Reinforcement learning and evolutionary algorithms for CHSH nonlocal games

Bachelor's Project Thesis

Jan Pastorek, jpastorek5@gmail.com
Supervisors: Mgr. Daniel Nagaj, PhD.

Abstract: Nonlocal games are a key concept in quantum information, utilized from complexity theory to certification of quantum devices. They involve two or more players that decide on a strategy, and then must stop communicating among themselves. They win the game if they provide properly correlated answers to questions. The typical example is the CHSH game, related to Bell inequalities that can be violated by quantum mechanics. In this game, quantum players have a higher winning probability than classical players. Actually determining the optimal winning probability is a difficult problem in general. In this paper, we investigate a variety of nonlocal games and search for optimal quantum strategies with the help of machine learning strategies (reinforcement learning). We find that ... To be added

https://www.overleaf.com/project/5fc387a73d4925e77906e3a6

1 Introduction

Every piece of information is always stored on some physical medium (paper, magnetic tape, brain ...). Since study of information is in our interest and all information is carried on some physical medium, we want to know how the physical medium behaves, so that we can fully utilize potential knowledge. In last decades, we have found that we can store information on the basis of quantum particle's spins, potentially giving us new ways of harnessing this physical medium options. Since then, we have already started utilizing knowledge from quantum phenomena in various quantum algorithms. Superposition is one such phenomenon that we have started mining. When a particle is in a superposition, it means it is in more than one state at once. From this phenomena we get qubits, where qubit is 0 and 1 at the same time, not anymore being in just one state. We need a vector instead of a simple bit to describe such a state (e.g. $\begin{bmatrix}1\\0\end{bmatrix}$). However, it gets much more complex when we describe many particles. Cleverly using superpositions of such states, finding ways to add desirable amplitudes for computational paths, while subtracting undesirable ones lets us get surprising algorithmic advances. Another important phenomena is entanglement. Entanglement is a state of particles that interacted and since this interaction, they can not be reduced to states of individual particles that interacted – the whole is more than just the sum of its parts.

By now we use superposition and entanglement in quantum algorithms. Grover's algorithm is provably optimal algorithm for searching an unstructured database or an unordered list that can search for items in $O(\sqrt{N})$. Shor's quantum factoring algorithm for integer factoring is able to factor in $O(\log N)$, while the best known classical one is exponential in N.

To understand, lets dig into it by first getting to know how physicists got there.

1.1 Short History of Quantum Mechanics

It all started with classical mechanics not being able to match our observations on both very large and very small scales. So, it could neither be used to predict exact trajectories of planets, [2] nor properties of atomic matter. [3] One of the main problems was ultraviolet catastrophe. Classical models predicted predicted that an ideal black body at thermal equilibrium will emit radiation in all frequency ranges, emitting more energy as the frequency increases. However, this would lead all matter to radiate all of its energy until it would reach absolute zero temperature, which was obviously an error at the heart of the theory.

Physicists were of course looking for ways to fix this, some took a wholly different path. It was just at this moment that Max Planck proposed that energy is absorbed and emitted in small packets called quanta. Hence the name 'quantum physics'. Calculations based on this hypothesis has matched data from observations of black-body radiation

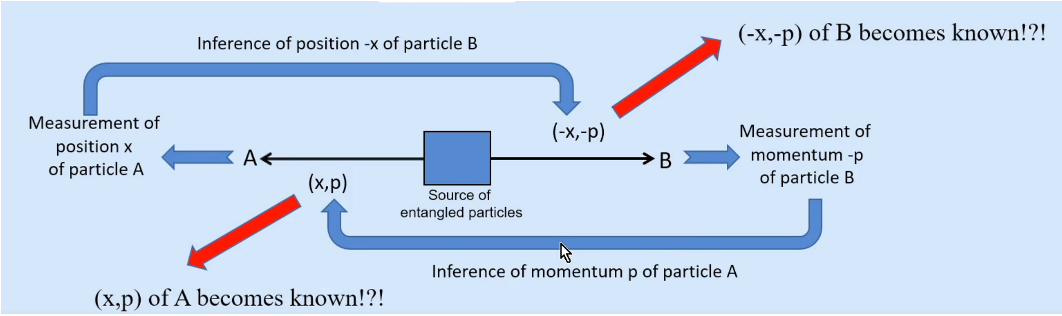

Figure 1.1: EPR paradox original [1]

and could eradicate the ultraviolet catastrophe at the expense of abandoning a classical description. Planck's discovery was coined as the birth of quantum mechanics. There were other observations that supported Planck's hypothesis such as the photoelectric effect. Einstein's Nobel prize came for explaining this effect by assuming that light is absorbed and reemitted in quanta as well.

This has led to the duality of light paradox. In classical mechanics, Young has shown that light behaves as a wave by his double slit experiment. The experiment shows how indeterminate and probabilistic our reality actually is.

