcome 01** For Bob, in order to reach the state of Alice cooperating while he defects with probability 1, he must play either of the following:

\[
\begin{pmatrix}
\theta_b \\
\phi_b \\
\alpha_b
\end{pmatrix} = \begin{pmatrix}
\theta_a + \pi \\
\pi - \phi_a \\
-\phi_a - \frac{1}{2} \pi
\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}
\pi - \theta_a \\
-\phi_a + \frac{1}{2} \pi \\
\phi_a - \frac{1}{2} \pi
\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}
\pi - \theta_a \\
\pi - \phi_a \\
-\phi_a + \frac{1}{2} \pi
\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}
\pi + \theta_a \\
-\phi_a + \frac{1}{2} \pi \\
\phi_a - \frac{1}{2} \pi
\end{pmatrix}
\]  
(2.109)

\(^{20}\)These conclusions have been drawn from computations with the MatLab and Maple software packages using the very same operators as defined in the original paper. Calculations proved that the concluding equations from this paper [CHH06] are not correct by violating the basic principle that the sum of the absolute amplitudes of the final state squared should add up to 1.

\(^{21}\)These counter-strategies also differ from the one example given in the paper [CHH06], where they give a perfect counter-strategy for Bob in order to have Alice cooperate while he defects.
Outcome 10 For Bob, in order to reach the state of Alice defecting while he cooperates with probability 1, he must play either of the following:

$$\begin{pmatrix}
\theta_b \\
\phi_b \\
\alpha_b
\end{pmatrix} = \begin{pmatrix}
\pi - \theta_a \\
\frac{1}{2} \pi - \alpha_a \\
-\phi_a
\end{pmatrix} \text{ or } \begin{pmatrix}
\theta_a + \pi \\
-\frac{1}{2} \pi - \alpha_a \\
-\phi_a - \pi
\end{pmatrix} \text{ or } \begin{pmatrix}
\theta_a + \pi \\
\frac{1}{2} \pi - \alpha_a \\
-\phi_a - \pi
\end{pmatrix} \text{ or } \begin{pmatrix}
\pi - \theta_a \\
-\frac{1}{2} \pi - \alpha_a \\
-\phi_a - \pi
\end{pmatrix}$$

(2.110)

Outcome 11 For Bob, in order to reach the state of mutual defection with probability 1, he must play either of the following:

$$\begin{pmatrix}
\theta_b \\
\phi_b \\
\alpha_b
\end{pmatrix} = \begin{pmatrix}
\theta_a \\
-\phi_a + \frac{1}{2} \pi \\
\frac{1}{2} \pi - \alpha_a
\end{pmatrix} \text{ or } \begin{pmatrix}
-\theta_a \\
-\phi_a + \frac{1}{2} \pi \\
-\alpha_a
\end{pmatrix} \text{ or } \begin{pmatrix}
\theta_a \\
-\phi_a - \frac{1}{2} \pi \\
\frac{1}{2} \pi - \alpha_a
\end{pmatrix} \text{ or } \begin{pmatrix}
-\theta_a \\
-\phi_a - \frac{1}{2} \pi \\
\frac{1}{2} \pi - \alpha_a
\end{pmatrix}$$

(2.111)

2.5.1 Unitary Actions

A single game of the Quantum Prisoner’s Dilemma is described by

$$|\varphi\rangle = J^1 (U_a \otimes U_b) J |00\rangle$$

(2.112)

The set of strategies involves quantum operations $U_a$ and $U_b$ which are local rotations with three parameters. The matrix representation of the corresponding unitary operators is taken to be

$$U(\theta, \phi, \alpha) = \begin{pmatrix}
e^{i\phi} \cos \frac{\theta}{2} & e^{i\alpha} \sin \frac{\theta}{2} \\
e^{-i\alpha} \sin \frac{\theta}{2} & e^{i\phi} \cos \frac{\theta}{2}
\end{pmatrix}$$

(2.113)

with $\theta \in [0, \pi]$, $\phi \in [0, \frac{\pi}{2}]$ and $\alpha \in [0, \frac{\pi}{2}]$. Hence, selecting strategies $U_a$ and $U_b$ amounts to choosing angles $\{\theta_a, \phi_a, \alpha_a\}$ and $\{\theta_b, \phi_b, \alpha_b\}$. This section will provide insight into why this specific matrix representation is chosen and prove that this representation is unconstrained in the sense that all possible mathematical end-states can be reached\(^{22}\).

Hypersphere

The combined action of the players $(U_a \otimes U_b)$ operates on a 2-qubit state and according to

$$(U_a \otimes U_b)(|\phi\rangle \otimes |\psi\rangle) = (U_a |\phi\rangle) \otimes (U_b |\psi\rangle)$$

(2.114)

\(^{22}\)This specific property is necessary in this game to allow both players to have perfect counter-strategies available in all cases. Otherwise, the game might have a local optimum as pointed out by Benjamin et al. [BH01b] as a comment on Eisert et al.’s [EWL99] two parameter representation.
can be considered as $U_a$ operating on the first qubit and $U_b$ operating on the second qubit in a mathematical sense. Therefore, in order to let $(U_a \otimes U_b)$ have full range, both $U_a$ and $U_b$ must have full range over a single qubit state. This means we cannot simplify a single qubit as mentioned in section about the Bloch sphere (section 2.1.4), but need to consider both the real and the imaginary axis in the amplitudes of a single qubit state.

To be more precise, a single qubit state $\varphi$ can be described by

$$|\varphi\rangle = \alpha |0\rangle + \beta |1\rangle$$

(2.115)

with $\alpha, \beta \in \mathbb{C}$ or alternatively by

$$|\varphi\rangle = (\alpha_{re} + i\alpha_{im}) |0\rangle + (\beta_{re} + i\beta_{im}) |1\rangle$$

(2.116)

with $\alpha_{re}, \beta_{re}, \alpha_{im}, \beta_{im} \in \mathbb{R}$. Any qubit state $|\varphi\rangle$ is subject to the normalization constraint that the inner-product with itself is equal to 1 ($\langle \varphi | \varphi \rangle = 1$), which requires:

$$|\alpha|^2 + |\beta|^2 = 1$$

(2.117)

$$\alpha_{re}^2 + \alpha_{im}^2 + \beta_{re}^2 + \beta_{im}^2 = 1$$

(2.118)

So, a single qubit state can be described by four real parameters which are constrained by equation (2.118). The set of all possible single qubit states can therefore be visualized as a four-dimensional hypersphere with radius 1. Note that for single qubit states we could pair points with the same observable properties (i.e. $|0\rangle \cong i |0\rangle$), but don’t want to use this simplification because we want to be able to construct every possible 2-qubit state.

**Unitary operators**

All systems in quantum computing evolve unitarily, which means that any operator on a quantum state must be unitary (as already mentioned in section 2.1.2). As a reminder:

**Corollary 2.5.1.** Every matrix representing a quantum operator is a unitary matrix.

All unitary matrices $U$ are subject to the rule

$$UU^\dagger = U^\dagger U = I.$$  

(2.119)

Here $U^\dagger$ denotes the complex conjugate transpose or Hermitian adjoint of $U$ and $I$ denotes the identity matrix. This rule also implies some other important properties:

---

23In this particular set-up the 2-qubit state is in full entanglement and can therefore not be factorized in two single qubit states, but for all our analytical purposes this is not relevant.
i $U^\dagger$ is also unitary;

ii $U^{-1} = U^\dagger$;

iii $|\text{det}(U)| = 1$, i.e. the determinant of $U$ lies on the unit circle in the complex plane;

iv the rows of $U$ are orthogonal;

v the columns of $U$ are orthogonal;

vi $U$ is norm preserving; $|Uv| = |v|$.

**Proposition 2.5.2.** The unitary operator $U(\theta, \phi, \psi)$ from (2.113) has full range with respect to the unit hypersphere $S_3$.

**Proof.** Let $U$ be a unitary matrix with full range

$$ U = \begin{pmatrix} u_1 & u_2 \\ u_3 & u_4 \end{pmatrix} $$

(2.120)

with $u_1, u_2, u_3, u_4 \in \mathbb{C}$ and thereby

$$ U^\dagger = \begin{pmatrix} u_1 & u_3 \\ u_2 & u_4 \end{pmatrix} $$

(2.121)

$$ U^{-1} = \begin{pmatrix} u_4 & -u_2 \\ -u_3 & u_1 \end{pmatrix} \cdot \frac{1}{u_1u_4 - u_2u_3} $$

(2.122)

Furthermore, let us restrict ourselves to the group of unitary matrices with $\text{det}(U) = 1$. This restriction gives us

$$ \text{det}(U) = \frac{1}{u_1u_4 - u_2u_3} = 1. $$

(2.123)

Now, according to property (ii) we get

$$ U^\dagger = U^{-1} $$

(2.124)

$$ \begin{pmatrix} u_1 & u_3 \\ u_2 & u_4 \end{pmatrix} = \begin{pmatrix} u_4 & -u_2 \\ -u_3 & u_1 \end{pmatrix} $$

(2.125)

24 This special group of unitary matrices, SU(2), finds wide application in the standard models of physics [AW05].
which reduces our unitary operator to

\[
U = \begin{pmatrix} u_1 & u_2 \\ -\overline{u_2} & \overline{u_1} \end{pmatrix}
\]  

(2.126)

Applying this operator to an arbitrary single qubit state \( |\varphi\rangle \)

\[
|\varphi\rangle = \beta |0\rangle + \gamma |1\rangle = \begin{pmatrix} \beta \\ \gamma \end{pmatrix} \tag{2.127}
\]

\[
U |\varphi\rangle = \begin{pmatrix} u_1 & u_2 \\ -\overline{u_2} & \overline{u_1} \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} u_1\beta + u_2\gamma \\ -\overline{u_2}\beta + \overline{u_1}\gamma \end{pmatrix}
\]  

(2.128)

shows us quite easily that given \( \beta \) and \( \gamma \), we are able to choose \( u_1 \) and \( u_2 \) such that we can reach any point in \( \mathbb{C}^2 \).