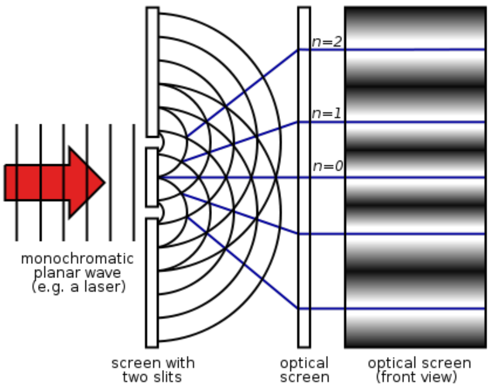

Figure 1.2: Double slit experiment [4]

When you prepare setup for experiment (see Fig. 1.2) in such a way that you prepare an optical screen and place another screen with two slits parallel to the optical screen. Then you begin the experiment by sending a light wave(e.g. a laser) in the direction of screens. While crossing the slits, the light will diffract and then interfere with the light passing through another slit causing pattern shown on the figure. At some places, the light will make constructive interference and producing "white" spots on the optical screen. At another places, the light will make destructive interference and produce "black" spots. However, if the light was a particle, it should not produce interference patterns like waves do. Therefore, Young concluded that light has a wavy nature.

However, while already since Newton, it was also useful to view light also as a stream of particles. Einstein's description supported this view. Basically two camps arose: One that was saying that light is a wave, the other saying that light is a particle. Surprisingly, experiments supported both views. The way light behaved depended on context.

Questions like: 'What is the true nature of light?', 'How to interpret our observations?' ,'Is everything observable?', 'Do our observations change reality?' bothered physicists for further decades. Many interpretations of quantum reality arose. Some have set limits to our knowledge of reality, others assumed that reality is just an illusion.

Stranger behaviour in the micro world was observed. In his experiments Heisenberg showed that we can't determine the position and the momentum of a particle, at the same time, with arbitrary precision. Physicists today argue this is because the momentum and position observables/quantities do not commute. Another aspect of this phenomenon manifests as the fundamental wave-particle duality of light. Once we attempt a photon's position with high precision (say with a small pinhole), its momentum will become uncertain, and the photon will necessarily diffract. The reason for this is that the wave-particle duality is a fundamental and inescapable law of nature and whereby, once a particle's position is attempted to be determined with high precision, say through a pinhole or a small slit, it will be diffracted and therefore its momentum will be rendered uncertain.

Throughout this time, Einstein has believed that quantum mechanics is an incomplete theory and was preparing his thought experiments to demonstrate it. In one of these, known as the EPR paradox, Einstein, Podolsky and Rosen argued that it could be in principle possible to determine the value of two non-commuting observables with certainty and without disturbing the system. They set up the following thought experiment (see Fig. 1.1):

1. Make an entangled state with two particles, whose momenta (p) and positions (x) are tied to each other – (e.g. by having a larger par-

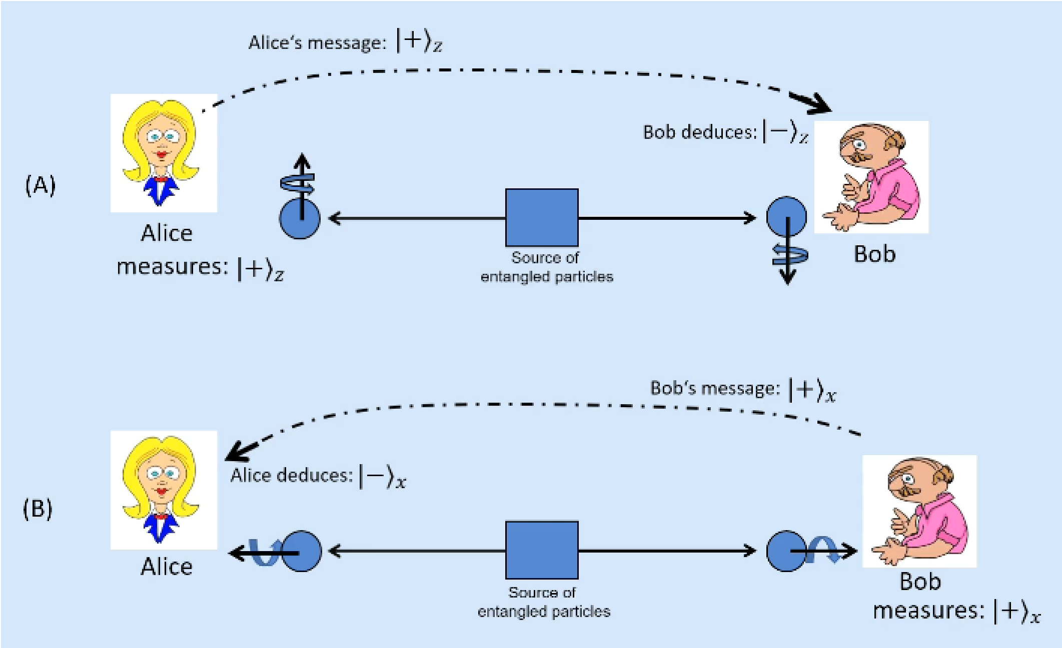

Figure 1.3: EPR paradox with spins [1]

ticle decay into two identical ones, moving in opposite directions).