We can now arbitrarily assign \( u_1, u_2 \in \mathbb{C} \) with the restriction that

\[
\det(U) = \overline{u_1}u_1 + \overline{u_2}u_2 = |u_1|^2 + |u_2|^2 = 1. \tag{2.129}
\]

In polar coordinates

\[
u_1 = r_1 e^{-i\phi} \tag{2.130}
\]

\[
u_2 = r_2 e^{i\alpha} \tag{2.131}
\]

restricting \( r_1, r_2 \in \mathbb{R} \) to

\[
|u_1|^2 + |u_2|^2 = r_1^2 + r_2^2 = 1 \tag{2.132}
\]

which has a solution in trigonometry:

\[
r_1 = \cos \frac{\theta}{2} \tag{2.133}
\]

\[
r_2 = \sin \frac{\theta}{2} \tag{2.134}
\]

yielding our unitary operator as

\[
U(\theta, \phi, \alpha) = \begin{pmatrix} e^{i\phi} \cos \frac{\theta}{2} & e^{i\alpha} \sin \frac{\theta}{2} \\ -e^{-i\alpha} \sin \frac{\theta}{2} & e^{i\phi} \cos \frac{\theta}{2} \end{pmatrix}
\]  

(2.136)

\[
\square
\]

We can now conclude that this unitary operation Alice and Bob are allowed to
perform on their qubit is functionally complete. In terms of a four-dimensional hypersphere with radius 1 representing all possible single qubit states, applying operator $U$ on a given single qubit state represents a rotation in this hypersphere. With $U$ having full range, any angle of rotation can be expressed.

Now, the set of Pauli matrices $\{I, \sigma_x, \sigma_y, \sigma_z\}$ is an orthonormal basis for $2 \times 2$ matrices in general:

\[
I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}
\] (2.137)

From these matrices, we define an orthonormal set of basis strategies, which we will label as the pure strategies in the Quantum Prisoner’s Dilemma.

\[
C = U(0, 0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I
\]
(2.138)
\[
D = U(\pi, 0, \frac{\pi}{2}) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} = i \cdot \sigma_x
\]
(2.139)
\[
Q_1 = U(\pi, 0, 0) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = i \cdot \sigma_y
\]
(2.140)
\[
Q_2 = U(0, \frac{\pi}{2}, 0) = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} = -i \cdot \sigma_z
\]
(2.141)

These pure strategies are quantum strategies and can be thought of as ‘conditional’ strategies. In full entanglement (choosing $J$ as noted in (2.99)), choosing strategy $Q_1$ for example, means that a counter-strategy of $D$ can be played in order to reach an endstate of mutual cooperation.

All these pure strategies each have their perfect counter-strategy, making this game some sort of a four-way rock-paper-scissors game. The game be viewed as such on basis of its four possible outcomes and the property that every strategy has four possible counterstrategies, each resulting in one of the four possible outcomes.\(^{25}\)

The Nash equilibrium in the rock-paper-scissors game is a mixed strategy of playing each item with equal probability of $\frac{1}{3}$. To the same extent, playing each of the pure strategies $\{C, D, Q_1, Q_2\}$ with equal probability is also a Nash equilibrium in the quantum Prisoner’s Dilemma (see [EW00]). This would mean that playing randomly is actually optimal when playing a single game versus an unknown opponent.

However, the quantum Prisoner’s Dilemma doesn’t have the symmetrical payoff

\(^{25}\)In contrast, Rock-Paper-Scissors has three possible outcomes: win, tie or lose. Given one of the three strategies in this game, we also have three possible counterstrategies, each resulting in one of the three possible outcomes.
structure of the rock-paper-scissors game and therefore is a bit different when considering repeated play, i.e. iterated quantum prisoner’s dilemma. The next chapter presents experimental results of multi-agent learners playing the quantum Prisoner’s Dilemma repeatedly.

Visualization of a single game

Conveniently, we can represent respective outcomes in a given game by means of the expected payoff for a player. Specific actions of both players lead to a probability distribution among the different measurable end states, which makes it possible to compute an expected payoff. For example, suppose that we would like to compute the expected payoff for Alice:

$$S_a = r \cdot Pr(|00\rangle) + s \cdot Pr(|01\rangle) + t \cdot Pr(|10\rangle) + p \cdot Pr(|11\rangle)$$

(2.142)

where $r$ is the reward, $s$ is the sucker’s payoff, $t$ is the temptation and $p$ the punishment. A standardized form has value assignment of 3, 0, 5, 1 respectively.

In order to plot the actions of a player along a single axis, a certain parametrization such that $U_a$ and $U_b$ depend on a single parameter $k \in [-2, 1]$ is needed.

$$U = \begin{cases} 
U(\pi + (k + 1)\pi, -(k + 1)\frac{\pi}{2}, -(k + 1)\frac{\pi}{2}) & \text{if } -2 \leq k \leq -1 \\
U(-k\pi, 0, 0) & \text{if } -1 \leq k \leq 0 \\
U(k\pi, 0, k\frac{\pi}{2}) & \text{if } 0 \leq k \leq 1 
\end{cases}$$

(2.143)

Even though this parametrization does not cover the whole set of actions for both players, we nevertheless are able to recognize all pure strategies and their ideal (pure) counter-strategy. These pairs of strategy and counter-strategy are given more precisely in table 2.5 and visually represented in figure 2.4.

<table>
<thead>
<tr>
<th>Alice</th>
<th>Bob</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>[00]</td>
</tr>
<tr>
<td>D</td>
<td>[10]</td>
</tr>
<tr>
<td>Q₁</td>
<td>[01]</td>
</tr>
<tr>
<td>Q₂</td>
<td>[11]</td>
</tr>
</tbody>
</table>

Table 2.5: The interaction between pure strategies

A player playing this game can use the knowledge of these interactions between strategies as a guideline for choosing his action for the next iteration. Let’s say this player chose strategy $Q_2$ during a previous iteration of the game and the measured
end state turned out to be mutual defection. He can now assume with reasonable probability that his opponent played a strategy in the neighbourhood of $C$. It would now be reasonable to alter his strategy more towards $C$ in order to reach the Pareto optimum of mutual cooperation. A next iteration of the game will give him more information on the action of his opponent, given that he didn’t radically change his strategy. Therefore, convergence to Pareto optimum of mutual cooperation should theoretically be obtainable. This is further investigated in chapter 3, where multi-agent learners play the quantum Prisoner’s Dilemma repeatedly.
Chapter 3

Experimental Setup and Results

This chapter presents the experimental results obtained by letting agents repeatedly play the Quantum Prisoner’s Dilemma. This chapter is split into two parts. The first part describes the performance of an agent adopting a simple learning-mechanism in playing against a player playing a fixed strategy, as well as its performance in self-play. The second part describes the performance of an agent adopting my quantum version of the Conditional Joint Action Learning-mechanism (QCJAL). The results from this experiment are compared with the result obtained by Banerjee et al. [BS07], who found Pareto-optimal behavior with a conditional joint action learner in the classical prisoner’s dilemma.

3.1 Experimental Methods

Operational quantum systems, as of today, have only been devised under specific circumstances in a laboratory setting and are highly experimental [DLX+02]. Consequently, all experiments described in this chapter are simulated on a ‘classical’ computer. Fortunately, the mathematical framework for the quantum prisoner’s dilemma, as described in section 2.5, is easily implementable for a couple of reasons.

First of all, the quantum state of the prisoner’s dilemma can be described fully by the amplitudes of its basis states, i.e. \{\ket{00} = CC, \ket{01} = CD, \ket{10} = DC, \ket{11} = DD\}, and these amplitudes correspond directly with the probability of observing this basis state.

And secondly, the game is expressed in linear algebra, which makes the matrix-based MatLab environment a suitable candidate for simulating the game. The build-in functionality makes it easy to implement the Quantum Prisoner’s Dilemma and, as Matlab is optimized especially for matrix operations, long iterations of the game are not too time consuming.

The MatLab code used in the experiments can be found in appendix A.
3.2 A Simple Learner

A first simulation of the Quantum Prisoner’s Dilemma is run with an agent playing against an invariable opponent using simple supervised learning. These experiments are meant as a pilot, in that it will give us some useful insight into the dynamics of the iterated game but probably won’t yield very interesting results.

The set-up for this problem will be as follows. Both Bob and Alice start a game with a (pseudo-)random assignment of the parameters $\theta$, $\phi$ and $\alpha$ describing their strategy.

$$U(\theta, \phi, \alpha) = \begin{pmatrix}
    e^{i\phi} \cos \frac{\theta}{2} & e^{i\alpha} \sin \frac{\theta}{2} \\
    e^{-i\alpha} \sin \frac{\phi}{2} & e^{i\phi} \cos \frac{\phi}{2}
\end{pmatrix}$$

(3.1)

constrained by

$$0 \leq \theta \leq 2\pi \quad \text{and} \quad 0 \leq \phi \leq \pi \quad \text{and} \quad 0 \leq \alpha \leq \pi \quad (3.2)$$

For the remaining iterations of the game this first strategy assignment for Alice will remain constant. It is now Bob’s task to figure out what Alice’s strategy could possibly be and decide on his perfect counter-strategy. As this game has no pure strategy Nash equilibrium, and in preparation of the next stage where both players are able to alter their strategies, we assume that Bob will strive for the Pareto Optimum of mutual cooperation.

Because of the quantum nature of this game, the only information Bob will gain after playing the game once is the actual outcome of that game. This outcome is a measurement of the quantum end-state where all possible outcomes are represented in a superposition. It is not possible for either of these players to find out the exact quantification of this quantum state. It is only possible to assume that the actual outcome of the game was strongly represented in this superposition. In order to simplify the experiment, we will have Bob assume that the actual outcome was fully represented in the quantum state. That is, Bob assumes that the probability of reaching the actual outcome is 1. This might seem as a rather blunt approach towards uncertainty, but on the other hand the probability landscape is quite fluent and some other choices in constructing the learning process can suppress the errors of this assumption quite reasonably.

We can discern two possible scenario’s from Bob’s point of view. Either the outcome is favorable; mutual cooperation, or the outcome any of the other three possible outcomes. In the first case, Bob should make a note that he reached his wanted outcome and after a specific number of consecutive positive outcomes Bob can terminate his learning process as a (near enough) solution has been found. This specific number can be viewed as a quality factor of the solution found. Also, when the wanted outcome is observed Bob won’t change his strategy.
In the other case, when an unwanted outcome is observed, Bob calculates the perfect counterstrategy to his chosen strategy that yields the given outcome. This can be considered as the assumed strategy chosen by Alice. Bob can now calculate a perfect counter-strategy from his point of view in order to reach mutual cooperation. This perfect counterstrategy is then used to change Bob’s strategy in a classical supervised way. Let \( \{\theta_n, \phi_n, \alpha_n\} \) be Bob’s strategy chosen during game number \( n \), let \( \theta_c, \phi_c, \alpha_c \) be the assumed perfect counter-strategy and let \( L \) be his learning rate. His strategy for the next game will then be:

\[
\begin{align*}
\theta_{n+1} &= \theta_n + L \cdot (\theta_c - \theta_n) \\
\phi_{n+1} &= \phi_n + L \cdot (\phi_c - \phi_n) \\
\alpha_{n+1} &= \alpha_n + L \cdot (\alpha_c - \alpha_n)
\end{align*}
\] (3.3-3.5)

Summing up, the parameters in this experiment are the maximum number of iterations allowed for Bob to find a solution, the amount of consecutive positive outcomes of the game in order to decide whether a strategy is stable and the learning rate for Bob. Naturally, these parameters are interwoven i.e. if we would chose a low learning rate then we should also chose for a larger number of consecutive positive outcomes before deciding the solution is stable as approaching an actual solution means increasing probability for finding successions.

Now, the experiments are run in two settings. First, agent Bob adopts the algorithm and plays against an invariable Alice. In the second experiment both players adopt the learning algorithm. In both scenario’s, the players decide on a random starting strategy.

The results of a general run when playing against an invariable opponent is shown in figure (3.1) and a general run of self-play is shown in figure (3.2). The iterated game terminates when 15 consecutive outcomes of mutual cooperation is reached otherwise it will run for 500 iterations.

A single instance of the Quantum Prisoner’s Dilemma can be described fully by the four probabilities \( Pr(CC) \), \( Pr(CD) \), \( Pr(DC) \) and \( Pr(DD) \) (this is explained fully in section 2.5). The simple quantum learner is ‘taught’, i.e. the agent is supervised, to play mutual cooperation \( (CC) \) and therefore it is descriptive enough for its performance to only consider the probability of this specific outcome\(^1\). The figures (3.1) and (3.2) show the probability of a game actually resulting in mutual cooperation as well as the actual outcomes observed by the agent throughout an iterated run of the game. The fluctuations in the probability of reaching mutual cooperation are a direct result of the agent(s) changing their strategy. Consider the following example.

---

\(^1\)Note that an agent playing the Quantum Prisoner’s Dilemma is never aware of the intrinsic quantum state, from which \( Pr(CC) \) can be calculated, but can only observe the actual outcome of an execution of the game.
3.2. A SIMPLE LEARNER

Example In figure (3.1) shortly before round 100 the probability for reaching mutual cooperation is roughly 75%, and so the probability for either of the other outcomes, \( \{CD, DC, DD\} \) sums up to about 25% (of which \( DD \) has the highest probability because it is mathematically closer to \( CC \) just as is \( CD \) to \( DC \), see probability equations (2.104-2.107)). However the actual outcome of the round to come turned out to be \( CD \); the invariable player playing cooperation while the learning agent plays defection.

This specific situation shows that the agent receives information of the game state, i.e. the quantum end state of a single game expressed in the probability distribution among outcomes \( \{CC, CD, DC, DD\} \). However, this information is not representative of the actual game state. The agent, however, acts on this case of misinformation and changes its strategy accordingly. As a result we see the probability for reaching mutual cooperation going down. In the next rounds, the agent receives representative information again and changes its strategy accordingly, ultimately converging to more than 95% chance at mutual cooperation.

This scenario in particular also sheds light upon why an agent can be unable to reach or sustain mutual cooperation in some iterations: it is simply a case of consecutive rounds of misinformation. In theory, given an endless iteration of playing the Quantum Prisoner’s Dilemma, the agent is bound to have an arbitrary number of consecutive representative outcomes and thereby finding convergence towards a high probability of mutual cooperation while on the other hand this is, sometime in the future, followed by a number of consecutive misrepresentative outcomes, leading the agent away from his goal. Therefore, there is only one stable (end) strategy, namely one of the four analytical solutions to the equation \( Pr(CC) = 1 \) (see equation (2.108)).

As can also be seen from these specific runs, the algorithm converges faster when playing against itself than when playing against an invariable opponent. This notion is further investigated by inspecting the average duration of a run. The resulting data is shown in table (3.1). This difference in performance between both settings

<table>
<thead>
<tr>
<th></th>
<th>( \mu )</th>
<th>( \sigma^2 )</th>
<th>Success Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invariable Opponent</td>
<td>182</td>
<td>24525</td>
<td>88.9%</td>
</tr>
<tr>
<td>Self-Play</td>
<td>105</td>
<td>9735</td>
<td>97.8%</td>
</tr>
</tbody>
</table>

Table 3.1: Average number of iterations and its variance before a solution is found with the percentage of runs that terminates with a solution. Statistical data accumulated by executing both experimental settings \( N = 5,000 \) times.
is significant\textsuperscript{2} enough that we can argue that playing against a player with the same intention of reaching the Pareto-optimal outcome of mutual cooperation helps in reaching this goal.

\textbf{Figure 3.1:} A simple quantum learner playing against an invariable opponent. The probability $\Pr(CC)$, which is the probability that the 2-qubit end state evaluates to $|00\rangle \equiv CC$, is shown as a solid line versus the left-hand $y$-axis and the actual outcomes are shown as dots labeled by the right-hand $y$-axis.

\textsuperscript{2}Intuitively, the results indicate a difference large enough to be significant instead of being a product of mere chance. For anyone interested, however, a simple Student $t$-test yielded a $t$-value of about 30.
3.3 Quantum Conditional Joint Action Learner

The experiments are run with two Quantum Conditional Joint Action Learners (henceforth QCJAL) repeatedly playing the Quantum Prisoner’s Dilemma against each other, i.e. self-play. The original Conditional Joint Action Learner proposed by Banerjee et al. [BS07] does not consider mixed-strategies in playing the Prisoner’s Dilemma. Therefore, we will constrain the QCJAL to playing the pure strategies proposed in section 2.5.1:

\[
C = U(0, 0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I 
\]
(3.6)

\[
D = U(\pi, 0, \frac{\pi}{2}) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} = i \cdot \sigma_x 
\]
(3.7)

\[
Q_1 = U(\pi, 0, 0) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = i \cdot \sigma_y 
\]
(3.8)

\[
Q_2 = U(0, \frac{\pi}{2}, 0) = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} = -i \cdot \sigma_z 
\]
(3.9)
In order to compare the performance of the QCJAL-algorithm to the results found by Banerjee et al. [BS07] the first two parts will consist of:

1. An explorative phase of $N = 400$, where the QCJAL agents select their actions randomly. This is followed by an exploitative phase, where the QCJAL agents always select the action with the highest expected payoff.

2. An explorative phase of $N = 400$, where the QCJAL agents select their actions randomly. This is followed by $\epsilon$-greedy exploration, where the QCJAL agents select a random action with probability $\epsilon$ and select the action with the highest expected payoff with probability $(1 - \epsilon)$, which is a mixed exploration/exploitation phase.

Banerjee et al. found Pareto optimal behavior in the first case (see figure (3.3(a))) but found the agents stuck in the Nash-equilibrium of mutual defection in the second case with $\epsilon = 0.10$ (see figure (3.3(b))).

![Graph](image)

(a) In case of $\epsilon = 0$.  
(b) In case of $\epsilon = 0.10$.

**Figure 3.3:** Variations in Expected Utilities of the CJAL-algorithm in the classical Prisoner’s Dilemma reproduced from Banerjee et al. [BS07].

In an effort to estimate the robustness of the QCJAL-algorithm, the last part of the experiments will consist of the two QCJAL agents playing in a noisy environment. That is, when an agent selects an action there will be some random noise on this action of which the agent is unaware. This noise on every action leads to a slight deviation from the pure strategies in that playing $C$ with some noise against $C$ with some noise results in a quantum end state of which $Pr(CC) \neq 1$. 
3.3. QUANTUM CONDITIONAL JOINT ACTION LEARNER

![Graph showing variations in Expected Utility](image)

(a) Overview of 10,000 rounds.

(b) Zoomed in on the critical point $N = 400$, where the explorative phase ends.

**Figure 3.4**: Variations in the Expected Utility of each pure strategy $\{C, D, Q_1, Q_2\}$ for both agents implementing the QCJAL-algorithm with $\epsilon = 0$.

### 3.3.1 Pure Exploitation

The results of the agents playing the Quantum Prisoner’s Dilemma according to the first scenario of an explorative phase followed by a purely exploitative phase is shown
in figure (3.4). In this figure we see the agents converging to a stable action pair of $(D, Q_1)$, which yields an outcome of mutual cooperation (see table (2.5)). In order to analyze how this conversion takes place, we should look specifically at the short period after the transition point of exploration to exploitation. Figure (3.4(b)) shows this critical phase in more detail.