2. Let the two particles fly very far from each other.

3. Measure one quantity on one particle and the other quantity on the other particle.

4. Knowing the position of one particle lets us infer the position of the other. Meanwhile, knowing the momentum of one particle would let us infer the momentum of the other. This way, we could infer missing information about the non-commuting observables – momentum and position, and thus get *complete* information about the quantum system.

To illustrate the EPR paradox in a more comprehensible way, we will move from continuous variables to discrete values, qubits. However qubits are computer science representations of particles' spins. For now, think of a spin of particle as rotation of particle around its own axis, like earth rotates around its own axis. Particle can rotate around its own axis in two directions, therefore, there are two spins, spin up and spin down. This feature of particles was found out by Stern and Gerlach in their experiment. [5] (see Fig. 1.4).

They found this feature by sending particles through magnet slit. They found out that sent particles will go either to upper part (spin up) or lower part (spin down) of the plate after they are influenced by magnets.

In quantum computation spin up is just state vector/ qubit $\begin{bmatrix}1\\0\end{bmatrix}$ and spin down is $\begin{bmatrix}0\\1\end{bmatrix}$

With knowledge of what a spin is, lets now focus on EPR through different lens. (see Fig. 1.3)

1. Make an entangled state with two particles, whose spins around two axis (z and x) are tied to each other.

2. Let the two particles fly very far from each other.

3. (A)lice measures spin on (z) axis on her particle and the other quantity the other particle. Bob deduces how is his particle rotating around (z) axis based on Alice's measurement findings.

4. (B)ob measures spin on (x) axis on his particle. Alice deduces how is her particle rotating around (x) axis based on Bob's measurement findings. Thus we get *complete* information about the quantum system.

Obviously, this thought experiment was somehow violating Heisenberg's uncertainty principle, aiming to determine the precise value of non-commuting observables. As a consequence of the EPR paradox, it must follow that either QM is not a complete theory (the wavefunction is not a complete description of reality, and some hidden variables are at play) or these observables cannot have simultaneous reality – or even worse, faster than light communication (FTL) is at work. FTL was in contradiction with Einstein's special relativity, where the velocity of light is limited to circa 300 000 m/s. Observables not having simultaneous reality was unthinkable for a deterministic universe. Hence, Einstein concluded that QM is not a complete theory. [7]

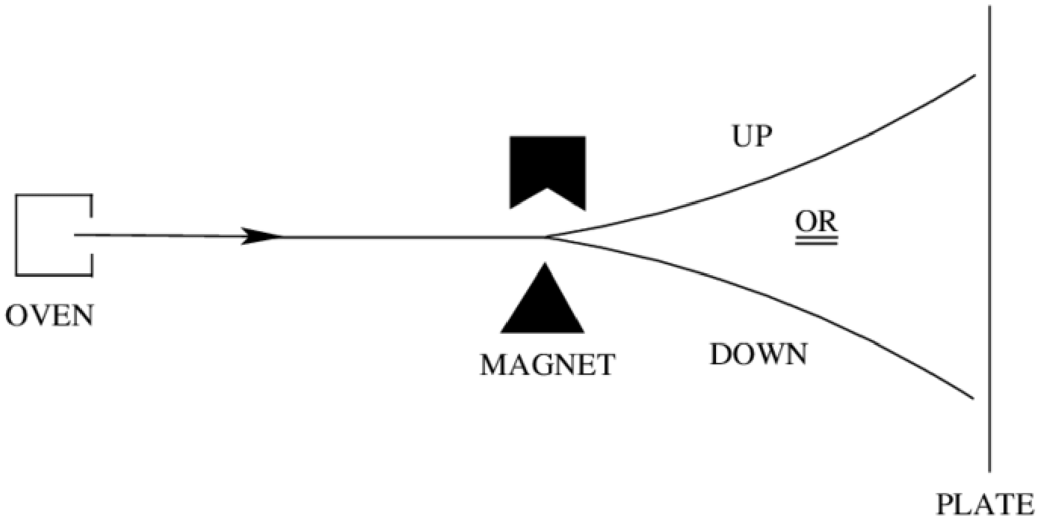

Figure 1.4: Stern-Gerlach experiment [6]

However, we believe Einstein was not correct. Based on observations and backed by theory, today we still claim that two observables/quantities cannot have simultaneous reality, and that the measurement on one observable makes the values of another observable indeterminate, if the observables are non-commuting. So, the way Einstein inferred knowledge about the second particle by observing the first was flawed.

Could not FTL be at work? It seems that entangled states do communicate at velocity faster than light, because even after we separate them far from each other, after measurement on one particle, the other collapses instantly as well.