As selecting any of the pure strategies $\{C, D, Q_1, Q_2\}$ can result in either of the four different outcomes $\{CC, CD, DC, DD\}$ with respective payoffs $\{3, 0, 4, 1\}$ for player A and $\{3, 4, 0, 1\}$ for player B, the expected utility of each strategy should be around 2 after the random explorative phase. This can be seen in figure (3.4(b)). However, as it is unlikely that each action-pair is played the same amount of time during exploration, both players have a specific strategy with the highest expected utility depending on the exact action selection during this explorative phase.

In the run depicted in the figure (3.4), we see that both players prefer the action $D$ and will play this strategy during the first few rounds of exploitation. When both players play defect, the payoff of the game is merely 1 for both players and consequently the expected utility of this action goes down. Now, a player will change his strategy when the expected utility for its current action goes below the expected utility of any other action. This happens to player B first: the expected utility for choosing action $C$ becomes a little higher than choosing any other action. Playing $C$ against $D$, is not lucrative for player B but it is for player A. We see a minor peak in the expected utility of playing $D$ for player A and see the expected utility of playing $C$ reducing for player B. Player B consequently switches to the next strategy with the highest payoff, which happens to be $Q_1$ and because of the payoff of 3 for each player, both $E_A(D)$ and $E_B(Q_1)$ increase and converge towards 3 in the long run.

### 3.3.2 Mixed Exploitation-Exploration

This section presents the results of two QCJAL-agents playing the Quantum Prisoner’s Dilemma adopting an $\epsilon$-greedy approach, see equation (2.84).

**$\epsilon$-greedy with $\epsilon = 0.10$**

The results of two QCJAL agents adopting a mixed exploitation-exploration approach is shown in figure (3.5). Here, the agents play a random strategy for the first 400 rounds followed by the agents individually selecting their best strategy (i.e. the one with the highest expected utility; exploitation) with a probability of $Pr = 1 - \epsilon = 0.9$ and selecting a random strategy (exploration) with a probability of $Pr = \epsilon = 0.1$.

Let us first take a look at figure (3.5(a)). We see a dominant strategy pair of $(Q_2, Q_2)$ emerging directly after the exploratory phase (Rounds > 400) yielding an outcome of mutual cooperation, namely $CC$. Now, both agents play a random action with $Pr = 0.1$ and therefore we can expect that a random strategy is played twice,
3.3. QUANTUM CONDITIONAL JOINT ACTION LEARNER

(a) Overview of 10,000 rounds.

(b) Zoomed in on the critical point $N = 4,500$, where player B switches its dominant strategy.

(c) Zoomed in on the critical point $N = 6,500$, where player A switches its dominant strategy.

Figure 3.5: Variations in the Expected Utility of each pure strategy $\{C, D, Q_1, Q_2\}$ for both agents implementing the QCJAL-algorithm with $\epsilon = 0.1$.

once for each agent, every 10 rounds, meaning that we can expect each agent playing one of the four strategies $\{C, D, Q_1, Q_2\}$ against the other agent’s dominant strategy about once every 40 rounds. Selecting a random strategy other than $Q_2$ will change the expected utility of this selected strategy. As a result, we see the expected utilities
$E_A(Q_1)$ and $E_B(Q_1)$ increase for both agents because playing $Q_1$ against $Q_2$ will deliver an outcome of defection for the player playing $Q_1$ and cooperation for the player playing $Q_2$, i.e. maximum payoff. Analogously, the expected utility for playing $D$ goes down as this yields reverse outcome, i.e. minimum payoff.

Because the individual payoff for the outcome of defection versus cooperation is higher than the individual outcome for mutual cooperation, the strategy pair of $(Q_2, Q_2)$ is not a stable strategy (as it was in case of pure exploitation). As a consequence, it is inevitable that sometime in the future the expected utility $E(Q_1)$ becomes greater than the expected utility $E(Q_2)^3$. This point is reached just before $Rounds = 4,500$. Figure (3.5(b)) zooms in on this transition phase.

A closer look at figure (3.5(b)) reveals that, even though the expected utility increase for playing $Q_1$ is roughly equal for both players, $E_B(Q_1)$ intersects with $E_B(Q_2)$ shortly before $E_A(Q_1)$ with $E_A(Q_2)$. Therefore, player B shifts his strategy to $Q_1$ before player A does in this particular scenario$^4$. As a result of the strategy shift by player B, playing strategy $Q_1$ is not an attractive option anymore for player A as it would lead to an outcome of mutual defection. Playing $Q_2$ is not lucrative either, as it would lead to playing cooperation against defection, resulting in minimum payoff. At this moment, the other two possible strategies for player A $\{C, D\}$ have a low expected utility and player A continues playing $Q_1$ (which has a slightly higher expected utility than $Q_2$) yielding an outcome of mutual defection for a couple of rounds. However, the expected utilities for playing $C$ or $D$ increase significantly.

Shortly after $Rounds = 6,500$, the expected utility of playing $C$ for player A becomes dominant. Playing strategy $C$ against strategy $Q_1$ of player B amounts to an outcome of player A defecting while player B cooperates. As a consequence, we see that this strategy is highly reinforced for player A as we observe a steep peak in the expected utility while in the expected utility of $Q_1$ and $Q_2$ declines for player B. This switch in strategies is shown in more detail in figure (3.5(c)). For the remainder of this iteration the agents end up playing $C$ against $Q_1$ leading to an outcome of player A defecting and player B cooperating. After about 500 rounds, player B adapts and starts playing $Q_2$ again, leading to an outcome of mutual defection. As a result, we see the expected utilities of both these strategies decreasing.

Analyzing this iteration as a whole, we see that close to half of the rounds was spent playing mutual cooperation followed by a period of strategy changes and adaptation to these changes. When we extrapolate figure (3.5), we see the expected utility of

$^3$To be more precise, it is only inevitable given an infinite amount of rounds, i.e. no termination. In practice, however, the probability for selecting a random strategy is high enough to have the expected utility of a strategy other than the current dominant strategy outperform the dominant strategy eventually within our sessions of 10,000 rounds.

$^4$This has to do with the actual outcomes of the exploratory phase and the specifics of the random strategies selected by the players in the mixed phase.
player B playing strategy $C$ on the rise, becoming the dominant strategy eventually. This would lead to a new period of playing mutual cooperation followed by a similar cycle of strategy changes as described above.

$\epsilon$-greedy with $\epsilon = 0.20$

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.6.png}
\caption{Variations in the Expected Utility of each pure strategy \{C, D, Q_1, Q_2\} for both agents implementing the QCJAL-algorithm with $\epsilon = 0.2$.}
\end{figure}

In order to further investigate the effect of the variable $\epsilon$, another set of experiments are run with $\epsilon$ set at 0.2. In this case, the agents explore more and exploit less. To be precise, an agent selects the strategy with the highest expected Payoff with a probability of $Pr = 1 - \epsilon = 0.8$ and selects a random strategy with probability $Pr = \epsilon = 0.2$. The results of such an experiment are shown in figure (3.6).

This figure shows the effect of increasing $\epsilon$ quite well. When we compare figure (3.6) with figure (3.5(a)), we see the shape of the graph of figure (3.5(a)) almost exactly represented in the first half of figure (3.6). Again, directly after the exploratory phase a dominant strategy pair of $(Q_1, D)$ emerges, amounting to mutual cooperation, followed by a similar period of strategy changes. Then in the second half of the figure, we see the agents converging to mutual cooperation again in the strategy pair $(D, Q_1)$. This strategy pair is not stable however, as we see a rise in the expected
utility of $C$ and $Q_2$ for agent A and B respectively. It can be expected to see a similar cycle of strategy changes again when we would allow the experiment to continue beyond 10,000 rounds. In general, an increase in exploratory behavior (increasing $\epsilon$) leads to a higher density of strategy changes within a fixed amount of rounds.

### 3.3.3 Noisy Actions

This section presents the results of two QCJAL-agents playing the Quantum Prisoner’s Dilemma in a noisy environment. The aim of these experiments is twofold. First of all, introduction of noise in the QCJAL algorithm gives us an indication of its robustness, but more importantly the QCJAL algorithm is only implementable when considering a discrete set of strategies for each agent to select. This is a slight simplification of the general approach [EWL99], where each agent utilizes a continuous strategy space: more or less analogous to mixed strategies in classical Game Theory\(^5\). For reference, equation (3.10) describes this continuous strategy space (see section 2.5 for an in-depth description).

$$U(\theta, \phi, \alpha) = \begin{pmatrix} e^{-i\phi} \cos \frac{\theta}{2} & e^{i\alpha} \sin \frac{\theta}{2} \\ -e^{-i\alpha} \sin \frac{\theta}{2} & e^{i\phi} \cos \frac{\theta}{2} \end{pmatrix}$$

with  
$$0 \leq \theta \leq \pi$$  \hspace{1cm} (3.11)  
$$0 \leq \phi \leq \frac{\pi}{2}$$  \hspace{1cm} (3.12)  
$$0 \leq \alpha \leq \frac{\pi}{2}$$  \hspace{1cm} (3.13)

As described earlier, the set of pure strategies $\{C, D, Q_1, Q_2\}$ are an orthonormal basis in this strategy space:

$$C = U(0, 0, 0) \quad Q_1 = U(\pi, 0, 0)$$
$$D = U(\pi, 0, \frac{\pi}{2}) \quad Q_2 = U(0, \frac{\pi}{2}, 0)$$

When we introduce noise, as described in equation (3.15), the agents can effectively play strategies which lie in between the aforementioned pure strategies. Playing such 'in between' strategies leads to a distribution of probabilities among the four possible outcomes and can thus be categorized as playing a mixed strategy:

$$\hat{U}_{\text{noisy}} = U(\theta + \Delta \theta, \phi + \Delta \phi, \alpha + \Delta \alpha)$$

\(^5\)A mixed strategy is a probability distribution over some set of pure strategies, e.g. playing 40% cooperation and 60% defection in a classical Prisoner’s Dilemma.
where
\[ (0, 0, 0) \leq (\Delta \theta, \Delta \phi, \Delta \alpha) \leq \kappa \cdot \left( \frac{\pi}{2}, \frac{\pi}{2} \right) \]
and
\[ 0 \leq \kappa \leq 1. \]

It should be noted that, because of the periodicity of the trigonometric functions in \( U(\theta, \phi, \alpha) \) and the fact that the probability for actually observing some pure state when measuring a quantum state is defined by means of squaring the amplitude of the respective quantum state (see equations (2.104-2.107), the following equivalences for the variables \( \{ \theta, \phi, \alpha \} \) follow directly:

\[
\begin{align*}
\theta & \equiv \min(\hat{\theta}, 2\pi - \hat{\theta}) \quad \text{with} \quad \hat{\theta} = \theta \mod 2\pi \\
\phi & \equiv \min(\hat{\phi}, \pi - \hat{\phi}) \quad \text{with} \quad \hat{\phi} = \phi \mod \pi \\
\alpha & \equiv \min(\hat{\alpha}, \pi - \hat{\alpha}) \quad \text{with} \quad \hat{\alpha} = \alpha \mod \pi
\end{align*}
\]

That is, the variables are symmetrical around \( \pi \) or \( \frac{\pi}{2} \) and are periodical in \( 2\pi \) or \( \pi \) for \( \{ \theta \} \) and \( \{ \phi, \alpha \} \) respectively.

The experiments are run in three set-ups, namely:

1. Both agents are subjected to noise of up to a quarter of the strategy-space each, i.e. uniform distribution of noise with \( \kappa = \frac{1}{4} \) on the action of each agent.
2. One agent is subjected to noise of up to half of the strategy-space, i.e. uniform distribution of noise with \( \kappa = \frac{1}{2} \) on the action of one agent.
3. One agent is subjected to noise of up to three quarters of the strategy-space, i.e. uniform distribution of noise with \( \kappa = \frac{3}{4} \) on the action of one agent.

Symmetrical noise with \( \kappa = \frac{1}{4} \)

The results of the first scenario, where both agents experience noise on their actions, is shown in figure (3.7). Here, we see the agents converging to an outcome of mutual cooperation similar to the results of the noise-free experiment of pure exploitation (figure (3.4)). Upon closer inspection, however, we see a difference in the expected utility of the dominant strategy pair.

In case of a noise-free environment, the expected utility for a mutual cooperation strategy pair converges to 3, a straightforward result as the pay-off for mutual cooperation is equal to 3 and an outcome of mutual cooperation is reached every time each agent chooses their dominant strategy. In case of a noisy environment, the expected utility of the dominant strategy pair stabilizes around 2.5. This is a direct result of the noise as both agents do select their dominant strategy, but it does not result in an outcome of mutual cooperation with probability 1. To be more precise: the agents
are naïve in that they consider themselves to have chosen a pure strategy based on its expected utility, but play a mixed strategy instead, thereby not fully enforcing the expected utility of their selected strategy as the outcome might be different than intended.

In the worst-case scenario of full noise, i.e. both agents playing \( U(\frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}) \), the result is equal probabilities \( Pr(CC) = Pr(CD) = Pr(DC) = Pr(DD) = \frac{1}{4} \). Therefore, the expected utility of the dominant strategy pair does not converge to 3. On the other hand, as long as the expected utility of the dominant strategy exceeds the expected utility of any other strategy, as it does in this scenario with the noise limited to a quarter of the strategy-space, the other strategies are not played and thereby not reinforced.

\[
U_{\text{noisy}} = U(\theta + \Delta\theta, \phi + \Delta\phi, \alpha + \Delta\alpha) \text{ where } (0, 0, 0) \leq (\Delta\theta, \Delta\phi, \Delta\alpha) \leq \frac{1}{4}(\pi, \frac{\pi}{2}, \frac{\pi}{2}).
\]

**Figure 3.7:** Variations in the Expected Utility for both agents implementing the QCJAL-algorithm with noise:

Asymmetrical noise with \( \kappa = \frac{1}{2} \)

The results of the second scenario, where one agent experiences noise on its action while the other agent does not, is shown in figure (3.8). Even though the total amount of noise in this set-up is equal to the total noise of the first set-up, there are some distinct differences between the results of these. We again observe a stable dominant strategy pair that leads to an outcome of mutual cooperation, namely \((D, Q_1)\), but
3.3. QUANTUM CONDITIONAL JOINT ACTION LEARNER

Figure 3.8: Variations in the Expected Utility for both agents implementing the QCJAL-algorithm with agent A subject to noise: $U_{\text{noisy}} = U(\theta + \Delta \theta, \phi + \Delta \phi, \alpha + \Delta \alpha)$ where $(0, 0, 0) \leq (\Delta \theta, \Delta \phi, \Delta \alpha) \leq \frac{1}{2}(\pi, \pi, \pi)$.

Notice a significant difference in the actual expected utility of these strategies. The expected utility of player A’s dominant strategy stabilizes around $E_A(D) \approx 2.75$, while the expected utility of player B’s dominant strategy remains stable around $E_B(Q_1) \approx 2.25$.

At first, this bias in expected utilities seems out of place as player A is the only player experiencing noise, but he receives a higher average payoff than player B as a result. Nevertheless, further analysis of this bias with respect to the strategy pair $(D, Q_1)$ gives us a mathematical explanation of this phenomenon.

Consider the (dominant) strategy pair of $D = U(\pi, 0, \frac{\pi}{2})$ versus $Q_1 = U(\pi, 0, 0)$. Now, because this strategy pair has the highest expected utility, both players will repeatedly select this strategy. However, player A experiences noise on his action and so will actually select a strategy $\tilde{U}(\tilde{\theta}, \tilde{\phi}, \tilde{\alpha})$ within the range defined in equations (3.17-3.18).

\[
(\pi, 0, \frac{\pi}{2}) \leq (\tilde{\theta}, \tilde{\phi}, \tilde{\alpha}) \leq (\frac{1}{2} \pi, \frac{\pi}{4}, \frac{3}{4} \pi) \quad (3.17)
\]

Normalized (see equation (3.16)):

\[
\frac{\pi}{2} \leq \tilde{\theta} \leq \pi \quad 0 \leq \tilde{\phi} \leq \frac{\pi}{4} \quad \frac{\pi}{4} \leq \tilde{\alpha} \leq \frac{\pi}{2} \quad (3.18)
\]
CHAPTER 3. EXPERIMENTAL SETUP AND RESULTS

As a result of these restrictions on the action of player A and the notion that the induced noise is uniformly distributed, we can estimate the expected outcome for a given game.

First of all, in case of full noise, we again observe equal probabilities for all possible outcomes as playing $\hat{U}(\frac{\pi}{2}, \frac{\pi}{4})$ versus $Q_1$ yields $Pr(CC) = Pr(CD) = Pr(DC) = Pr(DD) = \frac{1}{4}$, similar to the first scenario. However, when we estimate the distance from player A’s dominant strategy $Q_1$ to the other three pure strategies it should be noted that both $Q_1$ and $C' = U(0, 0, \frac{\pi}{2}) \equiv U(0, 0, 0) = C$ only differ from $D$ in the range of one variable, while for $Q_2 = U(0, \frac{\pi}{2}, \frac{\pi}{2}) \equiv Q_2$ the difference is in two variables. To be more precise: in order to change one’s strategy from $D$ to $C$, $\theta$ needs to be changed along a single period of the strategy space in the $\theta$-dimension. The same goes for changing the strategy from $D$ to $Q_1$ in the $\alpha$-dimension, but changing the strategy from $D$ to $Q_2$ requires changing both $\theta$ and $\phi$ along a single period in the strategy space. We can therefore argue that the actual played mixed strategies are skewed towards pure strategies $C$ and $Q_1$, with respective outcomes $DC$ and $DD$ when played against player B’s strategy of $D$. By definition, outcomes $CC$ and $DD$ deliver symmetrical pay-offs while $CD$ and $DC$ do not. Hence the difference in the expected utility of the dominant strategies.

The actual bias of the dominant strategy played by player A can be calculated by means of integration with respect to the range (equation (3.18)) of player A’s strategy from $D$.

\[
\begin{align*}
Pr(outcome|U_b = Q_1) = & \begin{pmatrix} Pr(CC|U_b = Q_1) \\ Pr(CD|U_b = Q_1) \\ Pr(DC|U_b = Q_1) \\ Pr(DD|U_b = Q_1) \end{pmatrix} \\
= & \begin{pmatrix} (-\sin(\frac{\phi}{2})\sin(\alpha))^2 \\ (\cos(\frac{\phi}{2})\sin(\alpha))^2 \\ (\cos(\frac{\phi}{2})\cos(\alpha))^2 \\ (-\sin(\frac{\phi}{2})\cos(\alpha))^2 \end{pmatrix}
\end{align*}
\]

Now, when we integrate these probabilities with respect to the noise-induced range, we get a biased end state as shown in equation (3.20).

\[
\int_0^\pi \int_0^\frac{\pi}{2} \int_0^\frac{\pi}{2} Pr(outcome|U_b = Q_1)d\theta_d\phi_d\alpha = \begin{pmatrix} \frac{\pi^4}{128} + \frac{\pi^2}{32} + \frac{\pi^2}{32} \\ \frac{\pi^4}{128} + \frac{\pi^2}{32} + \frac{\pi^2}{32} \\ \frac{\pi^4}{128} - \frac{\pi^2}{32} - \frac{\pi^2}{32} \end{pmatrix} \approx \begin{pmatrix} 0.65 \\ 0.03 \\ 0.14 \end{pmatrix}
\]

\[\text{Note that this can only be done because of uniformly distributed noise.} \]

\[\text{Although this can not simply be done linearly, a first-order estimator for } y = \sin(x) \text{ and } y = \cos(x) \text{ in the range } 0 \leq x \leq \frac{\pi}{4} \text{ is } y = x \text{ and } y = 1 - x \text{ respectively.} \]
Furthermore, this integral can be normalized with the range of the noise and multiplied by the respective payoffs, giving us the expected payoff (the theoretical expected utility) as given in equation (3.22), proving our argument.