It looked like some information was missing. However, no FTL was at work because the particles just did not send any other particles to communicate between them. They just inherently "knew" what state to be in. Further investigations into this strange reality was done by John Steward Bell, who in 1964 came up with a series of calculations and equations, called Bell/CHSH inequalities. [8] He calculated that if we assume that our universe is locally realistic (two physical systems can influence each other only when they are close enough), we will indeed come to contradictions. His conclusions were later demonstrated by Alain Aspect's experiments. [9] What is then the true nature of our reality? In order to not to come to contradictions, we had to abandon either realism or locality. If we abandon realism, it means that we abandon the idea that there exist objective, independent physical systems. But what are we living in, an illusion? Does not an illusion presuppose reality? On the other hand, if we abandon locality, in principle it means that physical systems can "interact" without sending any piece of information at any distance.

It might be useful for a while to let this discussion be, and turn to the "shut up and calculate" approach. We will see, how fruitful it will be, analyzing nonlocal scenarios involving communication: games. We will look back at what we did and return to interpretations later in Section 9

1.2 Nonlocal games

Indeed, Bell inequalities can be demonstrated using *nonlocal games* , where quantum players can win more than classical ones.

Nonlocal games are a key concept in quantum information, utilized from complexity theory to certification of quantum devices. They involve two or more players that win if they provide properly correlated answers to questions. The typical example is the CHSH game. We will later see that in this game, quantum players have a higher winning probability than classical players. Here (Fig. 1.5) is the game:

1.2.1 CHSH game

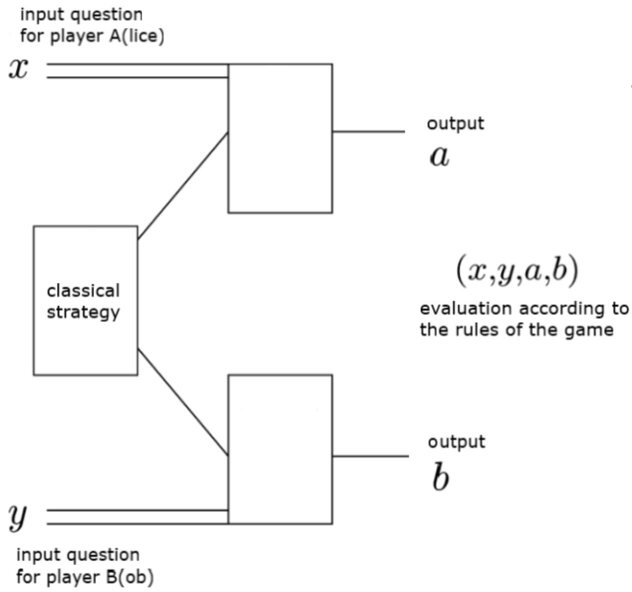

Figure 1.5: CHSH game, classic strategy

1. Alice and Bob are respectively given one random bit (questions x and y), and they each respond with one bit (answers a and b).
2. If either of them gets a question 0 ($x = 0$ or $y = 0$), they win if they answers agree ($a = b$).
3. If they both get 1 they win by giving opposite answers ($a \neq b$).

After agreeing on a joint strategy, Alice and Bob take a brief space voyage before getting their questions, and giving their answers. They will not be able to coordinate their answers after getting their input because they need to send their response back to Earth within a short time period. It's not just against the rules for them to communicate, it's physically impossible due to the finite speed of light.

What would make a good strategy for Alice and Bob? One possible strategy is to always output 1, regardless of input. Then they will win whenever either one of them receives a 0. If the inputs are given at random, they will win 75% of the time. Another strategy would be for Bob to always output 1, and for Alice to output 1 if her input is 1, and 0 otherwise. This also has a 75% chance of winning. It turns out that when the input is random, 75% is the best you can do classically. This can be easily calculated by going through all deterministic strategies. Since Alice and Bob have each 2 different possible inputs and each of them can provide on the basis of input 2 possible outputs, each has

$$2.2 = 4$$

possibilities. There is two of them, Alice and Bob, so there are just

$$4.4 = 16$$

different possible deterministic strategies.

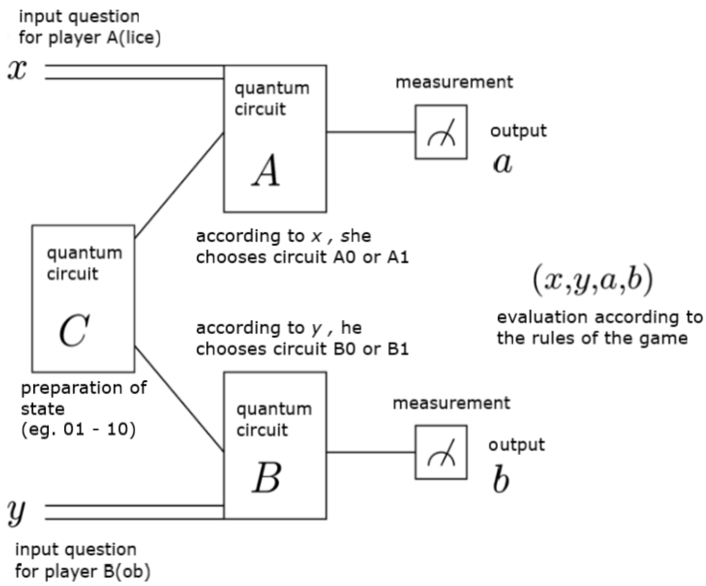

Figure 1.6: CHSH game, quantum strategy

However, in a quantum world (see Fig. **??**), Alice and Bob can prepare and share an entangled quantum state which they will bring with them on their voyage. They still cannot communicate, as the finite speed of light forbids communication. They can, however, coordinate their answers better and win with more than 75% probability. When Alice and Bob learn their input questions, they each apply a unitary gate to their half of their shared maximally entangled state. Then they measure their qubits, and answer 0 if their qubit was up and 1 if their qubit was down. We claim that they will do better than is physically possible classically. We will do the calculation in detail in Section 3.1.