\[
E(E_A(D_{\text{noisy}})) = (3, 0, 4, 1) \cdot \frac{1}{\frac{\pi}{2} \cdot \frac{\pi}{4} \cdot \frac{\pi}{4}} \left( \frac{\pi^3}{128} + \frac{\pi^2}{32} + \frac{\pi}{32} \right) \approx 2.75 \quad (3.21)
\]

\[
E(E_B(Q_1)) = (3, 4, 0, 1) \cdot \frac{1}{\frac{\pi}{2} \cdot \frac{\pi}{4} \cdot \frac{\pi}{4}} \left( \frac{\pi^3}{128} + \frac{\pi^2}{32} + \frac{\pi}{32} \right) \approx 2.29 \quad (3.22)
\]

Asymmetrical noise with \( \kappa = \frac{3}{4} \)

The results of the third scenario, where one agent experiences noise of up to three quarters of the strategy-space, is shown in figure (3.9). Again, we see a dominant strategy pair \((Q_1, D)\) emerging which leads to an outcome of mutual cooperation \(CC\). However, this dominant strategy pair is borderline stable. The expected utility of player A’s strategy lingers around \(E_A(Q_1) \approx 2\), which is the baseline for random play as \(\frac{1}{4}(\text{Reward} + \text{Sucker’s} + \text{Temptation} + \text{Punishment}) = 2\). As it turns out, the expected utility of this strategy remains slightly higher than the expected utility of the other actions. Strategy \(C\) and \(D\) are played shortly after the explorative phase, but are negatively reinforced. Therefore, the expected utility \(E_A(Q_1)\) remains dominant, but is considerably lower than the expected utility of player B’s dominant strategy \(E_B(D)\). This is a result of the biased consequences of noise, analogously to the results of the second set-up of less asymmetrical noise.

### 3.4 Chapter Summary

The results presented in this chapter cover two grounds. On one hand, experimental set-ups are devised to empirically analyze the dynamics of the Quantum Prisoner’s Dilemma, and on the other hand, the performance of the reinforcement learning algorithm CJAL is compared to its performance in the classical game.

Empirical analysis of the Quantum Prisoner’s Dilemma, by means of a simple supervised learning scheme, shows us that an agent playing the game can consistently deduce information about the intrinsic quantum state by considering the actual outcome of a game in relation to the strategy chosen by the agent. That is, an agent
can effectively approximate the strategy chosen by his opponent and use this information to his end, e.g. increasing the probability for reaching an outcome of mutual cooperation. This learning scheme is applied to a single agent playing against an invariable opponent and to two agents playing the game with each other, i.e. self-play. Even though information gained from playing against an invariable opponent should be more accurate than playing against a variable opponent, the algorithm performs significantly better in self-play.

The Conditional Joint Action Learning algorithm [BS07] has shown itself to be Pareto-optimal in self-play. An extension to this reinforcement learning algorithm, with respect to the quantum game, is devised on basis of four pure strategies. These conceived Quantum Conditional Joint Action Learners are shown to also be Pareto-optimal in self-play. Additionally, the quantum counterpart is shown to be more consistent and robust, in that the outcome of mutual cooperation prevails in all experiments. Introduction of more explorative (random) play has a less dramatical effect on the Pareto optimality of the algorithm than it has in the setting of a classical Prisoner’s Dilemma and introduction of noise on the strategy selection of the agent(s) further reinforces the notion of robustness.

The introduction of noise on the strategy selection of an agent also sheds light on
the actual positioning of the chosen pure strategies. These strategies are not placed in such a way that the (mathematical) distance from any one strategy to any of the other strategies is equal. Therefore, the noise induces a bias towards the closest two pure strategies. As a result, we should consider the four pure strategies to be positioned at the corners of a rectangle instead of on the corners of a tetrahedron. This particular notion is especially important when considering a continuous strategy-space instead of a discrete set of strategies.
Chapter 4

Discussion

4.1 Summary of Contributions

In section 2.5 an extensive account of the Quantum Prisoner’s Dilemma has been given. Even though the quantum prisoner’s dilemma has frequently been described in the literature ever since its introduction by Eisert et al. [EWL99], all descriptions seem to be lacking in some area. First of all, Benjamin and Hayden [BH01a] pointed out that the proposed scheme of the Prisoner’s Dilemma by Eisert et al. is constrained in the set of strategies available to the players. This comment is acknowledged by Eisert and Wilkens [EW00] and they propose enlarging the set of allowed operations, but omit an actual operator.

Chen and Hogg [CH06] propose the general single-qubit operator used in the description presented in section 2.5.1 (see equation (2.101)), but omit the reasoning why this in fact covers general single-qubit operations. Chen and Hogg present a further mathematical description of the game in their paper, but erroneously compute the probabilities associated with the four possible outcomes of the game, which also results in an incorrect presentation of perfect counter-strategies to known strategies.

The scheme presented in this thesis is intended as an improvement on all these accounts and should offer the reader, unfamiliar with quantum mechanics or quantum computing, a mathematically complete description of the quantum prisoner’s dilemma. Additionally, four pure strategies have been devised from an orthonormal basis of the strategy space. These four strategies give rise to the notion that in fact the dilemma ceases to exist and the quantum prisoner’s dilemma becomes an asymmetric four-way rock-paper-scissors game. That is, every strategy within the strategy space has a corresponding counter-strategy (see section 2.5) resulting in either of the four possible outcomes with a probability equal to 1, but other than rock-paper-scissors has a different payoff associated for each outcome making it asymmetric.

The main original work done in this thesis can be found in Chapter 3, where
the empirical results of a multi-agent learning framework in the quantum prisoner’s dilemma is given. This empirical study showcases the intricacies inherent to this quantum game. It is shown that an agent can estimate the actual strategy played by his opponent quite accurately and thereby converge to a Pareto-optimal outcome if there is an mutual inclination to do so. This is of course analogous to the classical prisoner’s dilemma, where mutual intention or negotiation can also result in Pareto optimal behavior, although on first sight this is not so obvious in a quantum game.

The proposed quantum adaptation of the conditional joint action learner (QCJAL) gives us results partially similar to those produced in the classical game. QCJAL will learn to converge to a Pareto-optimal solution if it plays a random exploration strategy followed by completely greedy exploitation. However, the quantum variant outperforms the classical conditional joint action learner when it mixes the greedy exploitation with random exploration. In the classical case it resorts to playing ‘defect’ when the greedy exploitation is mixed with a tenth of random exploration, while in the quantum case, periods of mutual cooperation are alternated with short periods of strategy changes. Even when the portion of random exploration in this mix is increased, the quantum conditional joint action learner still shows periods, albeit shorter, of mutual cooperation.

On a further note, when noise is introduced on the strategy selection of the quantum conditional joint action learner, an interesting notion of the positioning of the selected pure strategies reveals itself. Although mathematically, these pure strategies are placed at the limits of the strategy space, they are not positioned at an equal distance from each other. Consequently, outcomes of mutual cooperation and mutual defection are closer to each other, in the probability landscape, than to the other two outcomes of cooperation versus defection and defection versus cooperation, this also pertains the other way around. To be more precise, in the neighborhood of the perfect counter-strategy for reaching mutual cooperation, the probability for an outcome of mutual defection is significantly higher than the probability for defection versus cooperation and cooperation versus defection. This is also true for the neighborhood of the perfect counter-strategy for reaching an outcome of defection versus cooperation, where the probability for reaching cooperation versus defection is higher than the probability for reaching either mutual cooperation or mutual defection. This bias shows no effect in any of the experiments other than the introduction of noise.

4.2 Concluding Remarks

Quantum game theory offers some new perspectives on the prisoner’s dilemma. As a result of entanglement and allowing the full set of quantum strategies, the actual dilemma ceases to exist as now every strategy has a perfect counter-strategy. This change in dynamics could theoretically amount to oscillating behavior. For example,
let both players (Alice and Bob) be greedy in their approach to this game, i.e. both want the other to ‘cooperate’ while they ‘defect’, and let \((D, Q_2)\) (see table (2.5)) be the first strategy-pair chosen in a repeated game. Now, this leads to Alice cooperating while Bob defects and Alice will have the incentive to change her strategy to \(Q_1\) as \((Q_1, Q_2)\) leads to her cooperating and Bob defecting. Along this same line Bob will change his strategy to \(C\) in the following iteration after which Alice will change her strategy to \(D\) followed by Bob changing his strategy to \(Q_2\) again. However, this sequence of play is never observed in the experiments, simply because alternating between the temptation and the sucker’s payoff produces less payoff than playing mutual cooperation twice (a condition of the prisoner’s dilemma, see equation 2.78). Also, it is easy to understand that playing mutual defection is not a wanted outcome for both players either. Therefore, it can be concluded that the quantum prisoner’s dilemma proposed in this thesis is a repeated game which has mutual cooperation as the preferred outcome. In this light, it is no surprise that the quantum conditional joint action learner indeed converges to mutual cooperation, after a brief period of time, that is, where both agents need to figure out the current most prevailing strategy of the other agent. This notion also explains why playing mixed exploitation with exploration is characterized by recurring periods of mutual cooperation.

Originally, the prisoner’s dilemma has been devised in order to demonstrate that the Nash equilibrium need not be Pareto optimal. As the proposed quantum variant does not contain this dilemma between optimal and rational play anymore, one might wonder whether this specific quantum prisoner’s dilemma is still a meaningful game and whether the conceived quantum conditional joint action learner has any merits.