Actually, determining the optimal winning probability is a difficult problem in general. Brute force is in these cases literally impossible. Usually, approximation algorithms are quite successful in doing with such tasks. New approach for this optimalization task is needed, therefore we are going to look at this optimalization task with the help of universal approximator, neural network. Reinforcement learning, in particular. In short, we want to let our reinforcement learning agent learn paths to optimal winning probability.

The purpose of this work is to explore ways of approximating the optimal winning probability using reinforcement learning. We will use reinforcement learning to learn strategies that maximize the *quantum* value of a nonlocal game, going far beyond the toy CHSH one, where we know the answers.

2 Quantum Mechanics

> "If you are not completely confused by quantum mechanics, you do not understand it" - John Wheeler

"Quantum mechanics is simply this: it is a set of four postulates that provide a mathematical framework for describing the universe" (and possibly) "everything in it." [10]
These postulates state how to describe a quantum state, how it evolves, how to measure it and how to combine quantum states.

Postulate 1: State space

How do we describe physical system?

Every physical system has associated complex vector space. This vector space is to be called *state space* of a system.

System is *isolated* when it is not interacting with any other system.

If physical system is isolated, then the system is completely described by unit vector in this system's state space. This vector is to be called *state vector*.

We have already shown one such state in the introduction, state vector $\begin{bmatrix}1\\0\end{bmatrix}$. However, state vector can be any vector from complex vector space with length equal to 1. This length can be calculated by simply making sum of the squares of the amplitudes of the state vector. Amplitudes are coefficients of state vector. So for instance, $\begin{bmatrix}x_1\\x_2\end{bmatrix}$ has two amplitudes x_1 and x_2. The length of this state vector is $x_1^2 + x_2^2$ and this sum must be equal to 1 in order for a vector to be a state vector. This is known as *normalization constraint.*

$$x_1^2 + x_2^2 = 1$$

We have mentioned in the introduction that qubit can be both 0 and 1 at the same time. This is known as *superposition.* Superposition is just another term for linear combination. In our case qubit is a linear combination of $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$. To illustate superposition, we can have a state vector $0.6\begin{bmatrix}1\\0\end{bmatrix} + 0.8\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0.6\\0.8\end{bmatrix}$. It is a state vector because $0.6^2 + 0.8^2 = 1$ holds. The interpretation of this examplary superposition is that state can "collapse" after measurement to either state vector $\begin{bmatrix}1\\0\end{bmatrix}$ with probability 0.6^2 or state vector $\begin{bmatrix}0\\1\end{bmatrix}$ with probability 0.8^2. That is why quantum universe is probabilistic and why our qubit is 0 and 1 at the same time.

However we can also write state vectors in different notation - *ket notation*, which is used much more in the community. $\begin{bmatrix}1\\0\end{bmatrix}$ can be written as $|0\rangle$. $\begin{bmatrix}0\\1\end{bmatrix}$ can be written as $|1\rangle$. Superposition can be written like this: $0.6\begin{bmatrix}1\\0\end{bmatrix} + 0.8\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0.6\\0.8\end{bmatrix} = 0.6\,|0\rangle + 0.8\,|1\rangle$

Normalization constraint or that sum of the squares of the amplitudes of the any state vector must satisfy $x_1^2 + ... + x_n^2 = 1$, is just the same as dot product, which can be written in ket notation as $\langle\psi|\psi\rangle = 1$. This condition is called *normalization condition.*

Because of normalization constraint we can illustrate state of one qubit on a two-dimensional plane using just a simple unit circle. State vector can be any "point" on the simple unit circle. (see Fig. **??**)

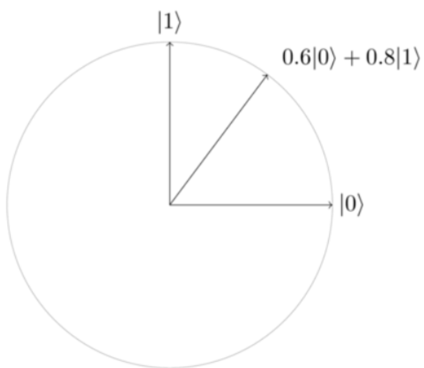

Figure 2.1: state vector on a unit circle [10]

Postulate 2: Dynamics of a system

How do we describe dynamics of isolated quantum system?

If we have an isolated physical(quantum) system, we can describe its evolution by a *unitary* matrix acting on the system's state space.

Postulate says that unknown state of the system $|\psi\rangle$ at time t_1 is in relation with the $|\psi'\rangle$ by unitary matrix U. For a matrix to be unitary, it needs to satisfy

$$U^\dagger U = I$$

Where,

$$U^\dagger = (U^T)^*$$

† is called *dagger operation.* This operation firstly transposes matrix and then makes complex conjugate(makes times (-1) its imaginary part) * of its terms. So,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^\dagger = \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}^T\right)^* = \begin{bmatrix} a & c \\ b & d \end{bmatrix}^* = \begin{bmatrix} a^* & c^* \\ b^* & d^* \end{bmatrix}$$

Unitary matrices have a special attribute, they preserve the length of vectors after multiplication. Why do we need to use just unitaries? It is because they preserve length of vectors, hence, sum of probabilities is also equal to 1. This can be written

$|\psi'\rangle = U|\psi\rangle$. U is not dependent on any of those states, though it could depend on times t_1 and t_2.

To illustrate this, have a state $|\psi\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and evolution described by matrix $U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. So our new state will be $|\psi'\rangle = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. So our state $|\psi'\rangle = |\psi\rangle$. That is because we used identity matrix as an example of unitary matrix. Identity matrix preserves the state.

If $U = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, then $|\psi'\rangle = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

There are some matrices that are commonly used in quantum computing, which are definitely worth remembering. These are foremost Pauli's Matrices, named after physicist Wolfgang Pauli. (see Fig. ??)

Gate is just another equal representation of matrices mostly used in when representing quantum circuits.

Operator is yet another shorter representation of commonly used matrices. We mostly use operators in ket notation.

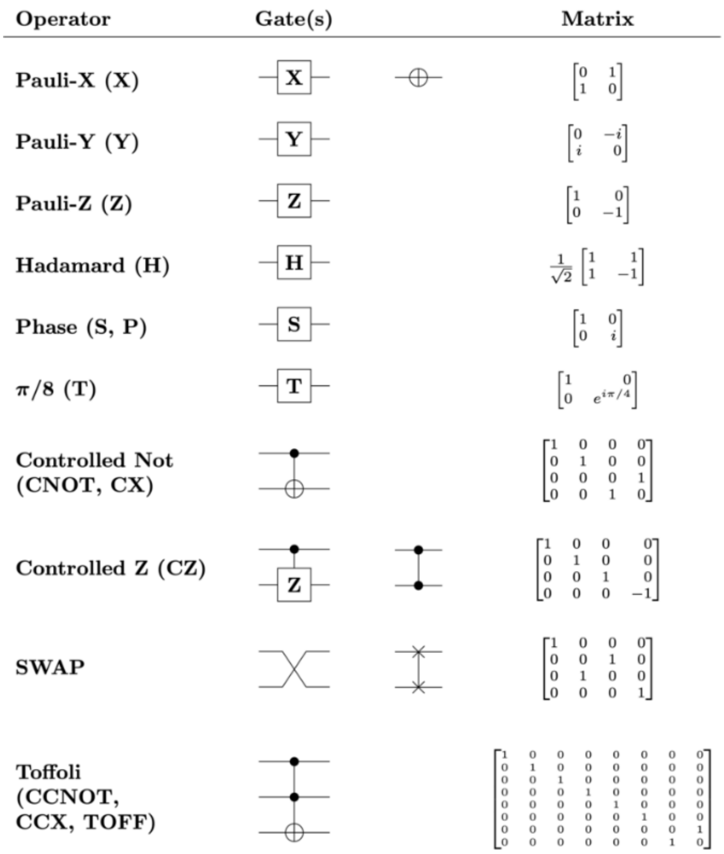

Figure 2.2: Commonly used matrices/gates [11]

Postulate 3: Measurement of system

How do we describe measurements on quantum systems?

Quantum measurements are described by a collection M_m of *measurement operators*. Each M_m is a matrix acting on the state space of the system being measured. The index m takes values corresponding to the measurement outcomes that may occur in the experiment.

If the state of the quantum system is $|\psi\rangle$ immediately before the measurement then the probability that result m occurs is given by

$$p(m) = \langle\psi| M_m^\dagger M_m |\psi\rangle \qquad 3.0$$

Posterior state is the state of the system after the measurement.

$$|\psi_m\rangle = \frac{M_m|\psi\rangle}{\sqrt{\langle\psi| M_m^\dagger M_m |\psi\rangle}} \qquad 3.1$$