As a matter of fact, Eisert and Wilkens have pointed out that this specific game with general unitary operations has no pure Nash equilibrium, but does have multiple Nash equilibria in mixed strategies [EW00]. For example, the aforementioned alternation between the outcomes of ‘defect’ versus ‘cooperation’ and the other way around amounts to a mixed Nash equilibrium, i.e. playing \(D\) and \(Q_2\) with a probability of \(\frac{1}{2}\) each\(^1\). The existence of these Nash equilibria in the Quantum game, which do not exist in the classical game, is a novelty.

The quantum conditional action learner, however, does not converge to any of these mixed equilibria, even though in earlier work on game theory it has been stated that in a game with more than one equilibrium, anything that attracts a (human) players’ attention towards one of the equilibria may make them expect and therefore realize it [Sch60]. In this context, a valid question remains whether the quantum conditional joint action learner is a feasible model of human behavior in the quantum prisoner’s dilemma. For one, the experimental results of this learner do not resemble

\(^1\) It should be noted that this mixed strategy can also be played in the classical game, but is not a Nash equilibrium because any player could benefit from simply changing to the dominant strategy \(D\).
4.2. CONCLUDING REMARKS

the results found by Chen and Hogg [CHH06].

On a more fundamental level there is the ongoing debate about the biological plausibility of the quantum prisoner’s dilemma itself. It might be a natural gametheoretic choice for phenomena taking place at a microscopic level where the laws of quantum mechanics reign. For example, Richard Dawkins [Daw89] is well-known for his study of evolutionary biology at the level of our genes. When considering the actual interaction of evolutionary biology at the level of our genes, it might be a natural game-theoretic choice for phenomena taking place at a microscopic level where the laws of quantum mechanics reign. For example, Richard Dawkins [Daw89] is well-known for his study of evolutionary biology at the level of our genes. When considering the actual interaction of evolutionary biology at the level of our genes, it might be a more appropriate description than its classical counterpart. It might even offer an explanation to why living organisms have a tendency to cooperate far more than classical game theory predicts.

In the light of quantum mechanics playing a role in brain processes, one of the most concrete and detailed descriptions of how this might take place can be found in the research by Beck and Eccles [BE92], later refined by Beck [ Bec01]. Beck’s research refers to particular mechanisms of information transfer at the synaptic cleft. It has been proposed that these quantum mechanisms therefore influence human thought processes and that we should consider it in the analysis of human behavior. The way in which these quantum processes are relevant for mental activity and in which way their interactions with mental states are conceived, however, remains unclarified and highly speculative.

The notion of quantum mechanics playing a role in the dynamics of our brain, has been further generalized by several philosophers, leading to theories commonly known as Quantum Mind theories. Quantum Mind theories attempt to explain human consciousness (and to a certain extent free will) by adopting quantum theory and quantum models. A notable theory in this field is the controversial proposal of Roger Penrose to relate elementary conscious acts to gravitation-induced reductions of quantum states [Pen89]. This proposal, ultimately, requires the framework of a theory of quantum gravity, which is still under development. Penrose’s approach moves away from mental causation as he proposes conscious acts to be non-computable. In an attempt to prove this non-computability he invokes the incompleteness theorem of Gödel, stating that any computable theory would be based on axioms which are fundamentally incomplete. This statement is generally regarded as highly debatable. Other well-known quantum approaches to human action and consciousness are Bohm’s theory of the relationship of mind and matter [Boh90] and Stapp’s quantum theory of mind-brain interface [Sta99].

---

2His BBC documentary Nice Guys Finish First is also quite a treat for any novices in game theory, particularly the prisoner’s dilemma.

3Neuronal transmission of action potentials in the brain takes place both chemically and electrically. Chemical transmission takes place at the synaptic cleft.

4The interested reader should consider the Stanford Encyclopedia of Philosophy’s article on
These theories on Quantum Mind and quantum mechanical effects in the human brain, however, have never been verified in any way. There is a widespread skepticism towards these theories in general. Philosopher David Chalmers, for example, stated that the motivation for Quantum Mind theories is: "a Law of Minimization of Mystery: consciousness is mysterious and quantum mechanics is mysterious, so maybe the two mysteries have a common source." Furthermore, Patricia Churchland is well-known for her criticism towards Quantum Mind theories [Chu96]. As none of these theories can productively be verified or contradicted, we should probably file them under pseudo-science.

On a further note, the quantization of the game might not even be the most plausible approach. Witte considers quantum game theory not as quantum games played by ‘classical’ players but as quantum players playing a classical game. His view is more philosophical as it is seated on the belief that any human player in a game will play the same game in his imagination several times before coming up with a choice of strategy. This limits the capacity of players to manipulate the quantum state of the game as they are the quantum state themselves.

In summary, quantum game theory has interesting implications, but it remains to be seen where these are applicable.

4.3 Further Research

This thesis restricts itself to a single reinforcement learner in one specific quantum game. It has only been tested in self-play and therefore it remains unknown whether other learning mechanisms might outperform the quantum conditional joint action learner. It might be fruitful to introduce other multi-agent learners to the realm of quantum games as well. Additionally, this learner could also be extended to other quantum games where it has yet to prove itself.

Another interesting approach might be to take Witte’s view to heart by devising different multi-agent frameworks where the agents reason quantum physically. To this end, quantum epistemic logic might be used to some success. However, this approach is highly hypothetical and, to my knowledge, no research based on this philosophy has been conducted yet.

Quantum Approaches to Consciousness, which offers an extensive (introductionary) review of this field.
Bibliography


Appendix A

Matlab source code

This Appendix contains a variety of Matlab codes used in the conducted experiments with the Quantum Conditional Joint Action Learner.

A.1 Quantum Prisoner’s Dilemma

A single instance of the Quantum Prisoner’s Dilemma:

```matlab
function it = singleIteration(A,B)
    zerozero = [1;0;0;0];
    J = (sqrt(2)/2).*[1 0 0 i; 0 1 i 0; 0 i 1 0; i 0 0 1];
    Jdag = inv(J);
    U = kron(unitOp(A(1),A(2),A(3)),unitOp(B(1),B(2),B(3)));
    endS = Jdag*U*J*zerozero;
    probs = abs(endS).^2;
    it = probs;
end
```

The measurement operation with respect to the amplitudes of the four possible basis states:

```matlab
function outcome = measure(probs)
    quart1 = probs(1);
    quart2 = sum(probs(1:2));
    quart3 = sum(probs(1:3));
    quart4 = sum(probs);
    quart4 = round(quart4);
    if quart4 > 1
        disp('ERROR probabilities do not add up to 1, but to: ')
        disp(quart4)
        outcome = 0;
    else
        toss = rand;
        if toss <= quart1
            outcome = 1;
        elseif toss <= quart2
            outcome = 2;
```
A.2 The ‘classical’ Prisoner’s Dilemma

The classical Prisoner’s Dilemma is simulated by introducing the quantum game without entanglement:

```matlab
function it = singleIterationNoE(A, B)
zerozero = [1;0;0;0];
U = kron(unitOp(A(1),A(2),A(3)),unitOp(B(1),B(2),B(3)));
endS = U*zerozero;
probs = abs(endS).^2;
it = probs;
end
```

A.3 Conditional Joint Action Learner

The Conditional Joint Action Learner as proposed by Banerjee et al. [BS07] playing the (iterated) Prisoner’s Dilemma:

```matlab
clear;
rewardsA = [3 0 4 1];
rewardsB = [3 4 0 1];
history = ones(4,1);
expectedUtilA = [0 0];
expecUtilB = [0 0];
plotexpA = zeros(2,10000);
plotexpB = zeros(2,10000);
conditionalsA = zeros(4,10000);
epsilon = 0.0;%0.1;

%number of times for random play:
N = 400;

for n = 1:10000
    expectedUtilA = [(rewardsA(1)*history(1)/(history(1)+history(2)) +
                      rewardsA(2)*history(2)/(history(1)+history(2)))...
                      (rewardsA(3)*history(3)/(history(3)+history(4)) +
                      rewardsA(4)*history(4)/(history(3)+history(4))];
```
expectedUtilB = \left( \frac{\text{rewardsB(1)} \times \text{history(1)}}{\text{history(1)} + \text{history(3)}} + \ldots \right) \\
\left( \frac{\text{rewardsB(3)} \times \text{history(3)}}{\text{history(1)} + \text{history(3)}} \right) \ldots \\
\left( \frac{\text{rewardsB(2)} \times \text{history(2)}}{\text{history(2)} + \text{history(4)}} + \ldots \right) \\
\left( \frac{\text{rewardsB(4)} \times \text{history(4)}}{\text{history(2)} + \text{history(4)}} \right) ;

% plot variables:
plotexpA(:,n) = expectedUtilA;
plotexpB(:,n) = expectedUtilB;
conditionalsA(:,n) = \left( \frac{\text{history(1)}}{\text{history(1)} + \text{history(2)}} \right) \ldots \\
\left( \frac{\text{history(3)}}{\text{history(3)} + \text{history(4)}} \right) ;

% explorative phase:
if n < N
    actionA = rand*pi;
    actionB = rand*pi;
    probs = singleIterationNoE([\text{actionA}; 0; 0], [\text{actionB}; 0; 0]);
    outcome = measure(probs);
    history(outcome) = history(outcome) + 1;
else % learning phase:
    actionA = 0;
    actionB = 0;
    if expectedUtilA(1) < expectedUtilA(2)
        actionA = pi;
    end
    if expectedUtilB(1) < expectedUtilB(2)
        actionB = pi;
    end
    if rand <= epsilon
        actionA = abs(actionA - pi);
    end
    if rand <= epsilon
        actionB = abs(actionB - pi);
    end
    probs = singleIterationNoE([\text{actionA}; 0; 0], [\text{actionB}; 0; 0]);
    outcome = measure(probs);
    history(outcome) = history(outcome) + 1;
end
n = n + 1;
end
t = 1:1:10000;
h = plot(t, plotexpA(1,:),’-k’, t, plotexpA(2,:),’--k’)