The probability of measurement outcomes must be equal to 1, thus, measurement operators must satisfy the *completeness equation*,

$$\sum_m M_m^\dagger M_m = I \qquad 3.2$$

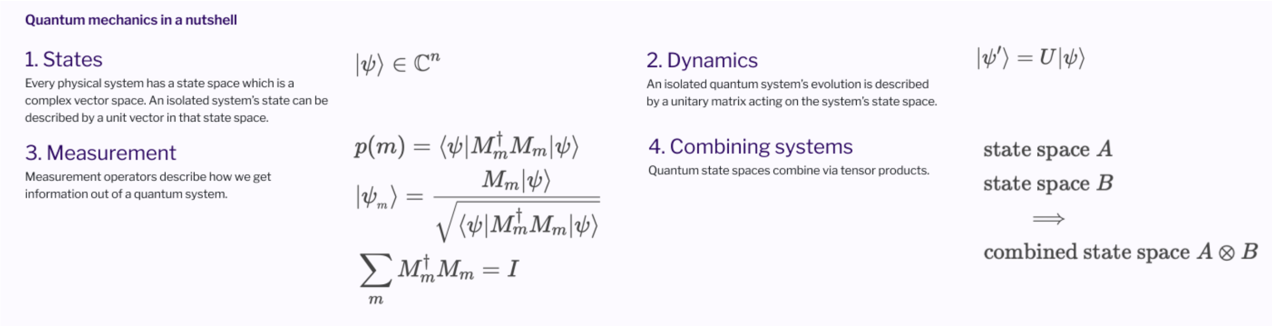

Figure 2.3: Quantum postulates [10]

Then,

$$I = \sum_m p(m) = \sum_m \langle\psi| M_m^\dagger M_m |\psi\rangle \qquad 3.3$$

Measurement operators are matrices made from the outer product one state with itself. *Measurement of qubit in the computational basis* is one type of measurement. Since this is a single qubit measurement, it can have two outcomes, and so two measurement operators.

$$M_0 = |0\rangle\langle 0|, M_1 = |1\rangle\langle 1|$$

Suppose we have the state

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \begin{bmatrix}\alpha\\ \beta\end{bmatrix}$$

Then, if we want to measure it in the computational basis and get probability of outcome 0. We will construct the corresponding operator M_m, where $m = 0$, by outer product of state $|m\rangle = |0\rangle$

$$M_0 = |0\rangle\langle 0| = \begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}1 & 0\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$$

and get the corresponding probability of outcome 0 by

$$p(0) = \langle\psi| M_0^\dagger M_0 |\psi\rangle = \langle\psi| \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} |\psi\rangle =$$

$$\langle\psi| \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} |\psi\rangle = \langle\psi| M_0 |\psi\rangle = \begin{bmatrix}\alpha\\ \beta\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\end{bmatrix} = |\alpha|^2$$

Posterior state, the state after measurement, is according to equation 3.1

$$|\psi_0\rangle = \frac{M_0|\psi\rangle}{\sqrt{\langle\psi| M_0^\dagger M_0 |\psi\rangle}}$$

We have already calculated the denominator, which is $|\alpha|^2$, therefore

$$|\psi_0\rangle = \frac{M_0|\psi\rangle}{\sqrt{\langle\psi| M_0^\dagger M_0 |\psi\rangle}} = \frac{M_0|\psi\rangle}{\sqrt{|\alpha|^2}} = \frac{\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\end{bmatrix}}{|\alpha|^2} =$$

$$\frac{\begin{bmatrix}\alpha\\0\end{bmatrix}}{|\alpha|} = \frac{\alpha}{|\alpha|}|0\rangle$$

Likewise, we would calculate probability of outcome 1.

Note that eq. 3.2 is also satisfied, since

$$M_0 = |0\rangle\langle 0| = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, M_1 = |1\rangle\langle 1| = \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}$$

$$M_0 + M_1 = I$$

Postulate 4: Combining systems

How do we describe state space of composite quantum systems?

State space of a composite physical system is described by tensor product of the state spaces of the component physical systems.

To illustrate this, have two physical systems, e.g. two qubits, described by their state space. $|\psi_1\rangle = |0\rangle$ and $|\psi_2\rangle = |1\rangle$ If we want to describe them as a part of composite physical system, we will use this postulate. Therefore,

$$|\psi_1\rangle\otimes|\psi_2\rangle = |0\rangle\otimes|1\rangle = \begin{bmatrix}x_{\psi_1}\\ y_{\psi_1}\end{bmatrix}\otimes\begin{bmatrix}x_{\psi_2}\\ y_{\psi_2}\end{bmatrix} = \begin{bmatrix}x_{\psi_1}.x_{\psi_2}\\ x_{\psi_1}.y_{\psi_2}\\ y_{\psi_1}.x_{\psi_2}\\ y_{\psi_1}.y_{\psi_2}\end{bmatrix} = \begin{bmatrix}0\\1\\0\\0\end{bmatrix}$$

We can summarize these postulates in Fig. **??**

3 Nonlocal games

3.1 CHSH

3.2 Complexity

4 Reinforcement learning

Reinforcement learning(RL) is a type of machine learning in which agents learn to perform actions by interacting with environment and getting reward from it based on their actions. "The learner and decision maker is called the agent. The thing it interacts with, comprising everything outside the agent, is called the environment." [12]

Environment can either punish the agent for undesirable actions (bad behaviour - negative reward) or reward the agent for desirable actions (positive reward).

For the successful training of agent, there needs to be the right balance between agent's exploration of unknown paths and agent's exploitation of already acquired knowledge from the previous experience.