A.4 Quantum Conditional Joint Action Learner

The Quantum Conditional Joint Action Learner playing the (iterated) Quantum Prisoner’s Dilemma:
APPENDIX A. MATLAB SOURCE CODE

%CIJL learner with four discrete strategies: CD Q1 Q2 considering a game
% with full entanglement
rewardsA = [3;0;4;1];
rewardsB = [3;4;0;1];
%history for Alice and Bob respectively
historyA = zeros(4,4) + realmin;
historyB = zeros(4,4) + realmin;
% CC CD Q1 Q2 
% DC DD Q1 Q2 
% Q1C Q1D Q1Q1 Q1Q2 
% Q2C Q2D Q2Q1 Q2Q2
outcomes = zeros(4,1);
outref = {'CC';'CD';'DC';'DD'};
linhist = zeros(rounds,1);
%reference table for possible outcomes:
% C  D  Q1  Q2
% CC 1 2 3 4
% CD 3 4 1 2
% Q1 2 1 4 3
% Q2 4 3 2 1
refs = [1 2 3 4;... 
  3 4 1 2;...
  2 1 4 3;...
  4 3 2 1];
N = 400; %number of explorative cycles
epsilon = -1; %0.25;
plotexpA = zeros(4,rounds);
plotexpB = zeros(4,rounds);
plotpayoffA = zeros(1,rounds);
plotpayoffB = zeros(1,rounds);
conditionalsA = zeros(16,rounds);
%introducing some (white) noise for Alice:
oise = [0;0;0];
for n = 1:rounds
    noise = [0;0;0];%0.75*[rand*pi;rand*pi/2;rand*pi/2];
    if mod(n,1000)==0
        disp(n)
    end
%plotting values for expected utility:
plotexpA(:,n) = expectedUtilA;
plotexpB(:,n) = expectedUtilB;
plotpayoffA(n) = transpose(outcomes)*rewardsA/n;
plotpayoffB(n) = transpose(outcomes)*rewardsB/n;
actionA = [0;0;0];
actionB = [0;0;0];
A.4. QUANTUM CONDITIONAL JOINT ACTION LEARNER

% conditional probabilities:

\[
\text{expectedUtilA} = \left[ (\text{rewardsA}(1) \times \text{historyA}(1,1)/\text{sum}(\text{historyA}(1,:)) + \text{rewardsA}(2) \times \text{historyA}(1,2)/\text{sum}(\text{historyA}(1,:)) + \text{rewardsA}(3) \times \text{historyA}(1,3)/\text{sum}(\text{historyA}(1,:)) + \text{rewardsA}(4) \times \text{historyA}(1,4)/\text{sum}(\text{historyA}(1,:))) \right] ; \ldots
\]

\[
(\text{rewardsA}(2) \times \text{historyA}(2,1)/\text{sum}(\text{historyA}(2,:)) + \text{rewardsA}(1) \times \text{historyA}(2,2)/\text{sum}(\text{historyA}(2,:)) + \text{rewardsA}(3) \times \text{historyA}(2,3)/\text{sum}(\text{historyA}(2,:)) + \text{rewardsA}(4) \times \text{historyA}(2,4)/\text{sum}(\text{historyA}(2,:))) ; \ldots
\]

\[
(\text{rewardsA}(3) \times \text{historyA}(3,1)/\text{sum}(\text{historyA}(3,:)) + \text{rewardsA}(1) \times \text{historyA}(3,2)/\text{sum}(\text{historyA}(3,:)) + \text{rewardsA}(2) \times \text{historyA}(3,3)/\text{sum}(\text{historyA}(3,:)) + \text{rewardsA}(4) \times \text{historyA}(3,4)/\text{sum}(\text{historyA}(3,:))) ; \ldots
\]

\[
(\text{rewardsA}(4) \times \text{historyA}(4,1)/\text{sum}(\text{historyA}(4,:)) + \text{rewardsA}(1) \times \text{historyA}(4,2)/\text{sum}(\text{historyA}(4,:)) + \text{rewardsA}(2) \times \text{historyA}(4,3)/\text{sum}(\text{historyA}(4,:)) + \text{rewardsA}(3) \times \text{historyA}(4,4)/\text{sum}(\text{historyA}(4,:))) ; \ldots
\]

\[
\text{expectedUtilB} = \left[ (\text{rewardsB}(1) \times \text{historyB}(1,1)/\text{sum}(\text{historyB}(1,:)) + \text{rewardsB}(2) \times \text{historyB}(1,2)/\text{sum}(\text{historyB}(1,:)) + \text{rewardsB}(3) \times \text{historyB}(1,3)/\text{sum}(\text{historyB}(1,:)) + \text{rewardsB}(4) \times \text{historyB}(1,4)/\text{sum}(\text{historyB}(1,:))) \right] ; \ldots
\]

\[
(\text{rewardsB}(2) \times \text{historyB}(2,1)/\text{sum}(\text{historyB}(2,:)) + \text{rewardsB}(1) \times \text{historyB}(2,2)/\text{sum}(\text{historyB}(2,:)) + \text{rewardsB}(3) \times \text{historyB}(2,3)/\text{sum}(\text{historyB}(2,:)) + \text{rewardsB}(4) \times \text{historyB}(2,4)/\text{sum}(\text{historyB}(2,:))) ; \ldots
\]

\[
(\text{rewardsB}(3) \times \text{historyB}(3,1)/\text{sum}(\text{historyB}(3,:)) + \text{rewardsB}(1) \times \text{historyB}(3,2)/\text{sum}(\text{historyB}(3,:)) + \text{rewardsB}(2) \times \text{historyB}(3,3)/\text{sum}(\text{historyB}(3,:)) + \text{rewardsB}(4) \times \text{historyB}(3,4)/\text{sum}(\text{historyB}(3,:))) ; \ldots
\]

\[
(\text{rewardsB}(4) \times \text{historyB}(4,1)/\text{sum}(\text{historyB}(4,:)) + \text{rewardsB}(1) \times \text{historyB}(4,2)/\text{sum}(\text{historyB}(4,:)) + \text{rewardsB}(2) \times \text{historyB}(4,3)/\text{sum}(\text{historyB}(4,:)) + \text{rewardsB}(3) \times \text{historyB}(4,4)/\text{sum}(\text{historyB}(4,:))) ; \ldots
\]

If \( n < N \)

\[
\text{stateA} = \text{ceil}(4 \times \text{rand});
\]

\[
\text{stateB} = \text{ceil}(4 \times \text{rand});
\]

switch stateA

  case 1
    \% A plays C
    actionA = [0;0;0];
  case 2
    \% A plays D
    actionA = [\pi/2;0];
  case 3
    \% A plays Q1
    actionA = [\pi/2;0];
  case 4
    actionA = [0;\pi/2];

end

switch stateB

  case 1
    \% B plays C
    actionB = [0;0];
  case 2
    \% B plays D
    actionB = [\pi;0];
  case 3
    \% B plays Q1


APPENDIX A. MATLAB SOURCE CODE

```matlab
105 actionB = [pi; 0; 0];
106 case 4
107     actionB = [0; pi/2; 0];
108 end
109
110 probs = singleIteration(actionA+noise, actionB); %+noise);
111 outcome = measure(probs);
112 outcomes(outcome) = outcomes(outcome) + 1;
113 linhist(n) = outcome;
114 % respective histories for the players are based on the actual outcomes of noisy values:
115 AassumesB = find(refs(stateA, :)==outcome);
116 BassumesA = find(refs(:, stateB)==outcome);
117 historyA(stateA, AassumesB) = historyA(stateA, AassumesB) + 1;
118 historyB(BassumesA, stateB) = historyB(BassumesA, stateB) + 1;
119
else
120
121 [utilA, stateA] = max(expectedUtilA);
122 if rand <= epsilon
123     stateA = mod(stateA + ceil(4*rand), 4);
124     if stateA == 0
125         stateA = 4;
126     end
127 end
128
129 [utilB, stateB] = max(expectedUtilB);
130 if rand <= epsilon
131     stateB = mod(stateB + ceil(4*rand), 4);
132     if stateB == 0
133         stateB = 4;
134 end
135
136 switch stateA
137     case 1
138         % A plays C
139         actionA = [0; 0; 0];
140     case 2
141         % A plays D
142         actionA = [pi; 0; pi/2];
143     case 3
144         % A plays Q1
145         actionA = [pi; 0; 0];
146     case 4
147         actionA = [0; pi/2; 0];
148 end
149
150 switch stateB
151     case 1
152         % B plays C
153         actionB = [0; 0; 0];
154     case 2
155         % B plays D
156         actionB = [pi; 0; pi/2];
157     case 3
158         % B plays Q1
159         actionB = [pi; 0; 0];
160     case 4
161         actionB = [0; pi/2; 0];
162 end
163
164 probs = singleIteration(actionA+noise, actionB); %+noise);
```
A.4. QUANTUM CONDITIONAL JOINT ACTION LEARNER

```matlab
outcome = measure(probs);
outcomes(outcome) = outcomes(outcome) + 1;
linhist(n) = outcome;
% respective histories for the players are based on the actual outcomes of noisy values:
AassumesB = find(refs(stateA,:) == outcome);
BassumesA = find(refs(:, stateB) == outcome);
historyA(stateA, AassumesB) = historyA(stateA, AassumesB) + 1;
historyB(BassumesA, stateB) = historyB(BassumesA, stateB) + 1;
end
n = n + 1;
end
t = 1:1:rounds;
p = plot(t, plotexpA(1,:), 'r-', t, plotexpA(2,:), 'g-', t, plotexpA(3,:), 'b-', t, plotexpA(4,:), 'c-')
```

93