The training of RL agents consists of agent-environment interactions. Agent gets the information about the state of environment and performs action. Agent's actions change environment and agent receives reward or punishment. (see Fig. 4.1)

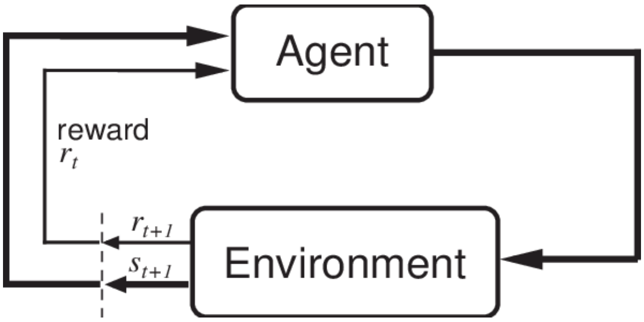

Figure 4.1: Reinforcement learning essence [12]

In the essence of reinforcement learning is the so called *Bellman equation* 4.0.

$$\underbrace{\text{New}Q(s,a)}_{\text{New Q-Value}} = Q(s,a) + \underset{\text{New Q-Value}}{\alpha}\, [\underbrace{R(s,a)}_{\text{Reward}} + \underset{\text{Discount rate}}{\gamma}\, \overbrace{\max Q'(s',a')}^{\substack{\text{Maximum predicted reward, given}\\ \text{new state and all possible actions}}} - Q(s,a)] \qquad 4.0$$

4.1 Markov Decision process

To comprehend Markov decision process, we need to firstly understand Markov chain, since it is its extension.

Random process is the sequence of random variables S_t. We call these variables states of the system in time t. If the set of all states is finite or countable, then $\{S_t\}_{t=1}^{\infty}$ is called chain. Then, the chain S_t has Markov property if

$$P(S_{n+1}|S_n, S_{n-1}, ..., S_1) = P(S_{n+1}|S_n)$$

We can describe mathematics behind RL through Markov decision process(MDP). MDP is powerful to describe decision making where a subset of outcomes are random. Therefore, agent does not have full control over remaining outcomes. MDP is a four tuple (S, A, R_a, P_a).

- S - is the set of all possible states (S_t is the state at epoch/time t)
- A - is the set of all possible actions (A_t is the action at epoch/time t)
- $R_a(s, s')$ is the expected immediate reward when state transition occurs from state s to state s' after action a. I
- $P_a(s, s') = P(s_{t+1} = s'|s_t = s, a_t = a)$ is the transition protability

4.2 Q-learning

4.3 Complexity

5 Optimalization

6 Implementation

For now just link `https://janpastorek.github.io/Bachelor-Thesis/CHSH%20-%20code.zip`

7 Analysis of obtained results

This 7.1 is the evolution of reward our agent gets through learning in 6000 epochs.

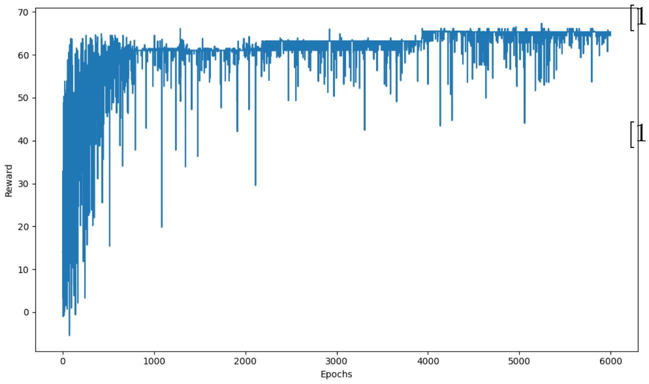

Figure 7.1: Reward

8 Discussion

9 Conclusions

References

[1] Marco Masi. *Quantum physics: an overview of a weird world: a primer on the conceptual foundations of quantum physics.* Independently published, 2019.

[2] Urbain J Le Verrier. Theorie du mouvement de mercure. In *Annales de l'Observatoire de Paris*, volume 5, 1859.

[3] Albert Einstein. The photoelectric effect. *Ann. Phys*, 17(132):4, 1905.

[4] Wikipedia, the free encyclopedia. Two-slit experiment light, 2021. [Online; accessed January 1, 2021].

[5] W. Gerlach and O. Stern. Der experimentelle nachweis der richtungsquantelung im magnetfeld. *Zeit. fur Physik*, 9:349–352, 1922.

[6] Stephen Adler. Where is quantum theory headed? *Journal of Physics: Conference Series*, 504, 01 2014.

[7] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777, 1935.

[8] John S Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.

[9] Alain Aspect. Proposed experiment to test the nonseparability of quantum mechanics. *Physical review D*, 14(8):1944, 1976.

[10] Andy Matuschak and Michael A Nielsen. Quantum country. *URL: https://quantum.country/qm*, 2020.

[11] Rxtreme. English: common quantum logic gates by name, circuit form(s) and matrices, December 2019.

[12] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction.* MIT press, 2018.

A Appendix

B Appendices